Information Processing and Management 58 (2021) 102535

Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/infoproman

hOCBS: A privacy-preserving blockchain framework for healthcare data leveraging an on-chain and off-chain system design

Ken Miyachi [a,b], Tim K. Mackey [a,c,d,e,f,*]

[a] BlockLAB, San Diego Supercomputer Center, San Diego, CA, United States
[b] IEEE, San Diego Chapter, San Diego, CA, United States
[c] Department of Anesthesiology and Division of Infectious Diseases and Global Public Health, University of California, San Diego – School of Medicine, San Diego, CA, United States
[d] Global Health Policy Institute, San Diego, CA, United States
[e] University of California, San Diego – Extension, San Diego, CA, United States
[f] S-3 Research LLC, San Diego, CA, United States

ARTICLE INFO

Keywords:
Blockchain
Distributed ledger technology
Healthcare
Privacy
Data storage
Secure computation
Health informatics

ABSTRACT

In the context of blockchain technology, "off-chain" refers to computation or data that is structurally external to the blockchain network. Off-Chain Blockchain Systems (OCBS) enable this information processing and management through distributed software architecture where the blockchain network interacts with off-chain resources. Hence, OCBS are a critical data governance component in the design of enterprise blockchain solutions, resulting in extensive research and development exploring the interplay between on-chain and off-chain storage and computation and efforts to evaluate their performance relative to other information management systems. Key features of OCBS' are their ability to improve scalability, reduce data storage requirements, and enhance data privacy, all extremely critical issues to enable broader blockchain adoption. These OCBS features map well to the needs of the healthcare industry, particularly due to the need to manage various types of medical, consumer, and other health-related data. However, different types of health data are also subject to stringent regulatory, security and legal requirements, a key factor limiting blockchain adoption in the sector. In response, there is a critical need to better align OCBS design features to different types of healthcare data management and their respective governance and privacy regimes. This article first reviews the characteristics of different constructs of OCBS. It then proposes a modular hybrid privacy-preserving framework

Abbreviations: BA, Business Associate; CHI, Consumer Health Information; CE, Covered Entity; CMS, U.S. Centers for Medicare and Medicaid Services; Dapp, Distributed Application; DAS, Direct Attached Storage; DHT, Distributed Hash Table; EHR, Electronic Health Record system; FIP, "Fit-for-Purpose" framework; FTC, Federal Trade Commission; GDPR, European Union General Data Protection Regulation ; HCS, Hyperconverged Storage; HIE, Health Information Exchange; HIPAA, U.S. Health Insurance Portability and Accountability Act; HITECH Act, Health Information Technology for Economic & Clinic Health Act; hOCBS, Hybrid Off-Chain Blockchain System; HOT-fit, Human-Organization-Technology Fit Framework; IEEE = Institute of Electrical and Electronics Engineers; IOC, Incentive Driven off-chain computation; IPFS, Interplanetary File System; IS Model, Information Systems Success Model; NAS, Network Attached Storage; ONC, U.S. National Coordinator for Health Information Technology; OCBS, Off-Chain Blockchain Systems ; P2P, Peer-to-Peer Network; PHI, Protected Health Information; PII, Personally Identifiable Information; PKI, Public Key Infrastructure; PoA, Proof-of-Authority Consensus; SMPC, Secure Multiparty-based Off-chain computation; TEEs, Trusted Execution Environments; WHO = World Health Organization; WGS, Whole Genome Sequenced

* Corresponding author at: University of California, San Diego, 8950 Villa La Jolla Drive, A124, La Jolla, CA92037, United States
E-mail address: tmackey@ucsd.edu (T.K. Mackey).

https://doi.org/10.1016/j.ipm.2021.102535
Received 27 July 2020; Received in revised form 23 January 2021; Accepted 24 January 2021
Available online 3 February 2021

leveraging off-chain and on-chain blockchain system design applied to three different reference models that illustrate how blockchain can enhance healthcare information management. Through this privacy-preserving framework we hope to liberate healthcare data by enabling sharing, sovereignty and enhanced trust.

1. Introduction

Privacy and security compliance have traditionally constrained the implementation and development of approaches for information system management of healthcare data. Traditional centralized database systems, siloed in physical locations, remain a popular and widely used approach in Electronic Health Record (EHR) system architecture, even as cloud computing and distributed systems have become more ubiquitous in other industries (Baniata, Anaqreh, & Kertesz, 2021; S. W. Smith & Koppel, 2014). These traditional system architectures are often protected by firewalls, encryption, and network isolation. Although secure, this static system architecture poses challenges to modernizing aspects of healthcare, such as encouraging patient-centric data stewardship, shared decision-making, and personalized medicine (Chu, Shah, Rouholiman, Riggare, & Gamble, 2018; Chu et al., 2016; Hoffman & Williams, 2011). Hence, there is an inherent need for an updated information system management framework, which enables consented sharing of healthcare data while maintaining security, privacy, and regulatory compliance.

Distributed systems are a type of system architecture where components are located on different computers across a network and communicate their actions by passing messages to one another. One type of distributed system architecture is called Peer-to-Peer (P2P) computing or networking where computers share files directly with one another across the Internet often associated with torrenting software. Although P2P networks significantly enhance file sharing, security and privacy are concerns (Sanjabi & Pommereau, 2010,). One solution to the privacy and security problems of P2P networks is blockchain technology, which was first proposed by Satoshi Nakamoto, the pseudo-anonymous person behind the cryptocurrency Bitcoin. Blockchain technology is a linked-list structure distributed over a P2P network that provided a systematic approach to maintain the order of transactions throughout a P2P network and avoid the double-spending problem of cryptocurrencies (Nakamoto, 2019).

Bitcoin, the first implementation of blockchain technology, orders transactions and groups them in a constrained-size structure named blocks. The nodes/peers of the network are responsible for linking the blocks to each other in chronological order, with every block containing the hash of the previous block to create a "chain" of "blocks" linked to each other (i.e. a "blockchain") (Kuo, Kim, & Ohno-Machado, 2017). Thus, the blockchain structure contains a robust and indelible record of all transactions. Blockchain began as applications enabling digital cryptocurrency transactions but have expanded into areas beyond financial transactions, such as government, energy, science, transportation, supply chain, media, copyright management, public auditing, and healthcare, wherein complex logic can be securely automated through smart contracts (Berdik, Otoum, Schmidt, Porter, & Jararweh, 2021; Chen et al., 2020; Jing, Liu, & Sugumaran, 2021; J. Li, Wu, Jiang, & Srikanthan, 2020; Mackey et al., 2019; Oham et al., 2021).

Healthcare services could significantly benefit from the evolution of traditional information management systems to more modern and participatory systems that utilize distributed networks (Berdik et al., 2021). Each year an estimated $7.2 trillion is spent worldwide on providing health services, and in the U.S alone, the 2018 bill for healthcare was $3.6 trillion (WHO, 2020). The massive spending in healthcare presents several opportunities to encourage innovation in healthcare data management and address the "Triple Aim"; improving the patient experience with care, improving the health of populations, and reducing the per capita costs of healthcare (Berwick, Nolan, & Whittington, 2008). Further, an explosion in the volume and diversity of healthcare data generation has highlighted the limitations of siloed traditional healthcare data storage systems.

Increasingly the ecosystem of healthcare data is not limited to medical records but includes a host of new sources of health and behavioral data collected outside the clinical setting, including proliferation of the Internet of Things (IoT) and mobile health (mHealth) technologies, as well as the need for remote data integrity checking (Hardin & Kotz, 2021; Zhao, Chen, Liu, Baker, & Zhang, 2020). For example, many wearable and mHealth devices now collect biometrics and consumer health data (e.g. heart rate, steps, sleep patterns, etc.) on devices owned by the individual (Dubovitskaya, Novotny, Xu, & Wang, 2019). This renaissance in digitization has led to increased attention around the applicability of privacy frameworks, such as the U.S. Health Insurance Portability and Accountability Act (HIPAA) and the European Union General Data Protection Regulation (GDPR), including debate about whether these regimes inhibit innovation and data sharing (Yuan & Li, 2019; Karampela, Ouhbi, & Isomursu, 2019). Significant health improvements could be made by opening up healthcare data for broader sharing, aggregation, and population health analysis, but maintaining privacy and autonomy is also critical (Kuo et al., 2017). However, the current state of institution-centric EHR information management systems introduces barriers that limit patient access to data, which can hamper efforts to enable patient engagement, data portability and information exchange (Hylock, 2019).

Blockchain technology has the potential to offer improvements for healthcare data management not possible with current information management system infrastructure (Mayer, da Costa, & Righi, 2019). The distributed nature of blockchain technology can enable transparent and fast access to personal healthcare data, promote data standardization, and enhance transfer and sharing of healthcare data by enabling privacy-preserving approaches using cryptography, private key management, and self-sovereign identity (Bernal Bernabe, Luis Canovas, Hernandez-Ramos, Torres Moreno, & Skarmeta, 2019). Distributed systems can also enable patient-centric data storage that utilizes an agreed upon governance model to share and distribute information under the rubric of data stewardship (Atzori, 2015). Labor intensive tasks of transferring healthcare data from different providers can be automated and mediated via smart contracts, significantly reducing both cost and time required in current systems (Kruse, Goswamy, Raval, &

Marawi, 2016). Evidencing the potential utility of blockchain in healthcare data management, several use cases have emerged, including EHR aggregation, privacy preserving algorithms for health systems data, integration of blockchain systems with the Internet-of-Things, and enhancing patient-provider directories (Dimitrov, 2019; Hussien, Yasin, Udzir, Zaidan, & Zaidan, 2019). However, there are also real-world challenges to implementing blockchain technology, including limitations on how to store and transfer data, manage permissions, and enable scalability (O'Donoghue, 2019; Attaran, 2020,).

Although blockchain technology has the potential to address contemporary healthcare information management challenges, there are inherent problems with translating traditional designs of public blockchain approaches in the context of health data storage and transfer. Specifically, the privacy and security requirements for the treatment of healthcare data are unique compared to other forms of data (Berdik et al., 2021). Additionally, there is a need for a scalable and high throughput system to enable processing of large volumes of data with relative low computation costs for the purposes of real-world clinical care. Importantly, traditional public blockchain characteristics may be incompatible with this type of dynamic treatment of healthcare data due to permission/access management, consensus mechanisms (often utilizing proof-of-work), and transparency that is distributed, necessitating new blockchain design approaches.

In response, this paper first describes the research objective of this study; the exploration of using core features and technology characteristics of Off-Chain Blockchain Systems (OCBS) as a blockchain-based design to address limitations of current healthcare information management systems. These blockchain technology features are then used to conceptualize a hybrid OCBS information system framework based on patient-centric and privacy-preserving approaches to managing three different types of healthcare data illustrated by a set of reference models. We conclude that a privacy-by-design and compliance-by-design approach is necessary for the optimal utilization of modern healthcare data and inclusive governance needs.

2. Research objective

In this paper, we first describe the current state of healthcare information systems, while also assessing ongoing challenges with traditional blockchain approaches to managing healthcare data. There is inherent friction when integrating traditional blockchain solutions with healthcare data storage and distribution due to issues related to scalability, and the security and privacy of the data stored and transferred. OCBS' mitigate this friction by integrating traditional data storage and security components that are needed to address challenges introduced by distributed ledger technologies.

Hence, the primary motivation of this study was to conceptualize, design, and evaluate a blockchain-based information management system framework that could simultaneously utilize OCBS system architecture while also preserving performance and security features required by existing healthcare data management systems, while also taking into consideration the compliance needs of different forms of healthcare data. This is necessary to ensure that the benefits of blockchain information system features can be scaled to the scope and diversity of types of health-related data, while also maintaining privacy and security, core tenets of healthcare data governance. These concepts were also crucial in breaking down data silos to ensure maximal utility of healthcare data for all stakeholders (Berdik et al., 2021).

In order to better understand how this interplay could work between traditional information systems and blockchain approaches, we first characterize OCBS system architecture and technology features, which were then used to conceptualize the specific design features of a on-chain and off-chain hybrid design architecture (hereinafter referred to as "hOCBS") mapped to key principles relevant to healthcare blockchains. We also describe an evaluation framework to assess the performance and security of the hOCBS compared to other traditional health information management systems. The aim of this study was to conceptualize a privacy-preserving blockchain system using OCBS characteristics and then translate it to three reference models of different types of healthcare data for assessment of real-world data application. Data types included:

1 Protected Health Information (PHI) defined under HIPAA (e.g., information contained in an electronic health record);
2 Consumer Health Information (CHI) (e.g., data generated outside of the clinical setting often by the consumer); and
3 Genomic data (e.g., whole genome sequencing data, data generated from a genetic screening test, etc.) (Demiris, 2016).

Specifically, OCBS' mitigate the issues associated with public blockchain systems integrating with healthcare information systems by enabling a modular and flexible system architecture that can interact with traditional data storage, while still maintaining the core value propositions around transfer, transparency, and immutability of blockchain technology. Some of the key benefits we aim to highlight in the conceptualization of our hybrid OCBS framework are the following:

- Ensuring that privacy-by-design and regulatory compliance is built into the software infrastructure of the proposed system. This would represent an improvement over traditional healthcare data storage and transfer systems that rely on other administrative forms of compliance.
- Focusing on integration with current healthcare information management systems as opposed to re-implementing all core components of healthcare data storage and transfer already available in off-chain systems. This is an improvement over other blockchain-based healthcare data management systems by reducing potential migration time. This also enables a tiered approach to converting to the new proposed architecture that is more practical.
- Enabling modularity in information system architecture as a critical improvement over other health data information systems in order to enable adaptive design responsive to different forms of healthcare data and their respective privacy and compliance requirements.

3. Related work

3.1. Limitations of current healthcare information management systems

One of the main challenges of current healthcare information management systems (e.g. EHRs, computer-based, client-server-based, cloud-based) is that the centralized and siloed storage of healthcare data in physical locations causes significant friction in sharing and transferring healthcare data (Ismail, Materwala, Karduck, & Adem, 2020; Norgeot, Glicksberg, & Butte, 2019). Furthermore, these siloed data systems create concerns about a single point-of failure, data fragmentation, system vulnerabilities, incompatible formats of data storage, and generally have failed to meet growing needs for data ingestion from non-clinical health data sources (such as Internet-of-Medical-Things, mHealth sources, sensors and wearables) that could improve the quality of diagnosis and care (Ismail et al., 2020). Another problem with these systems is the lack of active patient participation in facilitating access, management and sharing of healthcare data, as most of these systems only passively allow read access and do not have tools to facilitate data exchange (Dubovitskaya et al., 2019). Hence, current systems have yet to satisfactorily enable patient-centric and participatory information management approaches that are now in growing demand in modern medicine (Chu et al., 2016; Chu, Shah, Rouholiman, Riggare, & Gamble, 2018; Hoffman & Williams, 2011).

In response to these challenges, there have been concerted efforts by the U.S. Federal government under the Health Information Technology for Economic & Clinic Health (HITECH) Act and rules/guidance set by the National Coordinator for Health Information Technology (ONC) and the Centers for Medicare and Medicaid Services (CMS) to improve health information exchange (HIE) to reduce cost and enhance utilization (Joshua R Vest, 2010). However, progress towards more "meaningful use" required by HITECH and wide-scale adoption of HIE have been uneven, hampered by implementation and operational challenges including lack of interoperability across systems, data availability and quality, goal alignment and cooperation across different stakeholders, immaturity of current information systems, and an absence of high-quality tools and processes for exchanging data (Berdik et al., 2021; Hochman, Garber, & Robinson, 2019; Yeager, Vest, Walker, Diana, & Menachemi, 2017). Further, it is unclear if existing health information technology systems actually improve the quality of care or reduce costs, emphasizing the importance of other factors such as ensuring a combination of IT adoption and organization and technical innovation in driving productivity gains (Agha, 2014).

Given these challenges, development of health information systems that can facilitate a transition from institutional-centric to patient-mediated data sharing and that change incentives for parties in managing healthcare data are now actively being explored using blockchain technology (Gordon & Catalini, 2018). Though research and commercialization in the space of blockchain health information systems has increased, few studies or commercial solutions (where information is publicly available), specifically link the use of OCBS technology features for the purposes of concurrently addressing privacy and security with needs of data storage and scalability in a patient-centric manner as proposed in this paper. Next, we describe key components of blockchain systems, data storage, and cryptography that will act as the fundamental building blocks for our proposed hOCBS framework.

3.2. Components of blockchain systems

The fundamental technology behind all three reference models proposed in this paper is blockchain. Blockchain is a digital ledger technology comprised of near immutable, digitally recorded information in data structures called blocks. Each block is 'chained' to the next block using a cryptographic signature, hence the adoption of the term "blockchain." Consensus algorithms are used to ensure agreement in a distributed network of what operations are written to the blockchain, information which can be shared and accessed by anyone with the proper permissions (Ferdous, Chowdhury, Hoque, & Colman, 2020). The structure of blockchain technology is represented by a list of blocks with transactions in a particular order to establish data provenance. The fundamental data structure used in blockchain are linked lists. The core concepts behind blockchain technology are:

- **Node:** Any computer connected to the blockchain network is referred to as a node. A full node is a computer that can fully validate transactions and download the entire data of a specific blockchain.
- **Transaction:** smallest building block of information that can be stored in a blockchain system.
- **Block:** a data structure used for keeping a set of transactions which are validated through consensus algorithms and distributed to all nodes in the network.
- **Chain:** a sequence of blocks in a specific order
- **Consensus:** a set of rules and arrangements to verify and distribute blockchain operations
- **Digital Signatures:** A string generated by public key encryption and attached to an electronically transmitted document in order to verify the contents of the document.
- **Oracle:** In a blockchain network an oracle (human or machine) helps communicate data to a smart contract which can then be used to verify an event or specific outcome.
- **Smart Contract:** Software programs whose terms are recorded in an executable code instead of legal language. Smart contracts are automated actions that can be coded and executed once a set of conditions is met.

Our proposed hOCBS system utilizes these core components of blockchain technology as value propositions along with integration with traditional data storage and information management concepts that enable privacy and scalability.

3.3. Decentralized data storage concepts

OCBS systems enable users to store data external to a distributed system while enabling secure and standardized interaction between the distributed system and external storage. External data can be exposed to the distributed system upon meeting specific criteria, such as the owner of external data opting in and validating their identity. Data storage in OCBS systems can be categorized into three main groups: hardware storage, centralized cloud storage, and distributed storage. Hardware storage is often referred to as Direct Attached Storage (DAS). DAS includes types of data storage that are physically connected to a computer and is generally accessible to only a single machine. Some common devices in this category include:

- Hard Drives
- Solid-State Drives (SSD)
- CD/DVD Drives
- Flash Drives

Centralized cloud data storage is the storage of information such as files and databases shared between computing servers over a network. Centralized cloud storage systems are currently popular in healthcare data information management systems. However, these systems have raised issues regarding the restrictive nature of some services deployed on the cloud and broader concerns about security (single point of attack/vulnerability). Specifically, centralized cloud solutions have been vulnerable to issues such as data breaches and hacking as well as ransomware (Fernandes, Soares, Gomes, Freire, & Inacio, 2014). Some common types of centralized cloud data storage include:

- Networked Attached Storage (NAS)
- Storage Virtualization
- Hyperconverged Storage (HCS)

Decentralized storage is a system in which information is stored on multiple computers (called nodes) on a decentralized network (Andoni et al., 2019). Decentralized storage shares similarities with centralized cloud storage as users can request and receive data upon authorization and authentication on the network. Decentralized storage secures data stored in the network by automatically encrypting files, wherein only an encryption key can decrypt the data, ensuring security and proper access (Huang et al., 2020). Furthermore, through a process of sharding data, no single entity holding your information has the entirety of it. This, along with digital signatures and other traditional security techniques can ensure the security and privacy of data stored on decentralized networks. In comparison, centralized cloud storage keeps data in a central point, which can result in performance issues related to competing for bandwidth (van Steen & Tanenbaum, 2016). The nature of decentralized storage enables retrieval of data to be handled by nearby peers regardless of physical location. This can also result in higher transfer speeds due to utilizing local network bandwidth.

3.4. Cryptography concepts

Blockchain and traditional cyber security approaches heavily rely on cryptography to ensure security and privacy throughout digital storage and transfer systems (Shi et al., 2020). Our proposed hOCBS system utilizes many aspects of and layers of cryptography, along with a novel data storage and privacy-by-design infrastructure to create an optimal system for healthcare data storage and transfer. Some of the core concepts used are:

- Hashing
- Public Key Infrastructure (PKI)
- Asymmetric Encryption
- Digital Signatures
- Secure Multiparty Computation (SMPC)
- Trusted Execution Environments (TEE)
- Verifiable Computation

The value of combining certain cryptographic techniques with specific data storage types can enable highly available and scalable systems that maintain the security and privacy of traditional centralized data storage systems (Shi et al., 2020). Hashing, PKI, and asymmetric encryption are core cryptography techniques that are the foundation of blockchain technology (Yaga, Mell, Roby, & Scarfone, 2019). Digital Signatures is a broad scoped area of cryptography which underpins digital identity. In a healthcare information management system focused on data storage and transfer, it is integrally important to collect, verify, and confirm identity in the system due to the highly private nature and compliance requirements associated with data being accessed. SMPC, TEE, and verifiable computation are all security techniques used to mitigate the security and privacy issues that are introduced by distributed and decentralized systems and will be discussed in the context of our proposed hOCBS framework.

4. Characteristics of OCBS systems

4.1. Central tenets of OCBS

Examining the central tenets of OCBS systems starts with a discussion of the general technical objectives of off-chain systems, which focus on reducing computational load and data storage on blockchain networks (Eberhardt & Tai, 2017). Traditionally, data contents are locally stored directly on distributed nodes of a blockchain, leading to higher operating costs and hampering scalability (Yu, Li, & He, 2020). Hence, "off-chaining" is proposed as a method to enhance blockchain scalability and privacy, regardless of the permissiveness of a network (Eberhardt & Heiss, 2018).

Off-chain systems can enable scalability when the blockchain is used to reference or validate an off-chain data asset, without the need to store the data explicitly on the blockchain (Hardin & Kotz, 2021; Hepp, Sharinghousen, Ehret, Schoenhals, & Gipp, 2018). However, these off-chain systems lose their utility if they compromise fundamental properties central to blockchain advantages, such as enhancing data security, trust, and immutability (Eberhardt & Tai, 2018). Two of the primary challenges in this area are ensuring data integrity and data availability, as a blockchain loses an inherent advantage if it cannot ensure the trustworthiness of its associated data (Warren & Bandeali, 2017). Off-chain systems can also slow down blockchain computational output due to the unavailability of off-chain information that needs to be retrieved and verified before the computing phase (Eberhardt & Tai, 2017). Hence, existing research on OCBS focuses on experimentation to mitigate the degradation of data integrity and data availability (Poon & Dryja, 2016).

Blockchain systems also face challenges with scalability when managing large data sets, a key factor hampering broader adoption and moving blockchains into production in enterprise environments (Herrera-Joancomartí & Pérez-Solà, 2016). One of the reasons is the limited size of blocks on a blockchain, essentially making it difficult to store more complex data other than state of data, transaction history, registry entries, and hashes on a block (Sadhya & Sadhya, 2018). One solution is to use off-chain storage to store and access larger and higher complexity data through a hash pointer. For example, off-chain data may be linked with on-chain transactions that contain relevant metadata or state of data that takes up much of the on-chain data storage (Xu, Weber, & Staples, 2019). The on-chain metadata allows the secure proof and auditing mechanism that off-chain data was not tampered with or modified (Eberhardt & Heiss, 2018). Application developers can also determine what levels of availability, access and linkage they need between on-chain and off-chain data processes (Paik et al., 2019). For example, there may be different security credentials such as a public-private key pair protecting off-chain data (Lewison & Corella, 2016).

Critically, these off-chain systems must maintain certain functionality in order to be properly implemented on a blockchain through hash pointers. Off-chain systems also need to be properly backed up and distributed to avoid single points of failure, maintain stability with high accessibility, and offer unbreakable connectivity between the blockchain and off-chain storage sources during possible attacks (Eberhardt & Tai, 2017). The data stored on-chain must be immutable and unmodifiable and have a system to generate unique signifiers to integrate the blockchain network with off-chain storage locations. This is extremely important to maintain the deterministic nature of blockchain technology as off-chain data must always be retrievable given the input from the application layer or a smart contract (Eberhardt & Tai, 2017). Data stored off-chain may also be modified, though such modifications must be tracked in the on-chain network (Xiao, Zhang, Lou, & Hou, 2019). For example, if a row in an off-chain relational database table is modified, this should be recorded in the on-chain network and a new signifier should be generated wherein a new pointer is generated to the modified row.

4.2. OCBS design constructs

In order to strike a balance between maintaining utility of key blockchain features and also ensuring scalability and data integrity,

Table 1
OCBS design constructs.

Design Feature	Definition	Example
On-chain Transaction Execution [not OCBS but default for most blockchain systems]	Standard blockchain architecture wherein all the data is stored on the blockchain and distributed to connected nodes and all computation to write transactions, validate transactions, and distribute blocks is done through the blockchain network.	Bitcoin Ethereum Hyperledger Fabric
Off-chain Storage	Off-chain storage refers to storing data on an Off-chain node. An Off-chain node is an arbitrary computing machine not necessarily part of the blockchain network.	IPFS Swarm StorJ MySQL MongoDB
Off-chain Computation	Off-chain computation is an execution model where the state transition function is computed by an Off-chain Node and the resulting state then persists on-chain after verification of the computation of the state transition.	ZoKrates ZKSnarks ChainLink
Hybrid Off-Channing	Hybrid off-chaining, is the set of designs that combine off-chain state and off-chain computations in arbitrary ways and potentially in conjunction with on-chain processing.	Lightning Network Raiden Network Plasma

Table 2.
Summary of off-chain storage approaches.

Off-chain storage type	Description	Storage Method	Advantages
Swarm	Swarm is a distributed storage platform and content distribution service that stores data redundantly and distributed over multiple nodes, built as a native layer to the Ethereum Web3 stack It is primarily designed to store Ethereum's public records but also can store other types of files.	Swarm is DDoS-resistant, fault-tolerant, and censorship-resistant during large-scale operation. Swarm stores information in P2P networks separate from on-chain Ethereum storage in basic units called chunks, which are limited in size. Chunks of data are assigned a unique identifier known as a reference that allows clients to retrieve and access files.	Swarm is DDoS-resistant, fault-tolerant, and censorship-resistant during large-scale operation. Swarm philosophy is very anti-censorship and utilizes an incentive system to motivate peers to offer data storage, which can even penalize a peer for losing the hosted data of another party.
IPFS (Interplanetary File System)	IPFS is a distributed file management system with the goal of connecting computing devices with the same files. IPFS is a platform agnostic decentralized storage system and is linked to a cryptocurrency called FileCoin to incentive consumers to make storage available to the IPFS network.	IPFS identifies a file's content with its hash value while storing it in a Merkle directed-acyclic graph (DAG) for fast access (Benet, 2014). Hash values created on IPFS are unique, enabling distributed hash table (DHT) systems like Coral and Kademlia to take advantage of their ability to enable constant lookup times. This allows IPFS systems to act as a distributed repository for data that can be queried and linked based on a unique hash value through a DHT.	IPFS can transport large files, split into multiple parts called "chunks", each with by default, a maximum size of 256 kB. In place of the file on the DHT, a list of the chunk addresses is saved.
StorJ	StorJ is a network of distributed and encrypted data storage that splits a file into different "shards", which then get encrypted on the client site for uploading to different storage nodes. The hash value of the original file is used and supplemented with ordering information of the blocks. StorJ uses a DHT to locate all the shards and piece them together. These files are also encrypted before sharing and the person uploading it has their own private key to validate ownership.	As individual/private computers or servers can be used, the shards are stored redundantly on multiple machines. Further, the storage owner gets an incentive payment after checking the availability and integrity of the data from time to time. Splitting, hashing and encrypting has to be done by the publisher of the file, an approach that is generally computationally intensive for the client.	The developer of StorJ (Storj Labs) also runs a for-profit business that rents out its network to thousands of users and charges for the network usage. This represents a more centralized model and draws comparisons with traditional cloud storage services such as Dropbox and Google Drive. StorJ has recently developed an integration with IPFS where users can upload and store files to the Storj network through the IPFS system.

OCBS' utilize different design constructs (See **Table 1**). On-chain transaction execution is generally the default for standard blockchain architecture, where data is not only stored on-chain, but also enables on-chain transactions to occur without the need to interface with off-chain computational resources (Eberhardt & Heiss, 2018). Following, traditional on-chain storage and transaction environments, there are generally three design constructs for OCBS' including: (1) off-chain storage; (2) off-chain computation; and (3) hybrid off-chaining (Eberhardt & Heiss, 2018).

Hybrid off-chaining refers to an OCBS that combines both off-chain storage and off-chain computation as further discussed below. There are a number of different implementations of hybrid off-chain system due to all the potential combinations of off-chain storage and computational design features. This makes it difficult to describe hybrid off-chain systems in a general manner. However, some popular implementations of hybrid off-chaining include payment channel protocols such as the Lightning Network (Bitcoin) and the Raiden Network (Ethereum). Computation/State transitions are agreed upon by participant signatures and transactions are temporarily stored off-chain and committed to the network upon verification. Plasma is another hybrid off-chain system that combines off-chain state storage and off-chain computations by building a hierarchy of blockchains to achieve high scalability.

4.2.1. Off-chain data storage

Off-chain storage refers to any data storage that resides external to on-chain data. Therefore, traditional data stores such as DAS, NAS, storage visualization, centralized databases, cloud computing, and physical servers, are considered off-chain storage and can be incorporated into OCBS'. However, other decentralized data storage platforms are being developed in order to better align with blockchain technology different from traditional storage approaches. Off-chain storage is important for healthcare data storage for both privacy and scalability reasons. Many types of healthcare data are subject to regulatory requirements that require controlled and secure environments (e.g. HIPAA-compliant databases/storages), often not compatible with public storage (whether encrypted or not), therefore requiring off-chain storage (Pasquale & Ragone, 2014). By connecting off-chain healthcare data stores the friction regarding data transfer is significantly decreased. Furthermore, access and modification of data can be tracked in order to establish better data provenance.

There are a variety of open-source projects for decentralized data storage. Three popular decentralized storage systems are Swarm, IPFS and StorJ (Huang et al., 2020) (see **Table 2**). Overall, Swarm, IPFS, and StorJ have many similarities but also introduce different approaches to decentralized storage that can be adopted for different use cases, including healthcare. IPFS represents a platform agnostic approach to distributing data for purposes of linkage via hash values and is the most popular and mature of decentralized storage solutions and has been proposed for other blockchain-based data storage systems (Khalid, et al., 2021). Swarm enables data redundancy and anti-censorship incentivization for distributed storage, as well as being built into the base layer of the Ethereum Web3 stack. StorJ enables both distributed data storage and encryption on multiple personal and enterprise machines and offers consumer-friendly and dynamic usability that promotes adoption.

P2P file sharing was made popular by torrenting platforms such as LimeWire. The decentralized storage platforms outlined above improve on these P2P systems by incorporating incentivization mechanisms, along with encryption, security and integrity features. However, certain use cases require a centralized data store to be used. Depending on the type of data, storage needs, regulatory restrictions, accessibility requirements, and functional integration needed to a blockchain system, all of these off-chain storage approaches have different and unique capabilities that can be deployed.

4.2.2. Off-chain computation

Public blockchains present a host of potential privacy issues due to transactions being distributed to all participating nodes, which can also have a negative impact on the overall computational performance of the blockchain network (Eberhardt & Heiss, 2018). Private blockchains improve privacy issues by specifying authorization rules to join the network. However, if there is a single breach or mistake in authorization credentials, the improper participant could gain access to transactions distributed to participating nodes. Off-chaining has been suggested as a solution to address these limitations by offloading computational efforts and data-storage outside of the blockchain environment (i.e. processing and validation of transactions) (Eberhardt & Tai, 2017). However, this architecture introduces issues regarding the availability and immutability of data, which cannot be compromised as they are the defining aspects of blockchain technology.

Off-chain computation is required in OCBS systems to validate the integrity and correctness of off-chain data storage, as well as assisting with the scalability of blockchain-based healthcare data storage and transfer systems. Specifically, distributed file storage, such as IPFS, introduces new security and data integrity risks because the physical location of the storage is dynamic. Furthermore, processing large amounts of data may be computationally expensive and slow down on-chain activity. For example, querying large EHR, clinical, and genomic datasets may require significant computational logic and processing (Schadt, Linderman, Sorenson, Lee, & Nolan, 2010).

Off-chain computation can be used in a variety of ways and can integrate certain cryptographic techniques to mitigate issues that arise from using OCBS'. Off-chain computation is often used to perform state transition and computational logic to speed up and bypass on-chain verification and data distribution. These types of off-chain computations often cannot be verified by the same consensus algorithms used in on-chain computation (Eberhardt & Heiss, 2018). For example, when a payment channel is setup on the Bitcoin Network through Lightning, the transactions are not verified until the payment channel is closed, improving the speed in which the Bitcoin currency can be moved. While off-chain computation can reduce redundant processing, the mechanisms involved often introduce trust issues between parties, which are avoided in traditional on-chain transactions (Eberhardt & Tai, 2017).

Theoretically, in a system where one entity's computers/nodes are validating all transactions on a network, all the other participants must trust that entity to act in good-faith and that their validation mechanisms are correct and trustworthy. Hence, a number of

different off-chain computational models have been conceptualized to address this challenge while also addressing scalability and privacy (Eberhardt & Tai, 2017). Some popular approaches include Verifiable Off-chain Computation, Enclave-based Off-chain Computation, Secure Multiparty-based Off-chain Computation, and Incentive Driven Off-chain Computation, some of which will be used in our proposed reference models (see **Table 3** for summary).

Table 3.
Summary of off-chain computation approaches.

Off-chain storage type	Description	Technical Requirements/ Functions	Advantages	Example(s)
Verifiable Off-chain Computation	Verifiable Off-chain Computation utilizes cryptographic proofs to ensure the integrity and correctness of off-chain computations upon being written to the blockchain. Verifiable Off-chain Computation is a technique where an off-chain node (known as a Prover) executes a computation and then publishes the result of that computational output and generates a cryptographic proof attesting to the computation's correctness, which is then sent to the blockchain (Eberhardt & Heiss, 2018). An on-chain node (known as a Verifier) is designated to verify the proof and persists the result in case of success.	The underpinning cryptographic techniques are extremely complicated and there has been extended research resulting in many variants of verifiable computation schemes. Verifiable Off-chain Computation requires functionality such as *Non-Interactivity, Cheap Verification, Weak Security Assumptions*, and *Zero Knowledge* in order to integrate properly and preserve the value propositions of blockchain technology (Eberhardt & Heiss, 2018).	The major tenant behind Verifiable Off-chain Computation, is that a Cryptographic Proofing Process replaces the consensus algorithm for transaction verification.	Popular Off-Chain Computation projects include zk-SNARK, the underlying technology behind zCash. Zk-SNARK stand for zero-knowledge succinct Non-Interactive Argument of Knowledge and has strong privacy guarantees enabling fully encrypted transactions on the blockchain, while still being verified for correctness and integrity. ZoKrates, is a popular project which is the Ethereum (Ethereum is a popular public blockchain environment focused on smart contract functionality) implementation of zk-SNARK (Eberhardt & Tai, 2018).
Enclave-based Off-chain Computation	In this model Trusted Execution Environments (TEEs) are used to guarantee confidential and integral code execution of computational processes. A TEE is an isolated and secure processing environment that protects a computational execution from the rest of the system . Therefore, TEEs can still perform secure computation even if a computer has been accessed in an unauthorized manner or a bug occurs anywhere else in the computer.	To guarantee the enclave's authenticity, an attestation certified by a trusted external entity is attached to every message that is computed and distributed by the TEE. For example, private keys are often embedded into the TEE portion of the chip during manufacturing and every message may be encrypted with a private key. The private key ensures a message is coming from the TEE as the associated public key would not be able to properly decrypt any message signed by a different private key.	This approach enables verification of TEE execution and adds an extra security layer if a TEE computational result was intercepted .	An example of an Enclave-based Off-chain computation is Intel SGX, which is used as the Trusted Execution Environment (TEE) for the Hyperledger Sawtooth PoET algorithm.
Secure Multiparty-based Off-chain computation (SMPC)	This type of computation enables a set of nodes to compute functions on secret data in a way that none of the nodes ever has access to the data in its entirety .	SMPC is a subfield of cryptography that enables parties to jointly compute a function over their inputs while keeping those inputs private.	SMTP assures security and integrity of communication or storage of participants' privacy from each other (in a shared network) (Lindell, 1AD).	One Secure Multiparty-based Off-chain computation is used by a project called zkChannels, designed to build a modular privacy preserving transactional layer on-top of many blockchains created by Bolt Labs.
Incentive Driven off-chain computation (IOC)	IOC utilize systems that assume economical and rational behavior of blockchain participants who will strive to maximize their own utility (Eberhardt & Heiss, 2018).	System rules can be enforced by retaining deposits as a leverage against contravening activity and by financially rewarding desired behavior. For example, an entity may be required to deposit collateral in order to use their resources for off-chain computation. In the event error and malicious behavior occurs, the collateral would be lost to the system.	IOC systems use financial rules to ensure private and secure computation.	One incentive driven off-chain computation is called ChainLink, which is a decentralized Oracle designed to eliminate any one point of failure and incentive Oracle smart contracts that are highly secure, reliable, and trustworthy.

5. Design characteristics of healthcare blockchains

We now pivot to an assessment of how to apply blockchain technology concepts and central tenets of OCBS to health-related distributed data governance. Specifically, these concepts are mapped to our proposed hOCBS framework that emphasizes a patient-centered privacy-preserving approach to ensuring the resilience, provenance, traceability, and management of different forms of healthcare data. Factors such a scalability, accessibility, portability, identity validation, and maintaining the overall utility of data to effectuate healthcare transactions and processes, must be evaluated to determine optimal design features. Our modular framework also focuses on ensuring system architecture maximizes the benefits of OCBS in alignment with security and privacy legal requirements across different types of healthcare data. In conceptualizing our hOCBS framework, we adopt the "Fit-for-Purpose" (FIP) design framework by *Mackey et al.* Specifically, the FIP considers the following key design principles: (a) the general type of blockchain being developed (public, private, or consortium); (b) developing a data governance approach that includes identifying nodes in the blockchain network, their respective roles, and their decision-making process; and (c) deciding on a permissions structure for blockchain participants and assets (Mackey et al., 2019). Each of these hOCBS characteristics is described below and then discussed in the context of specific reference models.

5.1. Healthcare-centric data governance

Data governance defines decision-making and responsibility rights for a system or organization's use of data, executed based on agreed-upon models that describe the actions, actors, and circumstances to perform data-related processes (Abraham, Schneider, & Brocke, 2019). Data governance related to the use of healthcare data is of particular importance due to strategic business and regulatory requirements specific to the industry, including interoperability with other data systems (such as in and between EHRs, patient portals, and patient and provider directories), strict compliance with HIPPA and the management of PHI, and ensuring data maintains utility for its diverse set of users including patients/consumers, providers, healthcare administrators, and researchers (Hripcsak, et al., 2014).

It is also critical that healthcare data governance define decision-making and responsibility rights between internal and external organizations that may be involved in multiparty healthcare transactions. Healthcare delivery is not only complex, but also involves the coordination of care across multiple stakeholders, including different clinicians/providers, hospital systems, healthcare administrators, payers/insurers, and of course the patient and their families (D'Amore, Sittig, & Ness, 2012). Hence, distributed governance needs to ensure not only that information is communicated safely and securely, but also in a manner where robust identification and authentication schemes can validate identify of involved parties (Berdik et al., 2021).

Adding an additional layer of complexity is the policy environment that governs healthcare practice, coverage, and reimbursement, which can directly impact the availability and usability of healthcare data (Institute of Medicine (US) Committee on Regional Health Data Networks, Donaldson & Lohr, 1994). For example, national and state-level health and insurance policy often define different levels of healthcare coverage, which can lead to data silos, lack of health data portability, and challenges with interoperability in comparison to single-payer or universal healthcare systems (Institute of Medicine (US)National Academy of Engineering (US) Roundtable on Value & Science-Driven Health Care, 2011). All these factors need to be considered when devising good governance approaches to managing healthcare data regardless of technology utilized.

5.2. Blockchain system design for healthcare

Healthcare-related blockchains are often structured as private or consortium (also known as "hybrid") designs. This differs from public or "open" blockchain systems that allow public participation and have a wider user base to maximize decentralization, yet with no central authority. Public blockchain networks are the common design for many popular cryptocurrencies (e.g. the bitcoin blockchain). In contrast, private blockchain networks require participants or nodes to have their identity validated and pre-determined and are closed to the public (Zhuang, Sheets, Shae, Tsai, & Shyu, 2018). A private blockchain network, also known as a "permission-based blockchain", is preferred for healthcare use cases because of requirements for identity validation tied to levels of privacy and permissions to manage access to data (Pirtle & Ehrenfeld, 2018). By their definition, private blockchains have an owner or operator that determines the governance of the blockchain (such as who can join, permission structures, and rules about reading and writing to the ledger) and are inherently less decentralized than public blockchains.

Consortium-based blockchains combine elements of both public and private blockchains. In this sense, consortium blockchains are similar to private blockchains as they are not open to public participation, but instead participation is granted to a group of entities/organizations that meet certain pre-established credentials or criteria set by the consortium (Mackey et al., 2019). However, they are also similar to public blockchains as their operation is not limited to a single entity, but instead granted in a semi-decentralized manner by consortium members. Hence, consortium blockchains have their participation and authority more distributed and de-centralized compared to private blockchains and also allow pre-defined participation of nodes to smaller groups that share a common goal in the operation of the blockchain network.

Private and consortium blockchain approaches offer distributed networks that can enable more compatible health data governance that validates digital identity for the purposes of data storage, access and sharing (Azaria, Ekblaw, Vieira, & Lippman, 2016). For example, patients can leave and rejoin a blockchain system multiple times, for arbitrary periods, and always regain access to their history by querying, linking and decrypting on-chain transactions written to the ledger associated with their validated digital identity. As long as there are nodes participating in the network, the blockchain log is maintained along with the historical record and

provenance of that data. Hence, both private and consortium-based blockchain designs can enable shared healthcare data management tied to digital identity. These systems can also benefit from other key blockchain advantages of immutability, creating a central record of data management agreed to by all participants, an audit log, and enabling data access based on rules set by the network (Mackey et al., 2019).

Further enabling private and consortium blockchain designs in healthcare are OCBS characteristics, which ensure the correct balance between on-chain and off-chain storage and computation, while also enabling more dynamic permission structures. Other

Table 4.
Examples of healthcare data permission structures and roles adopted from MedRec.

Stakeholder	Role	Example of Implementation
Patients/Data Owners	The patient role is the primary "data owner" of information throughout the framework and choices about data access and availability are dictated by these preferences through smart contracts. The patient smart contract functions as an authorizer for patient data and an audit trail for participants in the system to locate their medical record history or other forms of healthcare-related data.	Relevant smart contracts contain references to Patient-Provider Relationships (PPRs) or other relationships with stakeholders in the system. These references represent all the participant's previous and current engagements with other nodes in the system. Patients, for instance, would have their timestamped data structure populated with references to all care providers they have been engaged with or anyone who has accessed their data. The patient smart contract also implements functionality to enable user notifications.
System Administrators	This global role maps participant identification codes to their blockchain wallet address used to identify nodes performing transactions in the blockchain network (e.g. healthcare administrators, staff involved in reimbursement, etc.). Mapping identification codes to blockchain wallet addresses allows integration into existing healthcare ID systems. Policies coded into the system administrator smart contract can regulate registering new identities or changing the mapping of existing ones. Identity registration can thus be restricted only to certified institutions, providers, and patients.	The system administrator will be connected to an Oracle Smart contract that will enable the verification of identification in third-party databases such as licensing databases, patient and provider directories, Medicare eligibility databases, etc.
Healthcare Providers	The provider role is associated with licensed healthcare providers and must be verified from external data sources (e.g. provider directories, National Practitioner Data Bank, etc.) in order to obtain access into the system. Providers in the system set the relationship status in their patients' smart contract whenever they update records or as part of creating a new physician-patient relationship. Accordingly, the patients can poll their smart contract and be notified whenever a new relationship is suggested, or an update is available. Patients can accept, reject or delete relationships, deciding which records in their history they acknowledge.	The Provider Smart Contract establishes a PPR between two nodes in the system when one node stores and manages medical records for the other. While we use the case of care provider and patient, this notion extends to any pairwise data stewardship interaction. The PPR defines an assortment of data pointers and associated access permissions that identify the records held by the provider. Each pointer consists of a query string that, when executed on the provider's database, returns a subset of patient data. The query string is affixed with the hash of this data subset, to guarantee that data have not been altered at the source. Additional information indicates where the provider's database can be accessed in the network, i.e. hostname and port in a standard network topology. The data queries and their associated information are crafted by the care provider and modified when new records are added. To enable patients to share records with others, a dictionary implementation (hash table) maps viewers' addresses to a list of additional query strings. Each string can specify a portion of the patient's data to which the third-party viewer is allowed access.
Other Third Parties that seek to access or provide Healthcare Data (e.g. Researchers, Population and Behavioral Health companies, etc.)	Given our framework is based on a Consortium blockchain design, other third parties such as research entities, non-profit organizations, and health and wellness companies seeking access to healthcare data will need to go through a screening process in order to be included as consortium members. Consortium members will acknowledge that any access to off-chain data will be governed by smart contract provisions dictated by the individual and relevant privacy regulations depending on the type of data.	Third parties will only be able to access data wherein the data owner has explicitly opted-in upon a third-party request to access data. Furthermore, they will only have specific rights in the blockchain network such as read only, single-use access upon an opt-in access grant from the data owner, verification through an external database and extending the on-boarding requirement process that can be agreed upon through network consensus.

fundamental technical components of this blockchain-based infrastructure critical in their integration with OCBS are validation, consensus and digital identity.

5.3. Healthcare data permission structures

Structuring data permissions on a blockchain is driven by privacy considerations, which in turn can originate from different legal interpretations. Generally, privacy includes the two fundamental concepts of anonymity and confidentiality. In blockchain terms, anonymity focuses on concealing the parties to a transaction, and confidentiality addresses the need to hide transaction details. Certain privacy regulations, including GDPR provisions, align with blockchain approaches that are geared towards improving data portability, data traceability, lawful access auditability, and consent management (Hawig, Zhou, Fuhrhop, Fialho, & Ramachandran, 2019). This is particularly true with private or consortium blockchain approaches, where transactions of digital records can be deleted or modified depending on the use of off-chain storage and consensus algorithms (Puthal, Mohanty, Yanambaka, & Kougianos, 2020). However, blockchain systems must also take into account other provisions of GDRP that may be less compatible with distributed information systems, such as having the "right to be forgotten" or automated smart contract execution that may limit a user's control and autonomy over their data (Mirchandani, 2019).

Hence, the first step in defining rules about privacy on a blockchain is by developing permission structures at a granular level with appropriate verification and authentication to identity of parties (Berdik et al., 2021). This includes the use of smart contracts with privacy features imbedded in their structural design, such as conferring access to a patient's personal information only to those granted lawful access (Dwivedi, Srivastava, Dhar, & Singh, 2019). Permission structures are also important when considering consortium blockchains where participants may be in "coopetition" (i.e. when enterprises collaborate in the business network to achieve a mutual objective, while also actively competing with each other) (Dagher, Mohler, Milojkovic, & Marella, 2018). In this case, not only is the patient data subject to privacy restrictions, but underlying business and transaction data between parties can also be deemed as highly sensitive and confidential. For this reason, user identity should be mapped to specific smart contract addresses in order to ensure proper use of the system and restrict user functionality based on their roles in the system. For example, a patient would only have the functionality to change, modify, or specify access to data tied to their validated digital identify.

As private and consortium blockchain designs are the most compatible approach for healthcare use cases, understanding how to formulate permissions structures as a critical interface between on-chain and off-chain data storage and computation while preserving privacy is critical (Cao, Sun, & Min, 2020). The multi-stakeholder nature of healthcare necessitates that multiple parties have access to data in order to effectuate optimal clinical decision-making, improve patient safety, and enable better population health outcomes. Hence, validated digital identity on blockchains using pre-defined roles that dictate data access privileges (both read and write) governed by smart contacts forms the basis for how we conceptualize a hybrid OCBS health information management system.

To further illustrate how this permission structure would work for different stakeholders, we describe blockchain permission-based roles in **Table 4** that are adopted from the blockchain solution MedRec operated by Beth Israel Deaconess Medical Center (Azaria et al., 2016). However, we note that the privacy-preserving emphasis of our framework necessitates placing the patient/consumer at the center of this permissions structure though validation of their digital identity as will be discussed in the reference models.

6. hOCBS health data framework reference models

6.1. Reference model framework design

We developed three reference models for our conceptual hOCBS framework in order to better illustrate the real-world utility and application of the framework to different forms of healthcare data. We focused on three distinct types of healthcare data due to their different purposes, sources of information, use and ownership, and legal treatment that include PHI, CHI, and genomic data. Importantly, all of these approaches utilize a hybrid off-chain architecture that combines on-chain storage and off-chain storage as well as off-chain computational approaches adapted to the specific needs of the healthcare data use case.

Different types of off-chain storage will be used based on the needs for distributed data storage, regulatory compliance, redundancy, and security. Different off-chain computation approaches will focus on verifying the credentials and identity of the data owner coupled with determining the integrity and correctness of data storage for purposes of on-chain and off-chain interaction. Issues related to ensuring scalability of healthcare blockchain systems and enabling efficient computation of blockchain-related information stored on-chain and off-chain are also important considerations incorporated into each of these reference models.

A core foundation of all three reference models is validating to the sovereign digital identity of the patient/consumer. Given this identity-centric approach, all reference models utilize a Proof-of-Authority (PoA) consensus mechanism. PoA is a consensus method that specifies a designated number of participants with the power to validate transactions or interactions on the network. As a general approach, the on-chain storage of each reference model will only act as a de-identified access log and provide linkage to off-chain data storage. The on-chain computation that occurs on these reference models will be used to validate the access logs and data linkage pointers to off-chain datasets.

Each of our three reference models starts with a depiction of the general blockchain architectural components and its specific off-chain functions and application layers. This includes the core functions of blockchain data, which may include certain transaction data/metadata or state-of-data information residing on-chain. The data layer represents where and what data resides on-chain and off-chain, as well as the interaction between these data assets. The core design features of the blockchain are also identified, including whether the blockchain is private or consortium. Finally, an application feature layer that consists of blockchain-enabled technology

applications (discussed below) are identified. A summary of the pros and cons of hOCBS design features is also summarized in Table 5. Additionally, a summary of the design principles of each of the reference models is provided in Table 6.

6.2. hOCBS blockchain application and features layers

We also identify the blockchain application feature layers that should be incorporated into our reference models in order to enable utility of a privacy-preserving health information management system, centered on patient/consumer digital identity verification. The specific blockchain application and feature layers adopted in our proposed hOCBS framework include smart contracts, digital wallets, and tokens. Smart contracts will dictate the flow of information creation, access, retrieval, and deletion, while also allowing records and information to be stored on-chain via a digital ledger and off-chain via OCBS (Chatterjee, Goharshady, & Velner, 2018). A digital wallet, a type of distributed application (Dapp) that holds verifiable credentials about an identity and enables the signing and submitting of a transaction through an off-chain construct (e.g. private keys), will authorize healthcare data-related transactions in a secure and patient-centric manner (Mikula & Jacobsen, 2018). Tokens, whether they be in the form of utility tokens (used for a specific purpose) or security tokens (ownership in an asset), can be used to encourage certain behavior on the network, such as enabling commoditization of healthcare data assets or encouraging positive health behavior change through token-based incentives (Dimitrov, 2019).

Our hOCBS framework envisions combining these different feature layers to address current challenges of fragmented healthcare data governance for the purposes of improving data portability, transparency, and patient-centered data stewardship. For example, if a patient moves from one provider to another, oftentimes there is no automatic way for healthcare records to be transferred and made visible to the patient in a consolidated way (i.e. patients may have to access separate provider patient portals for their full medical history). Similarly, healthcare providers often rely on multiple databases populated with different forms of patient information (Vazirani, 2019). However, these siloed databases can be too restrictive in allowing for appropriate sharing of data between provider parties, particularly in the absence of the patient serving as an intermediary to provide data access permission (Lokhande, Mukadam, Chikane, & Bhonsle, 2020). To address this, our framework puts the patient/consumer as the central intermediary within a distributed permissioned-blockchain network, allowing them to control the flow of data access per specific rules pre-determined in smart contracts that are also subject to applicable privacy requirements depending on the type of data.

Specifically, the terms of data access and sharing will be explicitly defined into the smart contract layer based on the health data classification and relevant schema (e.g. PHI, CHI, genomic data). This will enable the individual to act as the data steward for their own health information with associated visibility to the terms dictating how their data is shared, and who gained access to their data. This user-centered decision-making is also important to enabling dynamic consent management, where the user has the ability to adaptively change their "consent" to who can access their data, an approach that aligns with HIPAA and GDPR requirements (European Parliament, 2019). Importantly, permission preferences and consent versioning will be logged and recorded on-chain in a way that does not expose the underlining off-chain healthcare data itself and ensures its underlining security.

In the absence of blockchain and smart contract functionality, confirming what the individual has consented to in regard to data management falls back to paper and electronic-based consents that are static, or confusing terms of use and privacy statements dictated by platforms and providers that are not patient-centric or participatory. This can lead to data portability and sharing being delayed, mistaken, or potentially subject to greater risk of accidental disclosure, fraud or hacking.

6.3. hOCBS healthcare data reference models

6.3.1. Reference model 1: protected health information

The first reference model will focus on the most highly regulated form of healthcare data; Protected Health Information (PHI). The treatment of PHI is governed by the HIPAA Privacy Rule, which defines it as any information in a medical record that can be used to identify an individual, and that was created, used, or disclosed in the course of providing a healthcare service (e.g. diagnosis or treatment). In this sense, PHI focuses on information that identifies an individual patient, is linked to medical records and is created in the course of provisioning of healthcare services (National Academies, 2009).

Table 5.
Pros and cons of hOCBS features.

Design Feature	Pros	Cons
Distributed / Modular Data Storage and access/modification logs	Non-Siloed Databases, Increased Transparency of data to patient, Standardized Data Structures	Implementation Cost, Integration Issues, Attack Vectors of distributed ledger technology
Transfer and Access Control of Data controlled through Smart Contracts	Enhanced Security, Community Governance, Automated execution of rules	Smart contract bugs, Updating Smart contract features, incorrect/malicious information written to on-chain storage
Dual Traditional and Smart Contract Authorization/Authentication	Enhanced Security, Identity tied to specific functionality in the system	UI/UX of multi-factor authentication, added complexity to identity system
Data Integrity/Correctness Verification	Trust improved by automation/software, Enables integration to modular and distributed data storage	Bugs with data integrity/correctness, added computational processing to transfer data

Table 6.
Summary of design characteristics of reference models.

	Protected Health Information (PHI)	Consumer Health Information (CHI)	Genomic Data
Network Type	Consortium	Consortium	Consortium
Consensus Mechanism	PoA	PoA	PoA
On-Chain Storage	Access Logs, Linkage	Access Logs, Linkage	Access Logs, Linkage
On-Chain Computation	Validation of Access Logs	Validation of Access Logs	Validation of Access Logs
Off-Chain Storage	Central	IPFS	Modular
Off-Chain Computation	Verifiable Off-Chain Computation, TEE	SMPC, Incentive	Verifiable Off-Chain Computation, SMPC
Incentive	None (Regulatory)	Token	Token

HIPAA also focuses on two specific healthcare stakeholders in its compliance obligations; the Covered Entity (e.g. the healthcare provider, plan or clearinghouse, "CE") and the Business Associate (i.e. an organization that performs services on behalf of the Covered Entity and requires access or use of PHI, "BA"). Once deemed PHI, there are specific administrative, physical and technical controls and safeguards that must be implemented by HIPAA BAs and CEs for the purposes of data confidentiality, integrity, and availability (including the right for patients to obtain a copy of PHI). De-identification of PHI in order to relieve it from HIPAA obligations requires it to be stripped of 18 identifiers that constitute personally identifiable information (PII).

Given these regulatory obligations are focused on healthcare providers, administrators, and intermediaries, our first reference model (see Fig. 1) focuses on the management of PHI on a consortium blockchain model involving different organizations that are legally classified as Covered Entities and Business Associates, but also includes the patient at the center of this network, a party not traditionally included in this exchange process. Specifically, this consortium blockchain would invoke smart contracts dictated by

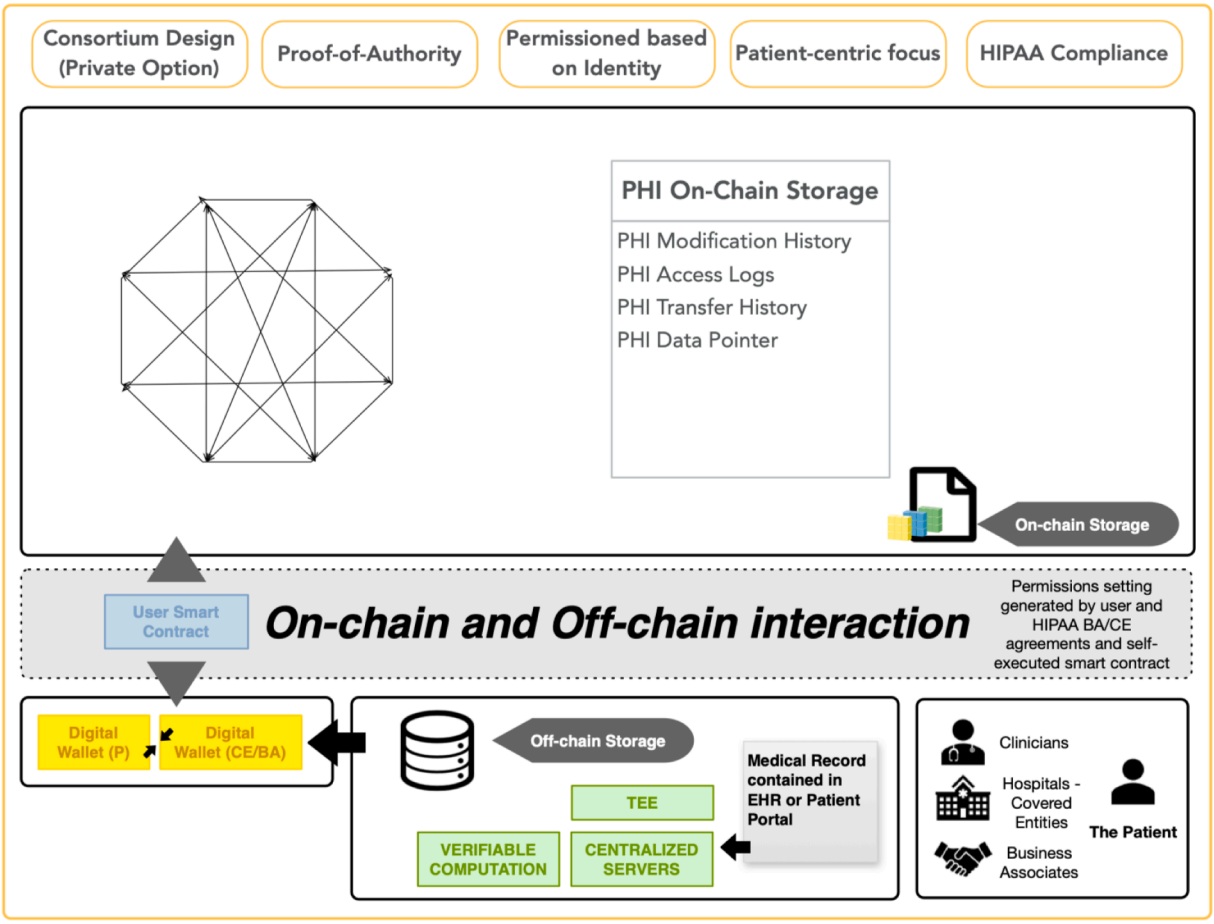

Fig. 1. PHI reference model
Description: The PHI reference model is the most restrictive and wherein the smart contract that is executed is specified by the data owner's pre-defined consent and legal requirements of the relevant privacy policy (i.e. HIPAA) which requests validation of identity of requestor, maintains private key and public key management, and grants or denies data access off-chain. Centralized Storage in Healthcare Facilities must still be implemented to maintain regulatory compliance.

patient data access privileges linked to each provider portal or EHR (retrieving medical records from these off-chain sources that hold personal health records) and populating it into a digital wallet linked to a patient's digital identity to enable greater health record portability facilitated by the patient's own participation.

Importantly, there are strict technical requirements for safeguards regarding where PHI is held. Distributed storage frameworks do not meet these criteria due to their non-specified physical location of storage. Therefore, the off-chain storage mechanism will remain located in healthcare facilities that generate PHI in an EHR or other sources. However, on-chain linkage is critical to enable the patient to specify access and transfer of PHI. Specifically, Verifiable Off-Chain Computation can be used to generate proofs regarding access granted to specific clinicians, healthcare facilities, and administrators. The verification process and data linkage from on-chain to off-chain resources can be done within TEEs to ensure security and integrity as has been suggested for blockchain applications using mHealth data (Hardin & Kotz, 2021). This will mitigate any improper access of the off-chain PHI data and will create an extra layer of security dictated by the patient.

The PHI reference model is the most restrictive with rules governing the smart contract specified by the data owner's pre-defined consent and also applicable legal requirements of HIPAA. Records will be shared by the patient's different providers who participate in the consortium blockchain but operate their own distinct EHRs, based on the permission rules set by the patient and any underlining HIPAA BA agreements, all executed through dynamic consent and smart contract logic/rules. The reference model requests validation of identity of the requestor and source (BA or CE), maintains private key and public key management, and grants or denies data access off-chain. Centralized storage in Healthcare Facilities must still be implemented to maintain data security and regulatory compliance per HIPAA. Permission rules may utilize off-chain storage databases to determine credentials and access of specific entities (e.g. for example making rules-based distinctions between a requestor who is validated as a licensed private health insurance provider versus Medicare provider).

The primary use case of this reference model is to enable patient-mediated and centered management of PHI in a secure environment with lower computational costs via hybrid on-chain and off-chain storage and computation. Data storage and computational costs are reduced as data continues to reside off-chain and consensus to write to the chain is based on POA tied to the patient identity. Costs can further be reduced by removing manual execution of HIPAA legal agreements via smart contract logic that facilities access and permissions of BAs and CEs with electronic consent by the patient. The real-world application focuses on enhancing portability of healthcare records subject to HIPAA requirements (the legislative intent of the HIPAA). As the current system is institutional-centric between HIPAA BAs and CEs, patients currently do not have decentralized health information systems that allow their active participation in accessing, sharing, and exchanging their healthcare data even though there are requirements for meaningful use. Hence, the proposed framework will allow healthcare data to remain in off-chain secure storage per HIPAA, but also enable patients to exchange data to a decentralized group of validated consortium healthcare providers and administrators in order to improve continuity and quality of care that is currently fragmented between providers, facilities and payers.

The technical blockchain and OCBS components of the PHI reference model were designed to enable patient control and access to their medical history from off-chain data sources and also have a verified history of all entities that processed or accessed their PHI in a HIPAA compliant fashion. The PHI reference model is the most rigid and similar to existing health information systems that utilize traditional data storage and transfer systems but emphasizes a blockchain-mediated form of health information exchange. As stated previously, it is necessary to utilize existing health information systems (e.g. institutional EHRs) given the regulatory compliance requirements needed to manage PHI. A summary of the key OCBS features of the reference model are provided below:

- **OCBS Feature Advantages:** Allows data storage to stay on current HIPPA compliant systems, while utilizing the security and auditability advantages of blockchain technology to enable health information exchange from existing EHR systems or patient portals with the reference model.
- **Potential Challenges:** Interoperability, integration and data standardization with existing health information systems are challenges with this model, requiring use of industry standardization of data formats (e.g. HL7 Fast Healthcare Interoperability Resources or IEEE P2418.6 standard for Framework of Distributed Ledger Technology Use in Healthcare and the Life and Social Sciences).
- **Security Considerations:** Security is added to ensure proper authorization and authentication of both access and modification of data. Furthermore, cryptographic techniques are used to validate information stored on the connected databases such as TEE and verifiable computation. Multiple layers of security are used to mitigate mistakes, and misuse of the system.

6.2. Reference model 2: consumer health information

The second reference model in our framework is designed for healthcare data that is subject to less stringent privacy requirements; Consumer Health Information (CHI). CHI generally refers to "information on health and diseases that is created for and directed to the general public", but also includes data that is generated and shared by consumers related to their health, lifestyle and well-being (Sherif, Pluye, Thoër, & Rodriguez, 2018). CHI-derived data is diverse, including aforementioned behavior, biometric and sensor data generated by IoT, wearables, connected medical devices, mHealth apps, consumer health portals, online health coaching platforms, blogs, online forums, social media posts, and other data sources that engage consumers in shared and interactive environments regarding their health.

The treatment of CHI is not explicitly governed by the HIPAA Privacy Rule if it does not include PII or is not involved in the course of care by a Covered Entity. In this sense, CHI data (generally any health-related data residing outside of the medical record or that is not transmitted to a healthcare provider for treatment) is much more expansive in scope and diversity than PHI, and is estimated to see

rapid increases in data volume (Demiris, 2016; S. Smith & Duman, 2009). Properly utilizing this data opens up many opportunities to advance individually tailored and personally driven healthcare solutions, such as precision medicine and behavior change and modification interventions. CHI also raises different privacy considerations regarding the treatment of PII that also includes health-related status or information but does not legally constitute PHI.

Hence, our second reference model focuses on the control, management, and sharing of CHI that is consumer-centric and subject to general privacy frameworks such as GDPR and the Federal Trade Commission Act. The second reference model maintains privacy as a key feature of the architecture, wherein data sharing can be opted into explicitly by the consumer for distributed sharing across a private blockchain or a trusted consortium or marketplace of vetted participants who agree to ethical principles of using CHI (see Fig. 2). Distinctly different from the PHI framework is that this data is generated primarily from consumer-initiated sources, though CHI may nevertheless be fragmented across different platforms, apps, websites, and other data stores. Specifically, this model envisions consumers managing their own CHI in a digital wallet tied to their validated identity governed by smart contracts.

Due to the diversity of types of CHI data, the utility of real-time and connected health data that can enable continuous and remote monitoring, along with the inherent value of data aggregation to identify broader trends in population health, we adopt IPFS for off-chain storage management. IPFS enables a fault tolerant, scalable, highly available distributed file store system that can be secured, modified, and deleted by the user. Sharing data for aggregation and big data analysis means that correct profiling of data is critical. Therefore, TEE will be used to execute any modifications to IPFS. Using this approach, CHI data can be retrieved from respective off-chain sources (APIs, consumer web portal profiles, etc.) and shared with participants in exchange for token payments with the consumer deciding what underlining data, metadata, or PII is shared per dynamic smart contract terms. Consumers could also share their

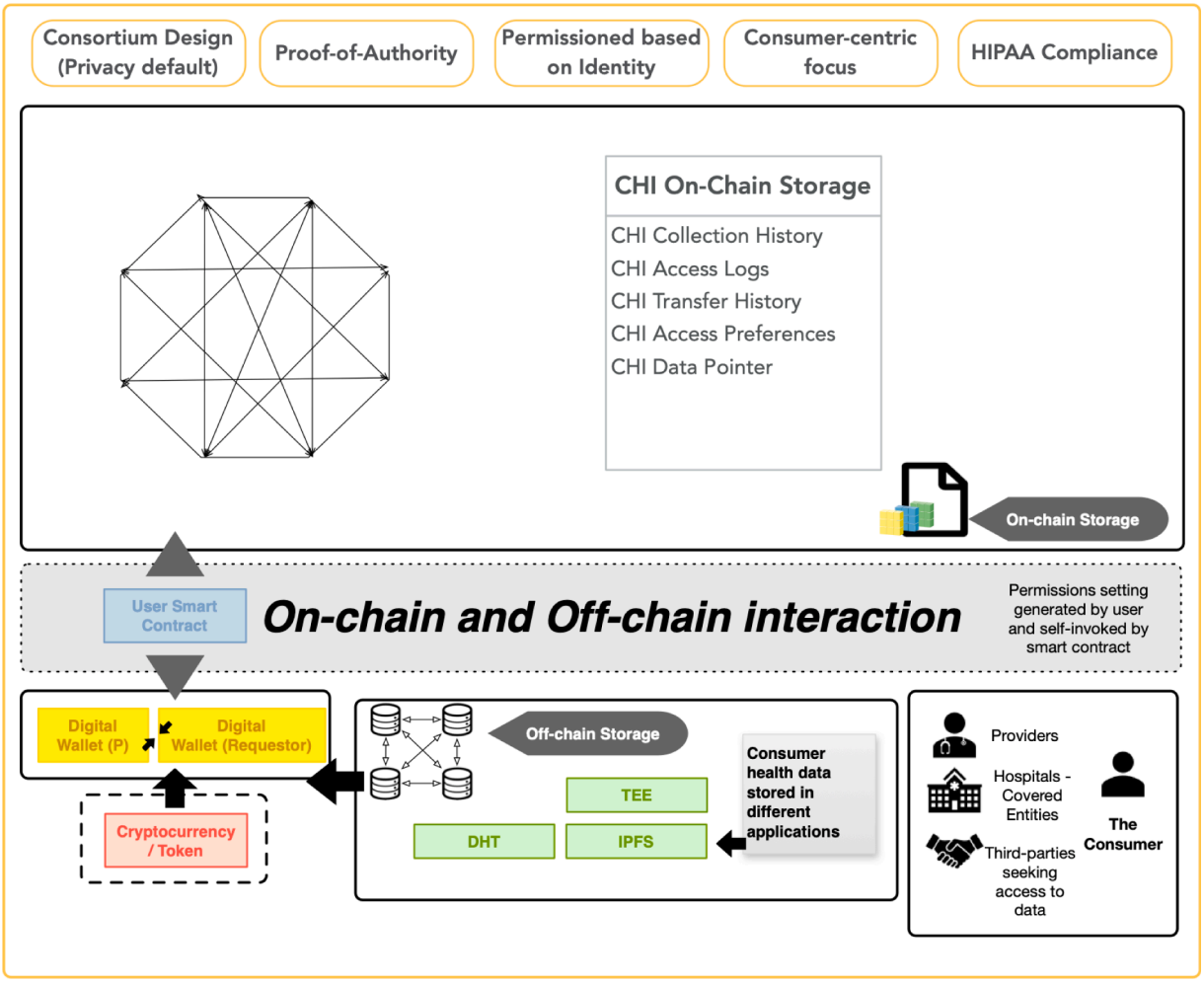

Fig. 2. CHI reference model

Description: The CHI Reference Model utilizes a decentralized storage mechanism such as IPFS. Off-chain computation is executed using a trusted execution environment (TEE) and using Distributed Hash Tables for fast access to data and verifying identity, and associated access and read privileges. This, along with dynamic security settings for CHI data access, controlled via a smart contract and a token-based incentive system enables and promotes sharing of CHI data while maintaining privacy and security of CHI data.

CHI to researchers through active consent written to the blockchain for clinical, epidemiological and health outcomes research.

The primary goal of this reference model is to manage CHI for the purposes of enhancing sharing in a privacy-preserving manner, while also helping to develop digital health tools that can provide more personalized care when augmenting clinical data. The blockchain and OCBS technical components of the CHI reference model were chosen to maximize user control and shareability of data as use is not as restrictive as PHI under HIPAA. Specifically, the CHI reference model is more flexible than the PHI reference model and is designed to optimize data transfer speeds throughout the system. Distributed data storage leveraging IPFS allows users to not only retrieve CHI, but also host their own content via hash values and content addresses in a platform agnostic environment similar to P2P networks. Validation of access to and exchange of CHI linked to a consumer's digital identity via POA can help ensure fast access, querying, and sharing governed by smart contract logic specific to a consumer's preferences.

Though generating a blockchain P2P distributed network for CHI might concern privacy advocates, the motivation for this approach would be to enable a privacy-preserving OCBS framework that allows consumer-driven autonomy over CHI in a decentralized manner. Currently, most CHI is subject to a process that is platform-centric, where data is stored and access controlled by platform owners and their complicated legal terms of use and privacy agreements that do not involve active participation from consumers who generate CHI (i.e. click thru agreements). Further, storage and computing costs for an IPFS-based CHI framework as proposed here would likely be minimal (Huang et al., 2020). Currently, most CHI is stored on central server-client architecture, limiting the speed and accessibility of CHI to both consumers and those who they may want to share it with (e.g. to their healthcare provider, researchers, health behavior companies, etc.). Enhanced scalability and relative lower costs of computing and storage of IPFS systems could reduce some of these barriers by decoupling CHI from origin servers, while at the same time ensuring content-addressing is validated to digital identity/host (Treiblmaier & Beck, 2019). Benefits in the context of security could also inure to such systems, insulating CHI from discontinued services and shut down websites where CHI may be lost to the consumer or subject to DDoS attacks.

In exchange for this autonomy and data resilience, consumers may be more willing to share CHI with outside parties that can put the data into action for better individual and population health level outcomes. Consumers would be able to change their preferences based on dynamic consent via smart contracts and could also be incentivized to share data through tokenomics models (which could also include incentives for health behavior interventions or gamification) in a transparent manner. Built into smart contracts in this framework is also required legal compliance to FTC and GDRP consumer-focused protections. This includes smart contract requests originating from the user for record deletion/removal (e.g. GDRP 'right to be forgotten'), rectification, or other revocation of access, all cryptographically logged and hashed on the blockchain.

Our CHI reference model focuses on the real-world use case of enabling consumers to gain greater control of all the disparate CHI data they generate on different platforms and applications while using blockchain and OCBS technology features to enable consumer-centric data stewardship. Due to the distributed nature and massive amounts of CHI data that are subject to less stringent privacy requirements than PHI, the main OCBS features of this reference model include:

- **OCBS Feature Advantages:** Distributed storage through IPFS that allows faster transfer while also enabling users to store their data only on their personal storage or devices while maintaining integration to the reference model. Enables opt-in governance to share information dictated by the consumer and utilizes distributed hash tables for fast lookups in the massive data storage network of CHI data.
- **Potential Challenges:** Availability and ensuring appropriate data access is the biggest hurdle in the CHI reference model due to the highly distributed and potentially sparse nature of data that largely still resides in platform-specific storage. Data schema standardization based on different types of CHI data will also be needed.
- **Security Considerations:** Security is implemented to ensure data is not shared without proper authentication and authorization from the consumer. Correctness verification is an important security aspect of the CHI reference model because the massive amounts and sparse distributed nature of storage. End users of CHI data must be ensured that data on distributed storage system has been collected correctly and not been modified due to a mistake or malicious source.

6.3. Reference model 3: genomics data

The third and last reference model focuses on a unique and growing form of health-related data; genomics data. Generally defined as data about a person's genes generated by sequencing techniques, this data is also protected by certain regulations (e.g. The Genetic Information and Nondiscrimination Act and HIPAA) and is a highly debated topic in relation to privacy. Genomic data is also diverse in both its data characteristics and uses, including array data, exome sequence data, and Whole Genome Sequenced (WGS) data, and can be collected in clinical, research, commercial, or consumer settings (e.g. direct-to-consumer genetic tests) (Navarro et al., 2019). As the volume of cost-accessible genomic data exponentially grows (due to the reduction of cost in WGS), concerns about what and how to store it, what phylogenetic/phenotypical data should be associated with it, deletion/expunging protocols, and whether physical DNA samples should be stored or destroyed, remain unresolved issues (Gutmann & Wagner, 2013).

Though privacy concerns remain a key topic of debate, the immense value of genomics data to advance precision medicine and pharmacogenomics has nevertheless been demonstrated by large scale collaboration projects, such as International Cancer Genome Consortium, which has amassed over 800 terabytes of data on cancerous genomes (Phillips et al., 2020). Several genomics companies are also attempting to use blockchain technology for purposes of enabling better management, sharing, and control of data in a privacy-centric fashion (Eman Ahmed, 2019; Thiebes, Schlesner, Brors, & Sunyaev, 2020). Importantly, genomics data requires an intermediary reference model that can handle large volumes of data (such as raw sequence data), that can facilitate off-chain computationally intensive analysis, and which enables sharing while emphasizing individual-centered privacy and ownership

(Ozercan, Ileri, Ayday, & Alkan, 2018). Hence, our third reference model (see **Fig. 3**) focuses on anonymized and secure multi-factor consent processes that enables trusted participants on a consortium blockchain to query and access genomics data upon explicit consent by the individual.

The genomics data reference model utilizes a consortium blockchain network made up again of pre-vetted participants (e.g. biotechnology companies, researchers, etc.) that uses smart contracts dictated by data owner access privilege preferences similar to our other reference models. Genomics data storage will be stored off-chain due to on-chain storage restraints and privacy issues. The exact location of the genomic data storage will be specified by the user upon data generation. For example, a genomics data owner may choose to store their genomics data in an off-chain private research or trusted third-party database, or they may choose to store their own genomics data themselves using DAS. A researcher can query the system with a set of characteristics of features of interest, and upon being identified by the system, a notification will go out to the data owner (if they opted-in) further requesting if they would like to share their genomics data without explicitly revealing their PII. This design is similar to our CHI reference model, with an added second layer of authentication, and the first layer always being encrypted and anonymized.

The primary goal of this reference model is to manage genomics data differently from PHI and CHI, by adding a second layer of authentication and using off-chain storage approaches that can accommodate larger volumes of data and higher off-chain computational demands. Aggregate analysis of genomics data can lead to powerful insights. However, the size of these datasets are often much larger as a whole genome sequencing data is approximately 2.9GB for an individual. Further, as regulations, next generation sequencing technology, and best practices change around genomics data management, a flexible framework is required. Therefore, modular off-chain storage, wherein the data owner can choose the off-chain storage medium to house their genomic data will be employed.

Here again, distributed storage such as IPFS can improve the scalability and availability of sharing, while a data owner could also store their data on their own servers/machines and share data upon explicit consent. All of the on-chain storage will be de-identified, while a verifiable off-chain computation approach will be used to prove a genomic dataset has not changed since generated. SMPC may also be used to perform computational analysis on the genomics data. This is beneficial for multiple reasons, as it enhances the security of any translations made to the genomic data as well as speeds up performance of retrieving and computing large genomic datasets that

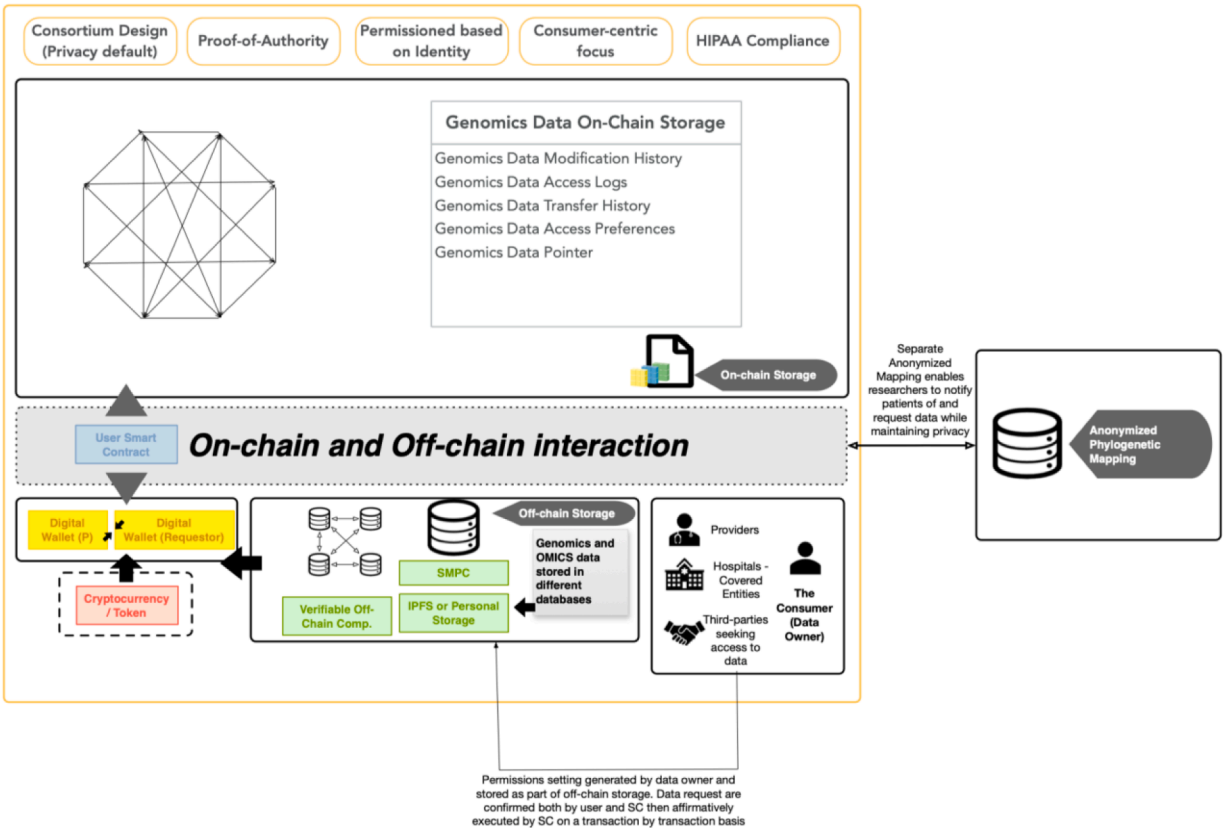

Fig. 3. Genomic data reference model

Description: The Genomic Data Reference Model has the most extensive modularity and dynamic settings based on user preferences. Off-Chain storage can either be in distributed storage, personal storage, or through a third-party service. A second layer of authentication is added via a phylogenetic off-chain storage mapping, that enables researchers to request access to data from owners without disclosing any private information. A token-based system, along with a consumer-centric security design promotes sharing of genomic data.

are stored.

A centralized off-chain database will store mappings of phylogenetic data associated with a de-identified code associated with a user. When specific types of phylogenetic features are queried, a TEE will perform a query on the centralized off-chain database and identify the codes associated with these features. This will invoke an explicit notification to the data owner regarding information about the querying party and details about the proposed usage of the data. The data owner can then decide whether or not to share their genomic data. The decisions and data transfer will be authenticated by a private key associated with the data owner.

The genomics data reference model is the most dynamic and modular reference model. The genomics reference model was also designed with the ability to change over time based upon the ongoing regulatory guidance surrounding genomics data. Similar to the CHI model, the genomics data reference model was designed to optimize data sharing and transfer mediated by the individual, important due to the extremely impactful nature of genomics data to inform precision medicine and population health issues. Key OCBS features of this reference model include:

- **OCBS Feature Advantages:**Distributed storage through IPFS which allows faster transfer while also enabling users to store their data on the storage modality of their choice while still enabling integration to the reference model. The model also utilizes SMPC and verifiable computation to ensure large genomics data maintains integrity and will use opt-in governance to share information.
- **Potential Challenges:** Ensuring correctness and data loss resiliency is the biggest hurdle with the genomics reference model due to the large file size of WGS.
- **Security Considerations:** Extra security is added to verify the owner approves sharing while maintaining anonymity. This is due to genomics data value having a high reliance on phylogenetic data.

7. Evaluation framework

7.1. Performance evaluation

In order to appropriately assess the potential utility of real-world deployment of our hOCBS framework, we propose an evaluation framework adopted from the literature on health information systems. Specifically, we adopt *Yusof et al.'s* Human-Organization-Technology Fit (HOT-fit) framework for purposes of evaluating the potential performance of our reference models when compared to traditional health information systems (Yusof, Kuljis, Papazafeiropoulou, & Stergioulas, 2008). The HOT-fit framework is an evaluation framework built on the Information Systems Success Model (IS Model) and the IT Organizational Fit Model (IT Model) (Yusof, Paul, & Stergioulas, 2006). Holistically, the IT Model is an evaluation based on a dynamic equilibrium of organizational components including business strategy, organizational structure, management processes, and roles and skills.

The HOT-fit Framework can be broken down into eight dimensions: System Quality, Information Quality, Service Quality, System Use, User Satisfaction, Organizational Structure, Organizational Environment and Net Benefits. Both the human and organizational aspects of the HOT-fit framework will be evaluated in the future through iterative testing, industry use cases, and semi-structured interviews with healthcare stakeholders and patient/consumer stakeholders who will utilize the user interface components of the reference model solutions, an approach that has been used in other blockchain prototype evaluations (Putz, Dietz, Empl, & Pernul, 2021). Specific measures to evaluate performance of the system will include: (1) the response time and availability of data in the hOCBS system compared to other healthcare data storage and transfer systems; and (2) quantitative measurement of storage and computing costs of the system compared to current stand-alone information management systems.

Anticipated outcomes include that the modular architecture of hOCBS' and its use of different combinations of OCBS technology features over a distributed network to store and transfer data will reduce the friction caused by extra-organization data transfer in a way that can be quantitatively measured by the costs of system operation, reduction in healthcare administrative-related costs (currently estimated at 34.2% of national health expenditures in the U.S.), and increased speed and access to permissioned data. The reliability and flexibility of the data should also be improved due to a coherent system requiring data standardization and encouraging interoperability. The security of the data will be built into the system with cryptographic techniques such as hashing, PKI, SMPC, and verifiable computation.

Relative computing costs are expected to be reduced due to lower requirements for systems integration based on reference models utilizing existing health information systems (e.g. PHI reference model will focus on integration with existing EHRs and patient portals) and decentralized platform agnostic applications and storage approaches (e.g. CHI and Genomics reference models using IPFS). In relation to analytical evaluation of transaction performance, we will also explore the use of novel approaches used on other blockchain frameworks (such as for permissioned blockchains on Hyperledger Fabric), including models developed to calculate transaction latency based on different network configurations for performance bottlenecks (Xu et al., 2021). Other non-technical components of the HOT-fit model will also be important to incorporate into this evaluation, including user experience evaluations to continuously iterate on the user interface and functionality of the system. Additionally, verifying the solution is privacy compliant will also likely require other evaluations, such as a HIPAA Security Risk Assessment or a GDPR Data Protection Impact Assessment (Campanile, et al., 2021).

7.2. Blockchain platform choice and evaluation

A variety of factors were considered when choosing the blockchain protocol and smart contract infrastructure that we will use to implement our proposed hOCBS framework, which included purpose, mode of accessibility, programming language, consensus

mechanism, ecosystem, and maturity. Ethereum was chosen as the blockchain protocol for future proof-of-concept (POC) development of our hOCBS framework due to its ability to excel in a variety of aspects of our decision framework. First, Ethereum has the largest and most mature smart contract blockchain ecosystem along with sophisticated tooling to enable developers to build complex and robust applications (Hu et al., 2021). Furthermore, Ethereum has an extensive network of experienced developers, allowing for robust development and maintenance of production services. The core development team along with its massive developer community has continued to make improvements in functionality, which is a great sign for long-term success of this open source blockchain platform.

Ethereum is also a very accessible software, and due to its open-source nature, it is relatively easy to create private Ethereum blockchains that are not connected to the public Mainnet. Solidity, the most popular programming language used to write Ethereum Smart contracts is a developer friendly language based off principles of many traditional programming languages such as JavaScript, Python, and C++. This enhances development speeds and enables application development for faster integration with other information systems as we propose. Some may argue that Ethereum was created for public blockchain applications while Hyperledger was designed specifically for more business and enterprise solutions. However, Ethereum has multiple commercial healthcare applications (e.g. projects supported by Consensys Health), enabling collaboration and re-usable code and healthcare specific smart contract functionality. Overall, Ethereum was determined to be the blockchain protocol that would enable more rapid development of POCs and offers the most mature and developed set of tooling to implement our proposed hOCBS system.

The inherent benefits of blockchain technology are improvements to both the system quality and information quality as previously described in this paper. Aligning with the HOT-fit evaluation model, Ethereum currently offers application features that can enable improvements in service quality, user satisfaction, system use, User Satisfaction, Organizational Structure, Organizational Environment and Net Benefits over other platforms. For example, projects such as chainlink, Uniswap, and the Raiden Network enable integration not currently supported by other protocols. However, we will also assess if other existing blockchain platforms (such as Hyperledger Fabric that is used for business blockchain development) can offer greater utility in scaling up a specific hOCBS reference model for business-to-business transactions, improve integration with other off-chain systems, or provide more modular architecture needed for management of specific types of healthcare data. The distributed technology space is consistently changing at a high-velocity and continuous evaluation will be conducted to ensure the best protocol is used for performance optimization.

8. Conclusion

This study sought to better understand and characterize OCBS' for the purposes of developing a privacy-preserving hybrid on-chain and off-chain blockchain framework to better manage different forms of healthcare data. We conclude that the relative complexity, regulatory requirements, and multiparty nature of modern health information systems would benefit from distributed governance facilitated by blockchain OCBS systems. The study proposes the hOCBS framework, which includes features that enable integration with off-chain storage in a secure, scalable, and privacy and patient-centric manner. Underpinning the conceptualization of the hOCBS is the fact that healthcare could significantly benefit from a privacy-by-design infrastructure that liberates current data silos currently not governed by the patient/consumer (H. A, K, D, SA, & A, 2020; Hylock, 2019). Some of the primary benefits for this architecture are enabling greater sharing and processing of healthcare data, reducing storage requirements to facilitate scale-up of health blockchain information systems, and developing dynamic privacy-persevering mechanisms, such as preserving anonymity and enabling dynamic consent management.

Conceptualization of this framework also considered key off-chain and on-chain storage tradeoffs, the regulatory and legal considerations of different types of healthcare data, and relevant storage and computational features in its adaptive design that are specific to PHI, CHI, and genomic data. Adaptability of the system was one of the core principles of the FIP framework used to conceptualize hOCBS, with a focus on ensuring modularity of blockchain design for healthcare. Specifically, a blockchain privacy-preserving framework needs to leverage the utility of different approaches for on-chain and off-chain storage while concomitantly adapting to different needs of data sharing, storage, access, and computing execution in order to ensure such data remains useful and beneficial for healthcare processes and parties involved.

Our framework also places special emphasis on a single stakeholder in this network; the patient or consumer who is the rightful subject of the data asset and associated digital identity. This is meant to align technology with patient-centered care approaches, defined as care provision that is consistent with the values, needs, and desires of patients and is only achieved when clinicians involve patients in healthcare discussions and decisions (Constand, MacDermid, Bello-Haas, & Law, 2014). Our framework makes a broader push for patient-centered values in a distributed OCBS governance system for healthcare data, features we argue are currently lacking in traditional healthcare information management systems (Hylock, 2019). We adopt this approach as any privacy-preserving system also needs to consider concepts of self-sovereign identity, which recognizes that users should control and manage elements of their own digital identity.

However, our framework also recognizes that in addition to control of identity, mechanisms need to be in place that ensure the right to privacy and anonymity, particularly given the sensitive and personal nature of healthcare data (Bernabe, Canovas, Hernandez-Ramos, Moreno, & Skarmeta, 2019). Though sharing of data is critical to improving the quality and performance of healthcare (both at the individual and population level), this invaluable digital asset should be managed and controlled by the patient/consumer, with technology facilitating this stewardship (Yue, Wang, Jin, Li, & Jiang, 2016). In this sense, blockchain technology provides a secure mechanism to track timestamped specified rights attributed to individual healthcare data all governed by smart contracts, where personalized access rules can be set by the individual themselves underpinned by regulatory frameworks meant to protect them.

It is our belief that a revolution to transform the utility of healthcare data will only occur when privacy protections are explicitly

linked to an individual's digital identity in a privacy-preserving fashion, and that blockchain systems using hybrid off-chain design approaches represent the optimal way to achieve this goal by establishing enhanced trust, transparency, and sovereignty in a distributed network. We also believe that an individual's healthcare data should be consistent and available across institutional boundaries, and the terms of its access strictly dictated by said individual. Without such principles purposefully imbedded in technology, data governance, and privacy policy, healthcare data will remain siloed, not able to reach its full potential to improve the health of all.

Declarations

Ethics Approval and Consent to Participate: Not applicable/Not required for this study. All information collected from this study was from the public domain.

Consent to Publish: Not applicable.

Competing Interests: Ken Miyachi is the principal owner of the blockchain startup company LedgerSafe. LedgerSafe has no financial relationship and no participation in this study. Tim K. Mackey is the CEO of the company S-3 Research LLC, a healthcare startup funded by the National Institutes of Health. S-3 Research LLC, has no financial relationships and no participation in this study.

Funding Statement: Authors received funding from the San Diego Supercomputer Center BlockLAB and greatly acknowledge this support.

Author Statement. KM and TKM jointly conceptualized, conducting data curation, formal analysis, investigation, project administration, resources, software, visualization, and jointly wrote the original and final draft including all editing. TKM provided funding acquisition and supervision of the study.

Data availability statement

All data associated with this manuscript is available in the references provided.

Author statement

KM and TKM jointly conceptualized, conducting data curation, formal analysis, investigation, project administration, resources, software, visualization, and jointly wrote the original and final draft including all editing. TKM provided funding acquisition and supervision of the study.

Acknowledgments

Authors thank the San Diego Supercomputer Center and LunaDNA PBC for their support and feedback associated with research conducted in this manuscript.

References

Hasselgren, A., Kralevska, K., Gligoroski, D., Pedersen, SA., & Faxvaag, A. (2020). Blockchain in healthcare and health sciences—A scoping review. *International Journal of Medical Informatics, 134*, Article 104040. https://doi.org/10.1016/j.ijmedinf.2019.104040.

Abraham, R., Schneider, J., & Brocke, vom, J. (2019). Data governance: A conceptual framework, structured review, and research agenda. *International Journal of Information Management, 49*, 424–438. https://doi.org/10.1016/j.ijinfomgt.2019.07.008.

Agha, L. (2014). The effects of health information technology on the costs and quality of medical care. *Journal of Health Economics, 34*, 19–30. https://doi.org/10.1016/j.jhealeco.2013.12.005.

Andoni, M., Robu, V., Flynn, D., Abram, S., Geach, D., Jenkins, D., et al. (2019). Blockchain technology in the energy sector: A systematic review of challenges and opportunities. *Renewable and Sustainable Energy Reviews, 100*, 143–174.

Atzori, M. (2015). Blockchain Technology and Decentralized Governance: Is the State Still Necessary? *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2709713.

Azaria, A., Ekblaw, A., Vieira, T., & Lippman, A. (2016). *Medrec: Using blockchain for medical data access and permission management* (n.d. IEEE: 2016 2nd International Conference on Open and Big Data

Yuan, Y. B., & Li, L. J. (2019). The Policy Effect of the General Data Protection Regulation (GDPR) on the Digital Public Health Sector in the European Union: An Empirical Investigation. *International Journal of Environmental Research and Public Health, 16*(6), 1070. https://doi.org/10.3390/ijerph16061070.

Baniata, H, Anaqreh, A, & Kertesz. (2021). PF-BTS: A Privacy-Aware Fog-enhanced Blockchain-assisted task scheduling. *Information Processing & Management, 58*(1).

Berdik, D., Otoum, S., Schmidt, N., Porter, D., & Jararweh, Y. (2021). A Survey on Blockchain for Information Systems Management and Security. *Information Processing & Management, 58*(1), Article 102397. https://doi.org/10.1016/j.ipm.2020.102397.

Bernabe, J. B., Canovas, J. L., Hernandez-Ramos, J. L., Moreno, R. T., & Skarmeta, A. (2019). Privacy-preserving solutions for Blockchain: review and challenges. *IEEE Access, 7*, 164908–164940. https://doi.org/10.1109/ACCESS.2019.2950872. publicationTitle":"IEEE.

Bernal Bernabe, J., Luis Canovas, J., Hernandez-Ramos, J. L., Torres Moreno, R., & Skarmeta, A. (2019). Privacy-Preserving Solutions for Blockchain: Review and Challenges. *IEEE Access, 7*, 164908–164940. https://doi.org/10.1109/ACCESS.2019.2950872.

Berwick, D. M., Nolan, T. W., & Whittington, J. (2008). The Triple Aim: Care, Health, And Cost. *Health Affairs, 27*(3), 759–769. https://doi.org/10.1377/hlthaff.27.3.759.

National Academies. (2009). *Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research.* USA: National Academies Press.

Campanile, L., Iacono, M., Marulli, F., & Mastroianni, M. (2021). Designing a GDPR compliant blockchain-based IoV distributed information tracking system. *Information Processing & Management, 58*(3).

Cao, Y., Sun, Y., & Min, J. (2020). Hybrid blockchain–based privacy-preserving electronic medical records sharing scheme across medical information control system. *Measurement and Control, 53*(7-8), 1286–1299. https://doi.org/10.1177/0020294020926636.

Chatterjee, K., Goharshady, A. K., & Velner, Y. (2018). Quantitative Analysis of Smart Contracts. In *Programming Languages and Systems, 10801* pp. 739–767). ESOP 2018: Programming Languages and Systems. https://doi.org/10.1007/978-3-319-89884-1_26.

Chen, Q., Srivastava, G., Parizi, R. M., Aloqaily, M., & Al Ridhawi, I. (2020). An incentive-aware blockchain-based solution for internet of fake media things. *Information Processing & Management*, Article 102370, 2020.
Chu, L. F., Shah, A. G., Rouholiman, D., Riggare, S., & Gamble, J. G. (2018). Patient-Centric Strategies in Digital Health. In *Digital Health, 13* pp. 43–54). Cham: Springer, Cham. https://doi.org/10.1007/978-3-319-61446-5_4.
Chu, L. F., Utengen, A., Kadry, B., Kucharski, S. E., Campos, H., Crockett, J., et al. (2016). Nothing about us without us"—patient partnership in medical conferences. *British Medical Journal, 354*, i3883. https://doi.org/10.1136/bmj.i3883.
Constand, M. K., MacDermid, J. C., Bello-Haas, V. D., & Law, M. (2014). Scoping review of patient-centered care approaches in healthcare. *BMC Health Services Research, 14*(1), 1–9. https://doi.org/10.1186/1472-6963-14-271.
Dagher, G. G., Mohler, J., Milojkovic, M., & Marella, P. B. (2018). Ancile: Privacy-preserving framework for access control and interoperability of electronic health records using blockchain technology. *Sustainable Cities and Society, 39*, 283–297. https://doi.org/10.1016/j.scs.2018.02.014.
Demiris, G. (2016). Consumer Health Informatics: Past, Present, and Future of a Rapidly Evolving Domain. *Yearbook of Medical Informatics*, (1), S42–S47. https://doi.org/10.15265/IYS-2016-s005. *Suppl*S 01.
Dimitrov, D. V. (2019). Blockchain Applications for Healthcare Data Management. *Healthcare Informatics Research, 25*(1), 51.
Dubovitskaya, A., Novotny, P., Xu, Z., & Wang, F. (2019). Applications of Blockchain Technology for Data-Sharing in Oncology: Results from a Systematic Literature Review. *Oncology, 98*(6), 1–9. https://doi.org/10.1159/000504325.
Dwivedi, A. D., Srivastava, G., Dhar, S., & Singh, R. (2019). A Decentralized Privacy-Preserving Healthcare Blockchain for IoT. *Sensors, 19*(2), 326. https://doi.org/10.3390/s19020326.
D'Amore, J. D., Sittig, D. F., & Ness, R. B. (2012). How the Continuity of Care Document Can Advance Medical Research and Public Health. *American Journal of Public Health, 102*(5), e1–e4. https://doi.org/10.2105/AJPH.2011.300640.
Eberhardt, J., & Heiss, J. (2018). Off-chaining Models and Approaches to Off-chain Computations. *Presented at the the 2nd Workshop* (pp. 7–12). New York, New York, USA: ACM Press.
Eberhardt, J., & Tai, S. (2017). On or Off the Blockchain? Insights on Off-Chaining Computation and Data. In *Service-Oriented and Cloud Computing, 10465* pp. 3–15). Cham: Springer, Cham. https://doi.org/10.1007/978-3-319-67262-5_1.
Eberhardt, J., & Tai, S. (2018). ZoKrates - Scalable Privacy-Preserving Off-Chain Computations. n.d. *2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*. https://doi.org/10.1109/Cybermatics_2018.2018.00199
Ahmed, Eman, & S, M. (2019). DNA Data Marketplace: An Analysis of the Ethical Concerns Regarding the Participation of the Individuals. *Frontiers in Genetics, 10*, 473. https://doi.org/10.3389/fgene.2019.01107.
European Parliament. (2019). *Blockchain and the General Data Protection Regulation: Can distributed ledgers be squared with European data protection law?*. Retrieved June 11, 2020, from.
Ferdous, M. S., Chowdhury, M. J. M., Hoque, M. A., &Colman, A. (2020, January 20). Blockchain Consensus Algorithms: A Survey. arXiv.org.
Fernandes, D., Soares, L., Gomes, JV., Freire, M. M., & Inacio, P. R. M (2014). *Security issues in cloud environments: a survey*. n.d.. *13* pp. 113–170) International Journal of Information Security volume
Gordon, W. J., & Catalini, C. (2018). Blockchain Technology for Healthcare: Facilitating the Transition to Patient-Driven Interoperability. *Computational and Structural Biotechnology Journal, 16*, 224–230. https://doi.org/10.1016/j.csbj.2018.06.003.
Gutmann, A., & Wagner, J. W. (2013). Found Your DNA on the Web: Reconciling Privacy and Progress. *Hastings Center Report, 43*(3), 15–18. https://doi.org/10.1002/hast.162.
Hardin, T., & Kotz, D. (2021). Amanuensis: Information provenance for health-data systems. *Information Processing & Management, 58*(2).
Hawig, D., Zhou, C., Fuhrhop, S., Fialho, A. S., & Ramachandran, N. (2019). Designing a Distributed Ledger Technology System for Interoperable and General Data Protection Regulation-Compliant Health Data Exchange: A Use Case in Blood Glucose Data. *Journal of Medical Internet Research, 21*(6). https://doi.org/10.2196/13665.
Hepp, T., Sharinghousen, M., Ehret, P., Schoenhals, A., &Gipp, B. (2018). On-chain vs. off-chain storage for supply- and blockchain integration. *It - Information Technology*, 60(5-6), 283–291. doi:10.1515/itit-2018-0019.
Herrera-Joancomartí, J., & Pérez-Solà, C. (2016). Privacy in Bitcoin Transactions: New Challenges from Blockchain Scalability Solutions. In *Modeling Decisions for Artificial Intelligence, 9880* pp. 26–44). Cham: Springer, Cham.. https://doi.org/10.1007/978-3-319-45656-0_3.
Hochman, M., Garber, J., & Robinson, E. (2019, August 14). Health Information Exchange After 10 Years: Time For A More Assertive. *Health Affairs Blog*. https://doi.org/10.1377/hblog20190807.475758/full.
Hoffman, M. A., & Williams, M. S. (2011). Electronic medical records and personalized medicine. *Human Genetics, 130*(1), 33–39. https://doi.org/10.1007/s00439-011-0992-y.
Hripcsak, G., Bloomrosen, M., Brennan, P. F., Chute, C. G., et al. (2014). Health data use, stewardship, and governance: ongoing gaps and challenges: a report from AMIA's 2012 Health Policy Meeting. n.d.. In *, 21. J Am Med Inform Assoc* (pp. 204–211) JAMIA
Hu, T., Liu, X., Chen, T., Zhang, X., Huang, X., Niu, W., Lu, J., Zhou, K., & Liu, Y. (2021). Transaction-based classification and detection approach for Ethereum smart contract. *Information Processing & Management, 58*(2).
Huang, H., Lin, J., Zheng, B., Zheng, Z., & Access, J. B. I. (2020). When Blockchain Meets Distributed File Systems: An Overview, Challenges, and Open Issues. n.d. *IEEE Access, 8*, 50574–50586
Hussien, H. M., Yasin, S. M., Udzir, S. N. I., Zaidan, A. A., & Zaidan, B. B (2019). A Systematic Review for Enabling of Develop a Blockchain Technology in Healthcare Application: Taxonomy, Substantially Analysis, Motivations, Challenges, Recommendations and Future Direction. *Journal of Medical Systems, 43*(10), 320–335. https://doi.org/10.1007/s10916-019-1445-8.
Hylock, R. H. (2019). A Blockchain Framework for Patient-Centered Health Records and Exchange (HealthChain): Evaluation and Proof-of-Concept Study. *Journal of Medical Internet Research, 21*(8), e13592. https://doi.org/10.2196/13592.
Institute of Medicine (US) Committee on Regional Health Data Networks, Donaldson, M. S., Lohr, K. N., & Institute of Medicine (US) Committee on Regional Health Data Networks. (1994). Health Databases and Health Database Organizations: Uses, Benefits, and Concerns. *Health Data in the Information Age: Use, Disclosure, and Privacy*. US: National Academies Press.
Institute of Medicine (US), National Academy of Engineering (US) Roundtable on Value & Science-Driven Health Care. (2011). Healthcare System Complexities, Impediments, and Failures. *Engineering a Learning Healthcare System: A Look at the Future: Workshop Summary*. US: National Academies Press.
Ismail, L., Materwala, H., Karduck, A. P., & Adem, A. (2020). Requirements of Health Data Management Systems for Biomedical Care and Research: Scoping Review. *Journal of Medical Internet Research, 22*(7), e17508. https://doi.org/10.2196/17508.
Jing, N., Liu, Q., & Sugumaran, V. (2021). A blockchain-based code copyright management system. *Information Processing & Management, 58*(3), Article 102518. Issue.
Vest, Joshua R, & G, L. D. (2010). Health information exchange: persistent challenges and new strategies. *Journal of the American Medical Informatics Association : JAMIA, 17*(3), 288–294. https://doi.org/10.1136/jamia.2010.003673.
Khalid, A., Iftikhar, MS., Almogren, A., Khalid, R., Afzal, MK., & Javaid, N. (2021). A blockchain based incentive provisioning scheme for traffic event validation and information storage in VANETs. *Information Processing & Management, 58*(2).
Kruse, C. S., Goswamy, R., Raval, Y., & Marawi, S. (2016). Challenges and Opportunities of Big Data in Health Care: A Systematic Review. *JMIR Medical Informatics, 4*(4), e38. https://doi.org/10.2196/medinform.5359.
Kuo, T.-T., Kim, H.-E., & Ohno-Machado, L. (2017). Blockchain distributed ledger technologies for biomedical and health care applications. *Journal of the American Medical Informatics Association : JAMIA, 24*(6), 1211–1220. https://doi.org/10.1093/jamia/ocx068.
Lewison, K., & Corella, F. (2016). Backing rich credentials with a blockchain PKI. *pomcor.com*.
Li, J., Wu, J., Jiang, G., & Srikanthan, T. (2020). Blockchain-based public auditing for big data in cloud storage. *Information Processing & Management, 57*(6), Article 102382. https://doi.org/10.1016/j.ipm.2020.102382.

Mirchandani, A. (2019). The GDPR-Blockchain Paradox: Exempting Permissioned Blockchains from the GDPR. *Fordham Intel. Prop. Media & Ent. L.J., 29*(1201).
Lokhande, S., Mukadam, S., Chikane, M., & Bhonsle, M. (2020). Enhanced Data Sharing with Blockchain in Healthcare. *ICCCE 2019, 570*, 277–283. https://doi.org/10.1007/978-981-13-8715-9_33.
Karampela, M., Ouhbi, S., & Isomursu, M. (2019). Exploring Users' Willingness to Share Their Health and Personal Data Under the Prism of the New GDPR: Implications in Healthcare. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 6509–6512). IEEE. https://doi.org/10.1109/EMBC.2019.8856550, 2019.
Mackey, T. K., Kuo, T.-T., Gummadi, B., Clauson, K. A., Church, G., Grishin, D., et al. (2019). Fit-for-purpose?" - challenges and opportunities for applications of blockchain technology in the future of healthcare. *BMC Medicine, 17*(1). https://doi.org/10.1186/s12916-019-1296-7, 68–17.
Mayer, A. H., da Costa, C. A., & Righi, R. D. R. (2019). Electronic health records in a Blockchain: A systematic review. *Health Informatics Journal, 1*(1), Article 1460458219866350. https://doi.org/10.1177/1460458219866350.
Mikula, T., & Jacobsen, R. H. (2018). Identity and Access Management with Blockchain in Electronic Healthcare Records. n.d. *Identity and access management with blockchain in electronic healthcare records*
Nakamoto, S. (2019). Bitcoin: A Peer-to-Peer Electronic Cash System. *Manubot, 3*(2), 99–111.
Navarro, F. C. P., Mohsen, H., Yan, C., Li, S., Gu, M., Meyerson, W., & Gerstein, M (2019). Genomics and data science: an application within an umbrella. *Genome Biology, 20*(1), 1–11. https://doi.org/10.1186/s13059-019-1724-1.
Norgeot, B., Glicksberg, B. S., & Butte, A. J. (2019). A call for deep-learning healthcare. *Nature Medicine, 25*(1), 14–15. https://doi.org/10.1038/s41591-018-0320-3.
O'Donoghue, O. (2019). Design Choices and Trade-Offs in Health Care Blockchain Implementations: Systematic Review. *Journal of Medical Internet Research, 21*(5), e12426. https://doi.org/10.2196/12426.
Oham, C., Michelin, R. A., Jurdak, R., Kanhere, S. S., & Jha, S. (2021). B-FERL: Blockchain based framework for securing smart vehicles. *Information Processing & Management, 58*(1), Article 102426.
Ozercan, H. I., Ileri, A. M., Ayday, E., & Alkan, C. (2018). Realizing the potential of blockchain technologies in genomics. *Genome Research, 28*(9), 1255–1263. https://doi.org/10.1101/gr.207464.116.
Paik, H. Y., Xu, X., Bandara, H., Lee, S. U., & Lo, S. K. (2019). *Analysis of Data Management in Blockchain-Based Systems: From Architecture to Governance* (p. 7). IEEE Access. https://doi.org/10.1109/ACCESS.2019.2961404.
Pasquale, F., & Ragone, T. A. (2014). *Protecting health privacy in an era of big data processing and cloud computing. 595*. 17 Stan. Tech. L. Rev.
Phillips, M., Molnár-Gábor, F., Korbel, J. O., Thorogood, A., Joly, Y., Chalmers, D., et al. (2020). Genomics: data sharing needs an international code of conduct. *Nature, 578*(7793), 31–33. https://doi.org/10.1038/d41586-020-00082-9.
Pirtle, C., & Ehrenfeld, J. (2018). Blockchain for Healthcare: The Next Generation of Medical Records? *Journal of Medical Systems, 42*(9), 172. https://doi.org/10.1007/s10916-018-1025-3.
Poon, J., & Dryja, T. (2016). *The bitcoin lightning network: Scalable off-chain instant payments*. lightning.network.
Puthal, D., Mohanty, S. P., Yanambaka, V. P., & Kougianos, E. (2020). *PoAh: A Novel Consensus Algorithm for Fast Scalable Private Blockchain for Large-scale IoT Frameworks*. arXiv.org.
Putz, B., Dietz, M., Empl, P., & Pernul, G. (2021). EtherTwin: Blockchain-based Secure Digital Twin Information Management. *Information Processing & Management, 58*(1).
Sadhya, V., & Sadhya, H. (2018). *Barriers to Adoption of Blockchain Technology*. AMCIS2018.
Sanjabi, S. B., & Pommereau, F. (2010). *Modelling, verification, and formal analysis of security properties in a P2P system*. n.d. (pp. 499–508) 2010 International Symposium on Collaborative Technologies and Systems
Schadt, E. E., Linderman, M. D., Sorenson, J., Lee, L., & Nolan, G. P. (2010). Computational solutions to large-scale data management and analysis. *Nature Reviews Genetics, 11*(9), 647–657. https://doi.org/10.1038/nrg2857.
Attaran, M (2020). *Blockchain technology in healthcare: Challenges and opportunities*. International Journal of Healthcare Management.
Sherif, El, R., Pluye, P., Thoër, C., & Rodriguez, C (2018). Reducing Negative Outcomes of Online Consumer Health Information: Qualitative Interpretive Study with Clinicians, Librarians, and Consumers. *Journal of Medical Internet Research, 20*(5), e169. https://doi.org/10.2196/jmir.9326.
Shi, S., He, D., Li, L., Kumar, N., Khan, M. K., & Choo, K.-K. R. (2020). Applications of blockchain in ensuring the security and privacy of electronic health record systems: A survey. *Computers & Security, 97*, Article 101966. https://doi.org/10.1016/j.cose.2020.101966.
Smith, S. W., & Koppel, R. (2014). Healthcare information technology's relativity problems: a typology of how patients' physical reality, clinicians' mental models, and healthcare information technology differ. *Journal of the American Medical Informatics Association, 21*(1), 117–131. https://doi.org/10.1136/amiajnl-2012-001419.
Smith, S., & Duman, M. (2009). The state of consumer health information: an overview. *Health Information and Libraries Journal, 26*(4), 260–278. https://doi.org/10.1111/j.1471-1842.2009.00870.x.
Thiebes, S., Schlesner, M., Brors, B., & Sunyaev, A. (2020). Distributed Ledger Technology in genomics: a call for Europe. *European Journal of Human Genetics, 28*(2), 139–140. https://doi.org/10.1038/s41431-019-0512-4.
Treiblmaier, H., & Beck, R. (2019). *Business transformation through blockchain*. Palgrave Macmillan.
van Steen, M., & Tanenbaum, A. S. (2016). A brief introduction to distributed systems. *Computing, 98*(10), 967–1009. https://doi.org/10.1007/s00607-016-0508-7.
Vazirani, A. A. (2019). Implementing Blockchains for Efficient Health Care: Systematic Review. *Journal of Medical Internet Research, 21*(2), e12439. https://doi.org/10.2196/12439.
Warren, W., & Bandeali, A. (2017). *0x: An open protocol for decentralized exchange on the Ethereum blockchain*. n.d. 0x.org
WHO. (2020). *Global Health Observatory (GHO) Data: Health Financing*. World Health Organization. http://www.who.int/gho/health_financing/en/.
Xiao, Y., Zhang, N., Lou, W., & Hou, Y. T. (2019). *Privacy Guard: Enforcing Private Data Usage Control with Blockchain and Attested Off-chain Contract Execution*. April 15 (pp. 610–629). European Symposium on Research in Computer Security 2020.
Xu, X., Sun, G., Luo, L., Cao, H., Yu, H., & Vasilakos, A. V. (2021). Latency performance modeling and analysis for hyperledger fabric blockchain network. *Information Processing & Management, 58*(1), Article 102436. https://doi.org/10.1016/j.ipm.2020.102436.
Xu, X., Weber, I., & Staples, M. (2019). Architecture for Blockchain Applications. *Architecture for Blockchain Applications* (pp. 113–148). Springer International Publishing. https://doi.org/10.1007/978-3-030-03035-3_7.
Yaga, D., Mell, P., Roby, N., & Scarfone, K. (2019, June 26). *Blockchain Technology Overview*. Gaithersburg, MD: National Institute of Standards and Technology. https://doi.org/10.6028/NIST.IR.8202. arXiv.org.
Yeager, V. A., Vest, J. R., Walker, D. M., Diana, M. L., & Menachemi, N. (2017). Challenges to Conducting Health Information Exchange Research and Evaluation: Reflections and Recommendations for Examining the Value of HIE. *eGEMs (Generating Evidence & Methods to Improve Patient Outcomes), 5*(1), 15. https://doi.org/10.5334/egems.217.
Yu, B., Li, X., & He, Z. (2020). *Virtual Block Group: A Scalable Blockchain Model with Partial Node Storage and Distributed Hash Table*. n.d.. *63* pp. 1524–1536) The Computer Journal
Yue, X., Wang, H., Jin, D., Li, M., & Jiang, W. (2016). Healthcare Data Gateways: Found Healthcare Intelligence on Blockchain with Novel Privacy Risk Control. *Journal of Medical Systems, 40*(10), 218. https://doi.org/10.1007/s10916-016-0574-6.
Yusof, MM, Kuljis, J, Papazafeiropoulou, A, & Stergioulas, LK. (2008). An evaluation framework for Health Information Systems: human, organization and technology-fit factors (HOT-fit). *International Journal of Medical Informatics, 77*(6), 386–398. Jun10.1016/j.ijmedinf.2007.08.011. Epub 2007 Oct 26. PMID: 17964851.

Yusof, M. M., Paul, R. J., & Stergioulas, L. K. (2006). *Towards a framework for health information systems evaluation*. n.d. Proceedings of the 39th Hawaii International Conference on System Sciences

Zhao, Q., Chen, S., Liu, Z., Baker, T., & Zhang, Y. (2020). Blockchain-based privacy-preserving remote data integrity checking scheme for IoT information systems. *Information Processing & Management, 57*(6), Article 102355. https://doi.org/10.1016/j.ipm.2020.102355.

Zhuang, Y., Sheets, L., Shae, Z., Tsai, J. J. P., & Shyu, C. R. (2018). Applying Blockchain Technology for Health Information Exchange and Persistent Monitoring for Clinical Trials. In *AMIA Annu Symp Proc* (pp. 1167–1175). AMIA.