

Measuring the Compliance Costs of Exchanging Part 2 Healthcare Claims Data Through Blockchain

SATYASAI MANDLEM*, JAMES CLAVIN*, and KARUNA JOSHI*, University of Maryland, Baltimore County, USA

Patient selections for keeping data confidential may differ between healthcare organizations, creating conflicts in confidentiality for how sensitive and demographic data is linked and merged. Validating that patient data exchange between organizations adheres to healthcare regulations, like the Health Insurance Portability and Accountability Act (HIPAA), is challenging and time-consuming and relies upon organizational due diligence to validate data upon receipt, or in the case of breaches, requires forensic examination to determine the scale of the breach. We address the need for timely compliance evaluation of substance abuse and mental health claims data exchanged between organizations by developing a novel approach integrating blockchain technology with semantic reasoners. The foundation of our methodology is an Ethereum blockchain integrated with a knowledge graph built from the clinical terms for sensitive data value sets maintained by the National Institutes of Health Value Set Authority Center for identifying health data protected by United States Title 42 Code of Federal Regulations (CFR) Part 2. When claims data is transferred, it is first validated by reasoning over the CFR Part 2 knowledge graph. Then the exchange is audited using the Ethereum blockchain to seek out sensitive patient data violating the rules. This paper presents our novel methodology in detail along with the results of sharing sensitive Part 2 data. The time and cost to detect and log out of compliance data transfers are measured and compared to the manual process equivalent. Our methodology can be used by organizations to ensure real-time auditable compliance for a secure and trusted health data exchange.

CCS Concepts: • **Applied computing** → **Health care information systems**; • **Computer systems organization** → *Peer-to-peer architectures*; • **Security and privacy** → **Information accountability and usage control**; *Economics of security and privacy*.

Additional Key Words and Phrases: HIPAA, 42 CFR Part 2, Byzantine Medical Faults, TEFCA, Automated Compliance

ACM Reference Format:

Satyasai Mandlem, James Clavin, and Karuna Joshi. 2026. Measuring the Compliance Costs of Exchanging Part 2 Healthcare Claims Data Through Blockchain. 1, 1 (February 2026), 19 pages. <https://doi.org/XXXXXXX.XXXXXX>

1 Introduction

Healthcare payments between U.S. payers and providers are done via electronic data interchange (EDI) formats mandated by the HIPAA. X12, a standards organization chartered by the American National Standards Institute, maintains the data standards [83]. Administrative healthcare data is a subset of the X12 format library, and forms the foundation upon which health care enrollment,

* Authors contributed equally to this research.

This research was partially supported by NSF award 1747724.

Authors' Contact Information: [Satyasai Mandlem](mailto:Satyasai.Mandlem@umbc.edu), satyasm1@umbc.edu; [James Clavin](mailto:James.Clavin@umbc.edu), jclavin@umbc.edu; [Karuna Joshi](mailto:Karuna.Joshi@umbc.edu), karuna.joshi@umbc.edu, University of Maryland, Baltimore County, Baltimore, Maryland, USA.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2026 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM XXXX-XXXX/2026/2-ART

<https://doi.org/XXXXXXX.XXXXXX>

eligibility, invoicing and payment is built. Claims data in X12 format are used by the Centers for Medicare and Medicaid Services (CMS) to determine the cost of health care funded by the federal government [15]. Adjudicated claims data supports a variety of research and operations, including predictive model development as well as new policy development and monitoring [32]. This data is therefore of very high quality and reliability and lends itself well to operations and research. As such, medical organizations share claims data by following the guidelines set forth in HIPAA, sometimes following Institutional Review Board oversight, and by adhering to a data use agreement that defines the variables and cohort of data that can be shared between organizations. As claims data in this case is static, organizations typically distribute claims data through secure file exchange in a machine readable format, such as comma-separated values.

Figure 1 demonstrates this data flow for 1 to n health care organizations sharing one patient's data over time. As can be observed, the patient may decide to allow each organization to share their patient data for research or operations by opting in or may decline by opting out. Each patient-organization decision may reflect the patient's trust in the organization at a given time, and may change over time. Downstream, a data repository gathers the data, and the receiving organization managing the repository may have separate data use agreements (DUA) with each sending organization, or one DUA covering data collection from all organizations. If a patient has opted out of sharing information, but previously opted in, the receiving organization data will have record of those choices over time. As the disparate sending organization's patient data is linked together, technical and business logic conflicts emerge. In the case of sensitive health information, such as mental health data, there is an additional level of authorization required for data to be exchanged, called a release of information consent form which the patient has signed. This authorization is enforceable by the Substance Abuse and Mental Health Services Administration (SAMHSA). Similarly, for demographic information gathered by different health organizations as the patient visits them or their website, a patient is presented with a variety of forms with differing input options for the same field. For example, input options for race or gender for one organization's web form by differ from another's. In addition, if the patient distrusts one organization more than another, they may choose to not disclose the information, and request that it not be shared downstream. While this decision may change over time as the patient gains trust and is presented with more opportunities to update the information, in general the data is considered restricted from further usage because of the restrictions as they existed at a given point in time.

Data sharing of claims between organizations for research or operations oftentimes involves establishing secure file exchange after a data use agreement is signed. For example, the Centers for Medicare and Medicaid Services provides a variety of data sets to the public, as well as secure access to claims data for research [16, 17]. During the drafting of data use agreements, parties attempt to mitigate the issues described above by stipulating the level of data that may be shared after merging and linking the different organizations' patient data sets. Any statistics generated from the merged data may be at an aggregate level, and not at the individual level, thereby obfuscating the question of whether or not the merged data included the patient-restricted data. Any data use agreement that allows for exchange of individual level data that has been linked and for which patients have expressed their willingness to share such information, become much more challenging to draft and finalize, especially if the agreements entail three or more parties. The challenge of which data is usable or restricted may eventually end with the researcher, as they translate the requirements from such DUAs into programming code.

Determining if patient restricted claims data is being utilized in accord with the DUA is currently a black box and difficult to quantify, unlike HIPAA EDI data transfers which are well-established and the backbone of health insurance payment in the U.S. During EDI the amount of claims information

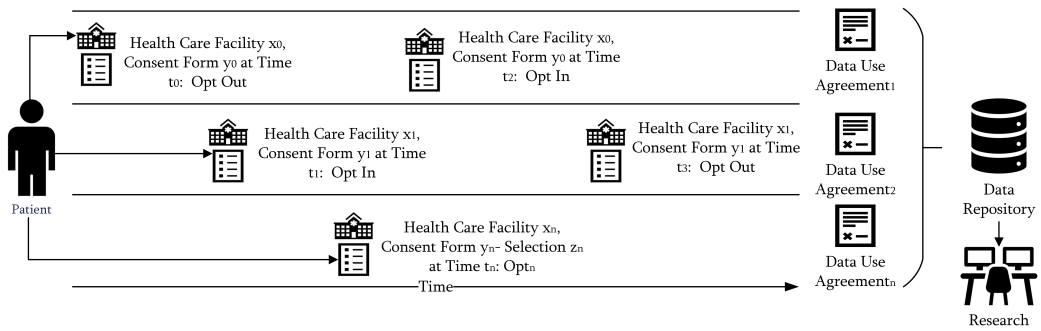


Fig. 1. Patient Privacy Consistency Model

being exchanged is always closely monitored, as the parties in such an exchange have incentive to insure the correct price is paid for each claim. After adjudication, when administrative claims are gathered and exchanged for research purposes, the volume of data shared between organizations often goes unmeasured, as it is perceived to be a strictly information technology measure. The workflow tends to focus on the correct data sets making it from point to point, the cost of storage, and the reliability of the communication channel. Providing insight into the data flows' adherence to patient-driven restrictions coming into and out of an organization over time would enable effective management of data streams.

One approach to determining how to merge data while maintaining patient confidentiality, is to evaluate the decision using distributed systems principles. Distributed systems consistency of data replication considers the total order of processes and transactions and whether or not they are either strictly consistent, or eventually consistent [76]. That is, states and processes that create and read data, distributed across several entities (servers) must function in a way such that each process or server observes the same state of the data at some point, even after the data is changed. This concept, combined with how claims data does not convey patient data restriction selections, enables us to consider a unique perspective. Interweaving claims and distributed systems principles in this way produces Byzantine Medical Fault Tolerance, or data shared between organizations that are accurate but not pertinent due to being restricted in some way from being shared, as directed by the patient [21]. Currently, there is a lack of a standard and openly available, well-established process to address this issue. Furthermore, as regulations safeguard access to sensitive data related to substance abuse records, as well as rules against information blocking patient access to data come into effect, the future landscape of access and auditing of patient data is set to change. Patients will expect that an organization will be able to locate their information and describe how it has traveled, if asked [60].

To address this challenge we focus upon claims data exchanges between healthcare organizations, and develop a means to identify SAMHSA-protected diagnostic data contained randomly within voluminous data sets, as claims data is exponentially larger relative to encounter data and patient demographic information. To distinguish substance abuse records from the broader spectrum of medical data, we have created a knowledge graph encompassing National Institutes of Health (NIH) clinical codes for these types of behavioral health data, as defined by the SAMHSA. Leveraging this knowledge graph and blockchain technology, we can effectively and efficiently identify records containing clinical codes for a variety of disorders. Subsequently, by scrutinizing the blockchain, we can ascertain what clinical terms and files contained the sensitive information, and assign a cost to the effort put forth by the blockchain to do so. This innovative methodology ensures adherence

Table 1. Part 2 Clinical Terms by Category

Clinical Terminology	Diagnosis	Finding	History	Lab	Medication	Procedure	Total
CPT ¹	0	0	0	0	0	48	48
HCPCS ²	0	0	0	0	0	59	59
ICD-10-CM ³	893	0	0	0	0	0	893
ICD-9-CM ⁴	522	0	0	0	0	0	522
LOINC ⁵	0	0	0	1311	0	0	1311
RXNORM ⁶	0	0	0	0	129	0	129
SNOMED-CT ⁷	1538	493	17	0	0	0	2048
Grand Total	2953	493	17	1311	129	107	5010

¹ Current Procedural Terminology,

² Healthcare Common Procedure Coding System,

³ International Classification of Diseases, 10th Revision, Clinical Modification,

⁴ International Classification of Diseases, 9th Revision, Clinical Modification,

⁵ Logical Observation Identifiers Names and Codes,

⁶ RxNorm,

⁷ Systematized Nomenclature of Medicine - Clinical Terms

to SAMHSA guidelines governing substance abuse and mental health data, thereby providing a comprehensive logging and compliance enabling solution for healthcare organizations exchanging claims data, verifying that the governing data use agreement was adhered to and providing a means of tracing any claim, and thereby patient data, having been exchanged.

2 Related Work

Prior work examining medical data compliance issues from an ontological and distributed systems perspective considered how best to represent and secure healthcare constructs such as electronic health records and patient data on cloud platforms [38, 44]. Research into Substance Use Disorder (SUD) data being shared out of compliance through knowledge graphs and Ethereum focused on encounter data, which captures the experience of the patient with the physician during evaluation of their health and development of a course of treatment and possible diagnosis [21, 81]. Managing health insurance through Ethereum “smart contracts” (business rules built into autonomous code) to handled claims settlement is possible, but should utilize knowledge graphs so the data can be audited later, as would be the case in reviewing the exchange of Part 2 data [19].

Table 1 categorizes all sensitive health terms (as defined by NIH) by their clinical terminology, and shows that there are over 5,000 terms to consider when attempting to identify Part 2 related claims in medical data where one or more of the terminologies was present, and that each terminology (with the exception of SNOMED-CT) focuses on one domain - be it procedures, diagnoses, lab results, or prescriptions[81].

2.1 Semantic Web

Reasoning about a specific domain using machine readable data has been the impetus behind the semantic web. Developing ontologies to capture knowledge in classes with attributes enables a machine to reason by inference or query [73]. Knowledge graphs have been used to extract knowledge about the security and privacy provisions included in HIPAA [27, 37–39, 44]. While exchanging sensitive data sets, organizations need to exchange information and execute queries with some assurance that they share a common meaning. The interoperability requirement is not just for the data itself but for having a common understanding shared about regulations across

organizations. Organizations that adhere to syntactic standards in coding clinical data rely upon some mechanism to determine semantics about their data. For example, the EDI 837 standard for claims enforces syntactic accuracy, but validity - both from a compliance and fraud perspective - depends on the application of a third party's rule set to the data to identify issues, in combination with subject matter expert analysis.

Software agents can use Semantic Web with medical datasets and policies to communicate and use each other's data and services effectively through languages such as Resource Description Framework (RDF) and Web Ontology Language (OWL) [46, 54]. Our prior work transformed encounter data files into the RDF format Terse RDF Triple Language (.ttl) to identify substance use disorder data [21, 47]. Our most fundamental requirement was that the ontological representation creates interoperability at both the syntactic and semantic levels to facilitate easy data exchange and compliance checking, while incorporating existing clinical terminologies and taxonomies, such as ICD-10 and SNOMED-CT, dynamically as coding systems mature and are refined.

2.2 Blockchain

Blockchain's application in healthcare has the potential to place control in the hands of patients, ensure data integrity, and foster secure data exchange between stakeholders [23, 31, 68]. Ethereum, a decentralized blockchain platform, has the ability to store and execute software, referred to as "smart contracts," that has business agreements and rules embedded in them. Ethereum smart healthcare contracts have been shown to be more cost effective than the equivalent business process and are able to address real-world issues, such as tracking patient COVID-19 data [6]. Because blockchains, such as Ethereum, are Byzantine Fault Tolerant (the strongest security in distributed systems), they can guarantee that data is both delivered correctly and is tamper-proof - attributes that would make them effective in combating waste, fraud, and abuse of digital protected health information (PHI) [20, 56, 86]. Blockchain has been utilized to detect counterfeit drugs by ensuring the integrity of the entire drug distribution cycle, from production to consumption [65]. Standardizing data exchange, blockchain has been shown to enhance interoperability, reduce costs, and improve security and compliance of PHI data sharing [23, 35, 85, 85].

2.3 Transaction Fee Calculation in Ethereum

Ethereum facilitates secure execution and verification of smart contracts within a peer-to-peer network [13]. These contracts are the byte code version of agreements between parties that are deployed to the Ethereum network, and become self-executing dependent upon the terms that were translated into the software. Smart contracts empower participants to transact without reliance on a central authority. Users engage in transactions through self-created Ethereum accounts, with senders signing transactions and expending Ether (ETH), the native cryptocurrency, to cover transaction processing fees. In Ethereum, a fundamental concept is the imposition of fees for every computational task resulting from a transaction. These fees are denominated in "gas," serving as a unit measuring the computational effort required. The gas price, measured in "gwei," signifies the amount of ETH allocated for each unit of gas. Notably, one gwei equals 1,000,000,000 Wei, with Wei being the smallest unit of ETH [43]. When initiating a transaction, the sender specifies a gas limit and gas price. The product of these values represents the maximum Wei the sender is willing to spend for the transaction. For instance, if the gas limit is set to 50,000 and the gas price to 20 gwei, the maximum expenditure would be 0.001 ETH [43]. This system provides participants with complete ownership and transparency over transaction data [28]. Open source versions of Ethereum allow for experimentation and development of new versions of the blockchain built for special purposes [62]. To assign a dollar value to gas expended, a Gwei-to-ETH calculator and an ETH-to-Dollar conversion tool can be used [3, 45].

2.4 Synthetic Data Generation

We have selected the synthetic patient data generator Synthea to create administrative claims data that contains diagnoses in SNOMED-CT code format [82]. Synthea's synthetic data generation takes into account regional variation in the United States, and offers a variety of configurations for tuning output, including a module designed for generating patient cohorts with sensitive health condition diagnoses on both encounters and claims. We are able to generate claims data wherein each claim has up to eight diagnoses, representing real-world combinations of clinical findings from a given encounter. Other data sets for generating synthetic patient data have been found appropriate for tuning various ML and AI based applications, proving the use of synthetic data as a valid means for system development [5, 18, 84, 87].

Table 2. Related Work, Opportunities, and Our Contribution

Existing Work	Identified Opportunities	Our Contribution
Blockchain & Semantic Web [19]	Limited regulatory compliance focus.	Integrate blockchain and semantic web to audit HIPAA Part 2 data exchanges. Measure the cost of data exchange in Ethereum and U.S. Dollars.
Electronic Health Record & Protected Health Information Ontologies [38, 44]	These works focused on cryptography of protected health information and creating synthetic health data ontologies as knowledge graphs.	Apply knowledge graphs and blockchain to dynamically determine compliance issues based upon multiple taxonomies of clinical terms; generate a detailed HIPAA audit log in Ethereum.
Value Sets for Part 2 Data [81]	The clinical terms are maintained by NIH as tabular data (spreadsheets), creating the opportunity for solutions that can find disorders based on the terms most efficiently.	Reduce interoperability friction and enable regulatory compliance checking portability via ontological query languages, that adapt as clinical terms are updated.
Synthetic patient claims data [82]	Synthetic claims data can be tuned to generate Part 2 terms, but consideration is not given to the additional layer of security and compliance needed with this kind of data.	For synthetic SUD claims measure the out of compliance data exchange between organizations as instantiated in an Ethereum network.

Table 2 summarizes related work, identifies opportunities, and describes our contribution based upon those.

3 HIPAA Part 2

42 Code of Federal Regulations (CFR) Part 2 safeguards the confidentiality of SUD treatment records, encouraging individuals to seek help by addressing discrimination and legal fears that may deter treatment-seeking [75, 79]. In response to significant changes in the healthcare system, regulatory modifications to adapt 42 CFR Part 2 with advances in integrated care models and electronic information sharing while maintaining essential privacy protections have occurred [42, 75]. Violation of HIPAA Part 2 carries severe financial and criminal penalties (see § 2.3 Criminal Penalty for Violation, under 42 U.S.C. 290dd-2(f), also sections 1176 and 1177 of the Social Security Act [42 U.S.C. 1320d-5, 1320d-6]) [1, 2, 26].

The healthcare industry maintains its unfortunate lead as the sector with the highest data breach costs, reaching an average cost per breach of USD \$10.93 million in 2023, an 8.2% increase from the previous year; since 2020 the average breach cost surged by 53.3%, emphasizing the persistent financial impact and regulatory challenges faced by the industry [14, 34]. The Office of Civil Rights tracks HIPAA violation cases, involving confirmed breaches and settlements, and has enforced monetary penalties ranging from \$100,000 in 2008 to \$3,536,500 in 2023 [40]. These penalties enable covered entities and associates to address violations without admitting liability, reflecting the dynamic evolution of HIPAA enforcement [40].

A streamlined response to data breaches, with identification and containment within 200 days, results in an average cost of USD \$3.93 million, whereas breaches extending beyond 200 days incur higher costs at USD \$4.95 million—a 23% difference. There exists therefore significant financial incentive for a shorter breach life-cycle, translating to a cost savings of USD \$1.02 million; there has been a trend towards faster resolutions over time [14, 34]. Breaches that cross the minimum threshold of 500 patients, as per the HIPAA Breach Notification Rule, 45 CFR §§ 164.400-414, require notification to the federal government [78]. Systems that can detect issues earlier can reduce the likelihood of a breach; and those that provide full audit trails allow for identifying the full impact of any data compromise.

Currently only one-third of companies discover data breaches through their own security teams and tools, highlighting a need for better data loss prevention. 67% of breaches were reported by a benign third party or by the attackers themselves. When attackers disclosed a breach, it cost organizations nearly USD \$1 million more compared to internal detection alone [14, 34].

To mitigate the risk of breaches, organizations should perform risk assessments, have HIPAA Security and Privacy officers, stand up employee training, possess legal teams for business associate agreements and DUAs, and create standard operating procedures for cyber incidents and auditing [72]. To determine if an organization's compliance posture is effective, industry standard certifications are available, such as the HITRUST Common Security Framework (myCSF). The HITRUST myCSF costs between \$70K and \$160K, depending upon the organization's risk profile, and takes between six months to a year and half. Certification must be renewed every one to two years, depending upon the certification level [22]. The overhead cost of HIPAA compliance varies - consultation fees, security tools, training, and vulnerability scanning are examples of potential expenses. Smaller businesses might pay about \$12,000 annually, but larger businesses might face recurrent expenses of hundreds of thousands of dollars [72].

4 Methods

To address HIPAA compliance issues, especially with safeguarding Part 2 data embedded in administrative claims data shared between organization, we have developed a system that captures Part 2 terms present in claims data and logs their occurrence on a blockchain, measuring the effort to do so in terms of time and cost. The architecture design shown in Figure 2 depicts how patient claims data from a transmitting organization flows through to a receiving organization within our system. The data in transit is secured using HTTPS for encrypted communication over networks, and SSH for secure remote access and command execution [25, 69]. Prior to secure transmission of patient claims, data is converted into an ontological format using existing libraries, then combined with the Part 2 knowledge data in a graph database (called RDFox) that we have created that contains the classes and attributes that semantically encapsulate what represents Part 2 data [47]. To secure data in RDFox, we use AES-256-CBC encryption for data at rest, TLS/SSL for data in transit, and strict access controls for authorized users [63, 64]. Data logged on the blockchain is encrypted or hashed to prevent exposure and is accessible via TLS/SSL only.

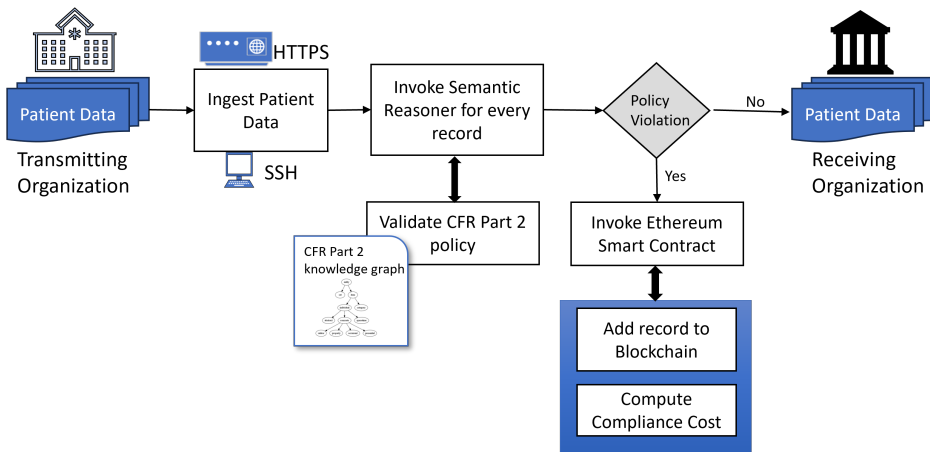


Fig. 2. System Architecture

Using the categorization from Table 1, we developed a knowledge graph to represent Part 2 Value Sets and Clinical types as shown in Figure 3 [52]. The graph captures any Part 2 patient experience using any clinical terminology system, including ICD, SNOMED, CPT, LOINC, and RxNorm. For example, Figure 4 demonstrates the graph populated with LOINC and SNOMED-CT terms for a Lab of Opioids and a Diagnosis of Mental Health Disorder. As administrative claims data contains billable information, including diagnoses codes in SNOMED-CT, derived from the clinical data present in the EHR, we can link the Part 2 graph with administrative data on the shared terminology; in this case SNOMED-CT. This enables the full exploration of claims data in unique ways, such as identifying heretofore unobserved associations as well as clustering of information. An example of such an insight into the patient experience and claims costs is presented in Figure 5, where one could extrapolate that a Mental Health Disorder has been linked to a claim; by extension the claim will contain patient demographic information, cost, and other pertinent attributes.

The graph model's associations between different classes of Part 2 concepts is shown in SPARQL Protocol and RDF Query Language (SPARQL) query format in Figure 6; the query returns all diagnoses, findings, medications, procedures, and labs for any Part 2 disorder from claims data loaded into a graph database. For our system, every claim has a policy violation check performed through SPARQL query, and in the event that there is Part 2 data contained in the transfer without the patient's consent, a smart contract is invoked to record the out-of-compliance exchange on the blockchain.

We instantiated an on premise, private version of OpenEthereum, as our blockchain; we visualized the blocks we added to it using a browser called BlockScout [10]. The open source version of Ethereum is customizable so that any healthcare organization can configure it based upon its infrastructure. The blockchain functions to create an audit trail, and for cost measurement of effort in identifying compliance issues [61, 62]. We developed a smart contract in Solidity and deployed it to our OpenEthereum network; the contract takes as inputs: the sending organization Ethereum address, the receiving organization Ethereum address, and a list of Part 2 clinical terms in SNOMED-CT. We measure the resources used in gas expended as measured in Gwei, or one-billionth of one ETH [28]. For data consistency and fault tolerance, our testnet utilized four nodes to replicate logging of Part 2 data being exchanged. This approach guarantees that all participating organizations come to consensus on the exchange of data and that it contained Part 2 information. This replication

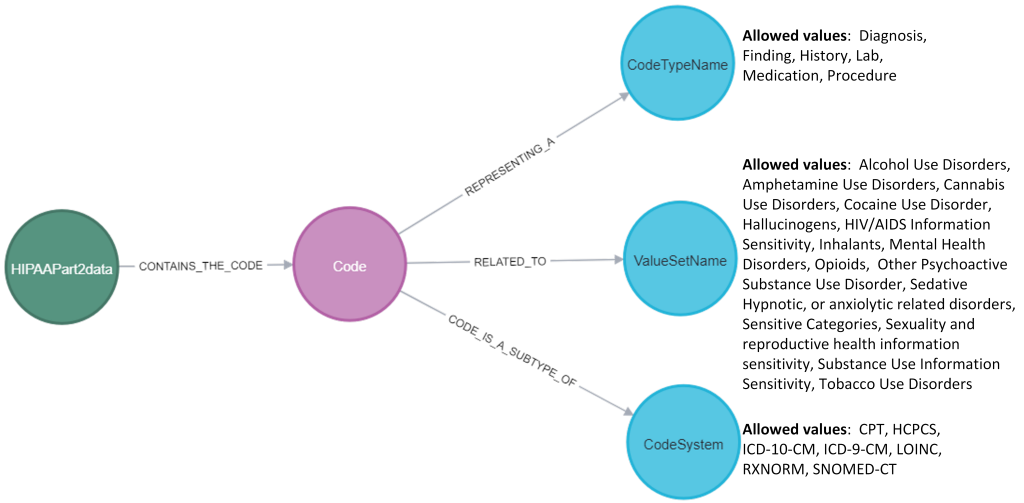


Fig. 3. Part 2 Knowledge Graph

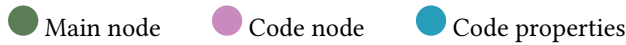


Table 3. Claims Data Input

Attribute	Value
Organizations	6
Claims Files	44
Average number of claims per organization	2,132,686
Total File Size	21.7 GB

adheres to the foundational principles of distributed systems for enforcing data reliability and strengthens our blockchain-based audit trail by creating backups of critical compliance logs across organizations [11, 57].

We produced healthcare claims using the Synthea data generator and then translated them into an ontological format via Python library [36, 47, 82]. We generated a set of 12M claims for 60,000 patients for six hospitals spread throughout the country, each located near academic medicine institutions, to account for geographic population variations and create cohorts that could be used in research between institutions (thereby triggering the need for Part 2 compliance). We relied on the synthetic data generator to create claims with Part 2 clinical terms in them based upon the demographic characteristics of the location for which the data were created. Table 3 shows the overall file count, size and average number of claims per organization. Each patient had zero or more claims, and each claim had eight mutually exclusive diagnoses fields contained in a set of 44 files.

For querying the data set of claims and determining if Part 2 data is included, we use RDFS given its in memory knowledge graph feature enabling better query performance [58]. Additionally RDFS adapts to new information in real time through forward chaining and has an API for loading and querying graph data [58]. As Figure 7 depicts, we developed an Elixir API for interacting with both RDFS and OpenEthereum, i.e. for both data loading/querying and smart contract invocation. The Elixir language provides significant fault tolerance and its web framework Phoenix is supported

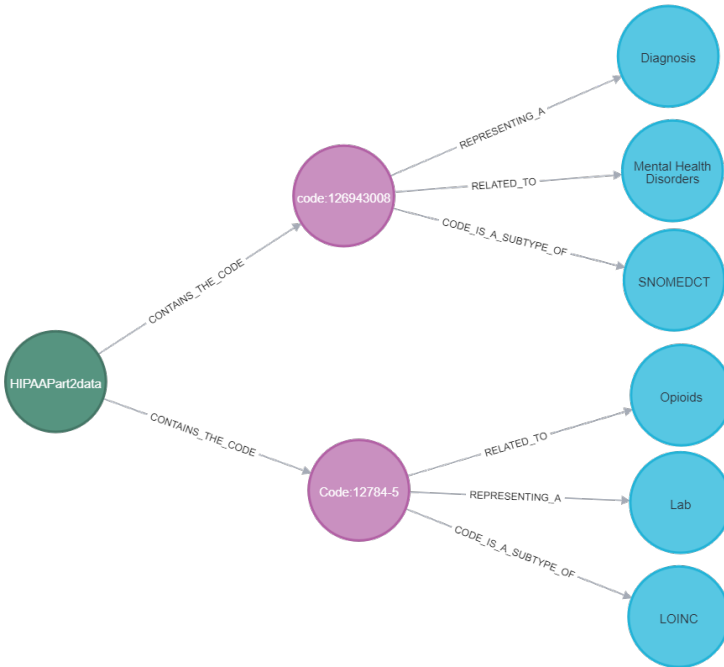


Fig. 4. Part 2 Knowledge Graph Example - identifying Opioid Labs and Mental Health Disorders Diagnosis

● Main node ● Code node ● Code properties

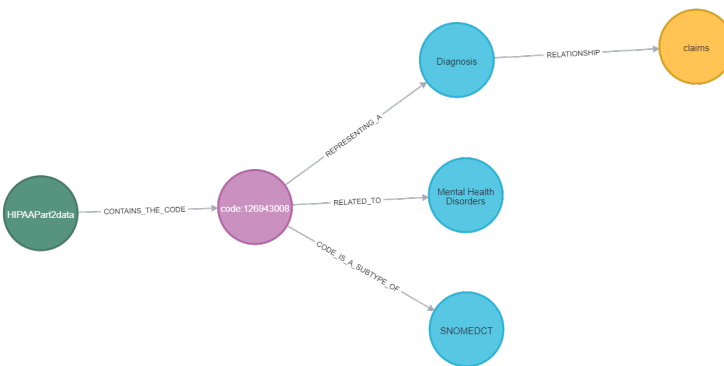


Fig. 5. Part 2 Mental Health Disorder SNOMED-CT code Linked to Claims

● Main node ● Code node ● Code properties ● Claim node

by an active open source community [41, 53]. We utilized Elixir Ethereum libraries to load and invoke a smart contract Application Binary Interface, and to query the OpenEthereum blockchain to gather cost information in gas expended [33, 51, 67]. Using a simple Elixir web client we developed a library to integrate with RDFox, thereby enabling on-demand loading of ontological data into the graph database, and execution of SPARQL queries against the newly loaded data [9]. Elixir benchmarking libraries enabled gathering performance metrics, inclusive of average time spent identifying Part 2 claims then logging them on the blockchain [7, 8]. By having a means to measure

```

PREFIX claims: <https://knacc.umbc.edu/synthea#>
PREFIX part2: <http://knacc.umbc.edu/hipaapart2_codes#>

SELECT ?claim ?diagnosisColumn ?diagnosis ?hasAttribute ?RELATED_TO ?REPRESENTING_A
WHERE {
  ?claim a claims:Claim .
  VALUES ?diagnosisColumn {
    claims:diagnosis1
    claims:diagnosis2
    claims:diagnosis3
    claims:diagnosis4
    claims:diagnosis5
    claims:diagnosis6
    claims:diagnosis7
    claims:diagnosis8
  }
  ?claim ?diagnosisColumn ?diagnosis .
  ?attribute a part2:Code ;
    part2:hasAttribute ?hasAttribute ;
    part2:RELATED_TO ?RELATED_TO ;
    part2:REPRESENTING_A ?REPRESENTING_A .
  FILTER (str(?hasAttribute) = ?diagnosis)
}
}

```

Fig. 6. SPARQL Query to Detect Part 2 Data in Claims

the effort in gas expended as well as time spent, we are able to calculate the cost of Part 2 HIPAA compliance auditing of Part 2 data exchanged.

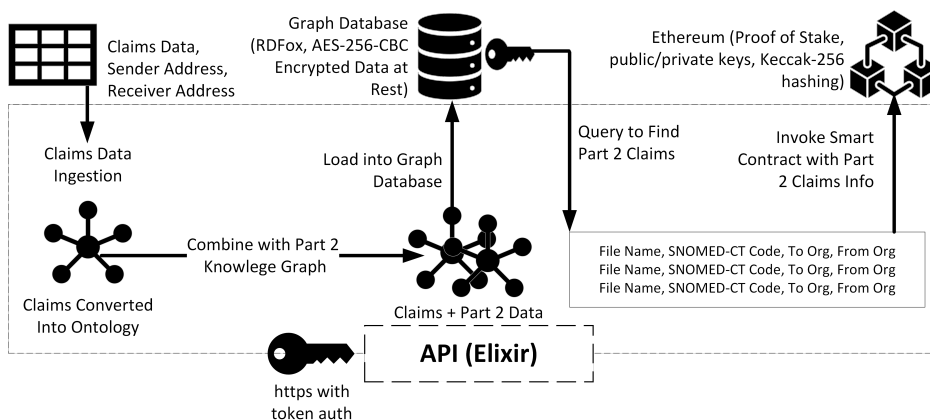


Fig. 7. API Components for Loading, Transforming, and Executing Part 2 Compliance Checks on Claims Data

The environment supporting this technology stack was a server running the Ubuntu 20.04.3 LTS OS, provisioned with 376GB memory and 72 Intel(R) Xeon(R) Gold 6140 @ 2.30GHz CPU's.

5 Results

Using the NIH value set of SNOMED-CT terms that identify claims that require patient consent, we evaluate claims data being shared between six organizations for approximately 60,000 patients [81]. We examine the data to detect one of 1,944 clinical terms in one of eight diagnosis fields for one of 44 files, giving us 1,944³⁵² possibilities for one or more Part 2 clinical terms to be present in the claims [81]. From this set of possibilities, our actual number of Part 2 terms generated totaled 204,744. Table 4 shows the distribution of this sample according to our knowledge graph's attributes of value set name - either diagnosis or finding in this case as we have no lab data.

Table 4. NIH Value Set Results

NIH Value Set Name	Diagnosis	Finding	Total
Alcohol	844		844
Cocaine Use Disorder	16		16
Mental Health Disorders	28,781	32,785	61,566
Opioids	2,298		2,298
Other Psychoactive Substance Use Disorder	122,518	17,502	140,020
Grand Total	154,457	50,287	204,744

In the event Part 2 data was shared a block with a single transaction listing the cause for compliance failure is added to the OpenEthereum blockchain via smart contract as shown in Figure 8. The block contains a transaction that has the sender organization, the compliance smart contract address, and the gas expended by the transaction. These values as well as the raw input to the smart contract are presented in Figure 8, where the hypothetical file name is `duke_to_wake_claim_0.ttl`, the diagnosis field is the first on the claim, and the SNOMED-CT codes vary across claims. The gas expended in this example is 68,577, which is 0.000068577 ETH or \$0.13 as of November 2023.

For all 44 claims files, the mode and median claim file size was 529MB. Table 5 captures the performance and cost metrics, showing that it took on average 3.17 hours to check data exchange point-to-point between all six organizations. Checking all files took 19.03 hours and produced 352 blocks of data in Ethereum. Converting the Ethereum gas expended into U.S. dollars with an exchange rate as of November 2023, the effort expended \$29.07 total, averaging \$0.08 per block, at a cost of \$1.53 per hour.

6 Limitations and Security Challenges

6.1 Synthetic Data

The decision to use Synthea for this iteration limited our validation to SNOMED-CT clinical terms; future work would use additional synthetic data sources with more terminologies, such as [5, 18, 84, 87]. Furthermore, the patient cohort size of 60,000 was selected based on determining the upper threshold for the available computing resources. Larger populations of patient data files should be achievable in scalable cloud environments.

6.2 Blockchain Security Threats

Public blockchains, characterized by their limited anonymity and open access to data, risk exposing sensitive healthcare information to unauthorized parties. While the transparency of blockchain

Transaction Details

Transaction Hash	0x32299b4f315610801d3eb8fe9689d1702d51557df40f93fb1936e4b8731a8f15
Status	Confirmed Confirmed by 43,666
Block	2765
Timestamp	3 months ago October-03-2023 02:49:22 PM +-5 UTC Confirmed within <= 0.03 seconds
From	0xf71B10B33f9a67Abd8D398399D565f9D73E70772
Interacted With (To)	0xbF5964aAdF1dE92c4e4402CC7a9a6de5Dba04178

Gas Limit	8,000,000
Gas Used by Transaction	68,577 0.86%
Raw Input	

```

duke_to_wake_claim_0.ttl Claim Part 2 SNOMED 153151000119100 Claim field: 1
duke_to_wake_claim_0.ttl Claim Part 2 SNOMED 191928000 Claim field: 1
duke_to_wake_claim_0.ttl Claim Part 2 SNOMED 199252002 Claim field: 1
duke_to_wake_claim_0.ttl Claim Part 2 SNOMED 228277002 Claim field: 1
duke_to_wake_claim_0.ttl Claim Part 2 SNOMED 248103001 Claim field: 1
duke_to_wake_claim_0.ttl Claim Part 2 SNOMED 268645007 Claim field: 1
duke_to_wake_claim_0.ttl Claim Part 2 SNOMED 268650001 Claim field: 1
duke_to_wake_claim_0.ttl Claim Part 2 SNOMED 268738002 Claim field: 1
duke_to_wake_claim_0.ttl Claim Part 2 SNOMED 60103007 Claim field: 1

```

Fig. 8. Transaction Details of Smart Contract

Table 5. Claims Data Part 2 Compliance Results

Attribute	Value
Average Compliance Check Time per Organization	3.17 hours
Total Compliance Check Time	19.03 hours
Total No. of Ethereum Blocks	352
Total Gas Expended	15, 400, 447
Average Gas Expended	43, 751
Total Cost ETH	0.015400447
Average Cost ETH	0.000043751
Total Cost USD	\$29.07
Average Cost USD	\$0.08
Hourly Cost USD	\$1.53

ensures that data recorded within each block remains unalterable—upholding the principle of immutability—this very transparency can inadvertently lead to severe privacy violations if sensitive patient data is not rigorously protected [74].

When one entity has multiple accounts or nodes on a blockchain network they may perform a Sybil attack, wherein one actor can influence the blockchain while seeming to be discrete, separate entities. To combat this, two mechanisms exist. First, and most straightforward, is to make the blockchain permissioned (i.e., private) - all actors are known and have proven their identities.

Second, in a permissionless (i.e., public) setup, it is key to have resources expended by participants when they run consensus algorithms so there is a cost; the most popular being Bitcoin's proof-of-work (PoW) and Ethereum's proof-of-stake (PoS). Despite this, some Sybil attacks simply try to slow the progress of the blockchain by executing a Distributed Denial of Service (DDoS) attack.

The 51% attack, or majority attack, is a type of Sybil attack where an attacker gains control over more than 50% of the blockchain network's computational power, they could potentially alter the blockchain by creating a new fork, or intercepting transactions. This would undermine the trust in the system designed to protect HIPAA Part 2 data by enabling the attacker to manipulate or prevent the recording of specific transactions [70]. However, since Ethereum transitioned from PoW to PoS, a successful 51% attack is unlikely because an attacker would need to own that much of the staked ETH in the blockchain [77]. Even with less than 50% control, strategic manipulation of the blockchain could still occur, compromising the integrity of the data stored [29, 50, 70, 74].

Smart contracts can introduce substantial security risks. For example, reentrancy attacks can allow malicious actors to repeatedly interact with a contract before its state is updated, leading to potential unauthorized data access or loss [12, 48]. This issue is further exacerbated by the immutability of smart contracts, which, while ensuring data integrity, also solidifies any vulnerabilities once the contract is deployed [24]. DDoS attacks can occur on smart contracts to keep them from being usable by network participants, or to drain funds from them [71]. As of late 2025, threat actors have found a new method for taking advantage of smart contracts - EtherHiding, wherein malware is distributed via smart contract, which is used to obfuscate its presence. Smart contracts can be used to set up Distributed Apps (DApps) and Distributed Autonomous Organizations (DAOs). DAOs are entities that live on the blockchain, whose governance and processes are determined by a community of users that agree to the rules of the entity, and memorialize those in code on the blockchain. For Ethereum, one of the more notable attacks occurred in 2016 when a venture capital DAO lost millions of dollars worth of its ETH when actors took advantage of a bug in the code to move funds out of the DAO [30, 71]. To resolve the issue, Ethereum was forked, with the "original" transactions with the DAO becoming Ethereum Classic, and the "main" Ethereum going back in time to move the funds that were taken to trusted accounts.

6.3 Economic Forces

Economic factors exacerbate these risks significantly. The high gas costs associated with deploying and executing smart contracts can lead to vulnerabilities related to cost inefficiencies and resource exhaustion [49]. These economic pressures can result in unintended service interruptions or financial losses. Additionally, attack vectors such as overflow and underflow attacks, short address attacks, and high gas cost Denial-of-Service (DoS) attacks highlight how design and execution flaws can lead to severe exploits [48, 66]. Overflow attacks may overwrite critical data, while short address attacks can mismanage token distributions. High gas cost DoS attacks can disrupt data processing and regulatory compliance by overwhelming the system with excessive gas fees. Addressing these vulnerabilities is essential for maintaining the security, reliability, and regulatory compliance of blockchain systems handling sensitive HIPAA Part 2 data.

In our research, we implement foundational trust principles to enhance the security of HIPAA Part 2 data exchanges. Our system integrates blockchain's safety and liveness principles with semantic web's reasoning capabilities to strengthen trust through consensus and cryptography. This approach ensures that the data used in decision-making aligns closely with expected outcomes, maintaining accuracy and filtering out non-compliant data. While our current system operationalizes these concepts, it also lays the groundwork for further incorporation of comprehensive trust models that apply within a specific context at a particular point in time [4, 55]. Additionally, models have begun to integrate the concept of zero-trust, where systems constantly re-verify participants. Blockchain's

principles reflect zero-trust, as, e.g., in Ethereum all transactions are verified and participants must establish identity through cryptography.

6.4 Real-World Implementation

The EDI claims processing sector, which has both commercial and governmental payers, would benefit from this Ethereum and knowledge graph-based system. This is because AI-enabled claims authorization is being leveraged to evaluate and deny claims faster than ever, while at the same time claims processors are vulnerable to cyber attack. The cyberattack on the Change Healthcare claims processor shows the weaknesses of legacy systems that now face increased regulatory scrutiny [80]. Bad actors are using previously compromised protected health data to create fake providers and synthetic claims, in one instance stealing \$900M from CMS by using the existing claim processing infrastructure [59]. Our Ethereum-based solution could integrate into the claims processing flow because it uses the same data standard (EDI 837 format) and would provide stronger security guarantees, not the least of which would be proven identities and validation of claim processing, with transparency for payers, providers, and patients (while being HIPAA secure). Our system would need to bridge to existing processes through API, or be set up as a claims processing service itself (requiring further development).

7 Conclusion and Future Work

We have developed an Ethereum-based blockchain and knowledge graph to detect Part 2 data in health care claims. Because the knowledge graph integrates all clinical terminologies (ICD, SNOMED, CPT, LOINC, RxNorm), cross-cutting questions about procedures, diagnoses, and lab results for any Part 2 can be asked of any claims data set where these terms are present.

Because we utilize Ethereum smart contracts to detect the Part 2 terms, we are able to measure the amount of effort taken to detect this sensitive data in both time and real dollars. The manual process equivalent is done by analysts at healthcare organizations on an ad hoc basis and is defined by the data use agreement's specification of the allowable variable lists and cohort requirements. Typically, an analyst earning \$100K per year that takes 8 hours to validate a data set against a data use agreement puts in \$385 worth of labor, or \$48 per hour. Practically, validating the data sets may exceed the available hours of the analyst for a year, if they are of the order of magnitude examined in our experiments. By comparison, the automated system described here represents 97% savings per hour in wages, provides guarantees that the data was validated, and is scalable.

In addition, the blockchain creates an immutable audit trail to meet compliance standards for organizations to maintain HIPAA compliance. By logging the claim field that contains the sensitive information onto the blockchain, there is a direct link to the patient for whom the claim is associated. Through this link, and because of blockchain's function as an immutable ledger, there is no ambiguity regarding when a patient's data was exchanged. To enhance security and mitigate the risk of malicious actions, we plan to incorporate anonymity preserving attributes into our Ethereum blockchain in the future, e.g., zero knowledge rollups which move processing of sensitive data offchain and log summary data on the main Ethereum chain. This approach will provide more controlled and secure transaction processes, ensuring a higher level of trust and integrity in the system, thereby enabling it to be used by legislative audit teams. Auditors that seek out fraud, waste, and abuse (FWA) would benefit from our system, as detecting FWA is a time-consuming and precise effort. The system could be measured against the timeliness of FWA findings (it will be more proactive), as well as the effort to validate such findings (it should require less manual effort).

Future research directions include the integration of generative AI. Specifically, translating a prompt into a knowledge graph query using available libraries, which would allow for a chat-like experience with the system, wherein users would ask about Part 2 data in natural language terms.

We also anticipate testing the system as a claims EDI processor, to compare its usage of Ethereum to settle claims, rather than the dollar as a medical claims fintech application.

References

- [1] Social Security Administration. 2023. *Section 1176: Penalties for rules and regulations*. [https://www.ssa.gov/OP_Home/ssact/title11/1176.htm#:~:text=\(D\)%20the%20amount%20described%20in,year%20may%20not%20exceed%20%241%2C500%2C000.&text=\(1\)%20Offenses%20otherwise%20punishable](https://www.ssa.gov/OP_Home/ssact/title11/1176.htm#:~:text=(D)%20the%20amount%20described%20in,year%20may%20not%20exceed%20%241%2C500%2C000.&text=(1)%20Offenses%20otherwise%20punishable) Accessed: November 21, 2023.
- [2] Social Security Administration. Year of Access. *Section 1171: Penalties for rules and regulations*. [https://www.ssa.gov/OP_Home/ssact/title11/1171.htm#:~:text=\(D\)%20the%20amount%20described%20in,year%20may%20not%20exceed%20%241%2C500%2C000.&text=\(1\)%20Offenses%20otherwise%20punishable](https://www.ssa.gov/OP_Home/ssact/title11/1171.htm#:~:text=(D)%20the%20amount%20described%20in,year%20may%20not%20exceed%20%241%2C500%2C000.&text=(1)%20Offenses%20otherwise%20punishable). Accessed: November 21, 2023.
- [3] Alchemy. 2023. *Gwei Calculator*. <https://www.alchemy.com/gwei-calculator> Accessed on: February 13, 2026.
- [4] Farag Azzedin and Mustafa Ghaleb. 2019. Internet-of-Things and Information Fusion: Trust Perspective Survey. *Sensors* 19, 8 (April 2019), 1929. <https://doi.org/10.3390/s19081929> Accessed: 2024-08-28.
- [5] Mrinal Kanti Baowaly, Chia-Ching Lin, Chao-Lin Liu, and Kuan-Ta Chen. 2019. Synthesizing electronic health records using improved generative adversarial networks. *Journal of the American Medical Informatics Association* 26, 3 (2019), 228–241.
- [6] Sai Batchu, Karan Patel, Owen S Henry, Aleem Mohamed, Ank A Agarwal, Henna Hundal, Aditya Joshi, Sankeerth Thoota, and Urvish K Patel. 2022. Using ethereum smart contracts to store and share COVID-19 patient data. *Cureus* 14, 1 (2022).
- [7] bencheeorg. 2023. Benchee: A benchmarking library for Elixir. <https://github.com/bencheeorg/benchee>. Accessed on: February 13, 2026.
- [8] bencheeorg. 2023. Benchee HTML: HTML formatter for Benchee, a benchmarking library for Elixir. https://github.com/bencheeorg/benchee_html. Accessed on: February 13, 2026.
- [9] Benoît Chesneau. Year Accessed. Hackney: Simple HTTP client for Erlang. <https://github.com/benoitc/hackney>. Accessed on: June 26, 2024.
- [10] BlockScout. 2022. BlockScout. <https://github.com/blockscout/blockscout>, Last accessed on 2022-11-20.
- [11] Silvia Bonomi, Antonella Del Pozzo, Álvaro García-Pérez, and Sara Tucci-Piergiorgianni. 2021. SoK: Achieving State Machine Replication in Blockchains based on Repeated Consensus. *arXiv preprint arXiv:2105.13732* (2021). <https://arxiv.org/abs/2105.13732> Accessed: 2024-09-04.
- [12] Michael Brenner, Kurt Rohloff, Joseph Bonneau, Andrew Miller, Peter Y.A. Ryan, Vanessa Teague, Andrea Bracciali, Massimiliano Sala, Federico Pintore, and Markus Jakobsson (Eds.). 2017. *Financial Cryptography and Data Security: FC 2017 International Workshops WAHC, BITCOIN, VOTING, WTSC, and TA*. Lecture Notes in Computer Science, Vol. 10175. Springer, Sliema, Malta.
- [13] Vitalik Buterin et al. 2014. A next-generation smart contract and decentralized application platform. *white paper* (2014).
- [14] Health Sector Cybersecurity Coordination Center. 2019. A Cost Analysis of Healthcare Sector Data Breaches.
- [15] Centers for Medicare Medicaid Services. 2021. *Definitions, Sources, and Methods*. <https://www.cms.gov/files/document/definitions-sources-and-methods.pdf> Accessed: Ocotober 12, 2023.
- [16] Centers for Medicare Medicaid Services. 2023. *CMS Chronic Conditions Warehouse*. <https://www2.ccwdata.org/web/guest/home/> Accessed on: February 13, 2026.
- [17] Centers for Medicare Medicaid Services. 2023. *CMS Related Sites*. <https://data.cms.gov/related-sites> Accessed on: February 13, 2026.
- [18] Edward Choi, Siddharth Biswal, Bradley Malin, Jon Duke, Walter F Stewart, and Jimeng Sun. 2017. Generating multi-label discrete patient records using generative adversarial networks. In *Machine learning for healthcare conference*. PMLR, 286–305.
- [19] Efthymios Chondrogiannis, Vassiliki Andronikou, Efsthathios Karanastasis, Antonis Litke, and Theodora Varvarigou. 2022. Using blockchain and semantic web technologies for the implementation of smart contracts between individuals and health insurance organizations. *Blockchain: Research and Applications* 3, 2 (2022), 100049.
- [20] James Clavin, Sisi Duan, Haibin Zhang, Vandana P Janeja, Karuna P Joshi, Yelena Yesha, Lucy C Erickson, and Justin D Li. 2020. Blockchains for government: use cases and challenges. *Digital Government: Research and Practice* 1, 3 (2020), 1–21.
- [21] James Clavin and Karuna P Joshi. 2023. Policy Integrated Blockchain to Automate HIPAA Part 2 Compliance. In *2023 IEEE International Conference on Digital Health (ICDH)*. IEEE, 307–314.
- [22] Cloudtcity Blog. 2023. Understanding the Cost of HiTRUST Certification in 2024. <https://blog.cloudtcity.com/hitrust-certification-cost-2024>. Accessed: December 8, 2023.

- [23] Gaby G Dagher, Jordan Mohler, Matea Milojkovic, and Praneeth Babu Marella. 2018. Ancile: Privacy-preserving framework for access control and interoperability of electronic health records using blockchain technology. *Sustainable cities and society* 39 (2018), 283–297.
- [24] Kevin Delmolino, Mitchell Arnett, Ahmed Kosba, Andrew Miller, and Elaine Shi. 2016. Step by step towards creating a safe smart contract: Lessons and insights from a cryptocurrency lab. In *Financial Cryptography and Data Security: FC 2016 International Workshops, BITCOIN, VOTING, and WAHC, Christ Church, Barbados, February 26, 2016, Revised Selected Papers 20*. Springer, 79–94.
- [25] Ivan Dolnák and Ján Litvik. Year. Introduction to HTTP Security Headers and Implementation of HTTP Strict Transport Security (HSTS) Header for HTTPS Enforcing. *University of Žilina, Faculty of Electrical Engineering* (Year).
- [26] Electronic Code of Federal Regulations. 2023. *Electronic Code of Federal Regulations (e-CFR): Title 42, Chapter I, Subchapter A, Part 2, Subpart A*. <https://www.ecfr.gov/current/title-42/chapter-I/subchapter-A/part-2/subpart-A> Accessed: November 21, 2023.
- [27] Lavanya Elluri, Aritrnan Piplai, Anantaa Kotal, Anupam Joshi, and Karuna Pande Joshi. 2021. A policy-driven approach to Secure Extraction of COVID-19 data from Research Papers. *Frontiers in Big Data* 4 (2021).
- [28] Ethereum Foundation. 2023. *Ethereum Gas*. <https://ethereum.org/en/developers/docs/gas/> Accessed: November 3, 2023.
- [29] Ittay Eyal and Emin Gün Sirer. 2014. Majority is Not Enough: Bitcoin Mining is Vulnerable. In *Proceedings of the 2014 ACM Conference on Computer and Communications Security (CCS '14)*. ACM, London, UK, 436–449.
- [30] Youssef Faqir-Rhazoui, Javier Arroyo, and Samer Hassan. 2021. A comparative analysis of the platforms for decentralized autonomous organizations in the Ethereum blockchain. *Journal of Internet Services and Applications* 12, 1 (2021), 9.
- [31] Jake Frankenfield. 2022. *What Is Ethereum and How Does It Work?* <https://www.investopedia.com/terms/e/ethereum.asp> Updated: September 27, 2022; Accessed: November 27, 2023.
- [32] Morgan Henderson, Fei Han, Chad Perman, Howard Haft, and Ian Stockwell. 2022. Predicting avoidable hospital events in Maryland. *Health Services Research* 57, 1 (2022), 192–199.
- [33] Hswick. 2023. ExW3: Elixir JSON-RPC client for Ethereum and Ethereum-compatible blockchains. <https://github.com/hswick/exw3>. Accessed on: February 13, 2026.
- [34] Ponemon Institute and IBM. 2023. 2023 Cost of a Data Breach Study: Global Overview. <https://www.ibm.com/security/data-breach>.
- [35] Yu Ji, Jun Zhang, Jun Ma, Cheng Yang, and Xin Yao. 2018. BMPLS: Blockchain-Based Multi-level Privacy-Preserving Location Sharing Scheme for Telecare Medical Information Systems. *Journal of Medical Systems* 42, 8 (2018), 147. <https://doi.org/10.1007/s10916-018-0998-2>
- [36] Alistair EW Johnson, Tom J Pollard, Lu Shen, Li-wei H Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G Mark. 2016. MIMIC-III, a freely accessible critical care database. *Scientific data* 3, 1 (2016), 1–9.
- [37] Karuna Pande Joshi, Yelena Yesha, and Tim Finin. 2016. An Ontology for a HIPAA compliant cloud services. *4th International IBM Cloud Academy Conference ICACON 2016* (2016).
- [38] Maithilee Joshi, Karuna Joshi, and Tim Finin. 2018. Attribute based encryption for secure access to cloud based EHR systems. In *2018 IEEE 11th International Conference on Cloud Computing (CLOUD)*. IEEE, 932–935.
- [39] Maithilee Joshi, Karuna Pande Joshi, and Tim Finin. 2019. Delegated authorization framework for EHR services using attribute based encryption. *IEEE Transactions on Services Computing* (2019).
- [40] HIPAA Journal. Year of Access. *HIPAA Violation Cases*. <https://www.hipaajournal.com/hipaa-violation-cases/> Accessed: November 21, 2023.
- [41] Sasa Juric. 2019. *Elixir in action*. Simon and Schuster.
- [42] Acting Deputy Assistant Secretary for Mental Health Kana Enomoto, Substance Use, and Secretary Sylvia M. Burwell. 2017. Dated: December 20, 2016. <https://www.govinfo.gov/content/pkg/FR-2017-01-18/pdf/2017-00719.pdf>. [FR Doc. 2017–00719 Filed 1–13–17; 11:15 am].
- [43] Preethi Kasireddy. 2022. *How Does Ethereum Work Anyway?* https://www.easygoing.pflog.eu/32_blockchain_P2P/ethereum_blockchain.pdf Published: September 27, 2022; Accessed: November 27, 2023.
- [44] Dae-young Kim and Karuna P Joshi. 2021. A Semantically Rich Knowledge Graph to Automate HIPAA Regulations for Cloud Health IT Services. In *2021 7th IEEE Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing,(HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS)*. IEEE, 7–12.
- [45] Kraken. 2023. *Kraken: Ethereum to USD Converter*. <https://www.kraken.com/convert/eth/usd> Accessed on: February 13, 2026.
- [46] Ora Lassila, Ralph R Swick, et al. 1998. Resource description framework (RDF) model and syntax specification. (1998).
- [47] Leroy Kim and contributors. 2021. Synthea RDF. <https://github.com/leroykim/synthea-rdf>. Accessed on March 19, 2023.

- [48] Xiaoqi Li, Peng Jiang, Ting Chen, Xiapu Luo, and Qiaoyan Wen. 2020. A survey on the security of blockchain systems. *Future Generation Computer Systems* 107 (2020), 841–853. <https://doi.org/10.1016/j.future.2017.08.020>
- [49] Daniel Macrinici, Cristian Cartofeanu, and Shang Gao. 2018. Smart contract applications within blockchain technology: A systematic mapping study. *Telematics and Informatics* 35, 8 (2018), 2337–2354. <https://doi.org/10.1016/j.tele.2018.10.004>
- [50] Daniel Macrinici, Cristian Cartofeanu, and Shang Gao. 2020. Smart Contract Applications within Blockchain Technology: A Systematic Mapping Study. *Blockchain Research Journal* 5, 2 (2020), 45–60. <https://example.com/smart-contract-applications>
- [51] Mana Ethereum. 2023. Ethereumx: Elixir JSON-RPC client for the Ethereum blockchain. <https://github.com/mana-ethereum/ethereumx>. Accessed on: February 13, 2026.
- [52] Satya Sai Mandlem. 2024. HIPAA Part 2 TTL File and Python Script. <https://github.com/Ebiquity/hie-synthetic-data/blob/SatyaSaimandlem-HIPAAPart2-ttl-file-and-pyghon-file/HIPAAPart2.ttl> Accessed: 2024-08-28.
- [53] Chris McCord, Jose Valim, and Bruce Tate. 2019. Programming Phoenix 1.4: Productive|> Reliable|> Fast. *Programming Phoenix 1.4* (2019), 1–325.
- [54] Deborah L McGuinness, Frank Van Harmelen, et al. 2004. OWL web ontology language overview. *W3C recommendation* 10, 10 (2004), 2004.
- [55] TJ Mitchell. 2019. *The Nature and Rationality of Trust and Trustworthiness*. Ph. D. Dissertation. The University of St Andrews. <https://research-repository.st-andrews.ac.uk/handle/10023/19059> Accessed: 2024-08-28.
- [56] Ahmed Musamih, Khaled Salah, Raja Jayaraman, Junaid Arshad, Mohamed Debe, Yousof Al-Hammadi, and Samer Ellahham. 2021. A blockchain-based approach for drug traceability in healthcare supply chain. *IEEE Access* 9 (2021), 6–14. <https://doi.org/10.1109/ACCESS.2020.3038992>
- [57] Kousik Nath. 2020. *Consistency guarantees in distributed systems explained simply*. <https://kousiknath.medium.com/sistensity-guarantees-in-distributed-systems-explained-simply-720caa034116> Accessed: 2024-09-04.
- [58] Yavor Nenov, Robert Piro, Boris Motik, Ian Horrocks, Zhe Wu, and Jay Banerjee. 2015. RDFox: A highly-scalable RDF store. In *The Semantic Web-ISWC 2015: 14th International Semantic Web Conference, Bethlehem, PA, USA, October 11-15, 2015, Proceedings, Part II* 14. Springer, 3–20.
- [59] Santul (The New York Times) Nerkar. 2025. U.S. Medicare Fraud Charges. <https://www.nytimes.com/2025/06/27/nyregion/us-medicare-fraud-charges.html>. *The New York Times* (27 June 2025). <https://www.nytimes.com/2025/06/27/nyregion/us-medicare-fraud-charges.html> Accessed 2025-10-20.
- [60] Office of the National Coordinator for Health Information Technology. [n. d.]. Trusted Exchange Framework and Common Agreement. Interoperability Standards Advisory. <https://www.healthit.gov/isa/trusted-exchange-framework-and-common-agreement> Accessed: March 25, 2023.
- [61] OpenEthereum. 2021. Configuring OpenEthereum. <https://openethereum.github.io/Configuring-OpenEthereum>, Last accessed on 2022-11-20.
- [62] OpenEthereum. 2021. OpenEthereum. <https://openethereum.github.io/>, Last accessed on 2022-11-20.
- [63] Oxford Semantic Technologies Ltd. 2024. Copyright Notice. © Copyright 2024, Oxford Semantic Technologies Ltd..
- [64] Oxford Semantic Technologies Ltd. 2025. Managing RDFox Servers: Server Parameters. <https://docs.oxfordsemantic.tech/servers.html#server-parameters>. <https://docs.oxfordsemantic.tech/servers.html#server-parameters> Accessed October 25, 2025.
- [65] Prateek Pandey and Ratnesh Litoriya. 2021. Securing e-health networks from counterfeit medicine penetration using blockchain. *Wireless Personal Communications* 117 (2021), 7–25.
- [66] Jack Pettersson and Robert Edström. 2016. *Safer Smart Contracts through Type-Driven Development*. Master’s Thesis. Örebro University, Örebro, Sweden.
- [67] POA Network. 2023. ExABI: Ethereum JSON-RPC bindings for Elixir. https://github.com/poanetwork/ex_abi. Accessed on: February 13, 2026.
- [68] Arumugam S Rajasekaran and Mohammed Azees. 2022. An anonymous blockchain-based authentication scheme for secure healthcare applications. *Security and Communication Networks* (2022), 1–10. <https://doi.org/10.1155/2022/2793116>
- [69] Chris Rapier and Benjamin Bennett. 2008. High speed bulk data transfer using the SSH protocol. In *Proceedings of the 15th ACM Mardi Gras Conference: From Lightweight Mash-Ups to Lambda Grids: Understanding the Spectrum of Distributed Computing Requirements, Applications, Tools, Infrastructures, Interoperability, and the Incremental Adoption of Key Capabilities* (Baton Rouge, Louisiana, USA) (MG ’08). Association for Computing Machinery, New York, NY, USA, Article 11, 7 pages. <https://doi.org/10.1145/1341811.1341824>
- [70] Ludovic Rembert. 2024. *51% Attack*. <https://www.privacycanada.net/51-attack> Last Updated on July 29, 2024.
- [71] Muhammad Saad, Jeffrey Spaulding, Laurent Njilla, Charles Kamhoua, Sachin Shetty, DaeHun Nyang, and David Mohaisen. 2020. Exploring the attack surface of blockchain: A comprehensive survey. *IEEE Communications Surveys & Tutorials* 22, 3 (2020), 1977–2008.

- [72] Secureframe. [n. d.]. *HIPAA Compliance: Cost and Time Savings*. <https://secureframe.com/hub/hipaa/cost-and-time-savings> Accessed: November 27, 2023.
- [73] Edward H Shortliffe, Edward H Shortliffe, James J Cimino, and James J Cimino. 2014. *Biomedical informatics: computer applications in health care and biomedicine*. Springer.
- [74] J. Smith, A. Johnson, and B. Lee. 2020. Exploring the Attack Surface of Blockchain: A Comprehensive Survey. *IEEE Communications* 58, 4 (2020), 123–135. <https://www.researchgate.net/publication/339634933> Accessed: [8/19/2024].
- [75] Substance Abuse and Mental Health Services Administration. 2023. *SAMHSA Confidentiality Regulations FAQs*. <https://www.samhsa.gov/about-us/who-we-are/laws-regulations/confidentiality-regulations-faqs> Accessed: November 21, 2023.
- [76] Andrew S Tanenbaum and Maarten Van Steen. 2007. *Distributed systems: principles and paradigms*. Prentice-Hall.
- [77] The Investopedia Team. 2025. *What is a 51% attack on blockchain? Risks, examples, and costs explained*. Investopedia. <https://www.investopedia.com/terms/1/51-attack.asp> Accessed: 2025-10-19.
- [78] U.S. Department of Health Human Services. 1996. Adverse Events, Near Misses, and Errors. <https://www.hhs.gov/hipaa/for-professionals/breach-notification>, Last accessed on 2023-12-22.
- [79] U.S. Department of Health Human Services. 2023. *HIPAA Part 2*. <https://www.hhs.gov/hipaa/for-professionals/regulatory-initiatives/hipaa-part-2/index.html> Accessed: November 21, 2023.
- [80] U.S. Department of Health and Human Services, Office for Civil Rights. 2025. Change Healthcare Cybersecurity Incident Frequently Asked Questions. <https://www.hhs.gov/hipaa/for-professionals/special-topics/change-healthcare-cybersecurity-incident-frequently-asked-questions/index.html>. <https://www.hhs.gov/hipaa/for-professionals/special-topics/change-healthcare-cybersecurity-incident-frequently-asked-questions/index.html> Updated as of March 14, 2025; last reviewed August 13, 2025.
- [81] U.S. National Library of Medicine. 2021. Value Set Authority Center (VSAC). <https://vsac.nlm.nih.gov/welcome>. Accessed on March 19, 2023.
- [82] Jason Walonoski, Mark Kramer, Joseph Nichols, Andre Quina, Chris Moesel, Dylan Hall, Carlton Duffett, Kudakwashe Dube, Thomas Gallagher, and Scott McLachlan. 2018. Synthea: An approach, method, and software mechanism for generating synthetic patients and the synthetic electronic health care record. *Journal of the American Medical Informatics Association* 25, 3 (2018), 230–238.
- [83] X12. 2023. *About X12*. <https://x12.org/about/about-x12> Accessed: October 12, 2023.
- [84] Liyang Xie, Kaixiang Lin, Shu Wang, Fei Wang, and Jiayu Zhou. 2018. Differentially private generative adversarial network. *arXiv preprint arXiv:1802.06739* (2018).
- [85] Ibrar Yaqoob, Khaled Salah, Raja Jayaraman, and Yousof Al-Hammadi. 2021. Blockchain for healthcare data management: opportunities, challenges, and future recommendations. *Neural Computing and Applications* 34 (2021), 11475–11490. <https://doi.org/10.1007/s00521-021-06197-y>
- [86] Peng Zhang, Douglas C Schmidt, Jules White, and Gunther Lenz. 2018. Blockchain technology use cases in healthcare. In *Advances in computers*. Vol. 111. Elsevier, 1–41.
- [87] Ziqi Zhang, Chao Yan, Diego A Mesa, Jimeng Sun, and Bradley A Malin. 2020. Ensuring electronic medical record simulation through better training, modeling, and evaluation. *Journal of the American Medical Informatics Association* 27, 1 (2020), 99–108.