Data redundancy of blockchain systems in construction projects

Weisheng Lu [1], Liupengfei Wu [1]* and Chen Chen [2]

[1]Department of Real Estate and Construction, The University of Hong Kong, Pokfulam, Hong Kong SAR, China, [2]Department of Building and Real Estate, The Hong Kong Polytechnic University, Kowloon, Hong Kong SAR, China

Industrial stakeholders have complained that current blockchain systems are too expensive, particularly in temporary endeavours like construction projects. However, while researchers have examined blockchain system structure among inter-firm organizations in construction, little research has considered the data redundancy of these systems. This research, therefore, provides insight by modelling data redundancy in construction project blockchain systems. We conduct a series of laboratory experiments on a Hyperledger Fabric blockchain system, discovering that the data volume of a blockchain system grows proportionally with the size of the files to be uploaded, the number of peer nodes in the network, and the frequency of blockchain operations in construction, regardless of the block size or how the peers are dispersed in different construction organizations. Beyond identifying the factors that determine data redundancy of a blockchain system, this research provides a basis for researchers to explore the optimization of blockchain storage and the impacts of blockchain system data redundancy in construction projects. In practical terms, the proposed data redundancy model in this research provides a reference for users in construction who aim to build blockchain systems.

1 Introduction

Several challenges face the construction industry, including poor collaboration, information-sharing and transparency, along with low productivity, lack of trust and late payments. In the digital transformation being undertaken to respond to these challenges, the adoption of blockchain is increasing (Penzes et al., 2018; Li et al., 2019). Compared with a traditional database, this distributed ledger technology offers increased decentralization, traceability, transparency, and immutability (Risius and Spohrer, 2017). Researchers and practitioners are actively investigating aspects of blockchain in construction, including procurement and supply chain (Tezel et al., 2021), design and construction (Lu et al., 2021a), operation and life cycle (Ye et al., 2018), smart cities (Chen et al., 2020), intelligent systems (Wu et al., 2022a), energy and carbon footprint (Rodrigo et al., 2021), and decentralized organizations (Perera et al., 2020).

As a distributed database, however, blockchain is bound to face the problem of data redundancy, defined as a condition in a database or data storage where the same set of data is stored in multiple places (Verma and Singh, 2018). This can occur unintentionally or intentionally. Unintentional or accidental data redundancy arises from inefficient coding or overcomplicated data storing processes (Fan et al., 2018). Its drawbacks include increased data discrepancy, corruption, database size, and cost. Intentional data redundancy may still

have these drawbacks but can also be used for data protection, reducing data corruption, promoting consistency, disaster recovery (Xenya and Quist-Aphetsi, 2019; Li et al., 2021), and increasing access speed (Verma and Singh, 2018). In storing data in multiple, decentralized, and distributed ledgers, blockchain is an intentional data redundancy technology (Xue and Lu, 2020).

Construction is the epitome of a project-based industry (Eriksson, 2013) in which the temporariness and one-off nature of projects lead to concerns about construction cost, including that of blockchain systems (Huang et al., 2015). In the big data era, the amount of data transactions submitted to blockchain is growing exponentially (Xu et al., 2023). In practice, construction practitioners may often upload construction data transactions to a blockchain system for safeguarding (Lu et al., 2021c). Construction projects are also subject to opportunistic behavior arising from their inherent uncertainties, so any project-oriented blockchain solution must improve information transparency (Yoon and Pishdad-Bozorgi, 2022). Without a means of measuring data redundancy level, the storage overhead and impacts of blockchain systems on such improvements cannot be accurately assessed in construction projects.

This research aims to model data redundancy in blockchain systems in a construction project setting. It does so by conducting a series of experiments on a Hyperledger Fabric blockchain system built in a laboratory environment. The research has two specific objectives. Firstly, it aims to analyze the relationships between the overall data volume (v) of a blockchain system and the size (s) of the construction files to be uploaded, the block size (b), the number of peer nodes (n) (also called consensus nodes, or consensus peers) in the network, the way the nodes are dispersed in different construction organizations (g), and the frequency (f) of blockchain operations in a construction project; all of which are considered to have an impact on the overall data size of a blockchain system. The second objective is to develop a model for construction practitioners to predict data redundancy in a blockchain system. The rest of this paper is organized as follows. The next section reviews data redundancy, blockchain basics, and its subject areas in construction. The following section proposes our research hypotheses, and the subsequent section describes the research methods. After that, the next section presents the data analysis, findings, and results. Finally, the discussion and conclusions are presented.

2 Literature review

2.1 Data redundancy

Data redundancy can be measured as the ratio of the total data volume of a ledger system (or say, a database system) to the original file size uploaded to the system (Huang et al., 2015). For example, when a file with an initial data volume (s) is uploaded to a database system, its data volume is increased to a certain size (P) due to the redundant process of the system. The data redundancy (R) can be calculated as $R = P/s$.

Data redundancy should not be treated as entirely negative and maybe unintentional or intentional. On one hand, unintentional data redundancy occurs owing to inefficient coding or overcomplicated data storing processes (Fan et al., 2018) that increase the size and complexity of the database in terms of data storage, writing, reading, transmission, or processing. Organizations may have to spend extra time and resources to store, process, and maintain such databases (Najafabadi and Azar, 2019). Data redundancy can also lead to data inconsistency, with the same data existing in multiple locations in different formats (Bakr and Lee, 2017). This may result in unreliable or meaningless data.

On the other hand, deliberate data redundancy strategies are often adopted. When data is replicated in multiple places, organizations can benefit from fast access and updates as the data may be available in a closer geographic location (Verma and Singh, 2018). In addition, storing the same data in two or more places can protect an organization from a single point of failure (SPOF) problem caused by events such as cyberattacks or electricity failure (Chervyakov et al., 2019). Organizations can also use data redundancy to check the accuracy and completeness of data by comparison so that relevant parties (e.g., customers, vendors) can have high data reliability (Verma and Singh, 2018).

In general, distributed systems adopt two types of data redundancy: replication and erasure code (Weatherspoon and Kubiatowicz, 2002). In replication, the distributed system replicates each data block into n copies and then distributes them to different network peers (Huang et al., 2015). Even if the $n-1$ copies are damaged, users can still recover the data. Erasure code transforms a message of k symbols into a longer message (code word) with n symbols such that the original message can be recovered from a subset of the n symbols. Some systems adopt both redundancy schemes simultaneously.

2.2 Blockchain systems

While a database usually organizes data into tables, a blockchain structures data into blocks that are strung together (Beck et al., 2017). Figure 1 shows a typical blockchain. The structure of a block consists of three sections, from top to bottom: block header, block data, and block metadata (Lu et al., 2021b). Block metadata contains the certificate and signature of the block creator. Block data contains a list of transactions arranged in order (Li et al., 2021). A transaction is the smallest unit of a work process involving one or more sequences of actions, e.g., revising, adding, or deleting something in a file (International Organization for Standardization, 2020). The block header comprises three fields written when a block is created (Hyperledger, 2020). The block number is an integer starting at 0 (the first, or genesis, block) and increasing by one for every new block appended to the blockchain. It also includes a hash value of all the transactions contained in the current block. This hash value is unique for each input, so if someone alters the transactions, the corresponding hash value will also change (Wu et al., 2022a). Each block also contains the hash value from the previous block header. Blocks are thereby linked to form an immutable ledger (Perera et al., 2020).

To form a blockchain system, blockchain copies need to be stored by different peers connected in a network, as shown in Figure 2. Blockchain systems may, according to network centralization levels, be public, private, or consortium (Wu et al., 2022a). Specifically, public blockchain is accessible to the public for use, while private blockchain has just one owner organization, and only preauthorized participants can
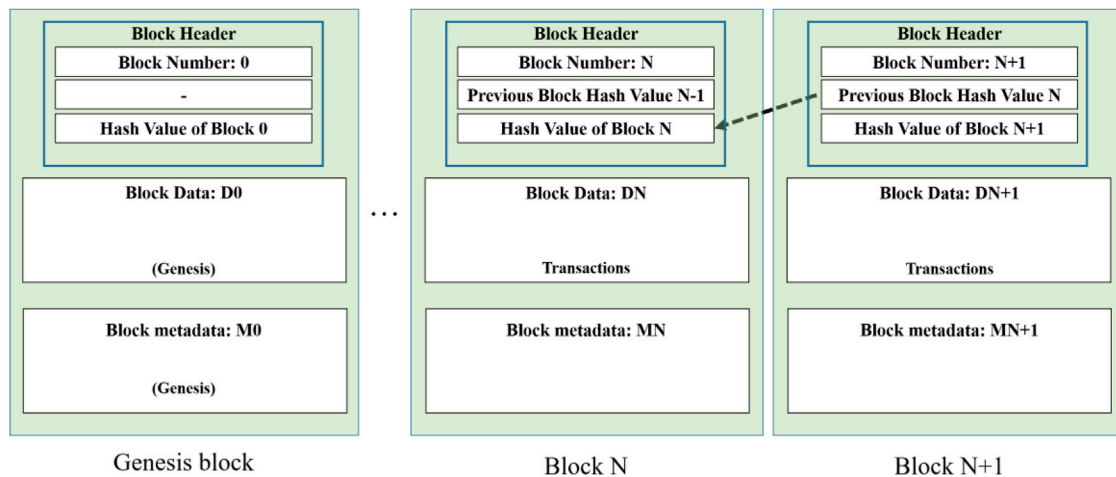
FIGURE 1
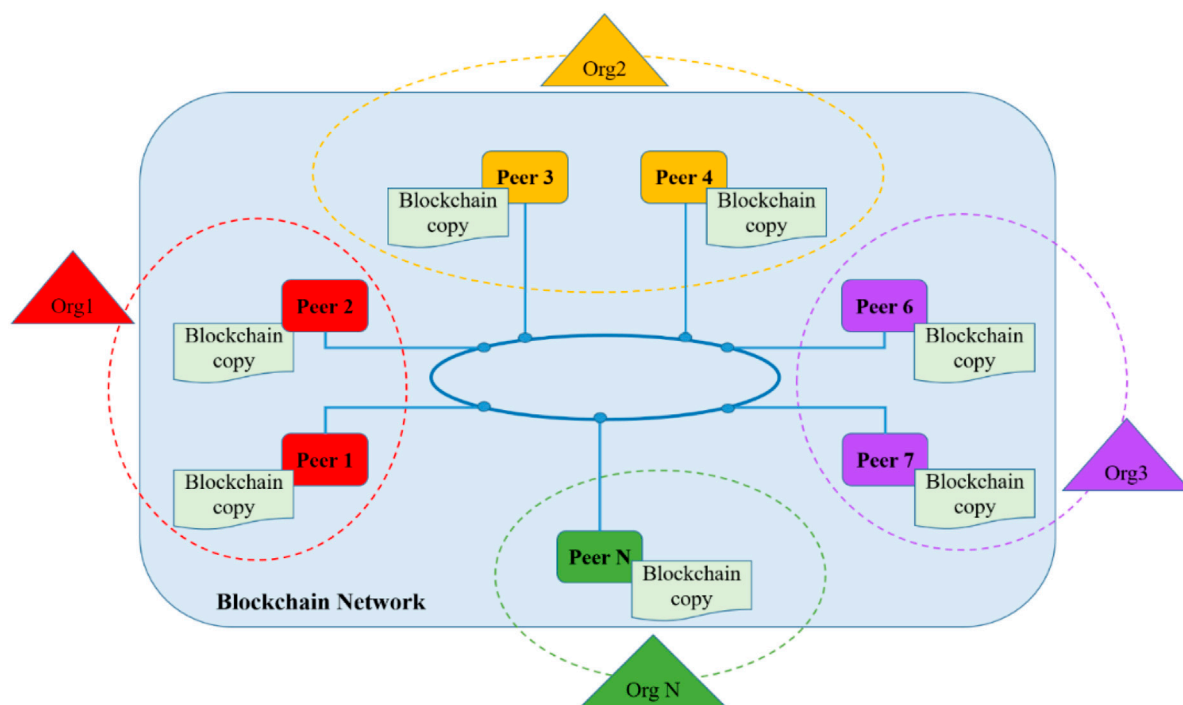An example of a blockchain (adapted from Li et al., 2021, with permission from Elsevier).

FIGURE 2
The physical distribution of a blockchain system—different peers and constituent project organizations.

perform particular activities (Lu et al., 2021c). Besides, consortium blockchain operates under a selected set of organizations (Hijazi et al., 2021). A distributed network like blockchain can help to avoid the SPOF problem, where the failure of one component of a system will make the entire system unable to perform its primary functions (International Organization for Standardization, 2020). The SPOF could be caused by electricity failure, Internet connection disruption, cyberattack, bad actors, or *force majeure*.

Blockchain operation involves several steps, as illustrated in Figure 3. First, a blockchain user proposes a transaction through an application (see Figure 3A). This transaction is then broadcast to the whole network for validation (see Figure 3B). This involves, among other things, checking whether the proposer is appropriate. Upon consensus of the network, validation is achieved, and the hashed transaction is included in a 'block', creating a tamper-proof record (see Figure 3C) (Lu et al., 2021b). Next, the block is sent to peers in the network (see Figure 3D) so that they can approve the order and

FIGURE 3
Blockchain in operation: **(A)** a transaction; **(B)** network validation; **(C)** block; **(D)** block publication; **(E)** correctness check; **(F)** block-adding; **(G)** transaction update; **(H)** transaction completion.
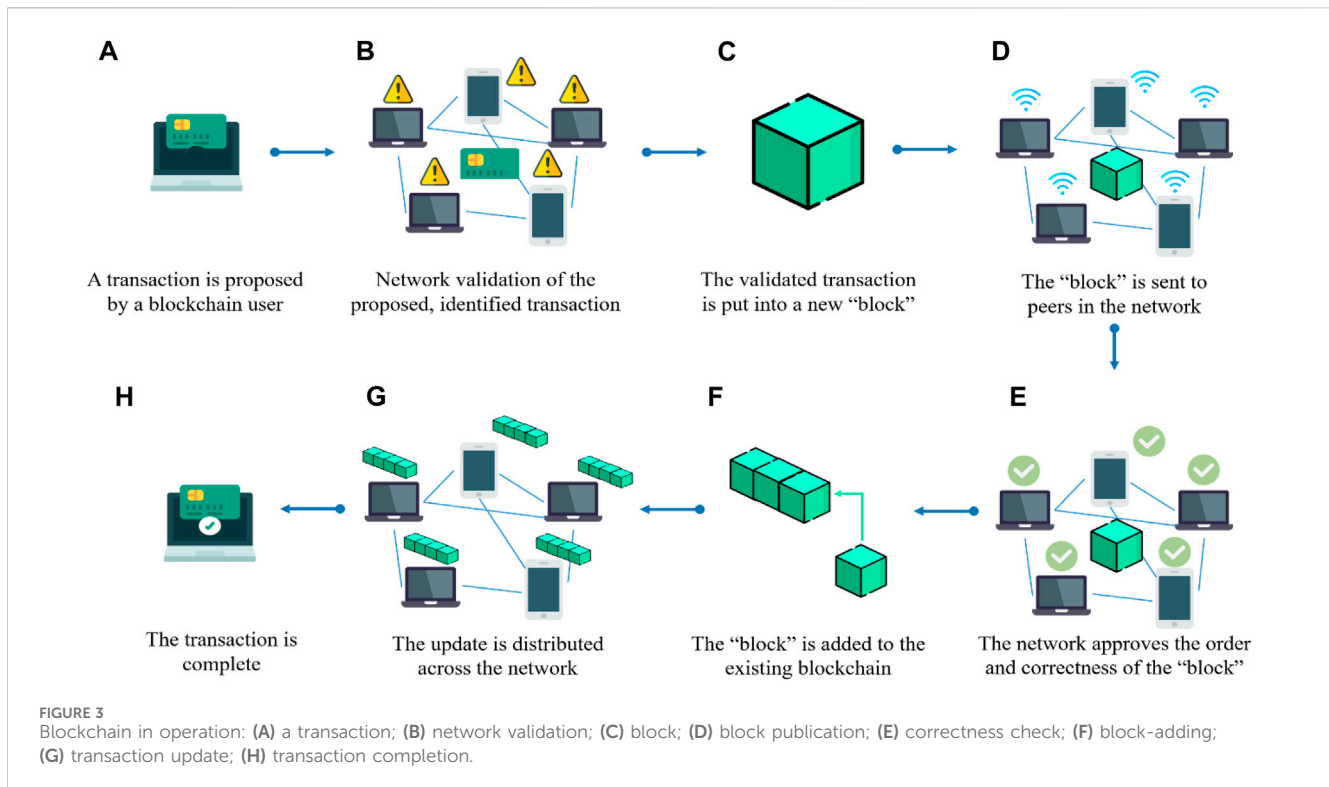
correctness of the block (see Figure 3E). After that, the blockchain is updated with the new block (see Figure 3F), and in this way, the whole distributed peer network has its up-to-date copy of the blockchain (see Figure 3G). Finally, the operation is concluded, and the completion notification is sent to the network (see Figure 3H).

Several factors can influence the data redundancy of blockchain. Firstly, file size plays a crucial role as larger files can impact storage and retrieval efficiency (Xue and Lu, 2020). Secondly, the number of peers and organizations involved in the blockchain network can affect data redundancy. More peers and organizations can enhance data availability and collaboration but may also increase network traffic (Tao et al., 2021). Additionally, block size, which determines the number of transactions included in a block, can impact data redundancy (Perera et al., 2020). A larger block size allows more transactions but can slow down the network. Lastly, the operation frequency of transactions in a project can influence data redundancy. Higher transaction frequency can increase network traffic and reduce performance but may improve data transparency and collaboration (Zhong et al., 2023). These factors must be carefully considered to optimize data redundancy and ensure efficient operations within the blockchain networks.

There are potential tools can address data redundancy in blockchain networks. One such tool is InterPlanetary File System (IPFS) (Tao et al., 2021), which enables decentralized storage and distribution of large files associated with blockchain transactions. IPFS reduces data redundancy by eliminating the need for centralized storage infrastructure and improving data availability. Another tool is Distributed Hash Tables (DHT), which stores and retrieves data in a peer-to-peer network (Byers et al., 2003), distributing blockchain data across multiple nodes to enhance

availability and reduce redundancy. Data compression techniques can also be employed to reduce the size of data (Jayasankar et al., 2021), reducing data redundancy and improving storage efficiency for blockchain. This can help optimize the use of storage resources and improve the overall performance of the blockchain network.

2.3 Blockchain in construction

In construction, the blockchain literature has grown dramatically in the past few years. Blockchain was first discussed in the context of Building Information Modelling (BIM) (Turk and Klinc, 2017) and smart cities (Coyne and Onabolu, 2017). Researchers have since provided sketches of potential use cases for blockchain in construction. For example, Das et al. (2021) point out the promise of blockchain as a complementary technology to BIM and the Internet of Things (IoT), which are constrained by trust and liability issues. Ye et al. (2018) propose the use of blockchain to store IoT-generated data in a transparent and secure environment and BIM as a tool for digitally processing construction project data. In a 2018 report, meanwhile, ICE comprehensively envisions the applications of blockchain in construction (Penzes et al., 2018).

The construction industry has become increasingly interested in blockchain technology (Xu et al., 2023). It is exciting to observe in construction the emergence of diverse and innovative blockchain applications, e.g., in operation management (Ye et al., 2018), decentralized organizations (Perera et al., 2020), smart cities (Chen et al., 2020), trust-building in the supply chain (Qian and Papadonikolaki, 2021), construction supply chain traceability (Lu et al., 2021a), procurement (Tezel et al., 2021), secure BIM and blockchain-based collaborative design (Tao et al., 2021), carbon

FIGURE 4
A conceptual research framework.
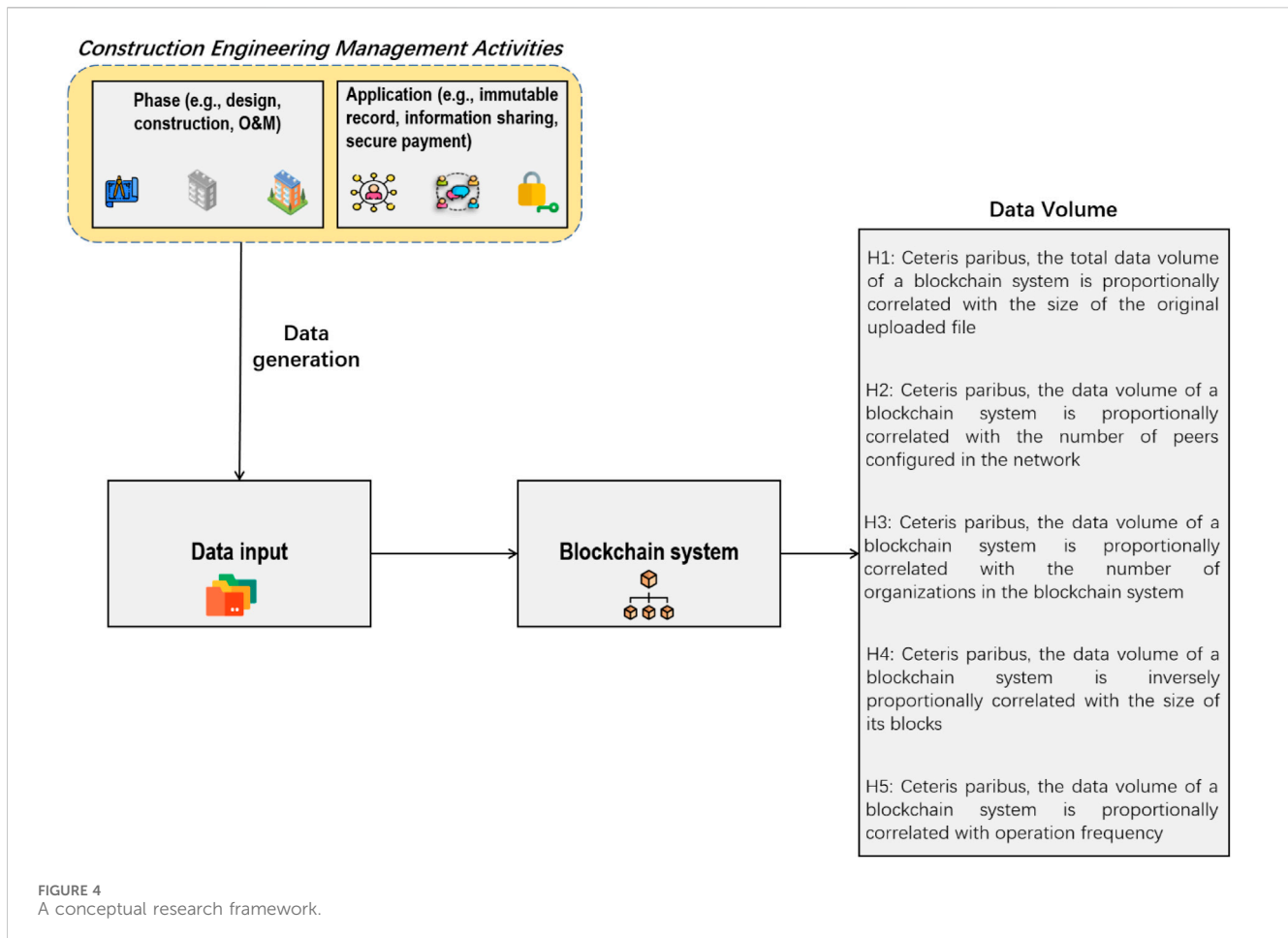
footprint (Rodrigo et al., 2021), privacy protection of modular housing production (Li et al., 2021), and confidentiality-minded design in digital collaborative environments (Tao et al., 2022). Other applications include supervision of offsite modular housing production (Wu et al., 2022a), accurate information sharing (Wu et al., 2022b), decentralized tendering (Ahmadisheykhsarmast et al., 2023), combined applications with internet of things (IoT), BIM, and edge computing (Zhong et al., 2023), on-site activity management (Wu et al., 2023), collaboration for fit-out operations (Jiang et al., 2023), and among many others. These studies are of great significance to blockchain research in the construction industry.

However, little research, if any, has considered the data redundancy of blockchain systems in construction. Blockchain systems are not cost-free. They involve capital investment, e.g., to develop the system, and operational cost, e.g., to maintain and upgrade it. Servers and storage are hosted by different project stakeholders and then dissolved when a project is completed, and this contributes a significant portion of the overall blockchain cost. Understanding the level of data redundancy in a blockchain, therefore, is a research topic that is of both academic and practical value.

3 Research hypotheses

In order to model data redundancy in blockchain systems in a construction project setting, a conceptual framework is proposed by the research team, as shown in Figure 4. Data inputs are records related to construction activities that are submitted to blockchain systems, including design, quality, progress, and safety records. Blockchain systems, built on various platforms, can record and share data (i.e., offering data transparency and traceability) to support different applications in construction. Five hypotheses are proposed to analyse the overall data volume of a blockchain system. These hypotheses are explained as follows.

First, it is intuitive that if all other things (e.g., node numbers, operation frequency) are equal, the overall data volume (v) of a blockchain system (see Figure 3G) will increase in line with the size (s) of the construction file to be uploaded (see Figure 3A). Therefore, Hypothesis one is derived:

H1: Ceteris paribus, the total data volume of a blockchain system is proportionally correlated with the size of the original uploaded file.

Further, if all other conditions (e.g., the files to be placed in the blockchain system, operation frequency) remain the same, it is legitimate to assume that the overall data volume (v) of a blockchain system will increase in line with the number of peers (n), as the chain of blocks will be agreed by all these peers and reside on them (see Figure 3G). Therefore, Hypothesis two is proposed:

H2: Ceteris paribus, the data volume of a blockchain system is proportionally correlated with the number of peers configured in the network.

Focusing on consortium systems, the peers connected in a network could be geographically dispersed in different project-

based construction organizations (*g*) (see Figure 2). Sometimes, this is because of the business structure of the organizations, e.g., the Digital Currency Electronic Payment (DC/EP) system in China's banking system (Lu et al., 2021b). In other cases, it is to prevent the SPOF problem by placing the nodes in different locations and forms (e.g., servers or cloud services). It is hypothesized that:

H3: Ceteris paribus, the data volume of a blockchain system is proportionally correlated with the number of organizations in the blockchain system.

An uploaded file will normally be divided into different transactions and blocks, whose sizes vary in different blockchain systems. Then, the transactions are added with headers (e.g., hash values), and other information to form blocks (*b*) (see Figure 1). The bigger the block size, the less information is added to the whole blockchain system. It is therefore hypothesized that:

H4: Ceteris paribus, the data volume of a blockchain system is inversely proportionally correlated with the size of its blocks.

In real-life construction practice, new transactions are frequently generated (e.g., to revise a file), agreed, and safeguarded in a blockchain system. Unlike a traditional server system that replaces the old transactions with new ones, a blockchain system will add new blockchains, as the old ones are immutable (see this transaction cycle in Figure 3). Therefore, the higher the frequency (*f*) of such blockchain operations, the higher the volume (*v*) of the data in the blockchain system. It is hypothesized that:

H5: Ceteris paribus, the data volume of a blockchain system is proportionally correlated with operation frequency.

4 Research methods

We employ a scientific control experiment to test our hypotheses. This is an experimental approach designed to minimize the effects of variables other than the independent variable (i.e., confounding variables) (Amir et al., 2018). This increases the reliability of the results, often through a comparison between control measurements (Amir et al., 2018). Based on this understanding, we develop a model to help construction practitioners predict data redundancy and plan a blockchain system.

4.1 Experiment setting-up

Three types of blockchain are available: public, private, and consortium (Perera et al., 2020). In a public blockchain, control is distributed among members of the public that join in its operation. Many cryptocurrencies (e.g., Bitcoin) use this type of blockchain. Private blockchain has just one owner organization, and only preauthorized peers can perform certain activities (International Organization for Standardization, 2020). Due to this restrictive nature, private blockchain tends to have better privacy and efficiency than public blockchain (Lu et al., 2021b). However, this controlled environment may reduce transparency and resistance to hackers (Perera et al., 2020). Consortium blockchain operates under a selected set of organizations, and only preauthorized peers are

TABLE 1 Summary of the test construction files.

File name	Type of file	Number of uploads (*f*)*
Test.gif	.gif	10
Test.docx	.docx	10
Test.jpg	.jpg	10
Test.pdf	.pdf	10

Note. One hundred uploads for Experiment 5.

allowed to conduct certain activities (Du et al., 2020). It has many of the advantages of the private blockchain while reducing the counterparty risk of private blockchain because it has more than one organization to manage the network.

There are different blockchain platforms based on which different blockchain systems can be developed. Lu et al. (2021c) compare the main blockchain platforms Bitcoin, Ethereum, Hyperledger Fabric, and R3 Corda in terms of their focus domain. Hyperledger Fabric is a generic, open-source platform for any industry (Perera et al., 2020) that can be used to develop both private and consortium blockchain systems, and that allows for transaction data to be accessed only by permissioned persons. Therefore, Hyperledger Fabric is attractive to construction organizations (Nawari and Ravindran, 2019).

We build a blockchain system based on Hyperledger Fabric as the testbed to test the hypotheses in five controlled experiments that involve changing file size, numbers of peers organized in a local area network (LAN), numbers of construction organizations, block size, and operation frequency. We organize the peers in a physical server only. In real-life blockchain systems, such peers are normally dispersed in different servers, and in different construction organizations. Each peer is initiated as a Docker container and then connected to the Fabric network using Docker Swarm. Docker swarm is a container orchestration tool, meaning that it allows the user to manage multiple containers deployed across multiple host machines. The "Proof of Stakeholder" consensus mechanism is adopted, which means the monotone logical expression 'AND' is used to specify that every endorsement peer from a construction organization is required to endorse transactions *via* the decentralized network. All the containers corresponding to each peer are run on the server. The physical server has 16 central processing units (CPUs) (Intel Xeon Silver 4110 @ 2.10 GHz) with 32GB RAM. The experiments are run on Ubuntu 20.10 long-term support (LTS) with the installed Hyperledger Fabric (version 1.4.2).

4.2 Experiment process

Step 1: In the experiments, we assume that a participant is asked to submit a file to the blockchain system as a typical blockchain operation in a construction project. Table 1 summarizes the test construction files used in this study. We select four types (i.e., gif, docx, jpg, pdf), each with 10 different sizes, as it is unclear whether different construction file types have different effects on file redundancy in a blockchain system. File content could include progress payment transactions, construction legal contracts,

inspection records, design ideas, or any meaningful construction information. The files are read into the network using the Base64 algorithm (Rahim et al., 2018). The default block interval, i.e., the amount of time that has elapsed since the last block (Hyperledger, 2020), is controlled as constant (1 s) in the experiments. In addition, with the exception of Experiment 4, a default block size of 98 MB is adopted. In each experiment described below, the data volume of the blockchain system is recorded by the research team after the completion of each operation.

Step 2: Hypothesis **H1** is tested by Experiment 1, which aims to understand data redundancy in a blockchain system by examining whether the total data volume of a blockchain system is proportionally correlated with the size of the original uploaded construction file. In this experiment, each test file is uploaded to the blockchain system 10 times for endorsement. During this process, the number of endorsement peers is controlled at 7, and all are under the same project-based construction organization.

Step 3: Hypothesis **H2** is tested by Experiment 2, which aims to understand data redundancy by measuring the impact of the number of peers on the overall data volume of the blockchain system. In this experiment, each test file is uploaded to the blockchain system 10 times for endorsement. The number of endorsement peers, all under the same organization, is adjusted from 1 to 10 correspondingly.

Step 4: Hypothesis **H3** is tested by Experiment 3, which aims to understand data redundancy by investigating the impact of the number of construction organizations on the overall data volume of the blockchain system. In this experiment, each test file is uploaded to the blockchain network 10 times for endorsement. During this process, the number of peers is controlled at 10. The number of construction organizations is adjusted from 1 to 10, therefore each construction organization contains 10 to one peer(s), respectively.

Step 5: Hypothesis **H4** is tested by Experiment 4, which aims to understand data redundancy by investigating the relationship between the overall data volume of a blockchain system and the block size. In this experiment, each test file is uploaded to the blockchain system 10 times for endorsement. During this process, the number of endorsement peers is controlled at 10, and all under the same organization. The block size is adjusted in 2 MB increments from four to 22 MB.

Step 6: Hypothesis **H5** is tested by Experiment 5, which aims to understand data redundancy by investigating the relationship between the overall data volume of a blockchain system and the operation frequency (file uploads). In this experiment, each test file is uploaded to the system 100 times for endorsement. During this process, the number of peers is controlled at 10 and peers are under the same organization.

5 Data analyses, findings, and results

5.1 Data redundancy in a construction blockchain system

The results of testing these five hypotheses are presented in Sections 5.1.1 to 5.1.5 accordingly. Section 5.2 then presents the data

redundancy model of the Hyperledger Fabric in construction projects. Lastly, the model validation results are given in Section 5.3.

5.1.1 Impact of the file size

The experimental results of Experiment 1 are summarized in this section and in Table 2. When corresponding variables (n, g, f, b) are under control, the total data volume of a blockchain system is proportionally correlated with the size (s) of the original uploaded construction file. However, as the size of the original upload file increases, the data redundancy decreases. Data redundancy occurs after converting original files to transaction proposals and endorsing transactions. Also, the ordering service contributes to data redundancy due to the introduction of block header and metadata. File type does not affect the results of the experiment.

The data volume added to the blockchain system in Experiment 1 is shown in Figure 5. It can be seen that the data volume added to a blockchain system grows linearly in line with the original file size (o) when other variables remain unchanged. Notice that the R-squared value is 1, which is a good fit of the line to the data. These results confirm that Hypothesis one is established.

5.1.2 Impact of the number of peers

Analysis results of Experiment 2 are summarized in Table 3. When corresponding variables (g, f, b, s) are under control, the total data volume added to a blockchain system rises with the increase in number of peers configured in the network (n). In addition, the data redundancy brought to the blockchain system increases when the number of endorsement peers configured in the Fabric network increases when corresponding variables (g, f, b, o) are controlled.

Figure 6 shows the data volume added to the blockchain system in Experiment 2. The data volume added to a blockchain system grows in a polynomial line with the number of peers (n) configured in the network. The R-squared value is at 1, which is a good fit of the polynomial line to the data. A polynomial trend line is the best fit because every time a new endorsement peer is added, a 1003-byte endorsement (not including data volume caused by adding a blockchain copy of an endorsement peer) is also added to the blockchain system. To sum up, the experimental results confirm that Hypothesis two is true.

5.1.3 Impact of the number of construction organizations

The analysis results of Experiment 3 are summarized in Table 4. When corresponding variables (n, f, b, s) are under control, the total data volume of a blockchain system stays constant with increases in the number of construction organizations (g) configured in the network. In addition, the data redundancy brought to the blockchain system remains the same when the number of construction organizations configured in the Fabric network increases while corresponding variables (n, f, b, s) are controlled.

Figure 7 shows the data volume added to the blockchain system in Experiment 3. It can be seen that the data volume added to a blockchain system does not grow or reduce when the number of construction organizations (g) configured in the network increases. Thus, the number of organizations configured in the network does not have any impact on the data volume of the blockchain. To sum up, the experimental results confirm that Hypothesis three is false.

TABLE 2 Total data volume added to the blockchain and data redundancy in Experiment 1.

File name	Original size (s) [KBs]	Average data volume added after each upload (p_{ave}) [KBs]	Total data volume added after ten uploads (p_{total}) [KBs]	Data redundancy (p_{total})/($s \times f$)
Test.gif-1	9.70	230	2,298	23.703
Test.gif-2	19.39	388	3,875	19.984
Test.gif-3	29.09	546	5,459	18.769
...	...	...	...	...
Test.docx-1	48.9	870	8703	17.796
Test.docx-2	97.8	1,669	16688	17.062
Test.docx-3	146.7	2,468	24678	16.822
...	...	...	...	...
Test.jpg-1	394.6	6517	65167	16.515
Test.jpg-2	789.2	12966	129663	16.430
Test.jpg-3	1,183.8	19414	194142	16.400
...	...	...	...	...
Test.pdf-1	1,567.0	25666	256664	16.379
Test.pdf-2	3,134.1	51281	512811	16.363
Test.pdf-3	4701.1	76886	768863	16.355
...	...	...	...	...

Note: $n = 7$, $g = 1$, $f = 10$, and $b = 98$ MBs. Please see Appendix A for full information.

FIGURE 5
Data volume added to the blockchain system in Experiment 1: **(A)** .gif file; **(B)** .docx file; **(C)** .jpg file; **(D)** .pdf file.
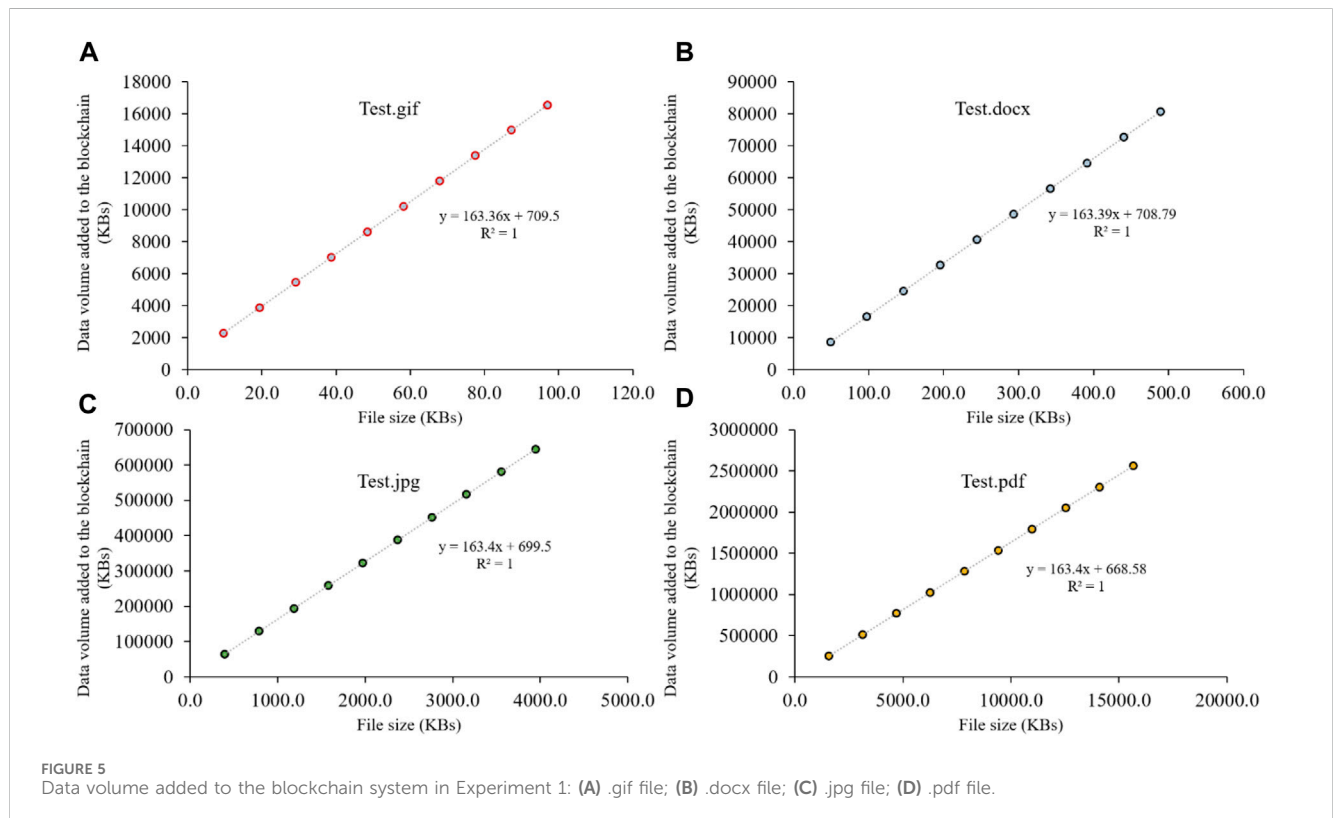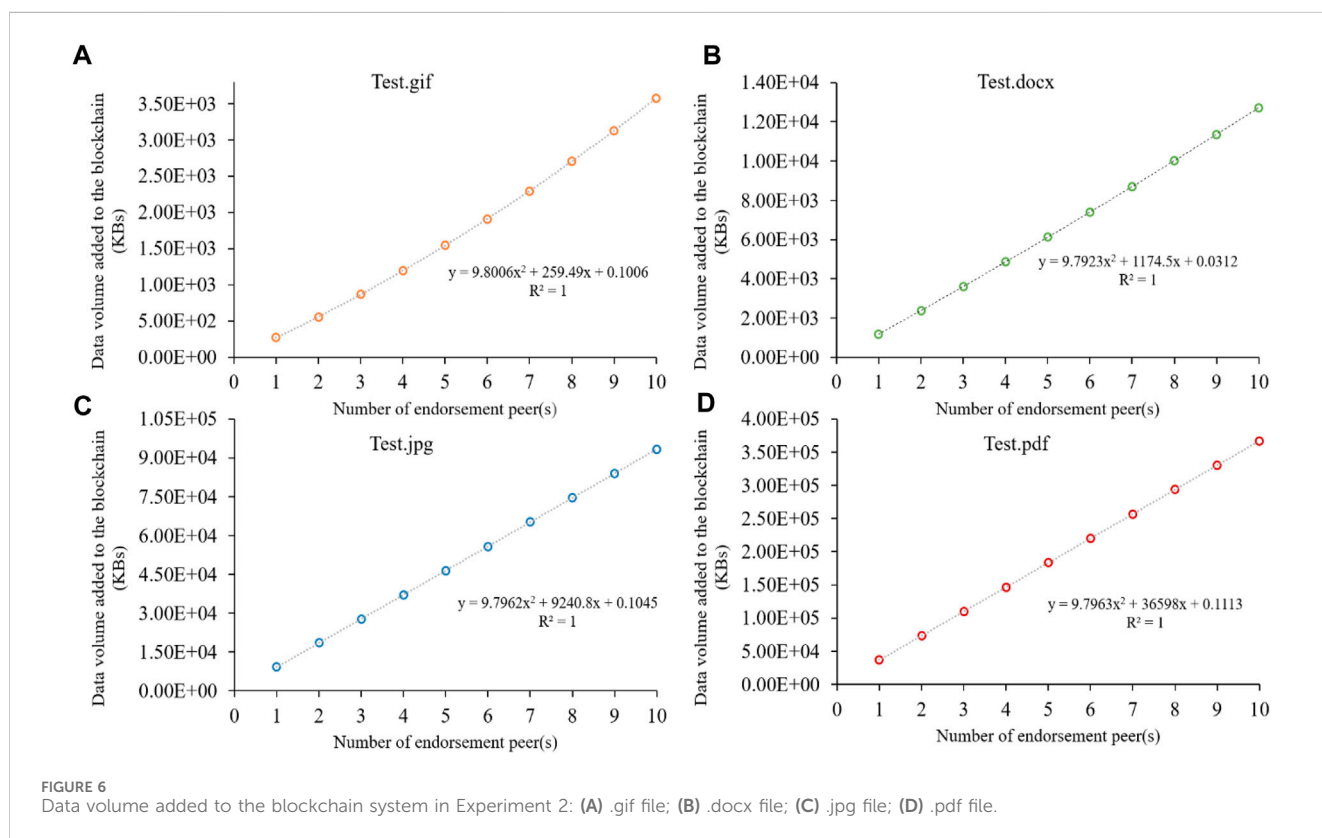
TABLE 3 Total data volume added to the blockchain and data redundancy in Experiment 2.

Number of endorsement peers (n)	Total data volume added after ten uploads (p_{total}) [KBs] (data redundancy) = (p_{total})/($s \times f$)			
	Test.gif	Test.docx	Test.jpg	Test.pdf
1	269 (2.778)	1,184 (2.422)	9251 (2.344)	36607 (2.336)
2	558 (5.760)	2,388 (4.884)	18521 (4.694)	73234 (4.673)
3	867 (8.940)	3,612 (7.385)	27811 (7.048)	109881 (7.012)
4	1,195 (12.324)	44855 (9.927)	37120 (9.407)	146547 (9.352)
5	1,543 (15.911)	6117 (12.509)	46449 (11.771)	183233 (11.693)
6	1910 (19.699)	7399 (15.131)	55797 (14.140)	219938 (14.035)
7	2,297 (23.689)	8701 (17.793)	65162 (16.514)	256663 (16.379)
8	2,703 (27.881)	10023 (20.495)	74554 (18.893)	293407 (18.724)
9	3,129 (32.275)	11364 (23.237)	83961 (21.277)	330171 (21.070)
10	3,575 (36.876)	12724 (26.019)	93388 (23.666)	366955 (23.417)

Note: g = 1, f = 10, b = 98 MBs, s = 9.7 KBs, or 48.9 KBs, or 394.6 KBs, or 1,567 KBs.

FIGURE 6
Data volume added to the blockchain system in Experiment 2: **(A)** .gif file; **(B)** .docx file; **(C)** .jpg file; **(D)** .pdf file.

5.1.4 Impact of block size

Analysis results of Experiment 4 are summarized in Table 5. When corresponding variables (n, f, g, s) are under control, the total data volume of a blockchain system stays constant with increases in the size of the block (b). In addition, the data redundancy brought to the blockchain system remains the same when the size of the block in the Fabric network increases while corresponding variables (n, f, g, s) are controlled.

The data volume added to the blockchain system in Experiment 4 is recorded, as shown in Figure 8. It can be seen that the data volume added to a blockchain system does not grow or reduce when the size of the block (b) configured in the network increases. Thus, the size of the block does not have any impact on the data volume of the blockchain. These experimental results confirm that Hypothesis four is false.

TABLE 4 Total data volume added to the blockchain and data redundancy in Experiment 3.

Number of organizations (g)	Total data volume added after ten uploads (p_{total}) [KBs] (data redundancy) $=(p_{total})/(s{\times}f)$			
	Test.gif	Test.docx	Test.jpg	Test.pdf
1	357.50 (3.687)	1,272.46 (2.602)	9338.85 (2.367)	36695.56 (2.342)
2	357.52 (3.688)	1,272.43 (2.602)	9338.85 (2.367)	36695.52 (2.342)
3	357.52 (3.688)	1,272.45 (2.602)	9338.84 (2.367)	36695.55 (2.342)
4	357.52 (3.688)	1,272.43 (2.602)	9338.86 (2.367)	36695.56 (2.342)
5	357.51 (3.687)	1,272.44 (2.602)	9338.86 (2.367)	36695.56 (2.342)
6	357.52 (3.688)	1,272.48 (2.602)	9338.86 (2.367)	36695.54 (2.342)
7	357.54 (3.688)	1,272.48 (2.602)	9338.86 (2.367)	36695.56 (2.342)
8	357.52 (3.688)	1,272.49 (2.602)	9338.85 (2.367)	36695.55 (2.342)
9	357.51 (3.687)	1,272.46 (2.602)	9338.85 (2.367)	36695.58 (2.342)
10	357.56 (3.688)	1,272.52 (2.602)	9338.88 (2.367)	36695.62 (2.342)

Note: $n = 10$, $f = 10$, $b = 98$ MBs, $s = 9.7$ KBs, or 48.9 KBs, or 394.6 KBs, or 1,567 KBs.
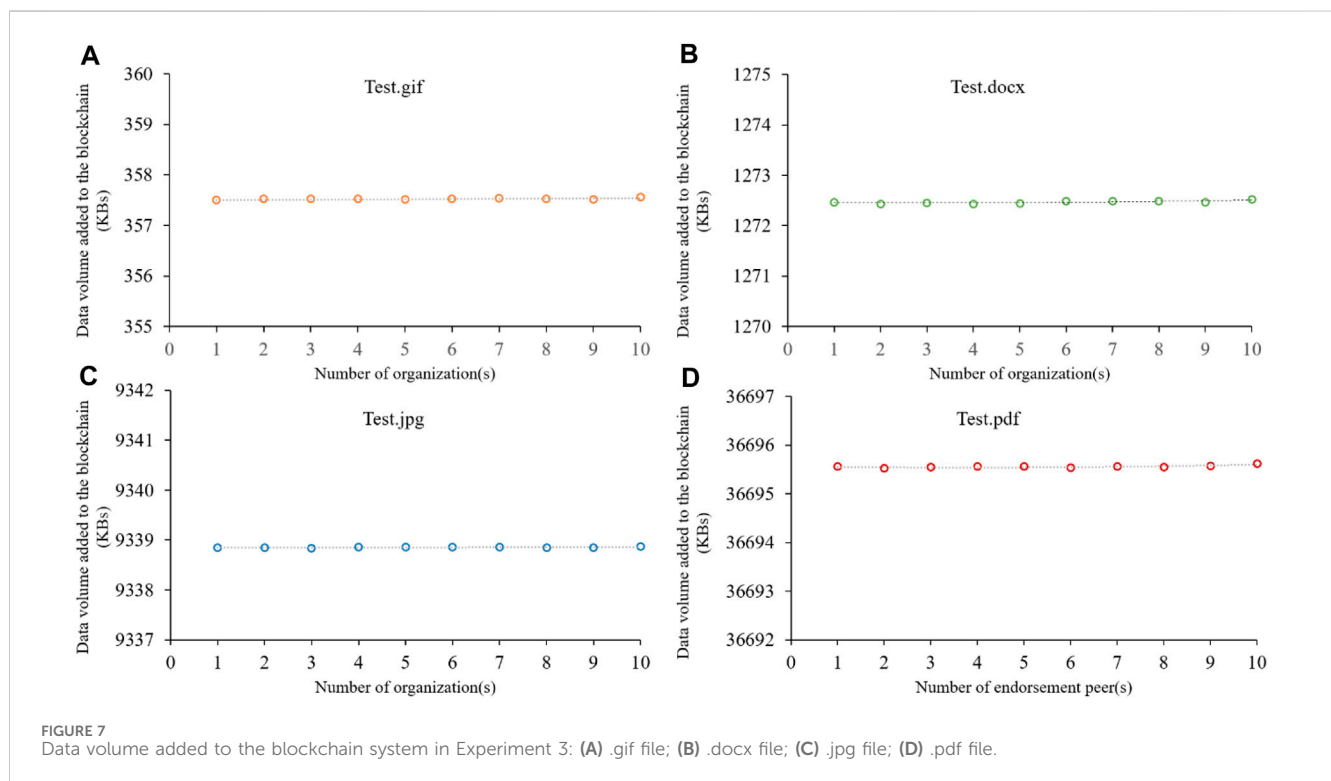

FIGURE 7
Data volume added to the blockchain system in Experiment 3: **(A)** .gif file; **(B)** .docx file; **(C)** .jpg file; **(D)** .pdf file.

5.1.5 Impact of the operation frequency of transactions in a construction project

The analysis results of Experiment 5 are summarized in Table 6. When corresponding variables (n, g, s, b) are under control, the total data volume added to a blockchain system is proportionally correlated with the operation frequency (f).
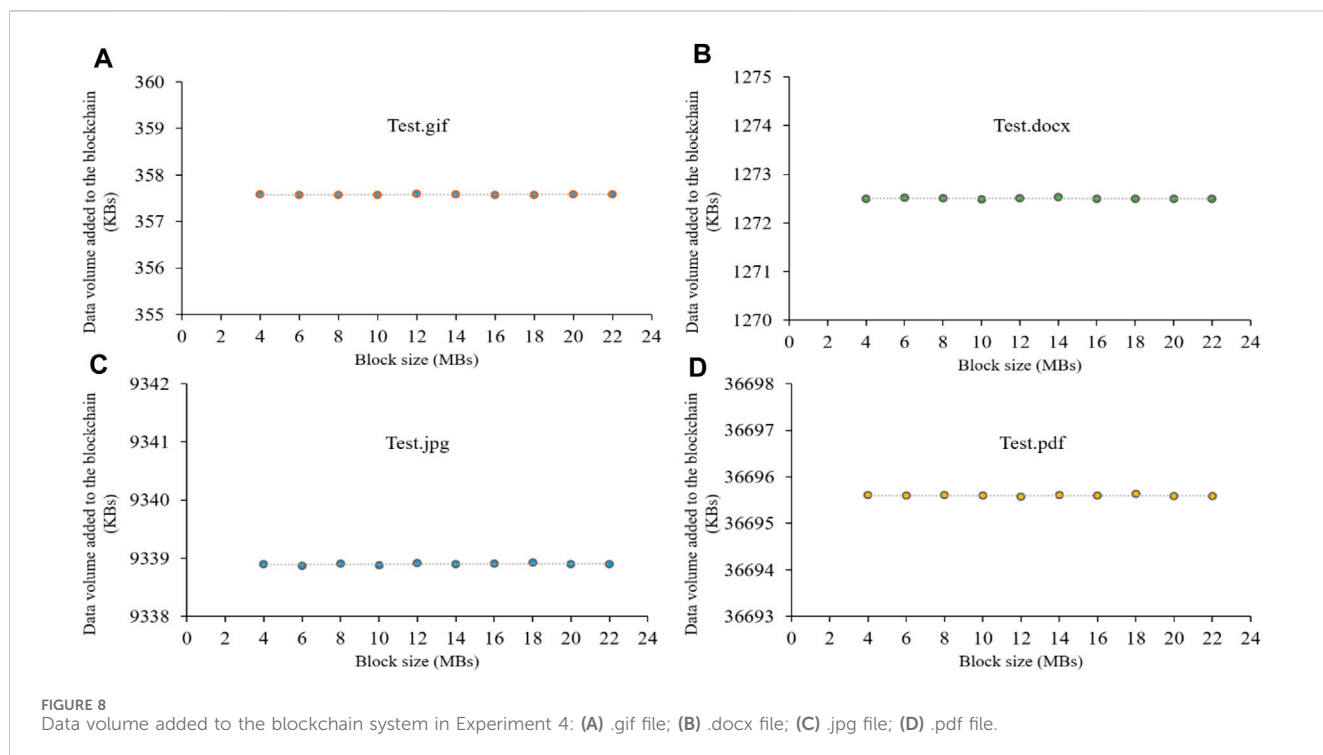
However, as the operation frequency increases, the data redundancy remains the same.

Figure 9 shows that the data volume added to a blockchain system grows linearly in line with the operation frequency (f) when other variables remain unchanged. Notice that the R-squared value is 1, which is a good fit of the line to the

TABLE 5 Total data volume added to the blockchain and data redundancy in Experiment 4.

Block size (b) [bytes]	Total data volume added after ten uploads (p_{total}) [KBs] (data redundancy) $=(p_{total})/(s \times f)$			
	Test.gif	Test.docx	Test.jpg	Test.pdf
4	357.59 (3.688)	1,272.50 (2.602)	9338.90 (2.367)	36695.61 (2.342)
6	357.57 (3.688)	1,272.52 (2.602)	9338.87 (2.367)	36695.60 (2.342)
8	357.57 (3.688)	1,272.51 (2.602)	9338.91 (2.367)	36695.61 (2.342)
10	357.57 (3.688)	1,272.49 (2.602)	9338.88 (2.367)	36695.60 (2.342)
12	357.60 (3.688)	1,272.51 (2.602)	9338.92 (2.367)	36695.58 (2.342)
14	357.59 (3.688)	1,272.53 (2.602)	9338.90 (2.367)	36695.62 (2.342)
16	357.57 (3.688)	1,272.50 (2.602)	9338.91 (2.367)	36695.60 (2.342)
18	357.58 (3.688)	1,272.50 (2.602)	9338.93 (2.367)	36695.63 (2.342)
20	357.59 (3.688)	1,272.50 (2.602)	9338.90 (2.367)	36695.59 (2.342)
22	357.59 (3.688)	1,272.50 (2.602)	9338.90 (2.367)	36695.59 (2.342)

Note: $g = 1$, $n = 10$, $f = 10$, $s = 9.7$ KBs, or 48.9 KBs, or 394.6 KBs, or 1,567 KBs

FIGURE 8
Data volume added to the blockchain system in Experiment 4: **(A)** .gif file; **(B)** .docx file; **(C)** .jpg file; **(D)** .pdf file.

data. These experimental results confirm that Hypothesis five is established.

5.2 Data redundancy model of the Hyperledger Fabric in construction projects

The data volume added to a blockchain system due to a construction transaction (e.g., file upload) is determined by four parts (Parts I to IV), as shown in Figure 10. Based on Experiment 1, it is found that the Part I (Value) contained in the proposal is directly affected by the original construction file size (s). Since we used the Base64 algorithm to read the test construction files during the experiments, Eq. 1 can then be used to calculate Part I (P_I).

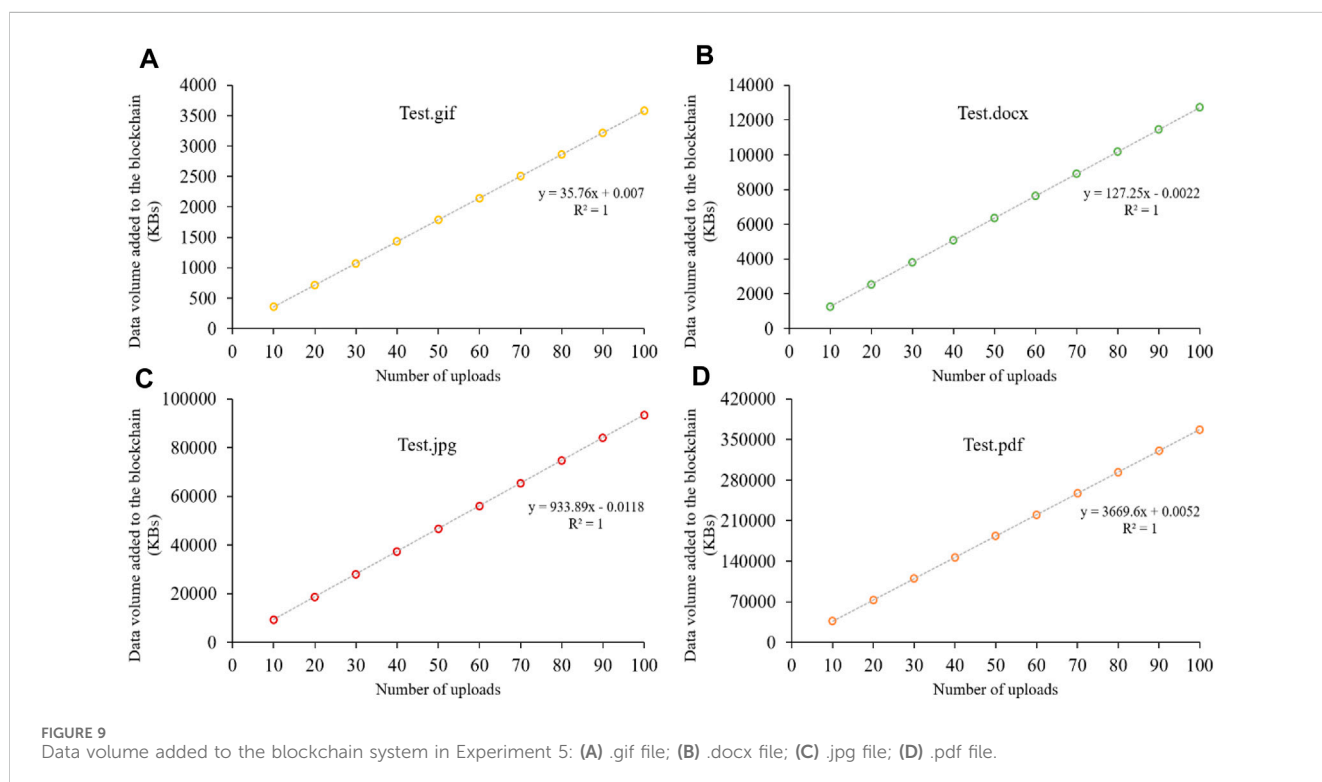$$P_I = \left\lceil \frac{s}{3} \right\rceil_{integer} \times 4. \tag{1}$$

In the Fabric blockchain network, Part II (Response) contained in a transaction is similar in size to the original construction file (s). Thus, the response (P_{II}) size due to a transaction can be obtained by using Formula Eq. 2 presented below. Based on Experiment 2, we discover that every time a new endorsement peer is added, an extra

TABLE 6 Total data volume added to the blockchain and data redundancy in Experiment 5.

Operation frequency (f)	Total data volume added after ten uploads (p_{total}) [KBs] (data redundancy) =(p_{total})/($s{\times}f$)			
	Test.gif	Test.docx	Test.jpg	Test.pdf
10	357.61 (3.688)	1,272.54 (2.602)	9338.93 (2.367)	36695.62 (2.342)
20	715.22 (3.688)	2,545.09 (2.602)	18677.85 (2.367)	73391.26 (2.342)
30	1,072.82 (3.688)	3,817.63 (2.602)	28016.78 (2.367)	110086.90 (2.342)
40	1,430.42 (3.688)	5,090.19 (2.602)	37355.72 (2.367)	146782.53 (2.342)
50	1788.04 (3.688)	6362.73 (2.602)	46694.66 (2.367)	183478.16 (2.342)
60	2,145.63 (3.688)	7635.28 (2.602)	56033.59 (2.367)	220173.79 (2.342)
70	2,503.24 (3.688)	8907.82 (2.602)	65372.52 (2.367)	256869.41 (2.342)
80	2,860.84 (3.688)	10180.36 (2.602)	74711.46 (2.367)	293565.03 (2.342)
90	3,218.44 (3.688)	11452.91 (2.602)	84050.39 (2.367)	330260.66 (2.342)
100	3,576.04 (3.688)	12725.46 (2.602)	93389.33 (2.367)	366956.29 (2.342)

Note: $g = 1$, $n = 10$, $b = 98$MBs, $s = 9.7$ KBs, or 48.9 KBs, or 394.6 KBs, or 1,567 KBs.

FIGURE 9
Data volume added to the blockchain system in Experiment 5: **(A)** .gif file; **(B)** .docx file; **(C)** .jpg file; **(D)** .pdf file.
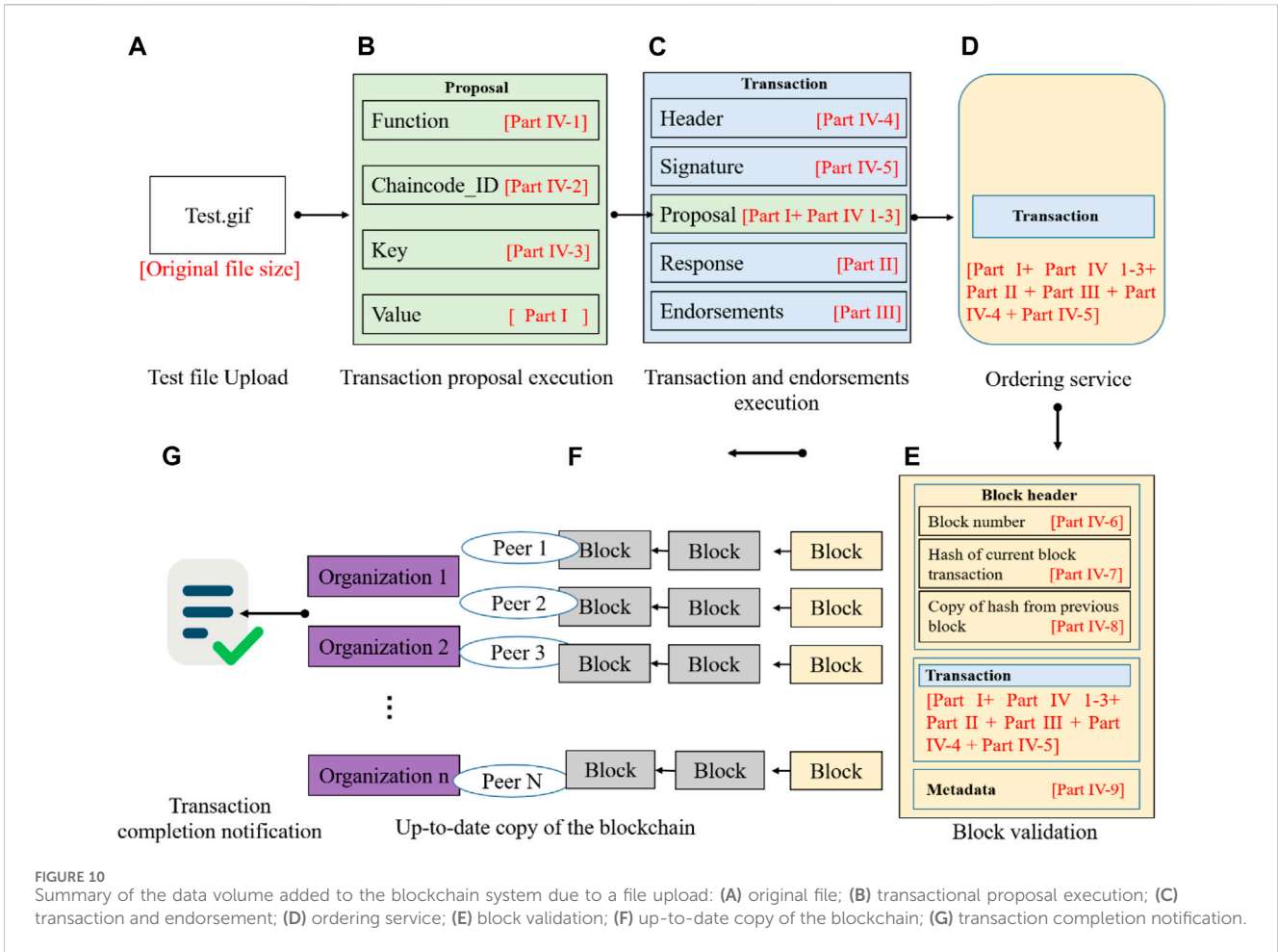
1003-byte endorsement is added to the data volume of an endorsement peer, on average, when a transaction is executed. Therefore, Part III (endorsements, $[P_{III}]$) included in a transaction can be calculated based on the number of endorsement peers configured in the network (n) from Formula Eq. 3.

$$P_{II} = s \qquad (2)$$

$$P_{III} = 1003 \times n. \qquad (3)$$

Part IV consists of nine sub-parts: the function (Part IV-1, $[P_{Iv-1}]$), chaincode_ID (Part IV-2, $[P_{Iv-2}]$), and key (Part IV-3, $[P_{Iv-3}]$) included in a proposal; the header (Part IV-4, $[P_{Iv-4}]$) and signature (Part IV-5, $[P_{Iv-5}]$) contained in a transaction; and the block number (Part IV-6, $[P_{Iv-6}]$), hash of current block transaction (Part IV-7, $[P_{Iv-7}]$), copy of hash from previous block (Part IV-8), $[P_{Iv-8}]$), and metadata (Part IV-9, $[P_{Iv-9}]$) produced from block ordering. According to Experiments 4 and 5, the data volume of these nine sub-parts is relatively

FIGURE 10
Summary of the data volume added to the blockchain system due to a file upload: **(A)** original file; **(B)** transactional proposal execution; **(C)** transaction and endorsement; **(D)** ordering service; **(E)** block validation; **(F)** up-to-date copy of the blockchain; **(G)** transaction completion notification.

constant. Therefore, the size of Part IV (P_{IV}) can be calculated by adding the data volume occupied by these nine sub-parts using Formula Eq. 4.

$$P_{IV} = P_{IV-1} + P_{IV-2} + P_{IV-3} + P_{IV-4} + P_{IV-5} + P_{IV-6} + P_{IV-7} + P_{IV-8} + P_{IV-9} \approx 3434.$$

(4)

The total data volume (P_{Total}) added to the blockchain system with n number of peers due to a transaction is

$$P_{Total} = (P_I + P_{II} + P_{III} + P_{IV}) \times n$$

(5a)

By substituting formulas Eqs 1–4 into Eq. 5a, we obtain Eq. 5b

$$P_{Total} = \left(\left\lceil \frac{s}{3} \right\rceil_{integer} \times 4 + s + 1003 \times n + 3434 \right) \times n.$$

(5b)

If the same transaction repeated f times, then the total data volume is

$$P_{Total} = \left[\left(\left\lceil \frac{s}{3} \right\rceil_{integer} \times 4 + s + 1003 \times n + 3434 \right) \times n \right] \times f$$

(6)

If we have multiple transactions with different sizes $o_1, o_2...o_n$, repeated $f_1, f_2...f_n$ times correspondingly, then the total data volume is (see Eq. 7)

$$P_{Total} = \left[\left(\left\lceil \frac{s_1}{3} \right\rceil_{integer} \times 4 + s_1 + 1003 \times n + 3434 \right) \times n \right] \times f_1$$

$$+ \left[\left(\left\lceil \frac{s_2}{3} \right\rceil_{integer} \times 4 + s_2 + 1003 \times n + 3434 \right) \times n \right] \times f_2$$

$$+ ...$$

$$+ \left[\left(\left\lceil \frac{s_n}{3} \right\rceil_{integer} \times 4 + s_n + 1003 \times n + 3434 \right) \times n \right] \times f_n$$

(7)

One can convert the unit from bytes to KBs by using Formula Eq. 8

$$P_{Total} [\text{Bytes}] = P_{Total} \times 0.00097656 [\text{KBs}]$$

(8)

The data redundancy (R) brought by multiple transactions can be calculated from

$$R = \frac{P_{Total}}{s_1 \times f_1 + s_2 \times f_2 + ... s_n \times f_n}$$

(9)

5.3 Model validation

In this section, we perform experimental measurements to verify the correctness of our data redundancy model. When corresponding
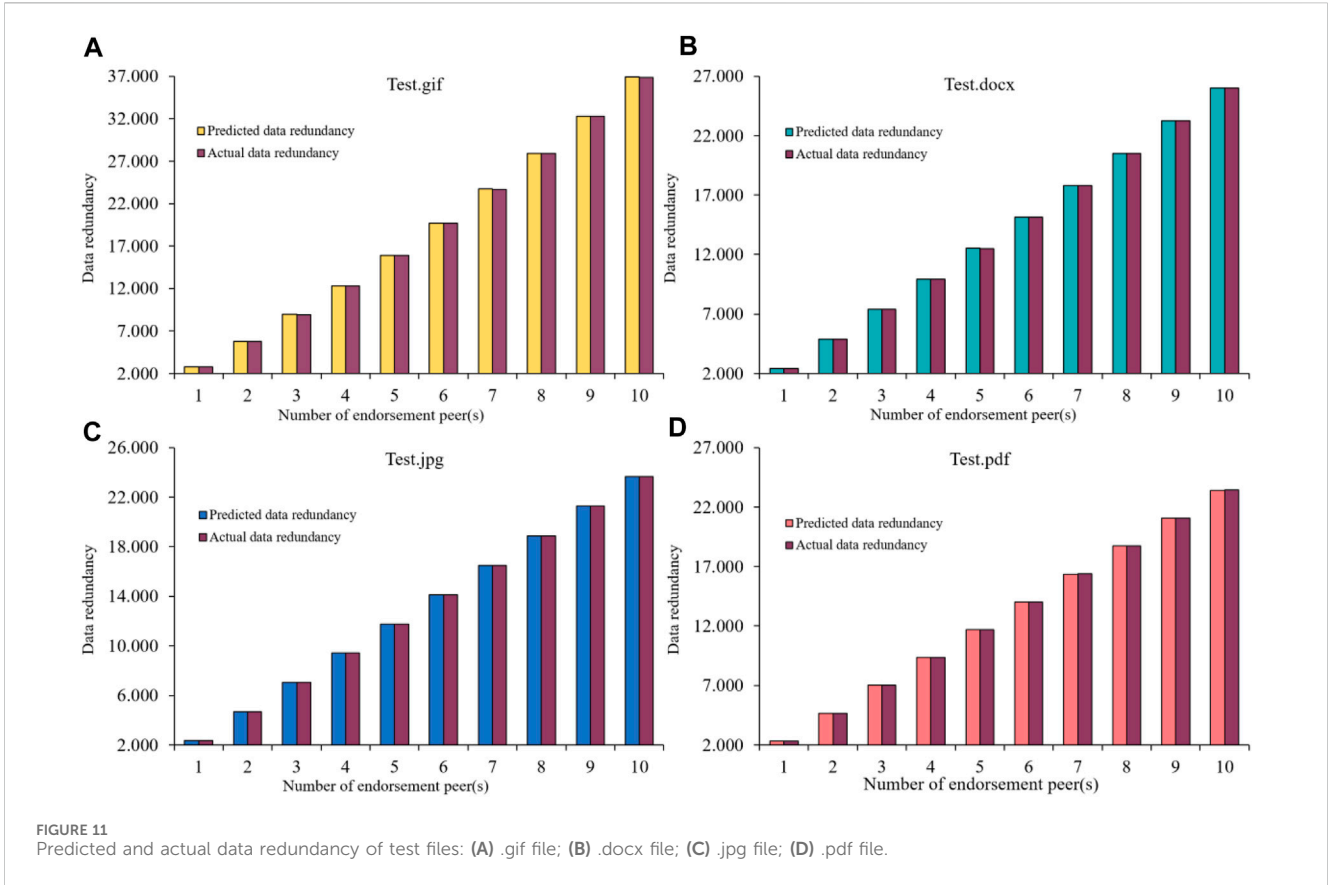
FIGURE 11
Predicted and actual data redundancy of test files: **(A)** .gif file; **(B)** .docx file; **(C)** .jpg file; **(D)** .pdf file.

TABLE 7 Model accuracy for different types of test files.

Test files	Model accuracy	Number of endorsement peers in the network									
		1	2	3	4	5	6	7	8	9	10
.gif	R_{Act}	2.778	5.760	8.940	12.324	15.911	19.699	23.689	27.881	32.275	36.876
	R_{Pre}	2.780	5.763	8.947	12.333	15.922	19.712	23.705	27.900	32.296	36.895
	θ	**99.93**	**99.95**	**99.92**	**99.93**	**99.93**	**99.93**	**99.93**	**99.93**	**99.93**	**99.95**
.docx	R_{Act}	2.422	4.884	7.385	9.927	12.509	15.131	17.793	20.495	23.237	26.019
	R_{Pre}	2.422	4.884	7.386	9.928	12.510	15.133	17.795	20.497	23.240	26.022
	θ	**100.0**	**100.0**	**99.99**	**99.99**	**99.99**	**99.99**	**99.99**	**99.99**	**99.99**	**99.99**
.jpg	R_{Act}	2.344	4.694	7.048	9.407	11.771	14.140	16.514	18.893	21.277	23.666
	R_{Pre}	2.344	4.694	7.048	9.407	11.771	14.140	16.514	18.893	21.277	23.666
	θ	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
.pdf	R_{Act}	2.336	4.673	7.012	9.352	11.693	14.035	16.379	18.724	21.070	23.417
	R_{Pre}	2.336	4.673	7.012	9.352	11.693	14.035	16.379	18.724	21.070	23.417
	θ	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**

Note: $g = 1$, $f = 10$, $b = 98$MBs, $s = 9.7$ KBs, or 48.9 KBs, or 394.6 KBs, or 1,567 KBs.
Note. θ is the accuracy of our data redundancy model.

variables ($g = 1$, $f = 10$, $b = 98$ MBs, $s = 9.7$ KBs [test.gif], or 48.9 KBs [test.docx], or 394.6 KBs [test.jpg], or 1,567 KBs [test.pdf]) are under control, Eqs 6, 9 are used to determine the predicted data redundancy (R_{pre}) of each test file. This is compared with the corresponding actual data redundancy (R_{act}) obtained from Experiment 2 (see Table 3). Figure 11A shows that the predicted

data redundancy is within one percent of the actual data redundancy error range, which indicates that the model has high accuracy. For the other three test files, the same conclusion can be drawn (see Figures 11B–D).

In summary, the predicted and actual data redundancy match well in the four types of test files. We take the data in Figure 10 as inputs and exhibit the accuracy θ in Table 7. We calculate and define the accuracy percentage as $\theta = R_{Act} / R_{Pre} \times 100\%$. As Table 7 shows, our model predicts data redundancy with an accuracy of 99.92% and higher for *Test. gif* files; 99.99% and above for *Test. docx* files; 100% for *Test. jpg* files, and 100% for *Test. pdf* files.

6 Discussion

This research measured the data redundancy of a blockchain system by conducting a series of laboratory experiments on a Hyperledger Fabric blockchain system in a construction project context. Five controlled experiments show that the data volume of a blockchain system grows proportionally with the size (s) of the construction files to be uploaded, the number of peer nodes (n) in the network, and the frequency (f) of blockchain operations in a construction project, respectively. We further discover that the data volume of a blockchain has little relationship with the block size (b) or how the peer nodes are grouped in different construction organizations (g). In real-life construction practice, prospective users may wish to build a blockchain infrastructure for different applications. Therefore, the data redundancy model developed in this study provides a structured methodology to help them predict data redundancy and better plan blockchain systems. In our study, this model predicts data redundancy with an accuracy of 99.92% or above for test construction files. Overall, this research is a meaningful step towards demystifying data redundancy of blockchain systems in construction. By referencing the model establishment method in this study, future studies can develop similar models for different blockchain platforms (e.g., Ethereum) to measure and quantify data redundancy in construction projects. This can help optimize data storage, improve information exchange, and reduce unnecessary duplication of data across different platforms, resulting in more efficient and cost-effective project management.

Blockchain presents enormous prospects and challenges for data management. From a technical point of view, blockchain components such as cryptographic algorithms, distributed ledgers, and consensus mechanisms are conducive to secure data storage. However, at present, the vast data storage volume on each peer is a primary bottleneck that restricts the expansibility of blockchain. To maximize the benefits of the proposed data redundancy model for blockchain in construction projects, stakeholders should consider several management recommendations:

1. Familiarize stakeholders with the data redundancy model: It is crucial to educate construction stakeholders about the proposed data redundancy model for blockchain. This includes explaining how it works and its potential impact on project management. This will ensure a clear understanding of the model's objectives and its relevance to their specific roles.

2. Assess project-specific data redundancy level: Each construction project may have unique data redundancy level based on its scale, complexity, and stakeholders involved. Stakeholders should assess and determine the appropriate level of data redundancy needed for their project, taking into account factors like data importance, accessibility, and security.

3. Implement the model during project planning phase: To maximize the benefits of the data redundancy model, it should be integrated into the project planning phase. Stakeholders should identify the specific data elements that involve redundancy and establish protocols to optimize data storage mechanism and control overall storage costs.

4. Monitor and evaluate the effectiveness of the model: Regular monitoring and evaluation of the data redundancy model's effectiveness are vital. Stakeholders should establish performance metrics to assess the model's impact on data storage, information exchange, and project outcomes. This feedback loop enables continuous improvement and optimization of the model's implementation.

Also, using the proposed model to determine data redundancy for construction stakeholders can have several practical implications:

1. The model allows stakeholders to accurately assess the amount of redundant data in their systems, optimizing data storage and reducing unnecessary duplication costs. This can lead to significant cost savings regarding storage infrastructure and maintenance.

2. The model enables stakeholders to evaluate the impact of data redundancy on system performance and efficiency. By analyzing the redundancy levels and their effect on data access and retrieval, stakeholders can fine-tune their systems to optimize performance and improve overall productivity.

3. The model helps stakeholders identify critical data elements that require higher redundancy levels, such as project specifications, contractual agreements, and financial records. By ensuring a higher level of redundancy for these essential data elements, stakeholders can mitigate the risk of data loss and ensure business continuity in the event of system failures or disasters.

4. The model facilitates data governance and compliance with regulatory requirements. By accurately estimating data redundancy, stakeholders can ensure compliance with data protection regulations and industry standards. This helps to maintain the integrity and privacy of sensitive data and builds trust with clients and partners.

Overall, using the proposed model to estimate data redundancy for construction stakeholders can result in cost savings, enhanced system performance, improved data management, and regulatory compliance.

Blockchain technology can benefit from using the IPFS to resolve data redundancy. IPFS can store and distribute large files associated with blockchain transactions, such as images, videos, and design specifications (Tao et al., 2021). This can improve data integrity, availability, and efficiency. Implementing methods or technologies like IPFS for resolving data redundancy in blockchain can significantly impact various factors influencing

TABLE 8 Comparison of this study with existing blockchain studies.

Item	This study	Tao et al. (2021)	Wu et al. (2023)	Zhong et al. (2023)
Data redundancy measurement	✓	✓	×	×
Data redundancy model	✓	×	×	×
Evaluation of the data redundancy model	✓	×	×	×

data redundancy. For instance, file size plays a crucial role in storage and retrieval efficiency, and using IPFS can enable the efficient handling of large files associated with blockchain transactions. The number of peers and construction organizations can affect data availability and collaboration, and leveraging IPFS for distributed storage can enhance these aspects by reducing reliance on centralized servers. Block size, another important factor, can impact the performance of blockchain networks, and utilizing IPFS technology can optimize block size by compressing files and reducing network traffic. Additionally, the operation frequency of transactions in construction projects can be streamlined through IPFS integration with smart contracts, automating data retrieval and distribution. By considering these factors, stakeholders can effectively leverage technologies like IPFS to resolve data redundancy in blockchain, leading to improved data management, collaboration, and overall efficiency in the construction industry. Other potential methods or technologies for resolving blockchain data redundancy include content-addressed storage, distributed data sharing, and file compression. These methods can reduce the need for centralized storage infrastructure, improve data retrieval and distribution, and reduce file and block sizes. Ultimately, resolving data redundancy in blockchain is crucial for the success of blockchain-based projects.

Compared with existing studies (e.g., Tao et al., 2021; Wu et al., 2023; Zhong et al., 2023), this study presents a novel approach by proposing a model that allows blockchain users in the construction industry to estimate data redundancy rather than merely reporting the system's data storage capacity, as shown in Table 8. This approach recognizes the challenges faced by the construction industry in managing large volumes of data while ensuring data accuracy and integrity. By estimating data redundancy, users can better understand the storage requirements and plan accordingly, ultimately improving data management and the efficiency of project delivery. This model leverages the benefits of blockchain technology, such as immutability and decentralization, while addressing the limitations of data storage capacity. By providing a more comprehensive understanding of data redundancy, this model offers a more effective and sustainable solution for managing data in the construction industry.

This research has some limitations. Firstly, the measurement of data redundancy in this study is based on the Hyperledger Fabric blockchain platform only. Other blockchain platforms are not evaluated. Secondly, the data redundancy model developed in this research is based on the Base64 file reading method. If other file reading methods are used, the model must be adjusted manually. Thirdly, this study lacks a cost evaluation framework with data redundancy, and cost is a critical criterion affecting user decision-making in developing blockchain systems. Finally, we conducted our experiments on a laboratory-built Hyperledger Fabric blockchain

system, and the model has not been tested in real-life construction practice.

Amid the global blockchain hype, potential users should carefully choose which work tasks to use on the blockchain and even whether they need it at all. As observed in this research, blockchain technology introduces data redundancy and sacrifices efficiency to enhance security. It generates additional and sometimes excessive costs. Thus, it is necessary to develop highly selective strategies based on a more thorough cost-benefit analysis when better empirical data is available. More than just using new software, blockchain implementation is about implementing new business technologies and philosophies. It can support the digitization of projects and provide solutions to many challenges, but before using blockchain, organizations need to analyze their existing business models. Hence, we need to study the relationship between project governance and blockchain systems and their impacts on project performance.

7 Conclusion

Scholars are actively exploring blockchain applications in the construction industry for benefits such as greater transparency, enhanced security, and improved traceability. Blockchain works on a data redundancy mechanism, but the literature has neither clearly stated the process by which data is redundant in blockchain nor empirically measured the degree of redundancy. Data redundancy and its associated cost are critical considerations for construction project stakeholders, who are often from different companies and naturally safeguard their costs and benefits. This research addresses this knowledge gap by conducting a series of experiments on a Hyperledger Fabric blockchain system built in a laboratory.

The research objectives were achieved by showing that: (1) data volume of a blockchain system grows proportionally with the size of the documents to be blockchained, the number of peer nodes in the blockchain network, and the frequency of blockchain operations, but has little relationship with the block size or how the peer nodes are dispersed in different construction organizations; (2) a data redundancy model for construction practitioners to predict data redundancy in a blockchain system. This research suggests that blockchain users consider data redundancy when planning a blockchain solution by paying attention to the factors that matter. Excessive data redundancy will also increase the computational burden and cause Internet connection 'traffic jams'. Thus, blockchain-based systems cannot be unlimitedly redundant. Our proposed data redundancy model based on our experimental results can help prospective users predict redundancy when planning a blockchain-based system. In this study, this

developed model predicts data redundancy with an accuracy of 99.92% or above for test construction files.

The limitations of this study provide opportunities for further investigation. Firstly, the research is based on the Hyperledger Fabric blockchain platform. More studies are desired to assess the data redundancy of other blockchain platforms. Secondly, the data redundancy model proposed in this study is based on the Base64 file reading method. Therefore, future investigations are encouraged to explore other file reading methods. Thirdly, the proposed model is a handy tool for predicting data redundancy of the blockchain, thereby not naturally considering the cost evaluation of data redundancy. Future research is expected to develop a cost assessment framework for data redundancy. Lastly, future research is expected to explore the dilemma of data redundancy and system storage optimization and find solutions.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Author contributions

WL: Writing–review and editing, Supervision. LW: Writing–original draft, Visualization, Validation, Investigation. CC: Writing–original draft, Software, Investigation, Formal Analysis, Data curation.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Ahmadisheykhsarmast, S., Senji, S. G., and Sonmez, R. (2023). Decentralized tendering of construction projects using blockchain-based smart contracts and storage systems. *Automation Constr.* 151, 104900. doi:10.1016/j.autcon.2023.104900

Amir, M., Mohsen, S., and Afsaneh, M. (2018). Simultaneous desorption and desorption kinetics of phenanthrene, anthracene, and heavy metals from kaolinite with different organic matter content. *Soil Sediment Contam. An Int. J.* 27 (3), 200–220. doi:10.1080/15320383.2017.1339666

Bakr, M. A., and Lee, S. (2017). Distributed multisensor data fusion under unknown correlation and data inconsistency. *Sensors* 17 (11), 2472. doi:10.3390/s17112472

Beck, R., Michel, A., Rossi, M., and Thatcher, J. B. (2017). Blockchain technology in business and information systems research. *Bus. Inf. Syst. Eng.* 59 (6), 381–384. doi:10.1007/s12599-017-0505-1

Byers, J., Considine, J., and Mitzenmacher, M. (2003). "Simple load balancing for distributed hash tables," in Peer-to-Peer Systems II: Second International Workshop, IPTPS 2003, Berkeley, CA, February 21–22, 2003 (Springer Berlin Heidelberg), 80–87.

Chen, R., Li, Y., Yu, Y., Li, H., Chen, X., and Susilo, W. (2020). Blockchain-based dynamic provable data possession for smart cities. *IEEE Internet Things J.* 7 (5), 4143–4154. doi:10.1109/JIOT.2019.2963789

Chervyakov, N., Babenko, M., Tchernykh, A., Kucherov, N., Miranda-López, V., and Cortés-Mendoza, J. M. (2019). AR-RRNS: configurable reliable distributed data storage systems for Internet of Things to ensure security. *Future Gener. Comput. Syst.* 92, 1080–1092. doi:10.1016/j.future.2017.09.061

Coyne, R., and Onabolu, T. (2017). Blockchain for architects: challenges from the sharing economy. *Archit. Res. Q.* 21 (4), 369–374. doi:10.1017/S1359135518000167

Das, M., Tao, X., and Cheng, J. C. (2021). BIM security: a critical review and recommendations using encryption strategy and blockchain. *Automation Constr.* 126, 103682. doi:10.1016/j.autcon.2021.103682

Du, M., Chen, Q., and Ma, X. (2020). MBFT: a new consensus algorithm for consortium blockchain. *IEEE Access* 8, 87665–87675. doi:10.1109/ACCESS.2020.2993759

Eriksson, P. E. (2013). Exploration and exploitation in project-based organizations: development and diffusion of knowledge at different organizational levels in construction companies. *Int. J. Proj. Manag.* 31 (3), 333–341. doi:10.1016/j.ijproman.2012.07.005

Fan, X., Wei, W., Wozniak, M., and Li, Y. (2018). Low energy consumption and data redundancy approach of wireless sensor networks with bigdata. *Inf. Technol. Control/Informacinès Technol. ir Valdymas* 47 (3), 406–418. doi:10.5755/j01.itc.47.3.20565

Hijazi, A. A., Perera, S., Calheiros, R. N., and Alashwal, A. (2021). Rationale for the integration of BIM and blockchain for the construction supply chain data delivery: a systematic literature review and validation through focus group. *J. Constr. Eng. Manag.* 147 (10), 03121005. doi:10.1061/(ASCE)CO.1943-7862.0002142

Huang, Z., Chen, J., Lin, Y., You, P., and Peng, Y. (2015). Minimizing data redundancy for high reliable cloud storage systems. *Comput. Netw.* 81, 164–177. doi:10.1016/j.comnet.2015.02.013

Hyperledger (2020) "Glossary," in *Hyperledger fabric*. Available at: https://hyperledger-fabric.readthedocs.io/en/latest/glossary.html.

International Organization for Standardization (2020). *Blockchain and distributed ledger technologies — vocabulary*. Available at: https://www.iso.org/obp/ui/#iso:std:iso:22739:ed-1:v1:en.

Jayasankar, U., Thirumal, V., and Ponnurangam, D. (2021). A survey on data compression techniques: from the perspective of data quality, coding schemes, data type and applications. *J. King Saud University-Computer Inf. Sci.* 33 (2), 119–140. doi:10.1016/j.jksuci.2018.05.006

Jiang, Y., Liu, X., Wang, Z., Li, M., Zhong, R. Y., and Huang, G. Q. (2023). Blockchain-enabled digital twin collaboration platform for fit-out operations in modular integrated construction. *Automation Constr.* 148, 104747. doi:10.1016/j.autcon.2023.104747

Li, J., Greenwood, D., and Kassem, M. (2019). Blockchain in the built environment and construction industry: a systematic review, conceptual models and practical use cases. *Automation Constr.* 102, 288–307. doi:10.1016/j.autcon.2019.02.005

Li, X., Wu, L., Zhao, R., Lu, W., and Xue, F. (2021). Two-layer Adaptive Blockchain-based Supervision model for off-site modular housing production. *Comput. Industry* 128, 103437. doi:10.1016/j.compind.2021.103437

Lu, W., Li, X., Xue, F., Zhao, R., Wu, L., and Yeh, A. G. (2021a). Exploring smart construction objects as blockchain oracles in construction supply chain management. *Automation Constr.* 129, 103816. doi:10.1016/j.autcon.2021.103816

Lu, W., Wu, L., and Xue, F. (2021c). Blockchain technology for projects: a multicriteria decision matrix. *Proj. Manag. J.* 87569728211061780. doi:10.1177/2F87569728211061780

Lu, W., Wu, L., Zhao, R., Li, X., and Xue, F. (2021b). Blockchain technology for governmental supervision of construction work: learning from digital currency electronic payment systems. *J. Constr. Eng. Manag.* 147 (10), 04021122. doi:10.1061/(ASCE)CO.1943-7862.0002148

Najafabadi, A. A. S., and Azar, F. T. (2019). Removing redundancy data with preserving the structure and visuality in a database. *Signal, Image Video Process.* 13 (4), 745–752. doi:10.1007/s11760-018-1404-8

Nawari, N. O., and Ravindran, S. (2019). Blockchain and the built environment: potentials and limitations. *J. Build. Eng.* 25, 100832. doi:10.1016/j.jobe.2019.100832

Penzes, B., Kirkup, A., Gage, C., Dravai, T., and Colmer, M. (2018). Blockchain technology in the construction industry: digital transformation for high productivity. Available at: https://www.academia.edu/38193166/Blockchain_Technology_in_the_Construction_Industry_ICE_pdf (Accessed December 12, 2021).

Perera, S., Nanayakkara, S., Rodrigo, M. N. N., Senaratne, S., and Weinand, R. (2020). Blockchain technology: is it hype or real in the construction industry? *J. Industrial Inf. Integration* 17, 100125. doi:10.1016/j.jiii.2020.100125

Qian, X. A., and Papadonikolaki, E. (2021). Shifting trust in construction supply chains through blockchain technology. *Eng. Constr. Archit. Manag.* 28 (2), 584–602. doi:10.1108/ECAM-12-2019-0676

Rahim, R., Nurdiyanto, H., Hidayat, R., Ahmar, A. S., Siregar, D., Siahaan, A. P. U., et al. (2018). Combination Base64 algorithm and EOF technique for steganography. *J. Phys. Conf. Ser.* 1007, 012003. doi:10.1088/1742-6596/1007/1/012003

Risius, M., and Spohrer, K. (2017). A blockchain research framework. *Bus. Inf. Syst. Eng.* 59 (6), 385–409. doi:10.1007/s12599-017-0506-0

Rodrigo, M. N. N., Perera, S., Senaratne, S., and Jin, X. (2021). Systematic development of a data model for the blockchain-based embodied carbon (BEC) estimator for construction. *Eng. Constr. Archit. Manag.* 29, 3311–3330. doi:10.1108/ECAM-02-2021-0130

Tao, X., Das, M., Liu, Y., and Cheng, J. C. (2021). Distributed common data environment using blockchain and Interplanetary File System for secure BIM-based collaborative design. *Automation Constr.* 130, 103851. doi:10.1016/j.autcon.2021.103851

Tao, X., Liu, Y., Wong, P. K. Y., Chen, K., Das, M., and Cheng, J. C. (2022). Confidentiality-minded framework for blockchain-based BIM design collaboration. *Automation Constr.* 136, 104172. doi:10.1016/j.autcon.2022.104172

Tezel, A., Febrero, P., Papadonikolaki, E., and Yitmen, I. (2021). Insights into blockchain implementation in construction: models for supply chain management. *J. Manag. Eng.* 37 (4), 04021038. doi:10.1061/(ASCE)ME.1943-5479.0000939

Turk, Ž., and Klinc, R. (2017). Potentials of blockchain technology for construction management. *Procedia Eng.* 196, 638–645. doi:10.1016/j.proeng.2017.08.052

Verma, N., and Singh, D. (2018). Data redundancy implications in wireless sensor networks. *Procedia Comput. Sci.* 132, 1210–1217. doi:10.1016/j.procs.2018.05.036

Weatherspoon, H., and Kubiatowicz, J. D. (2002). "Erasure coding vs. replication: a quantitative comparison," in *International workshop on peer-to-peer systems* (Berlin, Heidelberg: Springer), 328–337. doi:10.1007/3-540-45748-8_31

Wu, H., Li, H., Luo, X., and Jiang, S. (2023). Blockchain-based on-site activity management for smart construction process quality traceability. *IEEE Internet Things J.* 10 (24), 21554–21565. doi:10.1109/JIOT.2023.3300076

Wu, L., Lu, W., Xue, F., Li, X., Zhao, R., and Tang, M. (2022a). Linking permissioned blockchain to Internet of Things (IoT)-BIM platform for off-site production management in modular construction. *Comput. Industry* 135, 103573. doi:10.1016/j.compind.2021.103573

Wu, L., Lu, W., Zhao, R., Xu, J., Li, X., and Xue, F. (2022b). Using blockchain to improve information sharing accuracy in the onsite assembly of modular construction. *J. Manag. Eng.* 38 (3), 04022014. doi:10.1061/(ASCE)ME.1943-5479.0001029

Xenya, M. C., and Quist-Aphetsi, K. (2019). "Decentralized distributed blockchain ledger for financial transaction backup data," in 2019 international conference on cyber security and internet of things (ICSIoT), Accra, Ghana, May 29–31, 2019 (IEEE), 34–36.

Xue, F., and Lu, W. (2020). A semantic differential transaction approach to minimizing information redundancy for BIM and blockchain integration. *Automation Constr.* 118, 103270. doi:10.1016/j.autcon.2020.103270

Xu, J., Lou, J., Lu, W., Wu, L., and Chen, C. (2023). Ensuring construction material provenance using Internet of Things and blockchain: learning from the food industry. *J. Ind. Inf. Integr.* 33, 100455. doi:10.1016/j.jiii.2023.100455

Ye, Z., Yin, M., Tang, L., and Jiang, H. (2018). "Cup-of-Water theory: a review on the interaction of BIM, IoT and blockchain during the whole building lifecycle," in ISARC, Proceedings of the 35th International Symposium on Automation and Robotics in Construction (ISARC 2018), Berlin, Germany, (IAARC), 478–486.

Yoon, J. H., and Pishdad-Bozorgi, P. (2022). State-of-the-Art review of blockchain-enabled construction supply chain. *J. Constr. Eng. Manag.* 148 (2), 03121008. doi:10.1061/(ASCE)CO.1943-7862.0002235

Zhong, B., Pan, X., Ding, L., Chen, Q., and Hu, X. (2023). Blockchain-driven integration technology for the AEC industry. *Automation Constr.* 150, 104791. doi:10.1016/j.autcon.2023.104791