
AI translation · View original & related papers at
chinaxiv.org/items/chinaxiv-202308.00461

A Novel Hierarchical Digital Library Architecture Based on Blockchain Technology: Postprint

Authors: Gao Sheng, Zhu Jianming

Date: 2023-08-27T00:00:00+00:00

Abstract

[Purpose/Significance] Currently, the trustworthiness, interoperability, and scalability of digital resources in digital library systems have become major obstacles hindering the development of digital libraries. [Method/Process] Based on blockchain technology, a novel hierarchical digital library architecture is proposed, with detailed elaboration on the key functional components and core service modules of each layer, followed by a comprehensive comparison with existing related digital library architectures. [Results/Conclusion] Theoretical analysis demonstrates that the proposed architecture, by leveraging the distributed ledger structure, security and reliability, transparency and trustworthiness, and temporal immutability of blockchain technology, can effectively address the trustworthiness of digital resources in digital libraries, enhance the interoperability and scalability of the system architecture, and achieve secure circulation of heterogeneous digital resources.

Full Text

A Novel Hierarchical Digital Library System Architecture Based on Blockchain Technology

Gao Sheng, Zhu Jianming

School of Information, Central University of Finance and Economics, Beijing 100081

Abstract

[Purpose/Significance] Current digital library systems face major challenges regarding digital resource trustworthiness, interoperability, and scalability that hinder their development. [Method/Process] This paper proposes a novel hierarchical digital library architecture based on blockchain technology, detailing

the key functional components and core service modules of each layer, and comprehensively compares the proposed architecture with existing digital library architectures. **[Result/Conclusion]** Theoretical analysis demonstrates that the proposed architecture leverages blockchain's distributed ledger structure, security, transparency, and tamper-proof chronological sequencing to effectively solve digital resource trust issues, enhance system interoperability and scalability, and enable secure circulation of heterogeneous digital resources.

Keywords: digital library; blockchain technology; resource trustworthiness; security

1. Introduction

In today's society, digitalization, informatization, and intelligence have become dominant trends. In 2017, President Xi Jinping emphasized during the second collective study session of the Political Bureau of the CPC Central Committee: "We must promote the implementation of the national big data strategy, accelerate the improvement of digital infrastructure, advance the integration and open sharing of data resources, ensure data security, and accelerate the construction of a digital China." As a typical digital information system, digital libraries integrate library science, information systems, database management, multimedia, and human-computer interaction technologies [?]. M. A. Gonçalves et al. [?] applied the 5S model—streams, structures, spaces, scenarios, and societies—to formally define digital libraries. L. Candela et al. [?] proposed the DELOS reference model to characterize digital libraries from content, user, function, quality, policy, and architecture perspectives. C. B. Arms et al. [?] suggested that digital library architecture comprises repositories, handle systems, search systems, and user interfaces. Zhang Xiaolin et al. [?] constructed a standard description system for digital information resource construction from digital content creation, digital object description, resource organization, digital resource system services, and long-term preservation. Li Shuang [?] summarized design principles for digital library architecture: interoperability, scalability, standardization, hierarchy, and uniformity, with interoperability and scalability being particularly critical. Feng Guohe [?] reviewed early typical architectures including client/server two-tier, browser/web server/database server three-tier, client/server triangular, and distributed architectures. However, these architectures built on traditional infrastructure suffer from bottlenecks in inter-library interoperability, scalability, security, and service costs.

With emerging technologies like cloud computing and service computing, digital library architectures have evolved further. Existing research on cloud-based or service-oriented digital library architectures either integrates heterogeneous library resources through unified platforms—a typical centralized approach [?] with low service efficiency and high management costs and risks—or builds distributed digital libraries via P2P technology [?], where resource trustworthiness depends entirely on providers' credit, increasing trust establishment costs while inadequately addressing security. In 2017, S. Higgins reported [?] that the U.S.

government granted \$100,000 to study blockchain applications in public library systems. However, current blockchain-based digital library architectures remain scarce. In 2018, Zeng Ziming and Qin Siqi [?] constructed the first decentralized resource management architecture for smart library mobile visual search using blockchain, but this architecture lacked specific block structures and service processes adapted for digital libraries, with insufficient consideration for heterogeneous resource interoperability and service function extensibility.

To address these issues, we propose a novel hierarchical digital library architecture based on blockchain technology. On one hand, blockchain's distributed structure, security, transparency, and tamper-proof characteristics overcome problems of single-point failure, access bottlenecks, and data trustworthiness in centralized digital library data management, ensuring digital resource trustworthiness. On the other hand, customized unified block storage structures, cryptographic algorithms, and communication protocols greatly enhance interoperability and security of heterogeneous resources, while blockchain's dynamic nature and flexibility improve system scalability.

2. Related Work

Current digital library architectures vary due to different construction conditions and technologies. Numerous architectures have been developed worldwide [?, ?]. In 1995, R. Kahn and R. Wilensky [?] early proposed the K-W framework, which separated repositories and handle systems while using unified user interfaces to provide centralized services for heterogeneous resources within the same architecture. S. Payette and C. Lagozse et al. [?] proposed the flexible and extensible Fedora architecture for digital objects and repositories to address content storage and distribution. Specifically, Fedora encapsulates metadata and content within digital object models and provides service interfaces through repositories to improve heterogeneous resource interoperability. For interoperability issues, literature [?, ?] summarized common solutions including standardization (e.g., TCP/IP, Z39.50), middleware (e.g., OAI, Diens), normalized interactions (e.g., agent communication languages, high-level languages like SETL, PAISLey), and mobile function specifications (e.g., Lisp, Java applets). Projects like NDSTL and NCSTRL adopted these technologies.

In 2005, P. Knezevic et al. [?] noted that with Service-Oriented Architecture (SOA), P2P networks, and grid computing, future digital libraries would evolve from integrated, centralized systems toward distributed, dynamic, configurable alliances. Li Guangjian [?] summarized digital library architectures based on SOA, P2P, and grid technologies. Literature [?] studied SOA-based architectures. Yin Hong and Liu Wei [?] emphasized the importance of resource discovery in future systems. Overall, existing architectures focus on improving interoperability but rely entirely on resource providers for trustworthiness, increasing trust costs and inadequately addressing security. For instance, cloud-based digital libraries separate ownership from management, threatening resource security.

3. Blockchain Technology Fundamentals

Blockchain has gained widespread attention as Bitcoin's underlying technology, with applications extending from digital currency (Blockchain 1.0) and digital assets (Blockchain 2.0) to digital society [?]. We introduce blockchain to digital libraries to build a novel distributed architecture.

3.1 Blockchain Concept No unified definition of blockchain currently exists. In 2015, *The Economist* described blockchain as “the trust machine.” In 2016, China's Ministry of Industry and Information Technology defined blockchain from data structure and processing perspectives [?]: structurally, as a chronological chain guaranteeing tamper-proof distributed ledgers through cryptography; functionally, as a distributed infrastructure using chain structures for storage, consensus algorithms for updates, cryptography for security, and smart contracts for operations. Overall, blockchain is a distributed shared ledger integrating distributed ledger technology, cryptography, consensus algorithms, and smart contracts to enable tamper-resistant, non-forgable, traceable, and auditable data without trusted third-party endorsement.

3.2 Blockchain Structure Blockchain uses a chronological block chain structure (Figure 1 [Figure 1: see original paper]). Each block comprises a header and body, with variations across platforms (Bitcoin, Ethereum, etc.). China's Blockchain Technology and Industry Development Forum draft standard [?] defines data structures for blocks, transactions, and entities. Typically, block headers store the previous block's hash, Merkle root, nonce, etc., while bodies store historical transaction data. The current block's header links to the previous block's double hash, forming a chronological chain. The Merkle root enables rapid verification of transaction integrity, allowing specific transaction verification without synchronizing all data (“Simplified Payment Verification”), improving efficiency and scalability. Blockchain storage uses key-value pairs for convenient retrieval of transactions and blocks, stored either as files on disk or in databases. Bitcoin stores data in files with LevelDB indexes; Ethereum stores both data and indexes in LevelDB.

3.3 Key Blockchain Technologies Blockchain synthesizes distributed ledgers, cryptography, consensus algorithms, and smart contracts [?].

(1) **Distributed Ledger:** Different blockchain types use different ledger models: transaction-based (Bitcoin, Quantum) or account-based (Ethereum, Hyperledger).

(2) **Cryptography:** Digital signatures ensure transaction integrity and non-repudiation. Each user has a key pair: public keys generate addresses; private keys sign transactions.

(3) **Consensus Algorithms:** Based on participation and algorithms, blockchains are public, consortium, or private (permissioned). Public chains

use resource-based algorithms (e.g., Proof of Work in Bitcoin) with low performance and high overhead. Consortium chains use voting-based algorithms (e.g., Practical Byzantine Fault Tolerance [?]) for higher performance and scalability.

(4) Smart Contracts: N. Szabo first proposed smart contracts as computable transaction protocols [?]. V. Buterin [?] introduced them to Ethereum as Turing-complete programs controlling digital assets in secure, sandboxed environments.

4. Blockchain-Based Digital Library Architecture

Blockchain's technical features enhance interoperability and trustworthiness of heterogeneous digital resources while improving scalability and service quality. We designed an autonomous, hierarchical architecture (Figure 2 [Figure 2: see original paper]).

4.1 Infrastructure Layer This layer provides hardware and software facilities: computing devices (distributed servers for blockchain operations), storage devices (for block data, transactions, chain structures, contracts, metadata, and digital resources like documents, images, audio, video), and network devices (cards, switches, routers for data transmission). Infrastructure can be locally deployed (high cost, limited elasticity) or cloud-based (IBM Bluemix, AWS), using virtualization to abstract heterogeneous resources into unified, on-demand Infrastructure-as-a-Service (IaaS), reducing costs and improving utilization and scalability.

4.2 Data Layer All digital resource operations occur through blockchain transactions (acquisition, retrieval). Transaction information is broadcast to network nodes, periodically packed into blocks via consensus, and appended to the chain. Transaction structures include provider address, receiver address, index value, metadata, resource hash, and digital signature. Each node has a key pair; signature modules use ECDSA or RSA for integrity and non-repudiation.

To enhance interoperability, we designed a unified data structure for digital libraries (Figure 3 [Figure 3: see original paper]) based on resource types and block structures. The structure comprises block headers (linking blocks, verifying integrity via previous block hash) and bodies (recording transactions). Each resource record includes index value (for retrieval efficiency, extracted via information gain, mutual information, or cross-entropy), metadata (describing attributes using standards like MARC, AACR2, Dublin Core, displayed in HTML/XML), and resource hash (for retrieval and integrity verification). This unified representation enhances interoperability of complex heterogeneous resources.

4.3 Network Layer This layer provides P2P networking, routing, and validation. Bitcoin and Ethereum use TCP-based P2P; Hyperledger uses HTTP/2.0. Our architecture employs P2P networks where each node is both provider and

user, eliminating central points of failure and improving scalability. P2P topologies include structured (Chord, Pastry via distributed hash tables) and unstructured (centralized like Napster, fully distributed like Gnutella, hybrid like KaZaA). Routing enables node discovery and block/transaction broadcasting. Validation checks data structure, signatures, and consensus to prevent propagation of invalid data.

4.4 Consensus Layer This layer matches user needs with resources to generate new blocks, ensuring consistency and trustworthiness in distributed environments. Consensus algorithms fall into resource-based (for public chains) and voting-based (for consortium/private chains). Public chains use Proof of Work (low efficiency, high overhead) or Proof of Stake. Consortium chains use PBFT, QuorumChain, etc. [?].

Digital libraries have formed consortiums (CDLF, CALIS). Considering public chains' inefficiency, our architecture uses voting-based algorithms (PBFT, SBFT, Raft) [?]. With limited, fixed members and stable connections, block generation and validation occur only among consortium members via message voting rather than probabilistic consensus through computational resources, reducing time and costs while enabling permission controls for security. Since pre-selected nodes generate blocks while others only validate, economic incentives are unnecessary.

4.5 Application Layer This layer provides distributed applications: blockchain services (smart contracts, blockchain browser, system monitoring) and user services (authentication, resource retrieval, knowledge services).

Blockchain Services: Smart contracts digitize and automate library services (document delivery, navigation, push notifications), reducing fraud risks and enabling customization via Solidity, C++, Java. The blockchain browser provides Web-based querying by block hash, transaction hash, height, or address. System monitoring visualizes resource usage (CPU, processes, memory).

User Services: Authentication (password, certificate, biometrics) grants authorized access. Resource retrieval uses blockchain transactions: TF-IDF [?] extracts query features submitted as transactions; nodes compute similarity between index values and features using threshold or top-n strategies; matching records retrieve resources via hashes. Knowledge services include IP protection via digital signatures and data auditing/analysis via big data.

4.6 Management Layer This layer maintains the architecture through identity management (adding/updating/deleting consortium members, ensuring uniqueness and security via encryption, anonymity, differential privacy), access management (permissions via matrices, lists, capabilities; graded protection for varying security needs), key management (hierarchical wallets), interface management (inter-layer and external system APIs), and load management (task scheduling based on resource usage for performance optimization).

5. Architecture Comparison and Analysis

Table 1 compares our architecture with existing ones on architecture, interoperability, scalability, security, and resource trustworthiness.

SOA-based [?] and cloud-based [?] architectures integrate resources for unified services, offering good interoperability and scalability but suffering from centralized bottlenecks, single-point failure, and security risks. **P2P architectures [?]** manage resources distributedly with good interoperability and scalability, but resource trustworthiness depends entirely on providers, increasing trust costs without adequate security. **Zeng and Qin's decentralized mobile search architecture [?]** lacks specific block structures and service processes, with insufficient consideration for heterogeneous resource interoperability and extensibility.

Our architecture solves resource trustworthiness and reduces trust costs without third-party endorsement. Compared to existing architectures, it enhances interoperability and security through customized block structures, standardized cryptography, and protocols, while improving scalability via cloud infrastructure, consortium deployment, and load optimization. Future work will address block size 膨胀, performance optimization, and privacy protection to facilitate practical implementation.

References

- [1] Leonardo C, Castelli D, Pagano P, et al. Setting the foundations of digital libraries: The DELOS manifesto[J]. D-Lib magazine, 2007, 13(3/4): doi:10.1045/march2007-castelli.
- [2] Gonçalves M A, Fox E A, Watson L T, et al. Streams, structures, spaces, scenarios, societies (5s): A formal model for digital libraries[J]. ACM transactions on information systems, 2004, 22(2): 270-312.
- [3] Candela L, Castelli D, Ferro N, et al. The DELOS digital library reference model. Foundations for digital libraries (Version 0.98)[M]. Pisa: ISTI-CNR at Gruppo ALI, 2008: 1-215.
- [4] Arms C B, Overly E A. An architecture for information in digital libraries[J/OL]. [2018-09-03]. <http://www.dlib.org/dlib/february97/cnri/02arms1.html>.
- [5] Zhang Xiaolin, Zeng Lei, Li Guangjian, et al. Standards and specifications for digital library construction[J]. Knowledge Management Forum, 2004, 2(1): 52-54.
- [6] Li Shuang. Design principles and system composition of digital library architecture[J]. Library and Information Service, 2007, 51(7): 145-149.
- [7] Feng Guohe. Research on digital library architecture[J]. Journal of Intelligence, 2009, 28(7): 14-16.
- [8] Zhang Hongli. Research on distributed digital library framework model based on cloud computing platform[J]. Information Science, 2013, 31(3): 40-67.
- [9] Wang Wenqing, Chen Ling. CALIS digital library cloud service platform model[J]. Journal of Academic Libraries, 2009, 27(4): 13-18, 32.

- [10] Yu Xin, Wang Jingyi. Digital library cloud service platform architecture based on cloud computing technology[J]. *Information Science*, 2011, 29(7): 1049-1053.
- [11] Liu Chengshan, Wang Jieliang, Qin Chunxiu. A peer-to-peer cloud-based digital library architecture[J]. *Information Science*, 2013, 36(11): 114-117.
- [12] Podnar I, Luu T, Rajman M, et al. A peer-to-peer architecture for information retrieval across digital library collections[C]//Gonzalo J, Thanos C, Verdejo M F, et al. *ECDL' 06*. Berlin: Springer, 2006: 14-25.
- [13] Higgins S. Library Chain? US government grants \$100k for new blockchain research[EB/OL]. [2017-09-01]. <https://www.coindesk.com/librarychain-us-government-grants-100k-new-blockchain-research/>.
- [14] Zeng Ziming, Qin Siqu. Decentralized management system for smart library mobile visual search[J]. *Information Science*, 2018, 36(1): 11-15.
- [15] He Yanhui. Construction and development of Russian national digital library[J]. *Journal of Library Science in China*, 2014, 40(5): 35-51.
- [16] Zhang Xiaolin. Distributed digital library mechanisms[J]. *Journal of the China Society for Scientific and Technical Information*, 2002, 21(1): 63-70.
- [17] Kahn R, Wilensky R. A framework for distributed digital object services[J]. *International journal on digital libraries*, 2006, 6(2): 115-123.
- [18] Payette S, Lagoze C. Flexible and extensible digital object and repository architecture (FEDORA)[C]//Nikolaou C, Stephanidis C. *ECDL' 98*. Berlin: Springer, 1998: 41-59.
- [19] Paepcke A, Chang C, Garcia-Molina H, et al. Interoperability for digital libraries worldwide[J]. *Communications of the ACM*, 1998, 41(4): 33-42.
- [20] Knezevic P, Mehta B, Niederée C, et al. Supporting information access in next generation digital library architectures[C]//Türker C, Agosti M, Schek H J. *Peer-to-Peer, grid, and service-orientation in digital library architectures*. Berlin: Springer, 2005: 207-222.
- [21] Li Guangjian. Development trends of next-generation digital library system architecture[J]. *Library Work and Study*, 2006, 26(113): 1-5.
- [22] Hou Fuli. SOA-based digital library platform architecture[J]. *Information Science*, 2007, 25(6): 903-908.
- [23] Dong L, Xing C X, Lin J, et al. Personalized digital library framework based on service oriented architecture[C]//Buchanan G, Masoodian M, Cunningham S J. *ICADL' 08*. Berlin: Springer, 2008: 351-354.
- [24] Yi Fei, Long Chaoyang. Construction of digital library alliance information service platform based on SOA[J]. *Library Work and Study*, 2011(6): 4-7.
- [25] Yin Hong, Liu Wei. Next-generation library service system: Functional evaluation and vision[J]. *Journal of Library Science in China*, 2013, 39(5): 26-33.
- [26] Shao Qifeng, Jin Cheqing, Zhang Zhao, et al. Blockchain technology: Architecture and progress[J]. *Chinese Journal of Computers*, 2018, 41(5): 969-988.
- [27] Dinh T T A, Liu R, Zhang M, et al. Untangling blockchain: A data processing view of blockchain systems[J]. *IEEE transactions on knowledge and data engineering*, 2018, 30(7): 1366-1385.

- [28] Yuan Yong, Wang Feiyue. Blockchain technology development status and prospect[J]. Acta Automatica Sinica, 2016, 42(4): 481-494.
- [29] Ministry of Industry and Information Technology. China blockchain technology and application development white paper[R/OL]. [2017-04-08]. <http://www.cbdf.org.cn/bcweb/index/article/rsr-6.html>.
- [30] China Blockchain Technology and Industry Development Forum. Blockchain data format specification[R/OL]. [2017-12-27]. <http://www.cesi.cn/cesi/201712/3465.html>.
- [31] Szabo N. Smart contracts[EB/OL]. [2017-01-01]. <http://www.fon.hum.uva.nl/rob/Courses/InformationInS>
- [32] Buterin V. A next-generation smart contract and decentralized application platform[R/OL]. [2017-04-08]. <https://github.com/ethereum/wiki/wiki/White-Paper>.
- [33] Xue Tengfei, Fu Qunchao, Wang Cong, et al. A blockchain-based medical data sharing model[J]. Acta Automatica Sinica, 2017(9): 1555-1562.
- [34] Group D A. Descriptive metadata guidelines for RLG cultural materials[EB/OL]. [2017-04-15]. https://www.oclc.org/content/dam/research/activities/culturalmaterials/RLG_{{
- [35] Salton G, Buckley C. Term-weighting approaches in automatic text retrieval[J]. Information processing & management, 1988, 24(5): 513-523.
- [36] Wu Zongda, Xie Jian, Zheng Chengren, et al. Privacy protection framework for digital library user behavior preferences[J]. Journal of Library Science in China, 2018, 44(2): 72-85.

Author Contributions: Gao Sheng was responsible for overall framework, writing, and revision; Zhu Jianming was responsible for paper revision.

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv – Machine translation. Verify with original.