Research on scientific data sharing management mechanism based on blockchain technology

Maohu Tao
School of Mathematics and Statistics, Hubei University of Arts and Science, Xiangyang, Hubei 441053, China

Abstract: The shared management of scientific research data is conducive to promoting the circulation and reuse of scientific data and realizing the value-added of scientific data. With the development of scientific research and the exponential growth of data scale, the traditional data sharing management has gradually revealed its limitations. As a decentralized, secure and trusted distributed ledger technology, blockchain technology is highly compatible with the bottleneck that scientific data sharing management needs to break through, and helps to solve the long-standing problems of scientific data sharing management. This paper aims to explore the scientific data sharing management mechanism based on blockchain technology, in order to provide new ideas and methods for scientific data sharing management.

Key words: blockchain technology; Scientific data; Data sharing; Mechanism study

1.Introduction

With the deepening of scientific research and the explosive growth of data volume, the sharing and management of scientific data is conducive to promoting the circulation and reuse of scientific data and realizing the value-added of scientific data. The traditional scientific data sharing management mode has many problems, such as weak traceability ability, untraceable data use, difficult to guarantee data security, low credibility and many other challenges.As a decentralized, immutable, traceable distributed ledger technology, blockchain technology is highly compatible with the bottleneck that scientific data sharing management needs to break through, providing new possibilities for solving these problems.

The use of blockchain technology to achieve scientific data sharing and management has become the focus of research scholars, and a large number of research results have been produced.Lu Lina et al. constructed AgriDSM, an agricultural science data management model based on alliance chain. The model includes data acquisition layer based on agricultural Internet of Things, data storage layer based on encryption and cloud storage, data processing layer based on Fabric alliance chain and application layer based on agricultural science data sharing subsystems.Gu Jun et al. use the data recording mechanism of blockchain and choose Hyperledger Fabric blockchain framework is used as the foundation of the alliance chain, and the data storage method of the block is rewritten. Through the design of CA authentication, pre-submission, verification feedback, block packaging broadcast, ledger database update and other processes, the humanities and social science data sharing alliance chain model is built.Xu Shangying et al. incorporated the blockchain as a service (BaaS) architecture and thought methods into the science and technology service data management to build the data sharing system structure, and built the data sharing implementation mechanism based on the system structure and data classification structure.Hao Shibo et al.built a hierarchical blockchain architecture for scientific data sharing, deeply discussed the implementation mechanism of scientific data sharing blockchain interactive information, data blocks, consensus incentive mechanism, smart contracts and other aspects, and proposed the dilemma and countermeasures of scientific data sharing blockchain in terms of interactive data description, security, and consensus game.Wang Xianbindesigned the model from three aspects: consensus mechanism, block structure and smart contract, and proposed the overall model of scientific data sharing based on blockchain.Liao Wenjiediscussed the application concept of blockchain in scientific data supervision from the dimensions of data collection, storage, access, intellectual property protection and data alliance construction, and proposed optimization strategies and suggestions for the application of blockchain in scientific data supervision.Xia Xiaoliang et al.proposed a medical data classification encryption sharing scheme based on blockchain, combining attribute-based encryption and blockchain to achieve access control and data sharing of medical data.Lu Lina et al.built a blockchain based agricultural science data traceability model, providing a new research idea for agricultural science data traceability management;Zhang Xiaoting et al.designed a genetic data sharing model based on blockchain, using blockchain technology, cryptography technology and trusted hardware to ensure the privacy of genetic data; Smart contracts are used to deal with the sharing process between sharing parties, and a safe and feasible gene data sharing model is established. Jiang Xin et al.targeted the centralized storage, low sharing efficiency and privacy leakage of medical data.Zou Pengproposed a blockchain-based big data sharing model, starting from three aspects of the network's data reliable connection mechanism, data empowerment access

control mechanism, and data service customization mechanism, and carried out research on the key mechanism of the blockchain-based big data sharing model.Peng Yuan et al.proposed a hospital blockchain data sharing method based on the improved PBFT algorithm. By improving the master node selection strategy of the PBFT algorithm, and by setting the block verification stage and the verification threshold of generating formal blocks, the security of data sharing was improved.Sun Yuhong et al.built a data exchange incentive framework based on blockchain, using the data exchange incentive mechanism of historical contribution and current contribution, which can effectively motivate users to carry out positive data exchange in the long run.Peng Lu et al.proposed a secure office data sharing scheme using blockchain technology and attribute encryption technology;Wu Hao et al.proposed a data sharing scheme based on side chain and trust management model, and provided users with a trusted sharing environment and dynamic access control interface, and users dynamically decided to share data in the main chain or side chain according to a pre-set threshold.

2. Use blockchain technology for coupling analysis of scientific data sharing management

2.1 Principle and characteristics of blockchain technology

Blockchain technology as a decentralized distributed ledger technology, its core principles and main characteristics can be analyzed from many aspects. First, blockchain technology is built on a distributed network of nodes. Every node in the whole network participates equally and jointly maintains a distributed database that records all transaction information. This decentralized network structure can protect the entire system from the risk of single point of failure and greatly improve the fault tolerance and reliability of the system. Secondly, blockchain adopts a set of cryptographic algorithm-based accounting transaction rules, through distributed storage and computing power, forming a data chain structure that can only be added according to encryption rules, but can not be tampered with, once incorporated into the blockchain, these data will be extremely difficult to be tampered with, thus ensuring the credibility and immutability of the data. Furthermore, blockchain technology uses a consensus algorithm to ensure that all nodes in the entire network are consistent with the transaction data. Mainstream consensus mechanisms include proof of work (POW), proof of stake (POS), etc. Under the operation of these consensus mechanisms, there is a proper incentive mechanism in the network to ensure that most nodes can comply with the established rules, thus maintaining the operation order of the entire system. In addition, smart contracts write complex business logic into code, and then achieve the smooth execution of various transaction transactions through a highly automated way, greatly reducing the risk of human intervention, as well as the risk of operational error. Blockchain technology has the openness and transparency, open to everyone, anyone can participate in the node network, node data is fully transparent, open and available, this open sharing mechanism, not only to ensure the fairness of the system, but also to all node participants fair benefit.

In general, blockchain technology builds a decentralized, immutable, traceable data storage and circulation network through distributed ledger and cryptography principles, realizes the high trust of digital assets and transaction information, and realizes the efficient management, and its natural decentralization, open and transparent characteristics. It can be applied in the fields of data sharing, copyright protection, supply chain traceability and so on.

2.2 Coupling analysis of blockchain technology principles and characteristics with scientific data sharing management

2.2.1 Enhanced data security and tamper resistance

The application of blockchain technology in scientific data management can greatly enhance the security of data, as well as enhance the tamper resistance of data, mainly due to the immutable, decentralized and encrypted characteristics of blockchain technology itself.First, the blockchain is essentially a distributed data ledger, where all nodes jointly maintain the same copy of the data, and once the data is recorded on the blockchain, it is distributed to nodes across the network, forming multiple redundant copies. Moreover, any attempt to tamper with the data on a node will be different from other nodes, so as to be found and resisted, this mechanism based on multi-node consensus makes the data extremely difficult to tamper with once on the chain. Second, blockchain is able to encrypt and verify transaction data through cryptographic algorithms, in which each block contains the hash value of the previous block, and all blocks are related to each other through the hash value to form an immutable chain. At the same time, any tampering with the content of a block will cause all the block hash values after the block to become invalid, and thus be found and rejected by the network. This cryptographic protection mechanism greatly improves the integrity and security of data. Moreover, blockchain adopts distributed ledger technology, there is no centralized single point of failure risk, and all nodes participate in the maintenance of the network equally, there is no single authority node, any individual can not control and tamper with the data independently, fundamentally eliminate the risk of data being subjective manipulation, it can be seen that the whole system has natural fault tolerance and tamper-proof. In addition, blockchain also has traceability and audit capabilities. For every data generation and flow, it will be recorded on the immutable blockchain, forming a permanently traceable data background and path, if the data

is illegally stolen or abused, it can be traced from the blockchain to the data flow and responsible person, providing a strong guarantee for the supervision and audit of data use. It is worth mentioning that the above data security protection mechanism is not a simple implementation of a single technology, but requires the integration of cryptography, distributed systems, consensus algorithms and other advanced technologies, and continuous optimization and improvement in practice, and it is this multi-technology coupling application, to completely eliminate the single point of failure, subjective manipulation and other traditional centralized system faced with many security risks. It provides a higher level of technical guarantee for data security. In general, blockchain technology has brought a new paradigm for the security management of scientific data, which is inherently decentralized and immutable. Combined with the comprehensive application of advanced technical means such as cryptography and distributed consensus, it fundamentally improves the security and tamper-proof of data, and effectively protects the integrity and credibility of data assets. It provides reliable technical support for the efficient and safe circulation of data resources.

2.2.2 Enhanced data sharing and transparency

The application of blockchain technology in scientific data management is conducive to improving data sharing and transparency. First, the openness and transparency of the blockchain system itself lays an important foundation for the open sharing of data. All the data and algorithm logic of blockchain are basically completely open and transparent, and anyone can access the records and audit these records, so that the production, processing and circulation of scientific research data are under the sun, eliminating the problem of information asymmetry in the traditional centralized system. This transparent sharing mechanism is conducive to promoting open collaboration in the field of scientific research. Reduce data silos and waste. Second, blockchain can build a trusted peer-to-peer data sharing network, in which the participating nodes can share data resources securely with each other without intermediaries through encrypted identity authentication, and then cooperate with smart contracts to set conditions and rules for each data interaction, clearly agree on the rights and obligations of each participant, and promote benign interaction among multiple parties. This decentralized sharing paradigm not only eliminates the trust barriers between data institutions, but also avoids the time and cost overhead brought by intermediaries. Moreover, blockchain has excellent data traceability capabilities. The production, processing and circulation of data resources leave a permanent and auditable digital footprint. By tracing these immutable data, the source of the data, the change process and the responsible party can be accurately located. This data traceability mechanism is not only conducive to maintaining the real credibility of data, but also provides a basis for discovering and dealing with improper behaviors, and ensures reasonable compliance of data sharing behaviors. In addition, the blockchain system has an inherent incentive mechanism that can set reasonable digital token rewards for data providers and users, thereby promoting the efficient flow of data resources between different nodes. Then, by setting up cryptocurrency rewards, credit accumulation and other ways to motivate data contributors, a global data crowdsourcing network can be formed, so that researchers around the world can share the value dividend of scientific research data, thereby further enhancing the willingness and enthusiasm of scientific research data sharing. It is worth mentioning that the above data sharing and transparency mechanism is not achieved by blockchain alone, but relies on the integrated support of a series of advanced technologies such as distributed storage, privacy computing, digital identity, and smart contracts. These technologies are integrated with blockchain to form a new technical system that supports transparent data sharing, significantly enhancing the social value of scientific research data.

2.2.3 Data ownership and permission management

The application of blockchain technology in scientific data management is also conducive to the accurate control of data ownership and use rights. First of all, the encryption technology and smart contract means used by blockchain provide a solid technical foundation for the confirmation of data ownership. Each linked block of data generates a unique digital fingerprint and is permanently associated with that data, which is like embedding "digital DNA" into the data asset. In this way, the creator and ownership of each data can be accurately recorded, and with the help of the automatic execution mechanism of smart contracts, the ownership and disposal rights of different subjects are clearly defined. Second, the decentralized peer-to-peer network structure of blockchain can avoid the single trust risk in the traditional centralized system. This kind of different data rights holders participate in the network as equal nodes and reach consensus on data rights and interests through the consensus mechanism, fundamentally eliminating the subjective manipulation and abuse of power in the centralized system. This consensus-based distributed trust mechanism ensures the fairness and authority of data ownership and usage rights. Moreover, the blockchain network has natural traceability and audit capabilities, which can provide a complete ownership transfer record for data rights management. Every change in data ownership or use rights will be permanently recorded on the blockchain as a new transaction, and then, by tracing these immutable chains of evidence, the rights ownership and authorized use of any data asset can be accurately verified to ensure compliance and accountability. In addition, blockchain can achieve refined data usage rights control through digital identity authentication and access control policies. Typically, through authentication, each subject can have a unique trusted identity in the network,

and then through pre-deployed smart contract policies, different subjects can be granted different access rights to specific data on demand, achieving security level detailed permission control, and the implementation process of these policies is transparent and auditable. It is worth mentioning that the above data rights control mechanism can not be achieved by blockchain technology alone, but also needs to integrate the support of cryptography principles, identity authentication, access control, privacy computing and other related technologies to form a systematic comprehensive solution. At the same time, these technical solutions also need the policy support of laws and regulations to form a positive interaction between law and technology, in order to give full play to its role in promoting the protection of data rights and interests.

3. Design of scientific data sharing management mechanism based on blockchain technology

3.1 Data storage and authentication mechanism

In the scientific data sharing management mechanism based on blockchain technology, the data storage and verification mechanism is the key link to ensure the authenticity and integrity of the data, as well as the key link of immutable, which needs to be carefully designed by combining the characteristics of blockchain technology and integrating the characteristics of scientific data itself

First of all, scientific data itself has the characteristics of massive, multi-source heterogeneity, high value, etc. Therefore, it is necessary to pre-process the original data before storage, including format conversion, de-encryption, etc. to standardize it into a data format that can be efficiently stored in the blockchain system. At the same time, multi-level redundant storage strategies can be designed for the same data set to balance storage efficiency and fault tolerance. Secondly, the pre-processed data needs to be stored in blocks in the blockchain system according to a certain organization. This can use content-based data sharding method, the data is divided into independent data blocks, and each data block is encrypted and stored separately, and recorded on the blockchain, which can effectively reduce the amount of data in a single block, improve storage and verification efficiency. Then, the corresponding digital fingerprint needs to be generated for each data block as a cryptographic proof of its existence and integrity. Among them, the commonly used digital fingerprint algorithms include SHA256, SHA3, etc., and the generation process of digital fingerprints needs to introduce random seeds from a trusted random number generation mechanism to ensure the unpredictability of fingerprints. In addition, the generated digital fingerprint will be permanently associated with the original data block through the smart contract and linked to the card to ensure a dynamic and trusted mapping relationship between the fingerprint and the data. Any data tampering will result in changes to its fingerprint that can be detected by the blockchain network. In addition, in order to improve the efficiency of data verification, a multi-level index structure can be built in the blockchain system, such as building a data block index layer based on B+ tree, building an upper data fingerprint index layer based on Merkle tree, forming a multi-level index link. At the same time, it is necessary to quickly locate the target data fingerprint based on the index layer during verification. Then the verification can be completed by real-time comparison with the corresponding data block of the storage layer. In addition, while implementing data storage and verification, data privacy needs to be protected. For open data, it can be stored directly in plain text. For sensitive data, security preprocessing such as desensitization and encryption is required, and only authorized visitors can access the real content. On the whole, the data storage and verification mechanism organically combines blockchain technology with scientific data processing, makes full use of the immutable and traceable characteristics of blockchain, and solves the storage and verification problems of massive scientific data through reasonable technical design, laying the foundation for the safe and efficient circulation of scientific data resources.

3.2 Data rights and access control

In the blockchain-based scientific data sharing management mechanism, properly designing the data permission and access control mechanism is the top priority to ensure data security and use compliance, which needs to combine the decentralized characteristics of blockchain technology, and fully consider the sensitivity of scientific data resources and the diversity of use scenarios.

First of all, we need to establish a perfect data permission model. This can learn from RBAC (role-based access control), ABAC (attribute-based access control) and other mature permission models, and then set different access permission levels and different authorization policies for data resources according to data sensitivity and usage scenarios. Generally, the rights are divided into private, internal, partner, public and other levels, and the scope of data that can be accessed at each level is clearly defined. Secondly, the permission control policy is encoded into the blockchain network in the form of a smart contract, which can ensure the immutability of the permission policy, and at the same time can realize the automation of execution. The smart contract can be seamlessly integrated with the data storage module to automatically verify the identity and permissions of the visitor when the data is accessed, release the access request that meets the conditions, otherwise deny access, and prevent data leakage at the source. When implementing permission control, you need to introduce an identity authentication mechanism. An identity authentication scheme based on PKI (Public key Infrastructure) can be adopted to issue

digital certificates to each visitor and issue corresponding permissions based on the attributes of the certificates. The certificate key pair needs to be securely distributed and updated among nodes of the blockchain network, and at the same time, a third-party trusted digital identity authority can be introduced to ensure the authority of identity authentication. In addition, for public data that needs to be open and shared, a public access portal can be set up in the blockchain system for easy access by external users, but for private or internal data, strict identity authentication and authorization are required for visitors, and only after approval can they be consulted. In addition, for data usage scenarios, it is necessary to set up a regulatory mechanism for data flow, which requires the deployment of specialized regulatory smart contracts to record each data access, and thus form a traceable data flow audit chain. Once a violation is found, the data can be locked based on the record and the offending party can be severely punished. This greatly enhances the compliance and auditability of data usage. It is worth mentioning that many of the above core technical links have a large number of theoretical research data and practical case data sources as technical support, such as NIST's RBAC model, IBM's Hyperledger Fabric, medical data sharing project, etc., so as to ensure the feasibility and effectiveness of the data permission and access control mechanism.

3.3 Data transaction and incentive mechanism

In the scientific data sharing management mechanism based on blockchain, the design of a reasonable data transaction and incentive mechanism is crucial to promote the efficient flow of data resources, which needs to combine the decentralized characteristics of blockchain technology, and fully consider the fairness and security of data transactions and the return mechanism of participants' contributions.

First, a unified data trading market needs to be built. This requires the deployment of specialized data transaction smart contracts on the blockchain network, and then provide data providers and users with functional entrances to publish data resources and search transaction data, and achieve seamless docking and transaction compatibility between different data sets through standardized data service interfaces. Secondly, in the process of data transaction, the cryptocurrency mechanism native to the blockchain needs to be introduced. Relevant staff can design payment and settlement processes in combination with some mature digital currencies, such as bitcoin, Ethereum, etc. At the same time, for data users, it is necessary to use cryptocurrency for online payment and purchase the right to use the required data resources, while the data provider can obtain the corresponding monetary income, forming the free flow of data assets. In order to ensure the fairness and credibility of transactions, it is necessary to design a transaction supervision module based on consensus mechanism. This can be done by using classic consensus algorithms such as PoW (proof of work) and DPoS (Proof of equity interest) to ensure that the majority of nodes in the whole network reach agreement on each transaction data, and record it permanently and immutable in the blockchain ledger, effectively preventing forgery such as double payment. At the same time, it is also necessary to design a reasonable fee mechanism, which can charge an appropriate fee for each transaction, and this part of the fee will be included in the incentive pool as a return for network contributors (such as operation nodes, audit nodes, etc.). Through this decentralized fund allocation mechanism, all parties can be mobilized to participate in the enthusiasm of the network construction. For individuals or institutions that make outstanding contributions to data trading and network maintenance, additional incentive rewards can be set up, such as providing high-quality data resources, optimizing consensus algorithms and other behaviors to give corresponding rewards to encourage and encourage. It is worth mentioning that many of the above transaction links and mechanism design have a large number of practical cases for reference, such as Gartner's blockchain data market, Baidu's AI data transaction network, Opu ere project, AXA aviation insurance project, etc., which provide strong technical support and operational experience for the implementation of data transaction and incentive mechanism.

4. Conclusion

As an important data asset, the core of scientific data lies in the sharing and circulation of data to realize the value increase of scientific data due to reuse. However, in the sharing and circulation of scientific data, there are many challenges such as weak traceability ability, untraceable data use, difficult to guarantee data security and low credibility. As a decentralized, immutable, traceable distributed ledger technology, blockchain technology is highly compatible with the bottleneck that scientific data sharing management needs to break through, providing new possibilities for solving these problems. This paper focuses on how to use blockchain technology to promote scientific data sharing and management, analyzes the coupling between the principle and characteristics of blockchain technology and scientific data sharing and management, and further studies the scientific data sharing management mechanism based on blockchain technology from three aspects: data storage and verification mechanism, data permission and access control, data transaction and incentive mechanism. With the continuous development and improvement of blockchain technology, blockchain technology will become an important tool to promote the sharing and management of scientific data, promote the safe, reliable and efficient sharing of scientific data, and promote the vigorous development of scientific research.

Reference

[1] Zhao Xiande, Tang Fangfang. Zhao Xiande, Tang Fangfang. Blockchain Empowering Supply chain [M]. China Renmin University Press: 202201.

[2] Lu Lina, Yin Jufeng, Yu Xiao, et al. Agricultural scientific data sharing based on the alliance chain model building research [J]. Journal of library intelligence, 2022,66 (17) : 60-68. The DOI: 10.13266 / j.i SSN. 0252-3116.2022.17.006.

[3] Gu Jun, Xu Xin. Design and implementation of Humanistic and Social Science data sharing model: A case study of alliance link technology [J]. Journal of Information Technology,2019,38(04):354-367.

[4] Xu Shangying, Wang Qian, Chen Donglin, et al. Technology service data sharing mechanism integrating blockchain as a service [J]. Science and Technology Management Research,2021,41(06):116-123.

[5] Hao Shibo, Xu Wenzhe, TANG Zhengyun. Scientific data sharing block chain model and implementation mechanism research [J]. Journal of theory and practice of intelligence, 2018, 9 (11) : 57-62. DOI: 10.16353 / j.carol carroll nki. 1000-7490.2018.11.010.

[6] Wang Xianbin. Scientific data sharing model based on block chain research [J]. Journal of modern information technology, 2019, 3 (21) : 156-158. The DOI: 10.19850 / j.carol carroll nki. 2096-4706.2019.21.059.

[7] Liao Wenjie, Zhao Limei. Block chain technology in the application of the scientific data supervision idea [J]. Journal of library, lancet, 2021 (8) : 77-84. The DOI: 10.14037 / j.carol carroll nki TSGXK. 2021.08.016.

[8] Xia Xiaoliang, Qin Zhi, Wan Wonan, et al. Medical data classification encryption sharing scheme based on blockchain [J]. Chinese Journal of Applied Sciences, 2019,42(04):613-628.

[9] Lu Lina, Yin Lihong, Yu Xiao, et al. Model of agricultural science data source based on block chain build research [J/OL]. Information science, 1-13 [2024-08-01]. HTTP: / / http://kns.cnki.net/kcms/detail/22.1264.g2.20240729.1133.004.html.

[10] Zhang Xiaoting, Li Xiaofeng, Zhao He, et al. Genetic data security Sharing Model based on blockchain [J]. Computer Applications and Software, 2019,41(07):270-278. (in Chinese)

[11] [1] Jiang Xin, Wang Chundong. Sharing medical data security model based on block chain research [J/OL]. Journal of tianjin university of science and technology, 1-7 [2024-08-01]. HTTP: / / http://kns.cnki.net/kcms/detail/12.1374.N.20240328.1417.067.html.

[12] Zou Peng. Big data sharing based on the block chain model and the research and implementation of the key mechanisms [J]. Journal of information science and technology, 2024, 22 (5) : 39-41. DOI: 10.16661 / j.carol carroll nki. 1672-3791.2311-5042-5289.

[13] PENG Yuan, LIU Yun, JIANG Bright-colored. Block chain for sharing data security technology research [J]. Automation and instrumentation, 2024, (02) : 105-108 + 113. DOI: 10.14016 / j.carol carroll nki. 1001-9227.2024.02.105.

[14] Sun Yuhong, Zhang Yujian, Zhao Yue, et al. Data exchange based on the block chain incentive mechanism research [J]. Journal of network security and data management, lancet 2024 (02) : 23 to 28 + 36 DOI: 10.19358 / j.i SSN. 2097-1788.2024.02.004.

[15] Peng Lu, Ye Wei. Office data sharing Scheme based on blockchain and attribute encryption [J]. Journal of North China University of Technology, 2019,36(01):50-57.

[16] Wu Hao, Li Bei, He Xiaowei, et al. Based on the side chain and trust management model data sharing scheme [J]. Computer engineering and design, 2024, (01) : 24-31. DOI: 10.16208 / j.i ssn1000-7024.2024.01.004.

[17] Bi Da Tian, Cao Ran, Du Xiaomin. Research status and prospect of scientific data sharing [J]. Library and Information Work, 2019, 63 (24): 69-77. (in Chinese)

[18] Li Ying, Liu Jingbo. Scientific Data Management and Sharing based on DOI RAs mechanism: An overview of the First China-Japan-ROK DOI Project Cooperation Symposium [J]. Digital Library Forum, 2013, (08): 37-41.

Project information: Blockchain operation architecture and benefit distribution mechanism design of scientific data sharing Alliance (Project No. : B2021211)