Research on Scientific Data Curation Model of Institutional Knowledge Repository Integrated with Block Chain Technology

Nan Shi, Yanhui Shi

670118408@qq.com, Syh_sd@qq.com

Wuhan Railway Vocational College of Technology Wuhan, China, Jiangsu University of Science and Technology, Zhenjiang，China

Abstract—Building a data curation platform that can be used and saved for a long time is the core link and key of scientific research data curation service． Aiming at scientific data curation platform of institutional repository data standardiza- tion，dependability，security and privacy problem，using the block chain technology，such as security，justice and decen- tralized features，put forward chain technology into blocks of institutional repository of scientific data curation model． This model can realize the basic function of scientific data curation of institutional knowledge base，ensure the security of data storage，and improve the efficiency and utilization rate of scientific data sharing and intellectual property protection.

Keywords- block chain; institutional knowledge repository; data curation; scientific data; data management

1 Introduction

Since 1990s, the Open Access Movement (Open Access, OA) has been on the rise in the fields of library intelligence, editorial publishing, and news communications [1]. This initiative aimed at eliminating price barriers and licensing barriers between scientific data and promoting the sharing and widespread use of scientific data quickly gained widespread attention. Under the advocacy of open access, universities and scientific research institutions at China and abroad have taken the construction of institutional knowledge base as the focus of their scientific research information service. The original intention of setting up the institutional knowledge base is mainly to achieve two points: One is to realize the open access of scientific research results in the institutional repository, break the traditional data sharing barriers, and promote the exchange and reproduction of knowledge ; The second is that the scientific research results of the construction institution can be preserved for a long time, the academic influence of the institution and the display of the results can be improved, and the academic reputation, academic level and social value of the institution can be highlighted [2].

ICIDC 2023, June 02-04, Nanchang, People's Republic of China

DOI 10.4108/eai.2-6-2023.2334588

The institutional repositories have developed rapidly after being put forward, but behind the widespread attention and rapid development, the construction and application of institutional repositories have also encountered some problems: On the one hand, universities and scientific research institutions undertaking the construction of institutional knowledge bases are easily subject to their knowledge resource reserves, professional talent reserves, capital and technical constraints, and the scope of services for a single institution can easily limit the sustainable development of the institutional knowledge base. On the other hand, compared with traditional subject libraries, user awareness and user participation are not ideal, which is mainly due to the vague positioning of the institutional knowledge base and the imperfect service concept. In response to these issues, librarians need to assist scholars to improve data and metadata to promote sharing, actively help form the release of data results to the subject libraries of their respective fields, and build a data monitoring platform that can be used and stored for a long time [3].

The current data monitoring platform was insufficient in the sustainability and scalability of data organization [4]. There are many unresolved practical problems such as the risk of damage [5], modification, leakage or loss of stored data, insufficient data content disclosure and service methods [6], and prominent contradictions between data heterogeneity and data format standardization technology. Blockchain is in the ascendant Technology provides a possible solution to the above problems due to its security, fairness, and decentralization. This article attempts to integrate the relevant ideas of double chain blockchain technology into the data guardianship service, and proposes the corresponding model architecture to provide ideas for future data guardianship services.

2 The theoretical basis of blockchain technology and the feasibility of constructing the institutional knowledge base scientific data monitoring model

2.1 Theoretical basis of blockchain technology

Blockchain is a distributed database system involving multiple independent nodes. It can also be understood as a distributed ledger (DLT, Distributed Ledger Technology) maintained by these independent nodes. This is a kind of Data recording method based on decentralization and trustless thought. From this extension, it can be seen that blockchain technology is a technical solution that does not rely on any third party and conducts data interaction, verification, and storage through its own distributed nodes [7].

The reason why blockchain technology has attracted much attention is largely because this technical solution has fundamentally changed people's trust model. The traditional process of information exchange and value exchange must be carried out through an intermediary. As a third party in information exchange and value exchange, the intermediary builds a bridge for each node in the network that does not trust each other. Information and value are concentrated to the central node, after screening the central node are assigned to the target node. This centralized organizational form increases costs and reduces efficiency for the exchange of information and value. Once the central node is attacked, it will threaten the security of the

entire network. The blockchain technology provides a method that does not need to trust a single node and can create a consensus network to solve the basic problem in peer-to-peer communication which was Byzantine Failures.

The working principle of the blockchain (Figure 1) is to divide data into The same Block, the body of each block stores the Item, the header of the block contains the hash calculated by the Hash Function of the block header of the previous block Value, each block is closely connected to the previous block by the hash value of its block header to form a Chain [8].

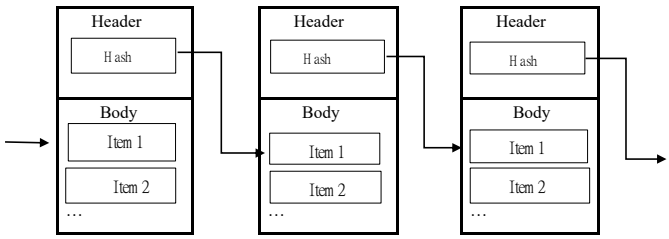

Figure 1. Schematic diagram of the working principle of blockchain

This block chain structure is completely time-series, and each block is given a time stamp that cannot be tampered with or forged when it is created. This allows the entire database have a complete and traceable history; each node in the network operates independently, and they enjoy the same rights and obligations, which ensures that there will never be a central node with special power in the network, which truly realizes decentralization.

2.2 Feasibility of Blockchain Technology for Building a Scientific Data Guardianship Model for Institutional Knowledge Bases

At present, the construction of institutional database has become an important area of university library management and service model innovation. A fully functional institutional repository can continuously expand library resources and provide professional scientific data management and knowledge sharing services. The use of institutional knowledge bases for data monitoring has achieved significant results but also has many problems. It is mainly reflected in the insufficient organization and standardization of scientific data submitted by various departments and scientific research personnel, the uneven quality of data submission [9], the difficulty of interconnection and intercommunication between scientific data, and the difficulty of information sharing and collaboration. There is a centralization problem. Most institutional repositories are centrally managed by a single institution, and data is tampered with, and users cannot verify the authenticity [10]. The lack of in-depth indexing of data content makes it impossible to provide one-stop retrieval services for scientific researchers. In the process of knowledge base construction and management, there are security and privacy issues, and intellectual property issues need to be protected urgently [11]. Blockchain technology is feasible to solve the problems in the institutional database. As a distributed database system involving multiple independent nodes, the blockchain itself is a database technology, and its applications, storage objects, and technical elements are the same as the institutional knowledge base. Scientific data has a data life cycle, and the time chain characteristics of blockchain technology

also fully match this expectation. It can be seen that it is completely feasible to use blockchain technology to solve the problems in the scientific data guardianship of institutional knowledge bases [12].

Integrating blockchain technology to build a scientific data guardianship model of institutional knowledge base has the following advantages. Firstly, the decentralization of the blockchain, which can make the constructed institutional knowledge base platform more secure and trustworthy, and make the collected scientific research data more accurate and standardized.Secondly, the technology of distributed ledger, block chain data structure, asymmetric encryption algorithm and smart contract in blockchain technology, which can ensure that the scientific data in the institution's database cannot be tampered with, and provide for the security of data storage in the institution's database. Thirdly, the decentralization of blockchain technology can greatly improve the speed of information dissemination and resource utilization efficiency in the network, which can provide great convenience to users for one-stop search and use. Fourthly, to use the national accounting and asymmetric encryption algorithms in the blockchain to provide intellectual property protection for users who upload scientific data. Only when the user provider decrypts permission can download and use information, and data security and privacy issues can be resolved . Therefore, the following describes the functional requirements and working principles of the scientific data guardianship model of the institutional repository based on the basic ideas of the blockchain.

3 Functional requirements of scientific data monitoring model

The scientific data guardianship service based on institutional knowledge base needs to follow a specific goal, follow a specific process, set up the corresponding functions, then set up functional modules and hierarchical architecture according to the functional requirements, and further elaborate the working principle and process of the model.

A complete scientific data monitoring platform should involve all aspects of the data life cycle, covering the cyclical process from the beginning of data generation to data storage and data utilization. The data monitoring model proposed by the DCC (Digital Curation Centre) globally divides the data life cycle into 4 stages, which are information description and representation, data preservation plan, organization observation and participation, data monitoring and preservation; And put forward the general work and specific work of data guardianship service involving data production, data organization, data sharing, data storage and other links, its service scope covers the entire cycle from scientific research project conception to scientific research results utilization[13]. With reference to the process division and function setting of DCC, this article divides scientific data monitoring into four basic functions: data collection and evaluation, data organization and processing, data storage and release, data sharing and utilization, each of which can be further detailed Divided into several interrelated specific functions, integrate the relevant principles of dual-chain blockchain technology, and organically integrate the basic modules that can achieve specific functions according to the workflow, and then obtain the overall scientific data monitoring model. The specific functions of the scientific data monitoring model constructed in this paper are shown in Table 1. Abbreviations and Acronyms.

Table 1 Functional requirements of the scientific data guardianship model of the institutional knowledge base

Basic Functions	Specific function	function description
Data Collection and Evaluation	File upload	Users upload files and related additional information, extract user and related responsible person information
	File screening	Preliminary screening of the document's readability, completeness, standardization and usage rights
	Document Quality evaluation	Assist authors to evaluate the maturity of scientific research results, the novelty of data content and academic value in the document, and formulate corresponding data management plans
	File Archiving	Back up, archive and create indexes of the evaluated files
Data organization and processing	Divide types division	According to the purpose, form, maturity, and openness of the data, and formulate corresponding data organization strategie.
	File format conversion storage	Convert file formats by data type and save them in a common file format for easy
	Data description	According to different disciplines, platforms and data types, set different metadata
	Establishing data associations	Establish scientific data associations based on the correlation between disciplines, institutions, projects, documents, and data to facilitate data sharing
Data Storage and Distribution	Data Storage Plan periods and data backup plans warehousing held by	Develop different data storage plans according to the data storage needs, set data retention Relying on the database of each joining institution and the public data registered users for distributed storage.
	Distributed Storage	Prevents data tampering and loss
	Data Release	Publish scientific data in accordance with the data sharing strategy and to consortium members, external repositories and associated data clouds according to author intent and disciplinary needs
Data Sharing and Utilization	Information Search	The data guardian platform provides various types of search methods to facilitate users to discover and browse the required information resources
	Data Access	Provide multi-functional data access interface and long-lasting data access methods, this part of the function can be based on the institutions Different flexible settings for platform portals
	Data Download	Open data download service for users within the shared scope and mark and count the citation relationship between scientific data, track the impact of the data, and adjust the data at any time

4 Construction and functional elaboration of scientific data guardianship model incorporating blockchain technology

Scientific data monitoring itself is a scientific data management concept and service project. It will continue to integrate new technologies within its original service concept to achieve better preservation, sharing and value-added scientific data. The above four basic functions and 15 specific functions need to be realized by different functional modules that integrate various technologies. Many of these links have been successfully implemented in reality. Therefore, in the subsequent model construction links, this article will focus on explaining the role of dual-chain blockchain technology in the scientific data guardianship service for institutional knowledge bases or institutional knowledge base alliances, and maintain scientific data guardianship to a large extent The flexibility and scalability of the platform in the selection of specific functions.

4.1 Overall model architecture

The institutional knowledge database scientific data guardianship model fused with blockchain technology is built for the institutional repository alliance. Combining the above-mentioned scientific data guardianship related functions and the relevant characteristics of blockchain technology, the model can be divided into 3 levels as a whole, respectively are the upper user

interface, the middle organization setting, and the lower alliance setting. The settings of the times are:

1）User interface. The user interface provides an interactive platform for all users within the service range, supports interoperability between PC and mobile terminals, can provide both web page version and App application user interface at the same time, and develops an API platform that can be embedded in other commonly used open-source software. The role of the user interface is mainly to guide users to submit, obtain and use scientific data as required, and to provide corresponding interfaces. The current user interface has 5 basic functions, which can be increased or decreased according to specific needs.

2 ） Institutional settings. The member institutions mainly composed of libraries and information centers of universities and scientific research institutes are the main body of the Institutional Repository Alliance. They are also the main advocates and executors of data guardianship services. University libraries and information centers focus on hardware facilities, staffing, The professional foundation can provide a guarantee and platform for the development of data guardianship services[14] The member institutions of the Institutional Repository Alliance can establish their own data organization platforms and data warehouses within a relatively unified technical framework and follow certain data storage protocols, and can share data with other members of the alliance in accordance with the corresponding data sharing agreements . The institutional setting layer is mainly responsible for the functions of data collection and evaluation, data description and processing in data monitoring. The realization of these two functions mainly uses information collection, information retrieval, knowledge organization, data mining, measurement analysis and visualization methods Finished.

3 ） Consortium setup. The data guardian platform of the institutional knowledge base federation is the operation layer of the blockchain, which contains a blockchain timing server, three independent blockchains that can synchronize their states through connectors, and a distributed storage server that can be invoked through instructions. The blockchain system in the federation setup is the core difference between the proposed data guardianship model and the traditional model. The design aims to abandon the drawbacks brought by the traditional database management system and build a more efficient, secure and legally valid data guardianship platform by integrating blockchain technology.

Taking the current mainstream distributed institutional knowledge base consortium as an example, the libraries and information centers of each member institution usually have certain data storage and organization capabilities, and can have portals for data distribution and access, so it is possible to establish an account blockchain (ABC) for the member institutions of the institutional knowledge base consortium. However, for the same institution, it can be both a publisher and a user of data, and in this way, two separate ABCs can be established for each member institution according to its role: ABC-S for data publishing and ABC-U for data using. The ABCs are only responsible for querying, keeping accounts, and creating blocks, and do not perform transactions. The two ABCs in the model are used to store the data published and used by each member organization, and to build blocks (publish, delete, maintain, and update scientific data) according to user operations, and to update the status in real time to ensure that the stored data is not tampered with.

Inter-agency or intra-agency user access to data resources, the Downloading, citing, and maintaining scientific data can be regarded as a transaction, which is not a monetary transaction in the traditional sense, but the sharing and utilization of scientific data as an information resource is itself a value exchange, and users need to follow the corresponding usage rules. This value exchange can also be regarded as a transaction because it can gain industry reputation, peer review, cooperation opportunities and even financial benefits. In the blockchain architecture, the transaction is performed by the TBC, which is the channel for transaction and settlement, and it does not keep the account information of both sides of the transaction, but is only responsible for building blocks and executing the transaction. In the scientific data guardianship service, different transaction acts such as accessing, downloading, referencing and maintaining scientific data can be done in the on-chain code or application system. TBC is responsible for responding to user requests, building blocks according to user instructions, automatically executing transaction acts (authorizing accessing, downloading and referencing scientific data), verifying payments (broadcasting usage records and ensuring ABC information synchronization), saving transaction records and synchronizing post-transaction status to ABC in real time, and presenting to users through the account server [15].

4.2 Data storage and access control

1) Data storage

In order to unify the data storage methods of each member institution in the institutional knowledge base consortium and facilitate quick data retrieval and utilization, the data guardianship model adopts an off-chain storage method, in which only data addresses are stored on the chain and the original data are stored in the underlying database after encryption and maintained by each member institution. When the data reaches the threshold, the member organizations store the data in the underlying database after symmetric encryption, and use the input timestamp, block length and the hash value of the previous block as the block header, and encrypt the data location index, plaintext access control policy and data Merkle root according to the access control tree, then pack the data into blocks and store them in the block body, and then upload them to the data warehouse [16]. The data storage structure is shown in Figure 2.

2) Access Control

Initialization: The data guardian platform of the institutional knowledge base consortium sets uniform security parameters, and the CA of the data guardian center performs the generation of master key mk and public parameters pp.

Identity registration: Apply for registration to the second level of the model through the user interface, and obtain the UID and the set of attributes SU corresponding to its real identity.

Key distribution: According to the key distribution algorithm KeyGen (mk, Su), CA calculates the private key of the registrant's attributes, attribute parameters, and thus the private key of the user, based on the set of registrant attributes $U \in SUSK$, sends it to the user over a secure channel for safekeeping.

Timestamp	Block length	Parent Block Hash Value
Data 1 Storage Location Index	Access Policy for Data 1	Data Merkle Root
Data 2 Storage Location Index	Access Policy for Data 2	Data Merkle Root
Data 3 Storage Location Index	Access Policy for Data 3	Data Merkle Root

Figure 2. Block body storage structure

Encrypted data: Data files need to be Quantitative evaluation and file classification generate file metadata and semantic metadata. All data files must follow certain data association rules to ensure the organization of scientific data. In order to ensure the security of data storage, the data uploader can formulate an access control strategy tree StrGen(Su)→Tcom for the data according to the attributes of the visitor, randomly generate a symmetric encryption key rs, and put the data into the organization database alliance distributed after symmetric encryption calculation In the underlying database of the warehouse, the symmetric encryption algorithm can be expressed as:

$$SEncrs(((d1 , d2 , d3 , ..., dn)) \rightarrow cph$$

Data upload: In the Institutional Database Consortium, the libraries and information centers of each member institution have the portal for data distribution and access, and users upload data to the consortium through the ABC-S blockchain of member institutions' accounts for data distribution. The blockchain node among member institutions encrypts the data index address add, symmetric encryption key rs, and broadcasts it to the blockchain according to the access control policy tree. At the same time, the node generates a mapping of the data index id to the data on the chain and puts it into the underlying database.

$$AddGen(cph) \rightarrow add\ SEncCOM(add, rs) \rightarrow CPH$$

Access cipher: The visitor finds the address of the data and the symmetric encryption key based on the data index id on the account blockchain ABC-U at the time the data was used by the member organization. If the visitor does not have access to this data, he/she cannot get the key rs, cannot decrypt the index address, and cannot access the original data. If the visitor's attributes satisfy the access control policy, he can decrypt the address add and the key rs, and then the system responds to the user's request through TBC and gives the information about the location of the retrieved data copy back to the user, and the user accesses the data to the underlying database of the institutional data repository consortium distributed storage [17].

4.3 Functional elaboration of the data guardianship model

In addition to the four basic functions of data custodianship described above, the scientific data custodianship model also has user management and user service functions, and this section develops these five functions and their corresponding parts of the model from a process perspective.

User management and service functions

1)User Interface - "User Registration/Login" interface: Provides an interactive platform for user management functions, provides an interface for collecting user information according to user metadata standards, collects information provided by users during registration and uploads it to the user database, and verifies the access rights of users when they try to log in to the system. The interface provides an interactive platform for user management functions.

Upload information for institutional settings - "User information database"Module inheritance.

2)Organizational settings - "User information database" module: To define user metadata standards (Alliance Standard User Metadata Option + Organization Extended User Metadata Option); to receive and respond to information and requests uploaded from the User Registration/Login interface, and to review, manage, and validate user information; to share information about our organization and registered users to the Alliance in accordance with the Organization Metadata Standard (Alliance Standard Organization Metadata) and the Alliance Standard User Metadata Option. Share organization and registered user information to the Alliance in accordance with the organization metadata standard (Alliance Standard Organization Metadata) and the Alliance Standard User Metadata option.

Shared information for affiliate setup - "Blockchain" platform inheritance.

Data collection and evaluation functions

1)User Interface - "File Upload / Data Maintenance" interface: provides an interface for uploading files, project information and other additional information, guides the user to sort and upload files according to their purpose, type and format, and performs a preliminary review of the form, format and size of the uploaded files. Add time stamps and user information stamps to successfully uploaded files.

Upload information for institutional settings - "Document Pre-Screening" module following Commitment.

2)Organization Settings - "Initial File Screening" Module: Receives and responds to information and requests uploaded on the "File Upload/Data Maintenance" interface, reviews the readability, completeness, and standardization of the newly uploaded data, checks the overlap ratio with the existing data content in the database, and screens the legitimacy of the uploaded data. Screening the legitimacy of the uploaded data by checking the overlap ratio with the existing data content in the database; blocking the data that do not pass the screening and providing feedback to the provider; the data that pass the screening will enter the quality evaluation process.

Upload information for inheritance in the "Document Quality Evaluation" module of Institutional Settings.

3)Institutional Settings - "Document Quality Evaluation" Module: Receive the information uploaded by the "Document Pre-Screening" module, and the corresponding data custodian in the institution will assist the author or project leader in evaluating the maturity, novelty and academic value of the research results in the document, and formulate the corresponding data management plan, and select the appropriate data sharing and storage strategy. The corresponding data custodian will assist the author or project leader in evaluating the maturity, novelty and academic value of the research results in the file, formulating the corresponding data management plan, and selecting the appropriate data sharing and storage strategy; the files that pass the quality evaluation will be backed up and archived according to the management plan, assigned file numbers, and created file index entries in the index list.

Upload information for inheritance from the "File Categories" module in Organization Settings; generate information for inheritance from the "File Descriptions" module in Organization Settings, the "Search Engine" module, and the "Organization Dataset Warehousing" module in Affiliate Settings. Search Engine" module and the Affiliate Settings - "Organization Data Set Storage" module.

Data organization and processing functions

1)Organizational Settings - "File Classification" module: Receives the information uploaded by the "File Quality Evaluation" module, classifies the data in the file according to its usage, form, maturity, and openness, and matches the data organization policies for different types of data; converts the data into a common file format according to the file format standard corresponding to the data type, and saves it into a common file format for easy storage. The data types are converted to a common file format for easy storage according to the file format standard corresponding to the data type.

Generate information to be inherited by the Organization Settings - "File Description" module and Affiliate Settings - "Organization Dataset Warehousing" module.

2)Organizational Settings - "Document Description" Module: Selects or develops metadata standards according to the different classifications of documents and the data they contain, and shares them with consortium members. Receive the information generated by the "Document Quality Evaluation" module and the "Document Classification" module for document description, which can be divided into two parts: the former describes the external information of documents according to the cataloguing rules and generates document metadata. The former describes the external information of documents according to the cataloguing rules and generates document metadata; the latter extracts knowledge elements and organizes the data content in documents according to the data association rules and generates semantic metadata to facilitate data sharing and utilization.

Generate information for institutional settings - "Search Engine" module and affiliate settings - "Blockchain" platform, "Institutional Dataset Warehousing " module.

Data storage and publishing functions

1)Federation Settings - "Institutional Dataset Warehousing" module: follow the data storage strategy defined by the "File Quality Assessment" module to match the data storage plan to the standard and indexed files received from the "File Classification" module, set the data retention period and data backup plan The standard and indexed files received from the "File

Classification" module are matched with the corresponding data storage plan, and the data retention period and data backup plan are set; the file storage and semantic dataset are constructed locally according to the file and semantic metadata standards of the "File Description" module, and the two datasets are linked by the set data association rules. The two datasets are linked by the set data association rules.

The storage of institutional data sets is the local storage of each member institution in the alliance. The part shared with the alliance needs to be organized and stored according to the unified metadata standard of the alliance, and can be called remotely through certain data sharing protocols; other non-shared data can be organized and stored by member institutions with their own policies.

2)Federation Setup - "Distributed Warehousing Federation" Module: This module actually follows the data sharing protocol of the Institutional Knowledge Base Federation

The federated distributed storage platform built by the member institutions, which combines the The shared parts of the built local data warehousing are integrated into a whole distributed data warehousing federation according to relatively unified data organization standards, data storage protocols, data sharing protocols, data access and transmission protocols[18].

The distributed data warehousing consortium itself may not provide storage space, and its main responsibility is to coordinate and coordinate the data resources stored by the consortium members, formulate, receive and update the data organization and storage specifications for the consortium members, optimize the data storage structure of the member institutions, and coordinate the data storage tasks of the member institutions when necessary to promote the storage capacity of the entire institutional knowledge base consortium Maximization.

3)Organizational Settings - "Data Publishing" Module: Assists authors or project leaders in publishing uploaded data resources in accordance with the data sharing policy, in preparation for subsequent data sharing and utilization. The basic questions embedded in the data sharing policy include whether the specified data resources are confidential (confidentiality level, confidentiality period, confidentiality scope), whether they are shared within the consortium, whether they are shared in full-text (full-text content, partial content, or only the title), whether they are shared outside the consortium (other institutional knowledge bases or institutional repository consortium, external disciplinary bases), and whether they are published as Linked Data. Whether to publish as Linked Data, etc.

Data sharing and utilization functions

1)User interface - "Information Retrieval" interface: provides an interface to information resources, receives search results from the "Search Engine" module and provides feedback to the user, provides a diverse, multilingual and limited search window For public resources, the search interface and search function of third-party search engines can be embedded.

Upload information for institutional settings - "Search Engine" module following Commitment.

2)Organizational Settings - "Search Engine" Module: Receives and responds to user requests uploaded from the "Information Retrieval" interface, invokes the consortium's shared document index and metadata warehousing, locates published data resources, and presents data requests to the "Distributed Storage Consortium" and present the search results to the user.

Generate information for the user interface - "Information Retrieval" interface, "Data Access and Download" interface to inherit.

3)User interface - "Data access and download" interface: for Users provide a data access portal to support online browsing, use, and access to data resources, and provide users with access to data in accordance with sharing agreements and user permissions.

Presenting the relevant access and download conditions. Based on the information provided by the "search engine" for locating digital resources, the corresponding data access and download links are generated and presented to users according to their information needs and access rights.

This interface is only a portal for exchanging information with users, and the specific process of sharing data resources is carried out by the federated setup - the "blockchain" platform.

4)Alliance Setup - "Blockchain" Platform: The blockchain platform acts as an information hub in the whole model, and the information exchange existing in all aspects of data uploading, storing, updating, sharing, accessing, and utilizing will be done in a blockchain way. Therefore, the blockchain platform is actually multifunctional.

5)Institutional Settings - "Statistical Analysis and Visualization" Module: The scientific data guardianship model can collect all open dynamics in the data life cycle for statistical analysis and visualization, facilitating researchers to have a true grasp of academic dynamics, thus promoting the scientific data It facilitates the further sharing and utilization of scientific data. User Interface - The "Statistical Analysis" interface is the user portal of this module.

5 Conclusion

Compared with the traditional data guardianship model, the model established in this research realizes five basic functions of user management and service, data collection and evaluation, data organization and processing, data storage and publication, and data sharing and utilization by integrating the dual-chain blockchain technology, which completely realizes the whole process of scientific data guardianship and management of the institutional knowledge base and solves the It solves the coordination problems among different institutions, users and standards in the knowledge base, as well as the problems of computation surge and data transmission and storage security brought by the concurrent user demands.

This data guardianship model incorporating dual-chain blockchain technology can realize self-service, timely and secure sharing of scientific data, and can fully respond to users' demands 24/7 without the need for alliance monitoring and agents in the whole process, which not only greatly improves the efficiency and utilization rate of data sharing, but also ensures data sharing through the tamper-proof timestamp of blockchain technology and smart contracts security and intellectual property protection.

References

[1] Zhong Yuan.Investigation and Analysis on the current situation of open access of Library and In formation Periodicals in China [J].digital Library Forum, 2015, (11): 64-68.

[2] Huang Xiaojin, HUANG Fumin, WANG Qian. Development status and Comparative Study of Institutional knowledge Base Alliance in China [J].Library Science Research, 2014, (12): 92-97.

[3] Yang Helin. A New Idea for the Construction of institutional library in American University Lib raries from the perspective of Data Monitoring -- Inspiration from DataStaR [J]. Journal of University Libraries, 2012, (2) : 23-28.

[4] Feng Jie, Si Li.Operational risk Investigation and Analysis of university Scientific research Data Institution Database Alliance [J].Libraries, 2019, (3) : 58-62, 68.

[5] Si Li, Chen Xuanning.Investigation and Analysis of the current situation of scientific research d ata institution database construction [J].Library, 2017, (4): 6-11.

[6] Cheng Jing, Liu Jiamei, Yang Qihong.Conceptual model and Operation Strategy of scientific res earch Data management System based on dissipative

[7] structure Theory [J].Modern Intelligence, 2008, 38 (1): 31-36.

[8] Chen Xiaojing.Design and Exploration of big data Platform of government affairs based on bloc k chain [J].Information Systems Engineering, 2018, (4): 127-128.

[9] CAI Weide, Yu Lian, Wang Rong, et al.Research on application System Development Method b ased on block chain [J].Journal of Software, 2017, (6): 1474-1487.

[10] Li Li, Zeng Yueliang.Research on data Governance Framework of Institutional Research Data K nowledge Base Alliance [J].Library Forum, 2008, 38(8) : 61-67.

[11] Jiao Tong, Shen Delong, Nie Tiezheng, et al.Blockchain database: a queriable and tamper-proof database [J/OL]. Journal of Software: 1-15. HTTPS: / / doi.org/ 10.13328 / J.cnki.jos. 005776, 2019-4-26

[12] Liu Guifeng, Pu Jingrong, Qian Jinlin.Analysis and Explanation of influencing factors of scienti fic research data sharing [J].Library Forum, 2008, 38(11): 10-17, 26.

[13] Lu Fangting.Research on the Application of block chain technology in institutional Knowledge Base [J].Library Work and Research, 2019, (4): 70-73.

[14] Cao Ran, Wang Qiong, Geng Qian, et al.Research on education and Teaching Reform driven by talent Demand in Data Monitoring [J].Journal of University Libraries, 2017, (2): 81-87.

[15] Zhou Shuyun, Wu Dan.Data Monitoring flow analysis of University Library based on informati on Life cycle [J].Journal of Shandong Library, 2016, (3): 26-29.

[16] CAI Weide, Yu Lian, Wang Rong, et al.Research on application System Development Method b ased on Block chain [J].Journal of Software, 2017, (6): 1474-1487.

[17] Wang Xiuli, Jiang Xiaozhou, Li Yang, et al.A data access control and sharing model using bloc k chain [J/OL]. Journal of Software: 1-9. https://: doi.org/10.13328/j．cnki．jos．005742，2019－06－16．

[18] Zhou Yao.Research on the Application of block chain technology in smart Library [J].Modern ti mes, 2019,39 (4): 94-102.

[19] Chen Meihua, Liu Wenyun, BI Yu, et al.Research on the Construction of American Institutional Knowledge Base Alliance and its enlightenment to China [J].Library, 2015, (11): 59-64.