EGQCY: A Smart Contract-Based Scientific Big Data System Approach for Incentive Sharing and Transaction on the Cost of Data Quality

Shuyi Yang
Beihang University,Yunnan Innovation Research Institute
Lusu Li
Yunnan University
Libo Feng (fenglibo@ynu.edu.cn)
Yunnan University

Research Article

Keywords: Scientific big data, blockchain, Smart Contract, data sharing and transaction, data incentive mechanism, the cost of data quality control

Posted Date: June 14th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3029457/v1

Additional Declarations: No competing interests reported.

EGQCY: A Smart Contract-Based Scientific Big Data System Approach for Incentive Sharing and Transaction on the Cost of Data Quality

Shuyi Yang[1], Lusu Li[2], Libo Feng[3*]

Abstract—Currently, scientific big data management is generally faced with the problems of scattered data resources, inconsistent data standards, and the inability to share and circulate data safely. Research personnel attaches great importance to whether sharing the first-hand property is secure under clear ownership and whether it can contribute to the large society. The isolation of the data management system is the obvious obstacle to collecting and managing across-disciplinary data. To a large extent, sharing and trading scientific big data is the primary purpose to realize the clarity of property rights, secure data sharing, and the value of the data assets step by step. We propose to construct a public platform for scientific big data management. The system is managed to unify and authorize the on-chain data, on which data sharing and trading is tracked throughout the process. Smart contracts are executed with vital functions and guarantee price matching in data transactions. We design the incentive mechanism which measures the incentive yield of data cost quality based on evolutionary game theory and data quality control theory (EGQCY), considering how the cost of data quality performs in controlling and impacting the rational release of the incentive yields in the sharing and trading process. The experiments found that the design of incentive yield and incentive coefficients only significantly affected the transition from low-quality data to medium-quality data. Both parameters converged to fixed values as the cost of data quality increased.

Keywords—Scientific big data, blockchain, Smart Contract, data sharing and transaction, data incentive mechanism, the cost of data quality control.

1.Introduction

The digital economy has become a new economic engine. With the deepening of the digital economy, blockchain is regarded as an essential means to support data to become a new factor of production. Due to the distinct characteristics of ensuring the uniqueness and non-replicability of storage objects and providing clear ownership of data, blockchain is not only a technology but also a set of thinking modes, action protocols, and application modes [1][2]. With the advent of big data, data management has risen more and more attention. Data assets are owned and controlled by the enterprise, and data management can be divided into three levels: data management, data resource management, and data asset management. Data asset management extends data management and data resource management. In addition, not all data are data assets. Only the data resources that can bring future economic benefits to the enterprise can be called data assets. Data asset management focuses on the storage and application layers and considers how to manage and share data as a unique asset securely. As a new type of database where multiple untrusted nodes work together to maintain a global state, blockchain can solve the data storage and sharing problems without mutual trust are considered very applicable to data assets management [3].

In information system research, the blockchain research agenda contains protocol layer research, research between protocol layer and application layer research, and application layer research. Based on the behavior of information systems, the design of information systems, and the economics of information systems, scholars summarized and analyzed the application degree and importance of blockchain applied to information systems. They emphasized the need to consider the support and

1.Beihang University Yunnan Innovation Institute, Kumming, China.
2.Beihang University Yunnan Innovation Institute, Kumming, China.
3.School of Software, Yunnan University, Kunming, China.
***Corresponding author**
Libo Feng, School of Software, Yunnan University, Kunming, China. Email: fenglibo@ynu.edu.cn

constraints of the blockchain protocol level to the application layer program [4]. From the perspective of information system behavior, many studies focus on the role of blockchain technology in transforming inter-organizational collaborative business processes and the openness of the ecosystem. Some pay attention to the sharing and value of data based on ownership and control and analyze the credit problems of blockchain applications in system construction from the individual and organizational levels. The consumers, managers, managers, and experts obtained beneficial feedback on applying blockchain technology to cross-organizational data interaction [5]. From the perspective of information system design, the centralized data management mode can no longer meet the needs of users and institutions for data value development and secure exchange. A fair and secure data-sharing environment is an urgent problem to be solved. Using distributed databases based on the blockchain and point-to-point trading networks with considerable data-sharing framework design research has been put forward. Such information system design research focuses on system performance and security testing. Among them, performance analysis focuses on a network share rate, data volume, and transaction failure rate. It attaches great importance to the data transmission speed and time consumption of balance. Security analysis mainly focuses on the security and reliability of data sources, the security of the data storage ledger, and the privacy security of transaction behavior [6][7]. From the information system economy perspective, blockchain's most apparent economic attribute to the big data management system is reducing the cost of using third-party intermediaries. The protocol layer provides a different logic of value creation and capture for the application layer, such as the blockchain's arithmetic competition that has developed a token market. However, the more profound economic attribute is reflected in the concealment of transactions, which involves blockchain's decision-making,efficiency, and incentive behavior, guarantees a secure consensus mechanism, and provides work proof[8][9].

The core contribution of this paper has three areas. (1) This study proposes the practices and rationality of building a smart-contract-based public platform for scientific big data management. Research further investigates the optimal software development path on how data is gathered and transformed into data assets on the blockchain, showcasing the functional design of the data authentication, sharing, and tracking processes. (2) Research explores the design method for using the smart contract to monitor the price matching process in data transactions following a reputation-based price priority matching algorithm. (3)This study combines evolutionary game theory and data control theory to propose an innovative model for measuring the incentive yield of data cost quality for data sharing and transactions (EGQCY).Then compares and analyses the evolutionary stabilization strategies under two sharing scenarios and further discusses how to control the data quality cost, design different return incentive rates, and promote users' active sharing and trading behavior.

This paper is structured as follows. Section 2 is related work. Section 3 introduces the background of the system design and the functional modules of the four core subsystems. Section 4 describes the four smart contract designs around the system's functional modules and explicitly showcases the smart contract's process for price matching in trading. Section 5 innovatively proposes an incentive model based on data quality and cost and tests the model in trading game scenarios. Their interrelationship is that Section 3 is a functional description of the whole system. Section 4 and Section 5 are the key technologies and algorithmic models used across the entire system design, which extends Section 3. Section 6 is the experiment and comparative analysis of the proposed data incentive model, which helps demonstrate the advancement and authenticity of the research. Section 7 concludes.

2.Related work

2.1 The advantages of blockchain in the service of big data. According to a recent market analysis report, the Big Data market size is expected to grow from $162.6 billion in 2021 to $273.4 billion by 2026, with the dramatic growth in data volumes driving the growth of the Big Data industry at a compound annual growth rate (CAGR) of 11.0% during the forecast period [10]. The security management of big data has been a serious challenge, specifically in the

form of challenges in data collection, data sharing, data storage, and data analysis. Related research outlines blockchain-based approaches and services for big data [11] revealing the security risks posed by existing cloud services for big data and the blockchain solutions as following points: (1) Data collection: the process of data collection is vulnerable to exposure and attack, and the consensus mechanism of blockchain provides an efficient and secure data sharing environment [12][13]. (2) Data sharing and transactions: The lack of authoritative and certifiable nodes and long response times in data sharing, blockchain's useless transaction filtering algorithm helps access data from the cache layer instead of the storage layer, helping to reduce response times and storage overheads, and smart contracts are used for authorization [14][15][16]. (3) Data storage: Blockchain is integrated with InterPlanetary File System (IPFS) to solve the file storage redundancy problem by implementing a decentralized platform to provide security for file storage systems. Data is stored in the cloud before using attribute-based encryption, and the hash of the data is stored in the blockchain to provide authenticity for users [17]. (4) Data management: A virtual shared ledger stores transaction history. Database transactions are recorded as blocks, and each interconnected using cryptographic hashes. Blockchain-based solutions integrate storage servers and cryptographic algorithms for reliable database access, and blockchain uses timestamping methods to overcome data tampering [18][19][20]. (5) Data training and learning: Sharing data with different attributes among multiple subjects to classify data types for machine learning comes with data privacy and security concerns. The Blockchain Consortium and Homomorphic Cryptosystem provide a trusted and secure training platform free from third-party intervention [21][22]. (6) Data privacy monitoring: protects personal privacy data while empowering governments and public administrations to monitor data from multiple nodes. The blockchain node network uses federation chain technology, which features node access, authority hierarchy, and consensus algorithms such as Delegated Proof of Stake (DPOS) or Practical Byzantine Fault Tolerance (PBFT). The supervising server strictly controls participating nodes in the federation chain, and only nodes that have accepted the block can access the blockchain data, which better maintains system privacy [23][24].

2.2 Main application areas of blockchain-based Scientific big data platforms. Applications of big data sharing in scientific research mainly include gene sequencing, data publication, citation, host data reuse, and scientific instruments. With the call of advocates of an open data-sharing culture and researchers in various fields, relevant government departments and large research institutions have introduced appropriate policies for sharing data resources. The aim is to encourage and even compel project leaders and paper authors to store supporting data related to research findings in a publicly accessible third-party database for centralized storage and management. Centralized data-sharing platforms in different disciplines and fields have emerged.

Representative platforms include the National Science Foundation (NSF) physiological data sharing platform PhysioNet, the web-based survey data platform Digital Coast, Dartmouth University's wireless data sharing platform Crawdad, the US National data.gov, the National Earth System Science Sharing Service Platform of the Chinese Academy of Sciences, and others. The data from these platforms are interdisciplinary, large in volume, and significantly impact other related fields and platforms [25]. Medical information sharing, closely related to scientific research data, is similarly faced with the need to ensure the accuracy and integrity of medical information throughout the sharing process. Medical institutions require sharing information on scientific research and development, and privacy and security hinder the sharing process, where data can be improperly manipulated and leaked. One study proposes a new business process for a blockchain-based healthcare information-sharing platform that leverages blockchain to store, share, and reliably verify healthcare information transactions, track records, and share between parties using a distributed network. A new consensus algorithm and a generic anonymous sharing model are also proposed to prevent manipulation and fraud while fully exploiting the value of medical information [26].

2.3 Smart contracts applied to big data management. Smart contracts effectively improve the quality and

accuracy of data collection and analysis of big data, allowing semi-structured and unstructured data to be regularly checked and filtered, reducing the proportion of invalid data in the database. The smart contract is essentially a digital protocol to facilitate, validate and execute one or more transactions. Writing rules achieve the translation of contractual terms between two parties into executable code, which is similar to a real-life physical contract. With smart contracts, transactions are only valid when the contractual agreement is met, thus storing the transactions in the blockchain [27]. Research proposed deploy smart contracts and multiple PoW consensus mechanisms to reduce the computational power consumption to investigate the validation and verification of node transactions. During transmission to delivery, vehicle information is optimized and stored in distributed immutable storage based on the designed smart contract data structure to simplify transactions and broadcast content [28]. Otherwise, smart contracts can be applied to digital identity verification, user information recording, digital asset sharing, and the automatic execution of contract terms. Smart contracts can verify the identity of users through code to achieve legal regulation for the authorized use and sharing of data on the blockchain. As such, smart contracts are seen as cracking the big data risk control puzzle [29]

In dealing with information sharing and transactions, game theory is widely used in the design of smart contracts, encompassing prisoner contracts, complicity contracts, and betrayal contracts [30]. Smart contract operation replaces trusted third parties as an auditable and fair payment protocol, providing an efficient data transaction. Chen et al.[31]propose a blockchain-based fair data exchange scheme that guarantees fairness and privacy protection of the transaction without a trusted third party and efficiently automates the exchange. Ma et al.[32] combined game theory with traditional reduced-gate computation to establish a game model based on a reputation mechanism and proposed a smart contract-based rational delegation computation protocol for three-party games. In terms of establishing and maintaining benign incentives, Liang et al. [33] designed a secure system of behavioral strategies based on social norms and reputation systems to motivate rational nodes to abandon malicious behavior for their benefit. Wang et al. [34] proposed the impact of social cloud reputation and structure on rational computation, which ensures that a party with a good reputation means that it is likely to cooperate with others. Kou et al. [35] proposed a new link prediction method. During the delegation calculation, a reputation mechanism is designed to increase the reputation of honest participants and decrease the reputation of malicious participants.

3. System design

The system design proposes to build an open big data management platform base on blockchain, which is government funding and has the nature of public service. There are three types of users:

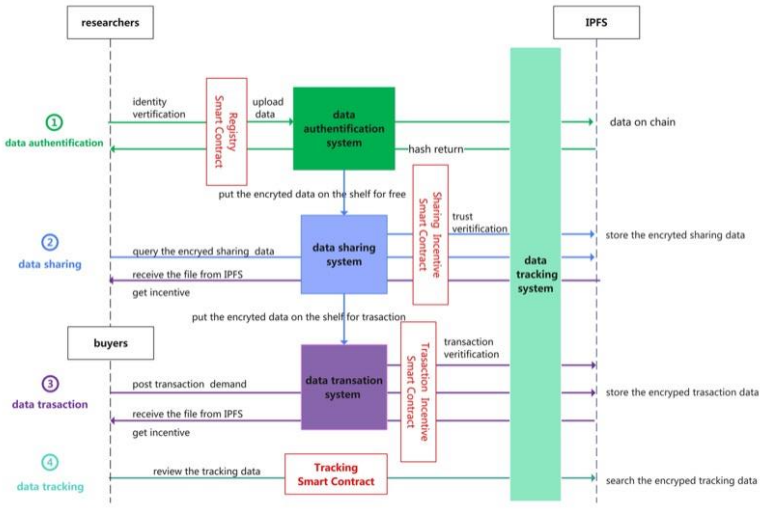


Fig.1.The

government, enterprise, and individual. The platform provides trustworthy data storage, rights confirmation, and assetization services in various fields. As the initial node provider, the government introduces leading units of various industries to form multi-consensus nodes, incubates the industry chain, and gradually realizes that industry nodes can join independently. It is foreseeable to see the scale operation share the cost of system construction, and node commercial operation is the critical step after the node consensus party and platform data volume reach a certain level. Through the on-chain tracking technology, this system not only protects the data owner's prior and subsequent interests but also closely follows the market feedback of the data demand side to explore the new growth curve of specific industries, with the profound value of reverse mining and reproduction of data. This study proposes the EGQCY data incentive model as one way to activate a self-organizing ecosystem of data producers/processors/consumers to realize a new digital economy model that maximizes the use of data value with accurate regulation.

Fig.1 presents the design framework of the data flow across the system, including the platform composition, the interaction logic of each system, and critical steps using smart contracts. The application is built based on a blockchain alliance chain with a fusion scenario of aggregation layer and distributed resource layer, applying efficient encrypted storage, hierarchical privacy supervision, and data security protection of the underlying blockchain technology to provide technical assurance. The core functions of the prototype system contain four sub-systems. Data aggregation, governance, and chain confirmation of scientific research data submitted by multiple consensus nodes are on the blockchain through the data authentication platform. The data tracking platform scans data asset identification, sharing, and trading process and provides data analysis support. At the same time, smart contracts are designed throughout the system to ensure fair execution of aggregated transactions and incentive yields and make the data more efficient and secure in circulation.

3.1Data authentication on blockchain. After the user completes the authentication, they can enter the system to register the data. The registry process is to upload the basic information of the data asset and define the core metadata, select the hosting form of the data asset, upload the proof of agreement, set the privacy configuration of the data, and apply it to the auditor. After passing the audit process, the system will publicize the asset and issue an on-chain contract. The role of the data authentication module is to ensure the originality and attribution of the file data and to avoid the problem of data being tampered with by malicious intrusion. The advantage is that the user can save the data packet and associate the hash value, user information data, and related attribute data to the blockchain network to achieve reliable proof of ownership. The specific process of data uploading to the blockchain includes the following：

(1) obtain the target file data to be stored by the user and then process the target file data using a hashing algorithm to generate a hash value.

(2) obtain the user's identity information and the basic attribute data of the target file data, package the user's identity information, the basic attribute data, and the hash value generated in the previous step to generate the stored evidence data package.

(3) digitally sign the packet and generate the corresponding data ID, store the data ID and packet to generate the authentication ID, and then block. Authentication information will be uploaded to the blockchain network.

*3.2 Data transaction.*The transaction demand buyers post matches the data uploaded and confirmed by

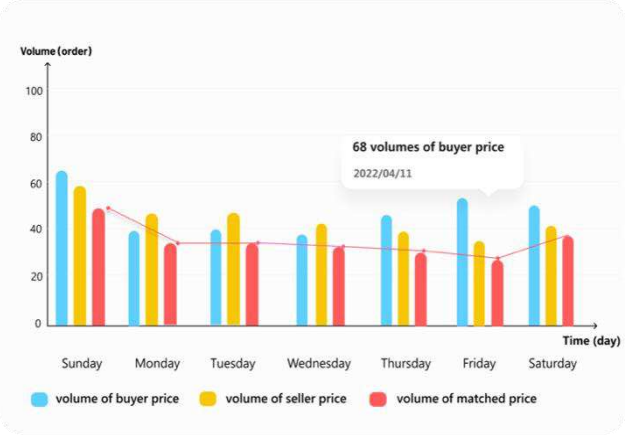
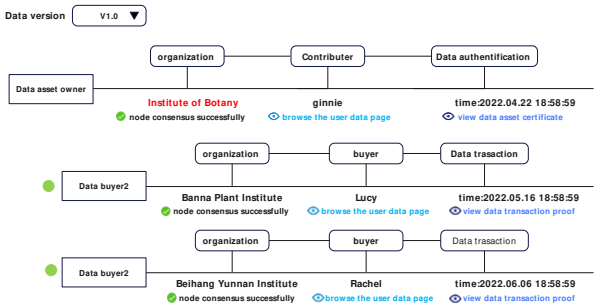


Fig.2. Price matching statistics for both sides of the transaction

sellers in the system according to industry and industry segmentation data latitude. Potentially tradable data is packaged and pushed to buyers and sellers, making information instantly available. The system generates a data cloud for buyers and sellers based on popular data tags, evaluating the amount of data aggregated based on buyers' and sellers' trading price demands. The popular data tag for buyers includes "uploaded data," "matched data," and "data in the process of matching data," and the popular tag for sellers includes "tradable data packages," "matched data," and "data in the process of matching." Through the statistical analysis, see Fig.2, the "volume of buyer price," "volume of seller price," and "volume of matched price" of buyers and sellers are provided. After the transaction occurs, the system uploads the transaction information to the chain, generates the digital asset transaction certificate, and generates the unique certificate fingerprint, transaction contract HASH, and platform public key address to ensure the authenticity and credibility of each transaction data on the chain.

3.3 Data tracking. After each share and transaction is completed, both parties can track ownership and use of updated versions of the digital asset. By searching



for a single digital asset ID or asset keyword, users can query basic information about the digital asset associated with the current account, including the asset storage block ID, block Hash, asset digital fingerprint, public key, uplink status, and uplink node. At the same time, the user can view the current version update list of the digital asset and view the version update of the same digital asset by the property owner, including the original property owner and the property acquiring party,see Fig.3. On the page of the asset version, both sides involved in digital asset sharing and trading can select the corresponding version and root data of digital assets generated by each party. The display includes the original data of contributors to access digital assets, multiple data buyers and their home page data, data link time, and certificate details of each sharing or transaction.

Fig.3. The interface of the data sharing and tracking details

4.Smart contract design

As an immutable code running on the blockchain, the smart contract can make users trust the system's interaction behavior with data due to its unchangeable contract terms. The functional logic of the smart contract is first agreed upon by all related parties and written and compiled offline. The written smart contract must be installed and instantiated by authorized members[36]. The Proof of rest (PoR) consensus mechanism and powerful smart contracts adapted to the federated blockchain scenario allow each participating node to take turns resting and creating blocks. The system automatically penalizes slack nodes, updates parameters, and increases or decreases the number of nodes, thus enabling the federated blockchain to renew itself [37]. Access control policies are implemented by smart contracts deployed in each access domain so that sidechains of different access domains can store access records externally. Side chains store records from external storage access records and maintain the integrity of the records [38].

In this paper, smart contracts ensure that users enjoy the protection of incentives while conducting data registration, data sharing, traceability, and the openness and transparency of the data transaction process. Formal verification of smart contracts performs to verify the contract format and security. First, the smart contract should meet the standard format requirements. After passing the format requirements, Move Prover is a formal verification tool to verify the security of a smart contract based on ascertaining whether a program conforms to a specification utilizing an automated theorem-proving solver in the formal verification domain. Move specification language is a specification language that describes how a program should run correctly through preconditions, postconditions,

invariants, etc. A compiler converts the Move program and specification into a boogie program, resulting in either the input program meeting the given specification or not. A specific path is given when it does not[39][40].

Fig.4 explains the execution data flow based on the four core smart contracts designed for the above system. The registration contract is to monitor the users for digital identity authentication, and the traceability contract tracks the key user data operation behaviors in the system that involve interaction with the chain. The incentive model proposed in this study is based on a high-level smart contract protocol. The incentive scheme is executed by specific incentive algorithm rules written in two contracts: transaction incentive contract and sharing incentive contract. Due to the transaction being regarded as a unique sharing behavior, we exemplify a specific smart contract for price matching. When the buyer and seller transaction is successful, the incentive algorithm of EGQCY controls the data quality and cost and measures the game behavior of the buyer and the seller before the transaction. The system explicitly selected the incentive coefficient to release the incentive yield properly. More importantly, this study emphasized that the incentive model proposed applies to the early stage of system construction. It is suitable when the node operation and the number of users have yet to be scaled up, so the release of incentives through the high-level agreement of smart contracts can prompt the occurrence of transaction and sharing behavior and drive the increase of the number of node users.

Transactions on the blockchain have a gas cost proportional to the internal operations of the respective function calls in the smart contract. Storing data on the blockchain is relatively expensive; thus, writing to the blockchain increases with the content size [41]. As a result, the deployment cost of a new smart contract is generally relatively high compared to the transactions generated by invoking the functions of that smart contract. The cost of updating the data meta-information in a smart contract increases with the number of requesters, and transactions of the data provider and the data requesters influence the cost of running a smart contract[27]. Based on the stage of this study, the deployment cost of smart contracts should be higher than the invocation cost. As a public data service platform, the government should bear the deployment cost of smart contracts and the invocation cost up to a specified number of times at the initial stage. When the system is mature, it is essential to explore how the platform can share the cost of specific smart contract invocations based on the data provider's and the requesting party's actual sharing and trading operations to assess how both parties can benefit from the process.

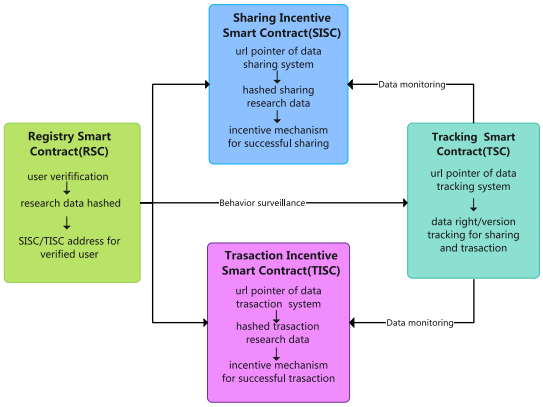


Fig.4.The main application of smart contracts

4.1 Smart contract of price matching in data transaction. In the data trading process, when the buyer initiates the transaction, the smart contract witnesses the price matching and negotiation between the buyer and the seller, see **Algorithm 1**.The data transaction contract ensures that the seller's offer is arranged in the system in the sell order of smallest to largest, which has an inverse proportion to the buyer's expected transaction price. The system also guarantees that the seller's original bid must be higher than its cost price, and the buyer's original bid must be lower than its predicted maximum price. A transaction can occur only when the seller's bid is lower or the same as the buyer's bid. The system also determines the volume of transaction data and analyzes historical transactions to match orders. When the above trading conditions are met, a set of best bid orders is selected from the above rules for inquiry. The buyer orders will remain pending until the best bid is matched by a sell order that completes the transaction. The smart contract monitors the data trading matching behavior of buyers and sellers throughout the process. It firstly protects the buyers'

right to price their data and eliminates unreasonable bidding in the trading market, promoting higher transaction rates and increasing the participants' revenue.

The operation of smart contracts for price matching follows the basic logic of a reputation-based price priority matching mechanism. Before entering into a transaction, the system determines whether the bid and demand prices of the trading data are within a reasonable range set by the platform. The system defaults to p1 ≤ P1, and Formula 1 is used to obtain the corresponding standard deviation for price selection.

$$\sigma = \sqrt{\sum_i^n (Xi - \mu)^2 / N} \qquad (1)$$

Assuming the seller user's tradable range is [P1, P4], the corresponding tradable seller order prices appearing in the data pool of transaction matchings are from small to large (P1<P2<P3<P4). The buyer's demand price range is [p1, p4], the corresponding tradable buyer order prices appearing in the data pool of transaction matchings are from large to small (p4>p3>p2>p1), while the default pricing range is [P1, p4]. However, the system prioritizes the highest buy order and the lowest sell order for matching (P1, p4), then the buy order is matched from large to small and the sell order from small to large in both directions, e.g. (P1, p3), (P1, p2), (P1, p1), (P2, p4) (P2, p3), (P2, p2), (P2, p3), (P2, p1) ....., and so on.After the buyer and seller orders are successfully matched, the seller first answers whether to agree to the transaction. If the seller agrees, the system hangs the transaction status waiting for the buyer to pay. If the buyer does not agree to the transaction, the system then looks for other buyers with higher prices to answer. When there is more than one order with the same seller's price in the data aggregation pool, the system will prioritize the sellers according to their credit score T, the data sharing quality score G, and the user's incentive yield Y. In industries with a high level of data security requirements, the confidence interval is generally in the plus or minus three standard deviations. According to the standard deviation algorithm, more than three sellers' data are used to calculate a standard deviation. The buyer's bid price must be within plus or minus 3 standard deviations to match the price and enter the next step of the process. Otherwise, the bid is not successful.

Algorithm 1:Price Matching In Smart Contract

Input: $p,\sigma,\mu[1,i],\ \beta[1,j]$
Output: M
1: trigger{dealMatching};
2: Function: dealMatching()//Perform price matching algorithm
3: if p∈ (u+3*sellingPriceStandard(P) ,u-3*sellingPriceStandard(P))then //Calculate the purchase and sale standard price
4: M=matchRank() //Match Price
5: for u ← 0 to i − 2 do
6: if verifyDuplicate(M[u],M[u+1])=true then // Verify whether it is repeated
7: Sort M[u],M[u+1] by G and T
8: end for
9: pushSellAndBuy//Push to both parties
10: Function: sellingPriceStandard(μ, β)//Perform initial selling price calculation
11: for x_i: P
12: run formula 1
13: end for
14: return σ;
15: Function: matchRank(P)//Match the buying and selling price and sort
16: for u ← 0 to i − 2 do
17: flag ← True
18: for z ← 0 to i − 2 − u do
19: if μ[z+1]<μ[z] then
20: swap(μ[z],μ[z+1]);
21: flag ← False
22: if flag = True return
23: for q ← 0 to j − 2 do
24: flag ← True
25: for w ← 0 to j − 2 − u do
26: if β[w+1]>β[w] then
27: swap(β[w],β[w+1]);
28: flag ← False
29: if flag = True return
30: for r ← 0 to j − 2 do
31: M[r] = [μ[r], β[r]]
32: return M;

5.The Incentive model of EGQCY for data sharing and trading

As a behavioral strategy for information sharing [42] [43] , evolutionary game theory is often used to explore the conditions of subsidies and penalties involved in the design of incentives in information sharing [44]. Studies have also discussed the risk of sharing costs for information-sharing parties under different penalty conditions [45]. While incentives have costs, few studies have focused on the impact of controlling data costs and quality on incentive yields in data sharing [46]. However, it is worth noting that current data quality and cost control theories are well-established[47][48].This section will integrate evolutionary game theory on data sharing incentive scenarios to design how to unlock reasonable

incentive yields from data sharing by controlling data quality and costs.

5.1Incentive models and algorithms of EGQCY. Data sharing and trading are based on game behavior. The data provider and data requester choose whether to exchange data according to data cost, quality, and benefit dimensions. The model design of EGQCY in this study defaults to data trading as a particular form of data sharing, where data trading occurs based on data-sharing behavior. When users start to register data, the system begins to record and collect parameters that affect subsequent data sharing and the incentive degree of both parties. At the same time, the process of data incentivization is not a single, uncapped process but rather a rational incentive based on evolutionary games that correlate user sharing and trading behavior data.

(1) S1 indicates that data sharing occurs, and S2 indicates that data sharing does not occur.

(2) The cost of uploading and acquiring shared data is C1 and C2 for data providers and requesters, respectively. C2 is the cost for the data requester to obtain the sharing data, that is, the cost for the data requester to register on the platform, real-name authentication, and data right confirmation, and thus C2 is the default fixed value. The average cost of converting the cumulative data sharing and transaction volume on the platform is C3. In this paper, the cost of C1 comes from the data provider and follows the statistical quality input cost equation [47]. See equation(2). G is the comprehensive evaluation index score of data quality, and the quality assurance and control costs of statistical data are collectively referred to as quality input cost (C investment). Internal quality loss cost and external quality loss cost are collectively referred to as quality loss cost (C loss), B represents the loss coefficient, and A represents the input coefficient. In particular, the trend line of total quality cost has an obvious downward trend with data quality improvement in a certain period. When both data quality and cost reach the M point, the total cost of data quality reaches the optimal benefit. When the data quality exceeds M, the total cost of data quality shows an upward trend.

$$\mathbf{C1}=\boldsymbol{\alpha G}^{\left(\frac{1}{\alpha+\beta}\right)} \qquad (2)$$

(3) The data quality assessment in this study is defined as the data quality requirement metric score to be achieved by the application in sharing and trading. The comprehensive data quality assessment model[49], the data set, the rule series, and the metric comprehensive assessment score are the triads for evaluating comprehensive data governance. The evaluation design of data quality in this study refers to the design of the G model, see equation 3.

$$\mathbf{G}=\sum_{i=1}^{n}\frac{\sum_{j=1}^{C} wj\frac{Nij}{Mij}}{C} \qquad (3)$$

D: Data sets have been authenticated and successfully up onto the chain.

Q: three data-sharing quality rules that apply to the data incentive mechanism customized for this study. Specifically, Q1 is the cumulative data sharing indicator of the platform according to each account after the data validation of a single account of the data provider in the platform, Q2 is the data being traded indicator, and Q3 is the comprehensive, measurable indicator of the platform data quality.

Q1 and Q2 are calculated as $\frac{Nij}{Mij}$ (number of records required by data quality/total number of records), combined with the weight value Wj corresponding to the three indicators given by the platform.

$\boldsymbol{Q1}=\frac{N1}{M}*\boldsymbol{w1}$, the number of times data was shared(N1)is divided by the number of times the dataset is browsed(M), and the weight of this part is 33.3%

$\boldsymbol{Q2}=\frac{N2}{M}*\boldsymbol{w2,}$ the number of trading data (N2) is divided by the number of times the dataset is browsed(M), the weight of this part is 33.3%

$\boldsymbol{Q3}=\frac{N3}{M3}*\frac{N4}{M4}*\boldsymbol{w3,}$ the calculation method of Q3 integrates the two subdivision indexes of data integrity and format standard, and the weight of this part is 33.3%. N3 indicates the number of records whose data has been overwritten, and M3 indicates the number of all records during data rights confirmation. N4 is the number of data items that comply with the standard specification for data items, and M4 indicates the number of data items that must comply with standards.

G: the score of the comprehensive evaluation index of data quality, see equation 4. The G score results from data set D after the rule combination of Q, which reflects the quality of data sharing. A is the sum of the quantity (count) of the data quality items

evaluated by corresponding indicators Q1-Q3, considered the rule coefficient, and the default coefficient in this study is 3.

$$G=\frac{Q1*w1+Q2*w2+Q3*w3}{A\ (A=3)} \quad (4)$$

(4). After the successful occurrence of each share, the user's sharing incentive yield is Y1. After each transaction's successful occurrence, the user's trading incentive yield is Y2, and the total incentive yield of the user in the system is Y=Y1+Y2.

5.2Data sharing assumptions based on EGQCY. The data authentication party who takes the initiative to share is identified to have the original property right of the data. Based on sharing data version and the time sequence of the sharing behavior, the smart contract guarantees the multi-level data sharing and traceability rights and interests of the data provider, in ownership and right of use, respectively. The quality of the sharing data is measured by recording the core factors such as the sharing times, transaction times, and cumulative shared data scale. Furthermore, thus, the data display priority and exposure rate are given according to the data sharing quality. Data providers who have data sharing behavior are incentives in the subsequent data transactions, and the credit incentive rules give discounts or reductions in transaction fees. Conversely, the system requires the party who received the free sharing data before having action to share data and who had transaction behaviors before obtaining the number of times to access the free sharing data.

The incentive yield scenarios discussed below are based on two identical parties with successful data-sharing behavior twice or less. It is important to note that incentive yields are a small part of the utility added to the utility gained from the sharing behavior to facilitate the sharing behavior. In the case of S1, four hypothetical situations can be excited:

S1.1 When the data provider takes the initiative to share, but the data requester does not initiate the sharing request, the incentive mechanism applies to both parties. It favors the data requester to promote successful sharing behavior. When the sharing occurs, it means that the data requester feels the incentive and actively shares, the data requester will get the revenue Y1=I* In(C1+C2), and the data provider will get Y1=I* In(C1).

S1.2. When the data provider does not take the initiative to share, while the data requester seeks to obtain the share, the incentive mechanism applies to both parties. It favors the data provider to promote successful sharing behavior. When the sharing occurs, it means that the data provider feels the benefits and actively shares. The data requestor will get the revenue Y1=I*In(C2), and the data provider will get the revenue Y1=I*In(C1+C2).

S1.3. When both the data provider and the requester are willing to share, the incentive mechanism will give full play to the utility and is the ideal scenario for data sharing. Both the data provider and the requestor receive revenue Y1=I* In(C1+C2)

S1.4. When the data provider and the requestor are unwilling to share simultaneously, the platform will not pay incentive yield to both parties.

In the case of S2, when it is detected that users who had sharing(transaction) behaviors in the system do not have sharing (transaction) behaviors in a certain period, two hypothetical situations can be excited:

1. When users have the behaviors of searching and clicking on sharing (trading) products, the platform will release revenue of Y2=I*In(C3) as an incentive.

2. When users successfully list sharing(transaction) data on the shelf, the platform will release the revenue of Y2=I*In(C3) as an incentive.

6. Experiment

In this section, the impact on data costs due to controlling data quality and hence data costs is observed by analyzing and comparing the differences in the incentive yields generated by three evolutionary incentive strategy scenarios for both sharing parties. The functional variables C1 and I that generate incentive yields between both sides of the game are measured and controlled.

6.1 Experiment environment and preparation. The experiment defines the range of data cost according to the evaluation standard of data quality. Based on the data quality composite score equation (4), considering the three quality evaluation dimensions Q1-Q3 of the data in the system：

$$G=\frac{Q1*w1+Q2*w2+Q3*w3}{A\ \ (A=3)} \quad (4)$$

We designed that the number of times high-quality data in this experiment was shared at least 10 times, traded at least 5 times, and viewed at least 20 times. The number of records specified by the system to complete coverage was 10, and the number of all records measured during data validation was 10. The number of data items complied with the standard specification is 10, and the number of data items that must comply with the standard specification is 10. Thus, the comprehensive quality score G for high-quality data is no less than 120, after $C1=\alpha G^{(\frac{1}{\alpha+\beta})}$ conversion to obtain the corresponding high-quality data cost of no less than 3900. We designed that the low-quality data in this experiment would be shared at most 2 times, traded at most 1 time, and viewed at most 5 times. The number of complete coverage records specified by the system is not more than 5, and the number of all records measured during data validation is 10. The number of data items that must comply with the standard specifications is not more than 5, and the number of data items must comply with the standard specifications is 10. Thus, the comprehensive quality score G of low-quality data is not higher than 11, and the cost of low-quality data is not higher than 100 after $C1=\alpha G^{(\frac{1}{\alpha+\beta})}$ conversion.

6.2 Result and analysis. As a result, the C1 range of 100-3900 was identified as medium-quality data dimensions with room for optimization and data quality improvement. **Table.1** shows the critical value of three experiment parameters under the S1 situation. By comparing the impact of data quality improvement on data cost, the experiment intends to find reasonable incentive yields to promote data sharing. It encourages the sharing parties to change sharing behavior from passive to active and the data providers to improve data quality and control data cost in the sharing process. In the case of S1, we selected three incentive coefficients I(1, 5, 10) to compare the incentive yields of both sharing parties. It suggested that the incentive yields of both parties reached reasonable levels and were significantly different when the incentive coefficient I took the value of 10.

Table.1 Experiment parameter

	Parameter value			
Situation	***C1***	***C1***	***C2***	***I***
S1	100	3900	300	1,5,10

Fig.5. Higher incentive yields for data requesters in S1.1

As seen from Fig.5, to effectively incentivize the data requestor to initiate sharing or trading demands, the data requestor gets as gain Y=I* In(C1+300), and the data provider gets Y=I*In(C1). When the cost of 100 to 1000 falls within the range of moderate quality data, the data requestor incentivizes revenue gains significantly greater than the data provider. The data provider's revenue gradually rises as the data cost increases. When the data cost exceeds 1500, the difference in revenue incentives between the two parties is less than 10. After reaching the critical point of 3900, almost both parties incentivize revenue to stabilize and slow down. As can be seen, incentives favor converting low-quality data into middle-quality data in sharing and trading. In other words, the incentives discourage the provision of high costs in exchange for data quality and instead focus on controlling data costs and rewarding proactive data sharing and trading behavior.

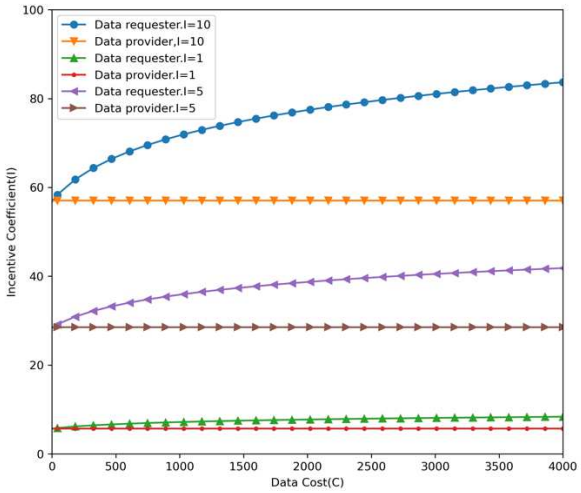


Fig.6.Higher incentive yields for data providers in S1.2

When the data provider has no desire to share and trade actively, the incentive mechanism will favor the data provider to promote this behavior. Fig.6 shows the data requestor receives an incentive yield Y=I*In(300), and the data provider gets an incentive yield Y=I*In(C1+300). The incentive yield for the data requestor is derived from the platform's

average data sharing and transaction costs, which is a fixed mean. The incentive yield is skewed to incentive the data provider more. When the data cost C1 is in the medium-quality data interval, the incentive yield for the data provider is higher than that for the data requestor. However, as the data cost gradually rises, the growth in incentive yields slows down, and the incentive yield line tends to level off when the threshold value is close to 3900. This scenario encourages data providers to improve data quality in exchange for incentive yields. At the same time, there is an upper limit on incentive yields, in line with the incentive model's requirement to control the quality of data costs.

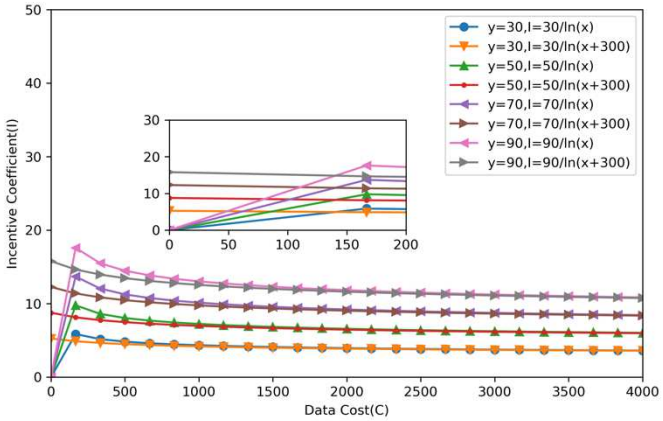


Fig.7 the relationship of incentive coefficient I and data cost C1

By analyzing several scenarios of incentive yields for both sides of a shared transaction as described above, two coefficient formulas for the value of the incentive coefficient I, I=Y/(In(x+300)) and I=Y/(In(x)) were developed. Fig.7 plots the relationship between incentive yield and data costs for the eight cases corresponding to the single sharing or trading benefit Y(30, 50, 70, 90) that both incentivized parties in the system may obtain. It can be found that the incentive yield coefficient and the data cost are only transiently positively related when the data quality is in the low-quality range, i.e., when the data cost is less than 170 intervals. Multiple corresponding values of the incentive coefficients exist when the data cost is the same. As the data cost rises, the incentive curve decreases and levels off. It suggests that the incentive coefficient correlates with the incentive yield and that the two do not effectively contribute to the increase in data cost. In the system design, the incentive coefficients only briefly facilitate the conversion of low-quality to medium-quality data.

6.3 Comparative test. The above experiment explores the data sharing incentive yield and coefficient generated by controlling data quality and cost in an evolutionary game. Further, it explores the system's reasonable range of incentive coefficient design I is the value between 0 and 20. Based on this scenario assumption, we found that the data incentive model of EGDSI proposed in another study [42] also compares the incentive yields in three cases but focuses on introducing the participation level of sharing members to analyze the stability of the evolutionary game strategy. However, it assumed that the data costs of both sharing parties are consistent, regardless of the impact of data quality on data costs. This study will eliminate the external influence factor of the participation degree of data-sharing members on the incentive mechanism. By comparing the algorithm of incentive mechanisms, we found the strength of the model is that we consider that the pull of data quality and cost between the sharing parties in the game process affects the final incentive yield on both sides. More importantly, the study finds that the design of the incentive coefficient is not random, and the incentive coefficient only plays a vital role in the process of data quality cost from low to medium. Therefore, this section will conduct comparative tests through control variables. Consistent with the above studies, the sharing return coefficient a=1 is adopted, the default data cost of both sharing parties is 100, and the value range of incentive coefficient I is between [0,20]. We respectively explore the two models' incentive yield trend with the same setting conditions of other parameters under the evolutionary game situation.

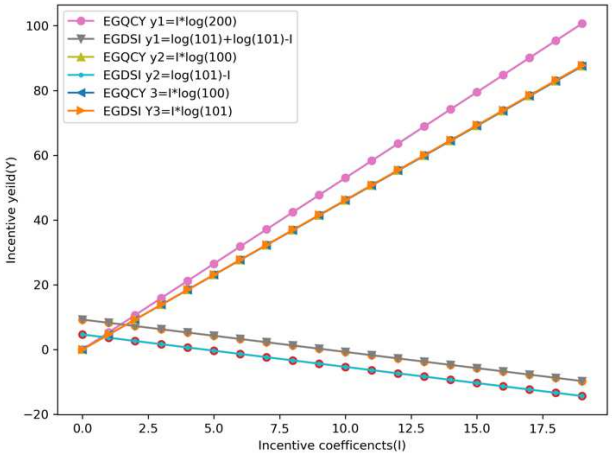


Fig.8 Comparing the relationship between incentive yields and incentive coefficients of EGQCY and EGDSI models

Fig.8 shows the comparative test results. Keeping other variables unchanged, when I is in the range of [0, 20], the incentive coefficient (I) in the data incentive model of EGQCY is proportional to the incentive yield (Y), while the incentive coefficient (I) in the EGDSI model is inversely proportional to the incentive yield (Y). The latter violates the objective authenticity of parameter setting in model design, indicating that the random set of the incentive coefficient does not positively affect incentive yields. This comparative test further confirms the research conclusion of this paper. In designing the incentive mechanism, it is necessary to attach importance to data quality cost control to accurately control the incentive coefficient's value range. Only in this way can the incentive coefficient, as an essential parameter in the incentive mechanism, have a positive, reasonable, and influential effect on incentive yield.

7 Conclusion

We designed the functional modules of the platform and explored the logic of data interactions, adopting four smart contracts to monitor the core function executions of data authentication, price matching, incentive sharing and transactions, and data traceability on the platform. We put forward a reputation-based price priority matching algorithm in data transactions smart contract design. In order to stimulate the user's motivation to share and trade data actively, this study innovatively proposes the incentive yield mechanism of EGQCY. It verifies the incentive yield in different situations and the role of incentive coefficients through experiments, which consider controlling the data quality cost to release the incentive yields to both sides. The experimental results found that: (1).When users provide medium-quality data, the data incentive mechanism plays the most significant role. The incentive yield released by the incentive mechanism is biased towards the playful side of sharing and trading. The incentive yield increases within a reasonable range with the cost of data, which is conducive to users improving data quality in controlling costs. (2). When a sharing user provides low-quality data, the incentive yield significantly differs for both sharing and trading parties. There is a shift in the incentive yield based on the cost of the data to promote active sharing or trading behavior by the passive party. (3). When users provide high-quality data, the incentives discourage high trading costs for data quality to give the high cost of data. Thus, the incentive yield will reach a constant value and remain the same. Future research in designing incentives for data sharing and trading behavior with data cost control should think beyond the single dimension of data incentive yields when dealing with users who provide data of moderate data quality cost and above. It is advisable to explore how the level of user engagement impacts data sharing and trading, considering the cost of data quality to design and validate the differences in incentive yields for different user sizes. Attention could be paid to how the conversion of credits to points drives sharing and trading behavior at the system level and how credit points are tied to reward revenue in exchange for the number of free transactions users can receive. Subsequent research can explore the incentives generated by the blockchain as an infrastructure level for node users. When the number of nodes and the data on the chain reaches a certain scale, the consumption of the blockchain by each running node in the blockchain's underlying consensus mechanism generates from the chain's consensus algorithm incentive[50]. In the near future, we can explore how the platform can share the specific invocation costs based on the actual sharing and transaction operations of data providers and requesters and how to translate the operating expenses of smart contracts through incentive mechanisms.

References

[1] Tang Daisheng, Xu Siyan, Meng Yan, Cao Jianfeng, “Industrial blockchain: An important breakthrough in China's core technology independent innovation,” *in Citic Press*, pp.22-29,2022.

[2] Schierstedt B, Göttel V, Klever L, “Blockchain economy: the challenges and opportunities of initial coin offerings” *Handbook of Digital Entrepreneurship,* pp.256-270,2022.

[3] Zhao Ming, Dong Dazhi, “Data asset management mechanism based on blockchain technology*,” Big Data,* vol. 7. no.4. pp.12, 2021. doi. 10.11959/issn.2096-0271.2021038

*[4] Rossi M, Mueller-Bloch C, Thatcher J B, et al, “Blockchain Research in Information Systems: Curre*nt Trends and an Inclusive Future Research Agenda,” *Journal of the Association for Information Systems*, pp.20,2019.

[5] Ballatore, Marta; Toumi, Dr. Mira; and Arena, Lise, "Blockchain-based data sharing system: an experimental analysis of behavioral features affecting inter-organizational cooperation" *ECIS 2022 Research Papers.* pp.166,2022. https://aisel.aisnet.org/ecis2022_rp/166

[6] Yang J, Wen J, Jiang B, et al. “Blockchain-Based Sharing and Tamper-Proof Framework of Big Data Networking”. *IEEE Network*, vol.34.no.4.pp. 62-67,2020.

[7] Nasonov D, Visheratin A, Boukhanovsky A. “Blockchain-based transaction integrity in distributed big data marketplace,” *International Conference on Computational Science. Springer*, pp. 569-577, 2018.

[8] Saleh, F. “Blockchain without waste: Proof-of-stake.” *New York University Press*,2018.

[9] Beck, R., Mueller-Bloch, C., & King, J. L. “Governance in the blockchain economy: A framework and research agenda.”, *Journal of the Association for Information Systems,* vol.19.no.10. pp.1020-1034,2018.

[10] “Big Data Market worth $273.4 billion by 2026,”2021. [Online]. Available: “https://www.marketsandmarkets.com/PressReleases/big-data.asp”

[11] Deepa N, Pham Q V, Nguyen D C, et al, “A Survey on Blockchain for Big Data: Approaches, Opportunities, and Future Directions,” 2020.

[12] Liu G, Dong H, Yan Z, et al. “B4SDC: A blockchain system for security data collection in MANETs.” *IEEE Transactions on Big Data*, vol.8.no. pp.739-752, 2020.

[13] Mohammad A, Vargas S, Čermák P. “Using Blockchain for Data Collection in the Automotive Industry Sector: A Literature Review,” *Journal of Cybersecurity and Privacy*, vol.2.no.2. pp.257-275, 2022.

[14] C. H. Liu, Q. Lin, and S. Wen, “Blockchain-enabled data collection and sharing for industrial IoT with deep reinforcement learning,” *IEEE Transactions on Industrial Informatics*, vol. 15, no. 6, pp. 3516–3526, Jun. 2019.

[15] Junqin Huang, Linghe Kong, Jingwei Wang, Guihai Chen, Jianhua Gao, Gang Huang, and Muhammad Khurram Khan. “Secure Data Sharing over Vehicular Networks Based on Multi-Sharding Blockchain.” *ACM Transactions on Sensor Networks*, 2023. https://doi.org/10.1145/3579035

[16] Z. Wang, Q. Chen and L. Liu, “Permissioned Blockchain-based Secure and Privacy-Preserving Data Sharing Protocol," *in IEEE Internet of Things Journal*, doi: 10.1109/JIOT.2023.3242959.

[17] C. Xu, K.Wang,P. Li,S.Guo, J.Luo, B.Ye,and M.Guo, “Making big data open in edges: A resource-efficient blockchain-based approach.” *IEEE Transactions on Parallel and Distributed Systems*, vol. 30, no. 4, pp. 870–882, Sep. 2018.

[18] X. Fan and Y. Huo, “Blockchain based dynamic spectrum access of non-real-time data in cyber-physical-social systems,” *IEEE Access*, vol. 8, pp. 64 486–64 498, Aril 2020.

[19] Paik H Y, Xu X, Bandara H M N D, et al. “Analysis of data management in blockchain-based systems: From architecture to governance.” *IEEE Access*, vol.7. pp.186091-186107, 2019.

[20] Zhu L, Wu Y, Gai K, et al. “Controllable and trustworthy blockchain-based cloud data management.” *Future Generation Computer Systems,* vol.91. pp. 527-535. 2019.

[21] Shen, Meng & Zhang, Jie & Zhu, Liehuang & Xu, Ke & Tang, Xiangyun. “Secure SVM Training Over Vertically Partitioned Datasets Using Consortium Blockchain for Vehicular Social Networks.” *IEEE Transactions on Vehicular Technology*. pp.1-1. 10.1109/TVT.2019.

[22] Waqas, M, Kumar, K, Laghari, AA, et al. “Botnet attack detection in Internet of Things devices over cloud environment via machine learning.” Concurrency Computat Pract Exper. vol.34.no.4.2022. doi:10.1002/cpe.6662

[23] Lv Z, Qiao L, Hossain M S, et al. “Analysis of Using Blockchain to Protect the Privacy of Drone Big Data.” *IEEE Network*, vol.35.no.1. pp:44-49, 2021.

[24] Xiaojie GUO, Jin LI, Zheli LIU, Yu WEI, Xiao ZHANG, Changyu DONG. “Labrador: towards fair and auditable data sharing in cloud computing with long-term privacy.” *Information Sciences*, vol.65.no.5. pp.112-125. 2022

[25] Song, Su. “An Effective Big Data Sharing Prototype Based on Ethereum Blockchain.” *Scientific Programming*, pp.1-14. 2022

[26] Du M, Chen Q, Chen J, et al. “An Optimized Consortium Blockchain for Medical Information Sharing.” *IEEE Transactions on Engineering Management*, vol.99. pp.1-13. 2020.

[27] Jaiman V, Pernice L, Urovi V. “User incentives for blockchain-based data sharing platforms.” *PLOS ONE*, pp.17. 2022

[28] Laghari, Asif Ali, Abdullah Ayub Khan, Reem Alkanhel, Hela Elmannai, and Sami Bourouis. "Lightweight-BIoV: Blockchain Distributed Ledger Technology (BDLT) for Internet of Vehicles (IoVs)." *Electronics 12*, no. 3 pp.677,2023.

[29] Ma J, Chen Y, Wang Z, et al. “A rational delegating computation protocol based on reputation and smart contract.” *Journal of Cloud Computing*, vol.10.no.1. pp.1-12. 2021.

[30] Dong C, Wang Y, Aldweesh A, McCrory P, van Moorsel A, “Betrayal, distrust, and rationality: Smart counter-collusion contracts for verifiable cloud computing.” *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security,* pp.211–227. 2017.

[31] Chen Y, Guo J, Changlou L, Ren W, “Fade: A blockchain-based fair data exchange scheme for big data sharing. Future Internet”, vol.11.no.11. pp.225,2019.

[32] Quanxing Z, Qiuxian L, Meimei F, “Anti-collusion delegation computation protocol of three-party game based on smart contract.” *Computer Engineer.* Vol.46.no.8. pp.124–131,2020.

[33] Xiao L, Chen Y, Lin W, Liu K, “Indirect reciprocity security game for large-scale wireless networks.” IEEE Transactions on Information Forensics and Security
. vol.7.no.4. pp.1368–1380,2012.

[34] Wang S, Tang X, Zhang Y, Chen J, “Auditable protocols for fair payment and physical asset delivery based on smart contracts.” *IEEE Access*, vol.7. pp.109439–109453,2019

[35] Kou, Huaizhen & Liu, Hanwen & Duan, Yucong & Gong, Wenwen & Xu, Yanwei & Xu, Xiaolong & Qi, Lianyong, “Building trust/distrust relationships on signed social service network through privacy-aware link prediction process.” Applied Soft Computing, vol.100.no.5.pp.106942,2020.
https://doi.org/10.1016/j.asoc.2020.106942

[36] C. Wu, J. Xiong, H. Xiong, Y. Zhao, and W. Yi, “A Review on Recent Progress of Smart Contract in Blockchain," *in IEEE Access*, vol. 10. pp. 50839-50863, 2022, doi: 10.1109/ACCESS.2022.3174052.

[37] Shen W, Huang X, Fu Y, et al. “Self-Renewal Consortium Blockchain Based on Proof of Rest and Strong Smart Contracts.” *Tsinghua Science and Technology*, vol.27.no.6. pp.964-972,2022.

[38] Yu X, Shu Z, Li Q, et al. “BC-BLPM: A multi-level security

access control model based on blockchain technology." *China communications.* vol.18.no.2. 2021
[39] Zhu Jian, Hu Kai, Zhang Bojun. "A review of formal Verification Methods for Smart Contracts." *Acta Electronica Sinica*, vol.49.no.4. pp.792-804,2021
[40] W. Nam and H. Kil, "Formal Verification of Blockchain Smart Contracts via ATL Model Checking," in *IEEE Access*, vol. 10, pp. 8151-8162, 2022, doi: 10.1109/ACCESS.2022.3143145.
[41] Sultana, Tanzeela, Ahmad Almogren, Mariam Akbar, Mansour Zuair, Ibrar Ullah, and Nadeem Javaid. "Data Sharing System Integrating Access Control Mechanism using Blockchain-Based Smart Contracts for IoT Devices." *Applied Sciences vol.*10.no.2. pp.488, 2020. https://doi.org/10.3390/app10020488
[42] Zhang B J, Guo Y C, Wang Z K, et al. "Research on data sharing incentive mechanism based on smart contract."Computer Engineering",vol.48.no.8.pp.37-44, 2022.
[43] Lixin Shen, Qin Yang, Yunxia Hou, Jinglin Lin, "Research on information sharing incentive mechanism of China's port cold chain logistics enterprises based on blockchain," *Ocean & Coastal Management*, vol.225. pp.106229,2022. ISSN 0964-5691, https://doi.org/10.1016/j.ocecoaman.2022.106229.
[44] Xing, X.H., Hu, Z.H., Wang, S.W., Luo, W.P., "An evolutionary game model to study manufacturers and logistics companies' behavior strategies for information transparency in cold chains." Mathematical problems in engineering. 2020.
[45] Tan J, Jiang G, Wang Z. "Evolutionary game model of information sharing behavior in supply chain network with agent-based simulation." *International Journal of Intelligent Information Technologies,* vol.15.no. pp.54–68,2019.
[46] W. Chen, Y. Chen, X. Chen, and Z. Zheng, "Toward Secure Data Sharing for the IoV: A Quality-Driven Incentive Mechanism with On-Chain and Off-Chain Guarantees," in *IEEE Internet of Things Journal*, vol. 7, no. 3, pp. 1625-1640, March 2020, doi: 10.1109/JIOT.2019.2946611.
[47] Tao Ran, Jin Yongjin, "Business Economics and Management," vol.3. pp.46-53,2009. doi 10.3969/j.issn.1000-2154.2009.03.008
[48] Eppler, Martin & Helfert, Markus. "A classification and analysis of data quality costs." *International Conference on Information Quality*.pp.311-323,2004
[49] Li Jingjing. "Data quality evaluation model and evaluation tool research." *Donghua University*, pp.40-55,2018. https://kns.cnki.net/kcms/detail/detail.aspx?dbname=cmfd201901&filename=1018839535.nh
[50] Li X, Liu Q, Wu S, et al, "Game theory based compatible incentive mechanism design for non-cryptocurrency blockchain systems." *Journal of Industrial Information Integration*, vol.31. pp.100426,2023

ORCID iDs

Shuyi Yang https://orcid.org/0000-0002-9180-4842
Lusu Li https://orcid.org/0000-0001-5799-8244
Libo Feng https://orcid.org/0000-0001-7804-3535

Declarations:

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

No external data was used in this research

Competing interest

The authors report there are no competing interests to declare.

Funding

This work was supported by the Science and Technology Plan in Key Fields of Yunnan under Grant 202103AN080001-001,202001BB050076, 202002AA100007, 202002AB080001-8,202202AD080002.

Author's contribution

Shuyi Yang: Conceptualization, Methodology,Incentive Algorithm design, System function module design, Comparative analysis and Visualization, Writing-Original draft preparation, Reviewing, and Editing. Lusu Li: Smart Contract design, Experiment and Visualization. Libo Feng: Supervision.

Acknowledgment

Not applicable

Author's information

Libo Feng received his Ph.D. from the School of Computer Science and Engineering, Beihang University, China, in November 2019. He is currently a lecturer with the School of Software, Yunnan University, and Yunnan Key Laboratory of Blockchain Application Technology. His research interests include Blockchain technology and application, network security, and the Internet of things technology.

Shuyi Yang is a research assistant at Beihang University Yunnan Innovation Institute, who graduated from the Chinese University of Hong Kong with a master's degree. She also serves as the product manager of the Blockchain Big Data Information Sharing Project at the Digital Economy Research Centre. The project aims to incubate applied blockchain technology products with scientific research. Her main research interests are information interface design, blockchain application, and Information sharing behavioral research.

Lusu Li is a research assistant at Beihang University Yunnan Innovation Institute and a graduate student at Yunnan University. He also serves as the java engineer of the Blockchain Big Data Information Sharing Project at the Digital Economy Research Centre. The project aims to incubate applied blockchain technology products with scientific research. He mainly engaged in blockchain technology development and integrated development of blockchain applications.