

Current Research Status, Construction Ideas, and Future Development of Data Element Circulation Trading Platform

Zixuan XIE

College of Economics and Management, Beijing University of Technology, Beijing 100124, China

E-mail: xiezixuan@emails.bjut.edu.cn

Xuefan DONG

College of Economics and Management, Beijing University of Technology, Beijing 100124, China

E-mail: dongxf@bjut.edu.cn

Junpeng BO

College of Economics and Management, Beijing University of Technology, Beijing 100124, China

E-mail: bojp@emails.bjut.edu.cn

Jian LI*

College of Economics and Management, Beijing University of Technology, Beijing 100124, China

E-mail: lijiansem@bjut.edu.cn

Abstract Data element circulation trading platforms are pivotal in unlocking the value of data, optimizing resource allocation, and enabling cross-industry data sharing. However, existing platforms face significant limitations in addressing the evolving demands of the data market and the diverse requirements of users, with their architectural frameworks remaining underdeveloped. This study proposes a comprehensive theoretical framework that integrates advanced technologies such as blockchain, non-fungible tokens, and federated learning to enhance the functionality and adaptability of these platforms. A comparative analysis with the Ocean Protocol platform is conducted to elucidate the theoretical contributions and practical implications of the proposed framework. Based on this analysis, the study offers strategic recommendations for advancing the design and implementation of data element circulation trading platforms, providing valuable insights to guide future research and practice in this critical domain.

Keywords data element; data element circulation trading platform; blockchain technology; digital economy

Received December 21, 2024, accepted January 9, 2026

Supported by National Natural Science Foundation of China (71932002); National Natural Science Foundation of China (72274010)

*Corresponding author

1 Introduction

Data elements have emerged as a vital resource and a key driver of productivity in the digital economy, playing an essential role in advancing China's strategy to integrate the digital and real economies^[1]. Data element circulation trading platforms are specialized online platforms designed to facilitate the exchange, sharing, and monetization of data assets among various stakeholders, such as individuals, enterprises, and government bodies in a safe and secure environment^[2]. As a core mechanism for realizing the value of data, data element circulation trading platforms facilitate data flow, amplifying its utility and contributing significantly to the development of the digital economy^[3].

Over the past decade, China has made substantial progress in establishing data element circulation trading platforms. According to a report published on OFweek Tele, by the end of 2023, the market for these platforms had expanded rapidly, with over 80 platforms operating nationwide^[4], ensuring that nearly every medium and large city hosts at least one. These platforms have evolved from basic intermediary models to comprehensive service platforms, offering a broader range of transactions and user-oriented services. Through continuous technological advancements, these platforms have optimized a feedback cycle of "transaction $\rightarrow$ measurement $\rightarrow$ data $\rightarrow$ value $\rightarrow$ further transaction", enabling sustained growth in transaction volumes and data scale. Nonetheless, as the data market matures and user demands grow increasingly complex, existing platforms face operational challenges. Current technical infrastructure often proves inadequate, and platforms struggle to meet the diverse needs of the market^[5].

Thus, this study examines the development and research landscape of data element circulation trading platforms in China and globally, identifies key limitations, and proposes an innovative platform framework tailored to contemporary market demands. The proposed framework integrates advanced technologies such as blockchain, a distributed and immutable digital ledger that securely records transactions in a decentralized manner^[6]; non-fungible tokens (NFTs), which are unique cryptographic assets on a blockchain that represent ownership of a specific digital item, such as a dataset^[7]; and federated learning, a decentralized machine learning approach that enables model training on data from multiple sources without exchanging the raw data itself, thus preserving privacy^[8]. An analysis with Ocean Protocol, a representative data element circulation trading platform, demonstrates the advantages of the proposed framework, particularly in data uploading and processing, privacy protection, pricing mechanisms, and domain-specific services. Based on these findings, the study provides several targeted recommendations for the future development and construction of data element circulation trading platforms.

The structure of this paper is as follows: Section 2 reviews the current development and research status of data element circulation trading platforms in China and internationally. Section 3 outlines the proposed platform architecture, incorporating blockchain, NFTs, and federated learning. Section 4 compares the proposed framework with Ocean Protocol, highlighting its innovations and strengths. Section 5 offers practical recommendations for the future construction

of data element circulation trading platforms. Finally, section 6 concludes with a summary of the study's findings, contributions, and limitations.

2 Literature Review

2.1 Research on Data Privacy and Security

Data privacy and security are foundational to the trust and functionality of data trading platforms, leading to the adoption of advanced cryptographic and architectural solutions. Early studies focused on blockchain as a tool for decentralized and secure data trading. For example, Hu et al.^[9] proposed a data trading system based on blockchain, which ensures data security and privacy through a decentralized mechanism and designs automated contracts to improve the transparency and fairness of data transactions. Gu et al.^[10] proposed the consortium blockchain-based data trading framework, which uses consortium blockchain technology to support data transactions in mobile crowdsourcing environments and enhances the security and transparency of data transactions through distributed trust and privacy protection mechanisms.

More recent research has shifted toward integrating multiple privacy-preserving technologies. Federated learning has emerged as a key technology, allowing collaborative model training without centralizing raw data. For instance, Liu et al.^[8] proposed a framework that combines federated learning with blockchain to create a secure and privacy-preserving data-sharing ecosystem, where only model parameters are exchanged, thus protecting sensitive user information. Similarly, to combat risks of data leakage from model parameters, differential privacy has been integrated. This technique adds statistical noise to data or query results to make it impossible to identify any single individual's contribution. Gao et al.^[11] proposed a crowdsourcing data trading mechanism with differential privacy, which ensures the privacy of data owners through privacy-preserving algorithms and provides effective privacy control options. Combining these technologies offers a multi-layered defense, a concept explored by A. Yazdinejad, who presented a hybrid model using blockchain and federated learning to secure data in industrial internet of things environments, ensuring both data integrity and confidentiality^[12].

2.2 Research on Data Pricing

Existing studies in this domain emphasize the importance of advanced pricing models in resolving key issues in data trading, including fairness, efficiency, and market optimization. By leveraging game theory, computational methods, and blockchain technology, these models contribute to the development of secure, equitable, and effective pricing mechanisms. For instance, Fernandez et al.^[13] constructed a pricing model for the data market based on game theory and flow control theory to optimize data circulation and pricing issues. Lu et al.^[14] proposed a new data pricing scheme to allocate data resources according to the amount paid by the buyer to ensure fair transactions and improve the allocation efficiency of data resources. Araz et al.^[15] analyzed the costs and benefits of data transactions through a computational pricing model and explored the impact of different pricing strategies on platform operational risk management.

Spiekermann^[16] explored the pricing trends and methods of data commodities and innovatively proposed a pricing model that took into account the cost of data storage and processing, helping platforms to effectively evaluate the market value of data commodities. Jones et al.^[17] studied how platforms could perform differentiated pricing based on data characteristics and innovatively proposed a price manipulation model for data intermediaries during transactions to optimize platform revenue and user benefits. Dai et al.^[18] proposed a blockchain-based data trading platform that innovatively incorporated transaction time and cost factors into the platform's pricing model to ensure a secure and efficient pricing mechanism.

2.3 Research on Industry Customized Services and Data Ownership

Existing studies in this domain have made significant advancements in addressing industry-specific needs and fundamental issues such as data ownership. In the realm of industry-customized services, Abbasi et al.^[19] designed an industrial internet of things data trading system, using blockchain technology to support the monetization of industrial data, and improving the security, transparency and revenue management of data transactions. Liu et al.^[20] proposed a spatial data transaction framework based on blockchain, which realized the secure transaction of spatial data through distributed ledger technology, enhanced the authenticity and privacy protection of data, suitable for geographic information services. Zhang et al.^[21] proposed a data trading framework for the industrial internet of things. Through a data usage control mechanism, they ensured the reasonable circulation of data and privacy protection, thereby improving the efficiency and security of data transactions in industrial environments. Oh et al.^[22] proposed a privacy valuation model for the platform, which enabled multiple parties in the IoT data market to find a balance between privacy protection and profit maximization, helping to enhance the fairness and participation of transactions. Xiong et al.^[23] proposed a blockchain-based "one-to-many" data trading solution for the platform. Through a distributed ledger, they achieved secure and fair data transactions, suitable for complex transaction scenarios with multiple buyers. Addressing the critical issue of data ownership, Dosis et al.^[24] explored the intersection between data ownership and privacy protection and proposed frameworks for assigning data ownership rights, considering both legal and economic implications. Boerding et al.^[25] offered a framework to conceptualize data ownership and the legal dimensions of access rights.

However, although relevant research has made some progress in technological innovation, its research perspective is still limited. These studies often focus on single-point breakthroughs at the technical level, but lack in-depth consideration of the overall construction of the platform and the systematic framework. This limitation has, to a certain extent, restricted the widespread promotion and application of technological achievements.

Existing research has not yet formed a comprehensive construction plan that can accurately connect with the technological frontier, fully consider the platform's systematic architecture, and integrate actual market needs. Based on this, this study, based on the development status and research status of platforms, innovatively proposes a construction architecture of a data element

circulation trading platform that can adapt to the current data market construction and user needs from a technical level, and compares it with the architecture of the current mainstream platform Ocean Protocol to demonstrate its scientific validity and innovation from a theoretical level. By proposing a new architecture design and theoretical demonstration, a more effective solution is explored in order to fill the gap in this research field and at the same time provides reference for the construction of Chinese data element circulation trading platform.

3 Comparative Analysis of Data Element Circulation Trading Platforms in China and Other Countries

3.1 Classification of Data Element Circulation Trading Platforms

According to the organizational form and construction entities of the platform, platforms can be divided into three types: government-led platforms, enterprise-led platforms and industry alliance platforms^[26] (Table 1). Specifically, government-led platforms, typically operated by state-owned capital holdings, benefit from access to extensive government-level data resources, ensuring high levels of credibility and authority. Enterprise-led platforms usually adopt a vertically integrated operational model, encompassing the entire industry chain from raw data collection to the provision of final data productization services. Industry alliance platforms, guided by government policies, are often co-managed by state-owned and private enterprises. These platforms emphasize alliance mechanisms to foster collaboration and focus on building and enhancing the data trading ecosystem within specific regions. Statistical analysis of platform types reveals that in China, government-led and enterprise-led platforms have developed in parallel, whereas industry alliance platforms are comparatively rare. In contrast, platforms in other countries tend to be dominated by enterprise-led and industry alliance models, with government-led platforms playing a relatively minor role. This divergence highlights differences in governance structures and policy orientations across regions.

Based on the platforms' business model, data element circulation trading platforms can be also classified into two categories: third-party data trading platforms and comprehensive data service platforms^[27] (Table 2). Third-party data trading platforms operate with a relatively straightforward business model, serving as intermediaries to connect data suppliers with data demanders and oversee the transaction process. These platforms typically offer services such as data sales, data purchases, and the release of data demands, without directly participating in the transactions. As a result, the professional expertise required for platform staff is comparatively low.

In contrast, comprehensive data service platforms have a more complex business model. In addition to serving as intermediaries to facilitate data exchange between suppliers and demanders, these platforms actively participate in transactions as data suppliers and service providers. This involvement requires data processing, which typically employs advanced technologies such as machine learning, data mining, and neural networks^[28]. A comparative analysis of Tables 1 and 2 reveals that most third-party data trading platforms in China are government-led,

Table 1 Classification of data element circulation trading platforms
(based on the platform's organizational form and construction entity)

Government-led	Enterprise-led	Industry alliance
Guiyang Big Data Exchange	Datatang	Zhongguancun Big Data Industrial Alliance
Beijing Big Data Trading Service Platform	Jingdong Wanxiang Platform	China Big Data Industry Ecology Alliance
Zhongguancun Shuhai Big Data Trading Platform	Juhe Data	Shanghai Big Data Alliance
Shanghai Data Exchange	Source of Big Data Trading Platform	Ocean Protocol (Singapore)
East China Jiangsu Big Data Trading Center	Tianyuan Data	GAIA-X (European Union)
Qiantang Big Data Trading Center	Tao Data	BiGDataOcean (European Union)
Zhongyuan Big Data Trading Platform	Shuduoduo	International Data Spaces (Germany)
Central China Big Data Exchange	CHECK AFDNET. CN	Mobility Data Marketplace (Germany)
Donghu Big Data	Merit Data Co. Ltd.	Open Data Space (Netherlands)
Harbin Data Trading Platform	Dawex (France)	Catena-X Automotive Network (Germany)
Xixian New Area Big Data Exchange	Dataplaza (Japan)	Data Market Austria (Austria)
Data.gov (USA)	Amass Insights (Japan)	
Open Data Institute (UK)	Qlikdatamarket (Sweden)	
MIDATA(Switzerland)	Acxiom (USA)	
Geospatial Data Exchange (Australia)	Azure (USA)	
MyData Finland (Finland)	Corelogic (Australia)	
	Data Republic (Australia)	
	Infosys Data Marketplace (India)	

whereas comprehensive data service platforms are predominantly enterprise-led. Similarly, in other countries, comprehensive data service platforms are largely enterprise-driven; However, their prevalence is significantly greater compared to third-party data trading platforms, reflecting a broader emphasis on integrated services within the global market.

Table 2 Classification of data element circulation trading platforms
(based on the platforms’ business model)

Region	Third-party data trading platform	Comprehensive data service platform
China	Guiyang Big Data Exchange	Datatang
	Beijing Big Data Trading Service Platform	Jingdong Wanxiang
	Zhongguancun Shuhai Big Data Trading Platform	Juhe Data
	Shanghai Data Exchange	Tianyuan Data
	East China Jiangsu Big Data Trading Center	Tao Data
	Zhongyuan Big Data Trading Platform	Shuduoduo
	Central China Big Data Exchange	Merit Data Co. Ltd.
	Donghu Big Data	iDataAPI
	Harbin Data Trading Platform	SHOWAPI
	Xixian New Area Big Data Exchange	
Other countries	Acxiom (USA)	Factual (USA)
	Corelogic (Australia)	MIDATA (Switzerland)
	Dawex (France)	Ocean Protocol (Singapore)
	Infosys Data Marketplace (India)	Open Data Institute (UK)
	Amass Insights (Japan)	Dataplaza (Japan)
	Data Market Austria (Austria)	MyData Finland (Finland)
		Open Data Space (Netherlands)
		International Data Spaces (Germany)
		Cloud Data Marketplace (Canada)
		DATANEXUS (Australia)

3.2 Comparative Analysis of Data Sources of Platforms in China and Other Countries

The data sources for data element circulation trading platforms can be broadly categorized into five types: 1) Government public data, which includes datasets provided by government entities for public use; 2) Enterprise internal data, generated within organizations and often proprietary; 3) Web crawler data, referring to structured information extracted from various websites by enterprises; 4) Data from data suppliers, contributed by external entities such as enterprises, research institutions, and individuals according to platform transaction rules; 5) Alliance partner data, encompassing data generated and accumulated through collaborations with enterprise partners^[28].

As shown in Table 3, platforms in China exhibit a diverse range of data sources, with government public data, enterprise internal data, and web crawler data serving as the primary contributors. This diversity enriches the platform’s data ecosystem, enabling broader data cov-

Table 3 Comparison of data sources of data element circulation trading platforms

Region	Name of platform	Data source
China	Guiyang Big Data Exchange	Government public data, internal data of the company, crawled network data, etc.
	Shanghai Data Exchange	Uploaded by data provider through the distributed platform
	Donghu Big Data	Government public data, internal data of the company, crawled network data, uploaded by data provider through the distributed platform, etc.
	Jingdong Wanxiang	Internal data of the company, cooperative enterprise data, etc.
	Tianyuan Data	Government public data, internal data of the company, Uploaded by data provider through the distributed platform, Cooperative enterprise data
Other countries	Factual (USA)	Government public data, crawled network data, uploaded by data provider through the distributed platform, etc.
	Acxiom (USA)	Uploaded by data provider through the distributed platform
	Dataplaza (Japan)	Uploaded by data provider through the distributed platform
	Refinitiv (UK)	Uploaded by data provider through the distributed platform
	Quandl (Canada)	Uploaded by data provider through the distributed platform

erage and enhanced usability. However, it simultaneously introduces significant challenges in ensuring data quality, security, and regulatory compliance, particularly given the heterogeneous nature of the sources. In contrast, platforms in other countries rely on more concentrated data sourcing, predominantly from data suppliers. This narrower scope simplifies data governance and regulatory oversight but may limit the variety and utility of the data available. These differences in sourcing strategies reflect varying regulatory environments, levels of digital infrastructure development, and market orientations. China's emphasis on integrating public and private data reflects a policy-driven approach aimed at maximizing the utility of national data assets, while the reliance on data suppliers in other countries indicates a more market-oriented approach with an emphasis on decentralized data governance. This contrast highlights the need for tailored strategies in platform development to address the distinct challenges and opportunities associated with diverse versus concentrated data sourcing models.

3.3 Comparative Analysis of Service Areas of Platforms in China and Other Countries

Based on the scope of services of the platforms, data element circulation trading platforms can be categorized into two types: multi-field comprehensive service platforms and single-field professional service platforms^[29] (Table 4).

Table 4 Comparison of service areas of data element circulation trading platforms

Region	Name of platform	Service areas
China	Beijing Big Data Trading Service Platform	Government, economy, humanities, transportation, etc.
	Zhongyuan Big Data Trading Platform	E-commerce, enterprise, life services, resources and energy, transportation and geography, financial services, medical care, etc.
	Juhe Data	Life services, financial technology, transportation, geography, recharge and payment, data intelligence, enterprise management, etc.
	Source of Big Data Trading Platform	Social, finance, e-commerce, automobiles, talent, real estate, medical care, enterprises, tourism, scientific research, consulting, reading, etc.
	CHECK AFDNET. CN	Financial stocks, recharge authentication, convenience, news articles, medicine and transportation, science and education, literature and art, etc.
Other countries	Corelogic (Australia)	Real estate, business, etc.
	Kuberre Systems (USA)	Business
	Qlikdatamarket (Sweden)	Communications, manufacturing, energy utilities, etc.
	Dataiku (France)	Insurance, marketing, supply chain and logistics, etc.
	DataStreamX (Singapore)	Geography

Multi-field comprehensive platforms address the increasing demand for cross-industry and multidimensional data by offering broad and efficient data collection and processing services. These platforms facilitate cross-sector integration and in-depth data exploration, significantly expanding the value boundaries of data applications. By catering to a wide array of industries, they support complex, interrelated data use cases and contribute to the broader utility of data in diverse economic and societal contexts.

In contrast, single-field professional platforms concentrate on specific industries or domains. Leveraging specialized expertise and deep domain-specific knowledge, these platforms provide precise and highly targeted data services. This focus enhances the accuracy and relevance of data for specific applications, offering robust support for decision-making, market analysis, and research within the targeted industry. Their specialization ensures the delivery of high-value insights tailored to niche requirements, positioning them as critical resources for domain-specific innovation and strategic planning.

3.4 Comparative Analysis of Pricing Methods of Platforms in China and Other Countries

As shown in Table 5, both Chinese and international platforms commonly employ flexible, on-demand pricing strategies. Depending on the nature of the traded data products, platforms can choose from time-based pricing models (e.g., by hour or by month/year), allowing for tailored cost structures that align with specific use cases. This approach not only acknowledges the intrinsic value of the labor and resources involved in data generation and processing but also incentivizes precise procurement based on actual demand, thereby minimizing resource idleness and waste^[30].

Table 5 Comparison of pricing methods on data element circulation trading platforms

Region	Name of platform	Pricing method
China	Tianyuan Data	Direct pricing, on-demand pricing mechanism
	Source of Big Data	Direct pricing, on-demand pricing mechanism
	Trading Platform	
	Datatang	Negotiated pricing, on-demand pricing mechanism
	Juhe Data	Direct pricing, on-demand pricing mechanism
	Zhongyuan Big Data	Direct pricing, on-demand pricing mechanism
	Trading Platform	
Other countries	Corelogic (Australia)	On-demand pricing mechanism
	Refinitiv (UK)	On-demand pricing mechanism
	Qlikdatamarket	On-demand pricing mechanism
	(Sweden)	
	Dataplaza (Japan)	On-demand pricing mechanism
	BDEX (USA)	On-demand pricing mechanism, negotiated pricing mechanism

For scenarios requiring customized or highly specialized data services, agreed pricing mechanisms are often utilized. In this model, the data demander articulates specific requirements, and the data provider proposes a tailored pricing plan, with both parties negotiating adjustments to finalize the terms. This model provides significant flexibility, enabling transactions to reflect the unique value of the data and services exchanged.

However, agreed pricing presents challenges due to its reliance on bilateral negotiation without standardized benchmarks. The absence of unified guidelines can introduce inconsistencies and, over time, may disrupt market order. Prolonged reliance on unregulated pricing practices increases the risk of market inefficiencies, including unfair pricing, opaque transactions, and reduced trust among participants. As the data trading market matures, the establishment of standardized pricing frameworks and regulatory oversight will be essential to ensure stability, transparency, and equitable practices, mitigating potential risks and fostering long-term market development.

3.5 Comparative Analysis of Transaction Models of Platforms in China and Other Countries

Current data element circulation trading platforms typically operate through three primary transaction modes: The big data analysis result transaction mode, the data product transaction mode, and the transaction intermediary mode^[31]. In the big data analysis result transaction mode, the platform processes specific datasets based on the requirements of the data demander and delivers the resulting analysis. This mode emphasizes the added value generated through data processing and is particularly suited for scenarios where actionable insights are prioritized over raw data access. The data product transaction mode, also referred to as the “data customization mode”, involves the development of personalized data products by the data provider. In this model, the provider tailors basic data products to meet the unique needs of the data demander, offering flexibility and precision for specialized applications. The transaction intermediary mode, by contrast, functions purely as a facilitator of data exchanges. Platforms operating in this mode neither store nor analyze data; Instead, they provide a channel for data suppliers and demanders to interact, manage the transaction process, and promote the efficient circulation of data elements.

As shown in Table 6, most platforms in both China and other countries adopt a hybrid approach, combining elements from multiple transaction modes. This integrative strategy enables platforms to leverage the strengths of different modes, enhancing their ability to address diverse user needs. For instance, the hybridization allows platforms to offer both raw data products and processed analytical insights while maintaining flexibility in transaction facilitation.

3.6 Analysis of Problems Existing in Chinese Current Data Element Circulation Trading Platform

A comparative analysis of data element circulation trading platforms in China and other countries reveals several critical challenges specific to Chinese platforms. While Chinese platforms have demonstrated rapid growth and diversification in response to increasing user demands and the expanding data market, they also face significant operational shortcomings. These include weak privacy and security mechanisms, unclear data ownership frameworks, a lack of specialized services, inadequate data pricing systems, and insufficient standardization in data management processes.

3.6.1 Weak Data Privacy and Security Mechanisms Undermine User Trust

Compared with platforms in other countries, Chinese data element circulation trading platforms benefit from highly diverse data sources, including government public data, enterprise data, and web-crawled data. This diversity significantly enriches the data ecosystem and broadens the scope of available resources. However, this advantage also brings heightened risks of privacy breaches, which are less pronounced in international platforms that often employ more advanced privacy-preserving technologies. For example, platforms abroad frequently leverage differential privacy algorithms, blockchain-based security frameworks, and rigorous encryption

Table 6 Comparison of transaction modes of data element circulation transaction platforms

Region	Name of platform	Trading model
China	Qiantang Big Data Trading Center	Big data analysis results trading model, data product trading model
	East China Jiangsu Big Data Trading Center	Big data analysis results trading model, data product trading model
	Data Bao	Big data analysis results trading model, data product trading model
	Data Galaxy	Big data analysis results trading model, data product trading model, trading intermediary model
	Baidu Smart Cloud	Data product transaction model, transaction intermediary model
Other countries	Azure (USA)	Big data analysis results trading model, data product trading model, trading intermediary model
	Corelogic (Australia)	Big data analysis results trading model, data product trading model
	Refinitiv (UK)	Big data analysis results trading model, data product trading model
	Qlikdatamarket (Sweden)	Big data analysis results trading model, data product trading model
	Dataplaza (Japan)	Transaction intermediary model

standards to mitigate such risks, ensuring stronger user trust and more secure data transactions. In contrast, many Chinese platforms lack robust privacy-preserving technologies and comprehensive protective measures, which limits their ability to conduct effective cross-source data mining and analysis. This deficiency reduces the potential value that could be extracted from diverse data sources and constrains the platforms' capacity to provide high-quality insights and services^[32]. Additionally, weak privacy protections erode user trust, discouraging data sharing and participation. Addressing these shortcomings requires the implementation of cutting-edge privacy-preserving technologies and stronger regulatory frameworks to bolster user confidence and facilitate secure data exchange.

3.6.2 Ambiguity in Data Ownership and the Need for Robust Ownership Confirmation Mechanisms

Compared to platforms in other countries, Chinese data element circulation trading platforms face significant challenges in clarifying and managing data ownership—A fundamental issue for building a transparent and efficient data market. The data shared on these platforms is diverse and often involves multiple stakeholders, creating a complex network of overlapping ownership rights^[33]. In contrast, many international platforms are advancing frameworks that define ownership more explicitly, leveraging legal, technical, and contractual mechanisms to address

these complexities.

The non-competitive and non-exclusive nature of data, combined with its unique cost structure characterized by high fixed costs and near-zero marginal costs, further exacerbates the problem. These features facilitate buyers to replicate and resell data at minimal cost, increasing the risks of data piracy. Such practices not only harm the interests of data owners but also disrupt market fairness and efficiency. While international platforms increasingly employ mechanisms like blockchain for ownership verification and digital watermarking to prevent unauthorized use, similar practices remain underdeveloped in China. The absence of clear and enforceable ownership mechanisms also discourages data owners in China from participating actively in the market due to concerns over potential losses. This hesitancy limits the optimal distribution and utilization of data resources, restricting innovation and the broader development of the digital economy^[34].

3.6.3 Insufficiently Specialized Services for Niche Markets

Chinese platforms exhibit a limited ability to provide specialized services tailored to specific industries or emerging data needs, a challenge that is less pronounced in many international platforms. Unlike platforms abroad that utilize advanced analytical tools and extensive domain expertise to anticipate and fulfill specialized demands, Chinese platforms often rely on instructions from data requesters to process and deliver results^[32]. While this approach can meet immediate demands, it has somewhat hindered a deeper exploration and utilization of the full potential of the data. This is particularly evident in specialized fields and cross-industry data mining, where the lack of advanced analytical tools and specialized knowledge restricts the ability to generate valuable insights and unlock the true value of data. To address these shortcomings, Chinese platforms must prioritize the development of professional services tailored to specific industries, enhance their understanding of emerging data demands, and invest in cross-domain analytical capabilities. By drawing on international best practices and integrating cutting-edge tools, Chinese platforms can improve their ability to deliver targeted solutions, unlock the full value of data, and better support innovation and economic growth.

3.6.4 Weak Data Valuation and Pricing Mechanisms

Chinese platforms face persistent challenges in establishing effective data pricing mechanisms, which are more advanced on international platforms. Globally, data pricing remains an unresolved issue, as no comprehensive or standardized pricing models have yet been widely adopted^[35]. However, international platforms are increasingly incorporating sophisticated approaches, such as algorithm-driven valuations and predictive analytics, into their pricing systems, while Chinese platforms lag behind. Unlike traditional commodities, the value and pricing of data are influenced by factors including variety, completeness, timeliness, and usage scope^[36]. Many international platforms address these complexities by leveraging advanced technologies, such as machine learning and simulation models, to establish dynamic pricing systems. In contrast, Chinese platforms often rely on static and simplistic pricing mechanisms that fail to

account for the nuanced characteristics of data, leading to inefficiencies and market distortions. Furthermore, the extraction of data value frequently requires computational tools, creating an information asymmetry between human users and automated systems. While international platforms are increasingly addressing this gap through intelligent valuation systems, Chinese platforms have been slower to adopt these innovations. Bridging this gap will require the integration of advanced pricing technologies and the development of transparent, data-driven valuation models to enhance market efficiency and user trust.

3.6.5 Lack of Data Management and Standardization, Uneven Data Quality

Chinese data element circulation trading platforms face significant challenges in data management and standardization compared to their international counterparts. These issues are evident in two primary areas: The low entry thresholds for trading entities and the absence of standardized data format specifications. Firstly, Chinese platforms typically employ basic registration systems, allowing users to participate in trading with minimal verification of identity or operational credentials^[30]. In contrast, international platforms often enforce stricter entry requirements, including thorough evaluations of user legitimacy, which help ensure higher levels of security and trust in the data ecosystem. The leniency of Chinese platforms results in inconsistencies in data quality, as initial reviews often fail to filter out unreliable or substandard data^[27]. Second, the absence of standardized data format specifications further hampers the efficiency of Chinese data market. International platforms increasingly adhere to global standards for data formats and structures, facilitating seamless interoperability and communication between trading entities. In comparison, many Chinese platforms operate independently, with fragmented systems that result in non-standardized data formats, inconsistent structures, and varied semantic interpretations. This lack of standardization disrupts collaboration and diminishes the overall efficiency of data transactions^[36].

4 A New Data Element Circulation Trading Platform

4.1 Existing Problems of the Platform Drive the Birth of New Systems

In the previous analysis, we systematically reviewed the developmental status of data element circulation trading platforms and critically examined their prevailing challenges. These issues, including weak privacy protections, ambiguous ownership frameworks, insufficient specialization, inadequate pricing mechanisms, and a lack of standardization, not only diminish the operational effectiveness of these platforms but also constrain the sustainable growth of the broader data market.

To address these critical pain points, this study proposes an innovative conceptual framework for constructing data element circulation trading platforms. The proposed framework seeks to create a secure, efficient, transparent, and user-centric trading environment by leveraging a comprehensive integration of advanced technologies. By aligning platform development with these principles, the study aims to contribute practical solutions to enhance platform functionality, improve market dynamics, and foster trust and engagement among stakeholders, ultimately

driving the healthy evolution of the data economy.

4.2 Overall Architecture

The overall architecture of the data element circulation trading platform (hereinafter referred to as “DECTP”) designed in this paper is shown in Figure 1, which mainly consists of the following modules:

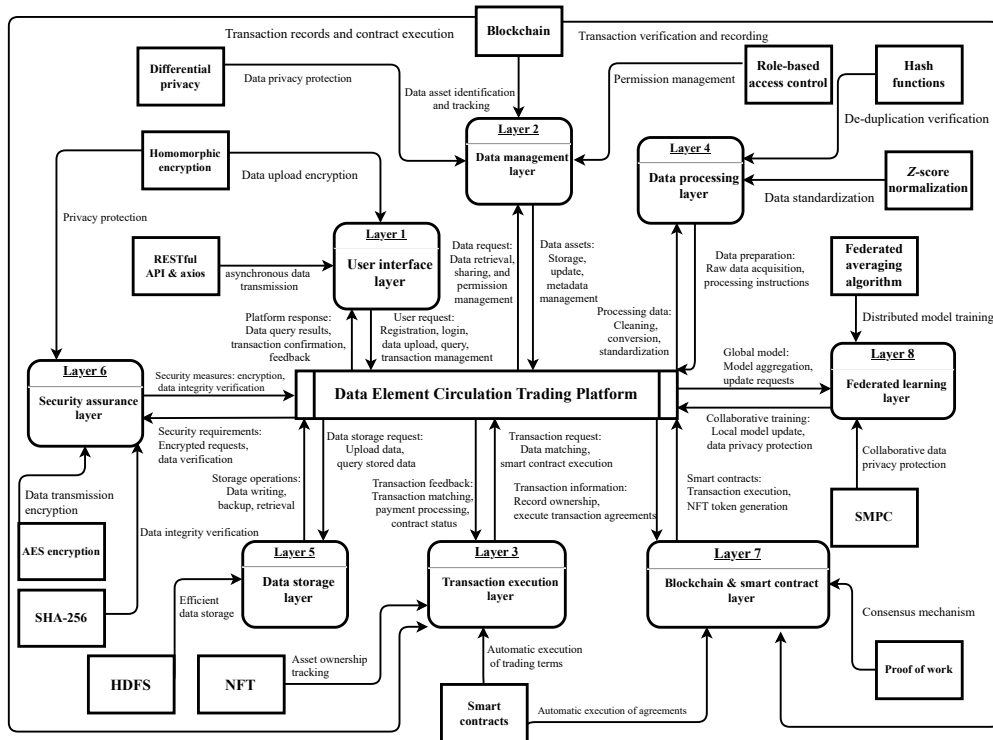


Figure 1 The overall architecture of the data element circulation trading platform

1) User Interface Layer

(i) Functional Overview

The user interface layer is the portal of the data element circulation trading platform, providing an entry point for users to interact with the platform. This layer supports data providers and demanders to perform registration, login, data upload, query, transaction management, and other related operations. This layer is designed to improve user experience, user participation and promote the wide application of the platform.

(ii) Implementation Mechanism

(a) **Front-end technology:** Use the React.js framework to build a responsive user interface to ensure the responsive design and cross-platform compatibility of the interface.

(b) **API communication:** The React.js front-end sends asynchronous requests to the RESTful API through Axios, such as data upload and query. The back-end service receives the request and performs operations through the microservice architecture, returning a response in

JSON format. The front-end updates the interface based on the response result. This method simplifies the data transaction and management process, and improves the interaction efficiency and user experience.

(c) Microservice architecture: Using the microservice architecture, the application is split into multiple independent service modules. Each module is designed around a specific business function and is independently deployed and expanded. It provides vertical and specialized services based on the industry and provides customized service modules for different industries, especially data analysis, visualization, and processing services^[37].

When a user uploads data, the system routes the request to the corresponding microservice module through the RESTful API based on the user's industry identifier, such as dedicated data processing services in the fields of finance and healthcare. Each service can be independently developed, tested, and deployed, and continuously expanded and optimized according to industry needs. For example, users in the financial industry may receive specialized risk control data analysis services, while users in the healthcare industry focus on data privacy protection and analysis.

(d) Homomorphic encryption: Homomorphic encryption is used when uploading data, allowing encrypted data to be directly operated without decryption. After the data uploaded by the user is encrypted, mathematical operations can still be performed in the encrypted state, ensuring that the data remains encrypted during the processing process to avoid leakage. The homomorphic encryption formula is as follows:

$$E(x + y) = E(x) \oplus E(y), \quad (1)$$

where $E(x)$ represents the encryption operation, and $\oplus$ is the encryption algorithm.

When uploading data, the platform first performs homomorphic encryption on the data. The platform uses homomorphic encryption to analyze and mine data while ensuring data privacy, thereby providing users with safer data processing services.

2) Data Management Layer

(i) Functional Overview

The data management layer is responsible for the full life cycle management of data assets, including data storage, update, sharing, and metadata management. This layer is designed to ensure the consistency, traceability, and security of data and provide a solid foundation for data transactions.

(ii) Implementation Mechanism

(a) Data asset identification and registration: Combining decentralized identifiers with blockchain technology allows data assets to obtain unique identification without a centralized agency^[38], and be recorded on the blockchain, ensuring the uniqueness and immutability of data identity and managing the identity and permissions of data.

When a user uploads data, the platform first generates a unique decentralized identifier (DID) for the data, which is recorded in the blockchain together with the metadata of the data

asset. Due to the distributed ledger characteristics of the blockchain, the creation and change of each DID will be recorded on multiple nodes, ensuring that the identity information of the data cannot be tampered with and is traceable. This mechanism not only provides protection for data ownership and tracking, but also effectively manages data access rights and usage records.

(b) Access control and permission management: Use the role-based access control model to manage permissions. By associating permissions with roles, the system is managed to ensure that only authorized users can access and use specific data assets^[39]. The RBAC permission allocation formula is as follows:

$$P(\text{permissions}) \subseteq A(\text{assigned roles}), \quad (2)$$

where P represents the permission set, and A represents the assigned role.

The platform predefines a set of roles for different types of users, such as data providers, data consumers, and administrators. Each role has a set of permissions P . When a user registers and logs in, the system binds the user to a specific role A . The user's actual permissions are a subset of the set of permissions associated with the role, that is, formula (2), which means that the user can only perform the operations granted by his role.

(c) Differential privacy: Differential privacy technology is introduced to introduce noise into data query or analysis results. Even if the attacker obtains the query results, he cannot accurately infer the information of a single data point, which is used to protect user privacy. The differential privacy formula is as follows:

$$\Pr[\mathcal{M}(D_1) = t] \leq e^\epsilon \cdot \Pr[\mathcal{M}(D_2) = t], \quad (3)$$

where D_1 and D_2 are two data sets, ϵ is a privacy operation.

Differential privacy blurs data by adding an appropriate amount of random noise to queries or statistical results. In two similar data sets D_1 and D_2 , the probability of the same query $\mathcal{M}$ result t appearing is controlled by the parameter ϵ . A smaller ϵ value means higher privacy protection, a larger amount of noise, and it is difficult to infer a single data point.

3) Transaction Execution Layer

(i) Functional Overview

The transaction execution layer is responsible for processing and executing the logic of data transactions, including data matching, transaction matching, smart contract execution, and payment processing. Blockchain-based smart contracts can ensure the automatic execution and immutability of transaction terms^[40]. This layer is designed to improve transaction efficiency and ensure the fairness and security of transactions. In particular, Figure 2 illustrates the data upload and NFT generation process supported by the transaction execution layer, showing how uploaded data are processed, stored, and then associated with metadata to generate the corresponding NFT through the blockchain and smart contract layer.

(ii) Implementation Mechanism

(a) Transaction matching algorithm: Optimize transaction matching based on the linear programming model to ensure efficient matching between data supply and demand^[38]. This

algorithm sets the objective function to maximize the matching weights of the supply and demand sides, thereby improving the success rate and overall efficiency of the transactions. The matching algorithm formula is as follows:

$$\text{maximize } \sum_{i=1}^n \sum_{j=1}^m w_{ij} x_{ij}, \quad (4)$$

where w_{ij} represents the data matching weight between the supplier and the demander, and x_{ij} represents the decision variable for whether the transaction is completed.

After the data provider uploads the data and sets the price, the platform will calculate the matching weights of the supply and demand sides based on historical transaction data, user demand, and market trends. The system then uses a linear programming algorithm to calculate and determine the optimal transaction pairing. The transaction results are updated in real time, which can improve the matching efficiency of subsequent transactions. In addition, the algorithm can be integrated with smart contracts to ensure that the transaction is automatically executed after it is reached, improving the transparency and security of the transaction process.

(b) Non-fungible tokens: Non-fungible tokens (NFTs) are digital assets that uniquely identify the ownership of data assets on the blockchain. After the user uploads the data, the platform generates the corresponding NFTs through smart contracts to ensure that each data asset has a unique identifier and is recorded on the blockchain. The ownership of each transaction is tracked through the transfer of NFT tokens^[41]. The transfer formula of NFTs is as follows:

$$T_{\text{nft}(i, \text{owner}_j)} = \text{Transfer}(\text{nft}_i, \text{owner}_j), \quad (5)$$

where nft_i represents the NFT of the i -th data asset, and owner_j is the current holder.

This formula indicates that the NFT token of the i -th data asset is transferred to the current holder owner_j . When a user uploads data, a smart contract creates a new NFT, records the detailed information of the data asset, and stores it on the blockchain. The transfer of each NFT is executed through a smart contract to ensure the security and transparency of ownership. On the platform, NFTs are not only used to identify the ownership of data assets but can also be used as asset certificates in transactions to promote the secure exchange and market circulation of data. This mechanism can enhance users' trust in the platform and improve the transparency and security of data transactions.

(c) Dynamic pricing model: Data quality is calculated by scoring indicators such as data completeness, accuracy, and timeliness. Data prices are dynamically generated based on multiple factors such as data demand, supply, and quality. The pricing results will be executed through smart contracts and recorded on the blockchain^[42]. The pricing algorithm is as follows:

$$P = \alpha \cdot \frac{D_q}{D_s} + \beta \cdot R, \quad (6)$$

where P is the data price, D_q represents the demand, D_s represents the supply, R is the data quality score, α and β are the adjustment coefficients.

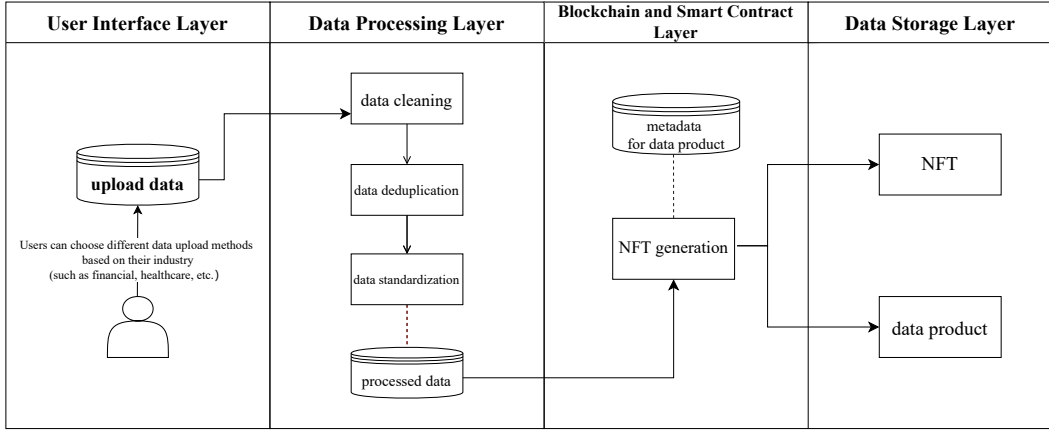


Figure 2 Data upload and NFTs generation process

When users upload data on the platform, the system will automatically evaluate the quality of the data, including indicators such as completeness, accuracy, and timeliness. The evaluation results will generate a comprehensive quality score R as an important factor in price calculation. The platform monitors market demand and supply in real time and collects the current demand and supply of data by analyzing users' transaction histories, market trends, and industry dynamics. The data price is calculated using a dynamic pricing algorithm, which adjusts α and β according to market conditions, reflecting market changes and user feedback. The calculated data price P will be automatically executed through a smart contract, that is, after the two parties to the transaction reach an agreement, the smart contract will process the transaction according to the currently calculated price to ensure the execution and recording of the transaction. Each price update and transaction execution result will be recorded on the blockchain to ensure the transparency and immutability of price changes and enhance user trust.

4) Data Processing Layer

(i) Functional Overview

The data processing layer provides automated ETL and is responsible for pre-processing data, including data cleaning, standardization, conversion and fusion, reducing errors caused by manual operations, improving data quality and interoperability, and ensuring data availability and reliability during transactions. This layer is designed to provide high-quality data for the platform and support diverse data analysis and application requirements.

(ii) Implementation Mechanism

(a) Data cleaning and deduplication: Use hash functions to detect and delete duplicate data^[43], ensuring data accuracy and consistency. When a user uploads data, the system applies a hash function to generate a hash value for each data item $h_k(x)$. The hash value is a unique representation of the data content and can effectively identify identical data items. The deduplication algorithm formula is as follows:

$$H(x) = (h_1(x), h_2(x), \dots, h_k(x)), \quad (7)$$

where $h_k(x)$ is the hash value of the data item, which is used to identify and remove duplicate data.

The system first compares the hash value of the newly uploaded data with the hash value of the existing data in the database. If the same hash value is found, the data is considered to be duplicated and the system will automatically delete the duplicate data to ensure the accuracy and consistency of the database.

(b) Data standardization: z -score standardization is used to enable unified processing of data from different sources, ensuring the compatibility and interoperability of data within the platform. The standardization formula is as follows:

$$z = \frac{X - \mu}{\sigma}, \quad (8)$$

where X is the data value, μ is the mean, and σ is the standard deviation.

The system first calculates the mean μ and standard deviation σ of all uploaded data, then converts each data item X to obtain the standardized value z , and stores the standardized data in the platform to ensure the compatibility and consistency of subsequent data processing.

5) Data Storage Layer

(i) Functional Overview

The data storage layer is responsible for securely and reliably storing and managing data assets within the platform and supports efficient data upload, retrieval, and backup. This layer is designed to provide a scalable storage solution that supports persistent storage and high-performance access requirements for large-scale data.

(ii) Implementation Mechanism

(a) Distributed storage technology: Use HDFS distributed database systems to achieve safe storage and efficient access of data^[44]. When users upload data, the platform will split large files into multiple 128 MB or 256 MB small blocks and store these data blocks dispersedly on multiple nodes in the cluster. HDFS copies each data block across multiple nodes (three copies by default) to ensure that even if some nodes fail, the data is still available. At the same time, HDFS supports large-scale parallel processing and can quickly read and write data. When a user requests data, the system will quickly obtain the data directly from the corresponding node according to the storage location of the data, which can significantly improve access efficiency.

(b) Data backup and recovery: The platform regularly performs data backup and automatically performs backup tasks by scheduling the Apache Airflow system to ensure regular snapshots of all data and reduce the risk of data loss. The CAP theory is also used to ensure the consistency of the storage system, availability and partition fault tolerance.

6) Security Layer

(i) Functional Overview

The security layer provides all-round data security protection, preventing data leakage, tampering, and unauthorized access through encryption technology and data integrity verification. This layer is designed to ensure the security of the data element circulation and trading platform

and the confidentiality, integrity, and availability of data.

(ii) Implementation Mechanism

(a) AES encryption: Use the AES encryption algorithm^[43] to ensure the confidentiality of data during transmission and storage. The encryption formula is as follows:

$$C = E_k(M), \quad (9)$$

where M is the plaintext, C is the ciphertext, E_k is the encryption operation, and k is the key.

When a user uploads data, the system first generates a key k . Next, the platform encrypts the user's plaintext data M with AES to generate ciphertext C . This process uses symmetric key encryption, that is, the same key is used for encryption and decryption. The platform uses a key management system to generate, store, and distribute keys. Only authorized services can access the keys to ensure data security. When data is transmitted over the network, all data is transmitted in ciphertext to prevent man-in-the-middle attacks and data leaks. At the same time, all data stored in the database is also stored in ciphertext to ensure that even if the database is attacked, the data content cannot be accessed.

(b) Data integrity verification: Use the SHA-256 hash algorithm to verify data integrity. The formula is as follows:

$$H(M) = \text{SHA-256}(M), \quad (10)$$

where $H(M)$ represents the hash value of data M .

During data upload or transmission, the platform will perform SHA-256 hash calculation on the data M to generate a hash value $H(M)$, which uniquely represents the content of the data. During data transmission, the system will also transmit the hash value of the data. After receiving the data, the receiver will calculate its hash value again and compare it with the transmitted hash value. If the two are the same, it means that the data has not been tampered with during transmission. If the hash values are inconsistent, the system will trigger an alarm and refuse to accept the data to ensure the integrity and reliability of the data.

7) Blockchain and Smart Contract Layer

(i) Functional Overview

The blockchain and smart contract layer provides decentralized transaction recording and automated contract execution functions, ensuring the transparency, immutability, and efficient execution of transactions. This layer is the core of trust in the data element circulation transaction platform and is designed to build a trusted transaction environment without third-party intervention.

(ii) Implementation Mechanism

(a) Distributed ledger technology: Use blockchain to record every data transaction and update to ensure that transaction data cannot be tampered with. Based on the working principle of Bitcoin^[45], the Proof of Work (PoW) mechanism is used to ensure the security of the ledger. The consensus algorithm formula is as follows:

$$H(\text{nonce} \parallel \text{data}) < \text{target}, \quad (11)$$

where nonce is a random number, data is transaction data, and H is a hash function.

When a user submits a transaction, the transaction data will be broadcast to the nodes in the blockchain network. Each node will collect these transactions and package them into a block. The node needs to find a random number nonce so that the hash value of the block meets certain conditions, that is, the formula (11), data is the transaction data in the block, H is the hash function, and target is the difficulty set by the network. Multiple nodes (miners) will calculate nonce at the same time. The node that first finds the condition will add the block to the blockchain and broadcast it to the network. This process ensures the validity and legitimacy of the transaction, prevents double payment, and ensures the security and fairness of data transactions.

(b) Smart contract execution: Use Solidity to write smart contracts, which contain logic such as transaction protocols and data access control, and can define transaction rules and conditions. When the contract conditions are met, the smart contract will be automatically executed on the blockchain^[46]. For example, when a user completes a payment, the smart contract will automatically confirm the transaction, update the data ownership, and generate the corresponding NFT. All contract execution records will be permanently stored on the blockchain, ensuring the transparency and immutability of the contract^[38], and enhancing users' trust in the platform.

(c) Non-fungible tokens (NFTs): Each data asset uploaded by a user will generate an NFT, which uniquely identifies the ownership of the data asset on the blockchain^[40]. In data transactions, NFTs are automatically transferred according to the transaction agreement. During the transaction, the smart contract will handle the transfer of NFTs and record the change of ownership to ensure the security and transparency of the transaction. Each transfer of NFTs and change of ownership will be recorded on the blockchain, making the transaction history traceable.

(iii) Role and Benefits of Integration

The integration of blockchain and NFTs serves as the trust foundation of the entire DECTP framework. In a data market fraught with challenges like ambiguous ownership, lack of transparency, and difficulty in enforcing agreements, this layer provides critical solutions.

(a) Role of Blockchain: The blockchain acts as a decentralized, immutable, and transparent ledger for all platform activities. Firstly, by removing the need for a central intermediary, it creates a trustless environment where participants can interact securely. All transaction rules and records are publicly verifiable, which significantly reduces information asymmetry and the potential for fraud^[12]. Secondly, it provides a verifiable and permanent audit trail for every data transaction, from initial upload to final sale. This is crucial for resolving disputes, ensuring regulatory compliance, and allowing buyers to trace data provenance to assess its quality and legitimacy. Thirdly, smart contracts automate the execution of transaction agreements, which eliminates manual intervention, reduces operational costs, and guarantees that transaction terms

are enforced fairly and consistently for all parties^[47].

(b) Role of NFTs: NFTs are leveraged to solve the critical problem of data ownership and usage rights verification. Each dataset or data product is tokenized as a unique NFT, transforming an intangible data asset into a manageable and transferable digital object. First, an NFT provides cryptographic proof of ownership for a data asset, which is easily verifiable on the blockchain. This directly addresses the ownership ambiguity that hinders data market growth. Furthermore, by representing data as tradable digital assets, NFTs enhance market liquidity, which also enables novel economic models^[48].

8) Federated Learning Layer

(i) Functional Overview

The federated learning layer uses a distributed learning model to achieve collaborative training and model updates of multi-party data while protecting data privacy. This layer is designed to overcome the problem of data silos and achieve the sharing and utilization of data resources.

(ii) Implementation Mechanism

(a) Federated average algorithm: Each client independently trains a local model and aggregates it on the server to generate a global model^[49]. The formula of the federated average algorithm is as follows:

$$w_{t+1} \leftarrow \sum_{i=1}^k \frac{n_i}{n} w_t^i, \quad (12)$$

where w_t^i is the model weight of the i -th customer at time t , n_i is the amount of data on the client, and n is the total amount of data.

Each client trains its model independently on local data. Each client uses its own data set to update the model weights w_t^i and calculate the corresponding local model. At the end of the training cycle, the client sends its local model weights to the server. The server does not directly access any raw data, but only obtains the model weights. It calculates the new global model weights w_{t+1} through formula (12), and aggregates the weights of each model according to the amount of data from each client to generate a new global model. The updated global model will be sent back to each client to continue the next round of local training.

(b) Secure multi-party computing (SMPC): SMPC is used to enhance privacy protection in federated learning. When participants conduct collaborative training, SMPC ensures that the parties do not need to share original data but instead complete the cooperation by exchanging encrypted intermediate calculation results^[50]. Participants encrypt their local model parameters and exchange parameters under a secure multi-party computing protocol. With this encrypted data, participants can perform necessary calculations and generate updated model parameters. After the training is completed, the participants will obtain an encrypted global model update, which can ensure data privacy while achieving model improvement. As a result, the platform can not only protect the privacy of users but also conduct effective model training while ensuring data security, thus promoting the safe sharing and use of data.

4.3 Comparative Study with Ocean Protocol Platform

Ocean Protocol, as a leading example of decentralized data element circulation trading platforms, has successfully enabled secure and transparent data transactions using technologies like blockchain and smart contracts. Its technical framework not only facilitates the market-driven circulation of data but also ensures data privacy through robust protection mechanisms. This makes it one of the most advanced practices in the field of the data economy today. The data element circulation trading platform (DECTP) proposed in this paper shares some common architectural features with Ocean Protocol, particularly in its use of blockchain and smart contracts. The comparison with Ocean Protocol is intended to assess the similarities and differences in the technical implementation and innovation of the two platforms. It also aims to highlight the specific improvements and optimizations that DECTP offers in certain areas. This comparison will demonstrate the strengths of DECTP in key aspects such as data privacy protection, pricing mechanisms, and industry-specific services, while also validating its practical applications and innovative contributions in various scenarios. The overall architecture of Ocean Protocol, as outlined in its technical white paper^[51] serves as the foundation for the comparison and is illustrated in Figure 3.

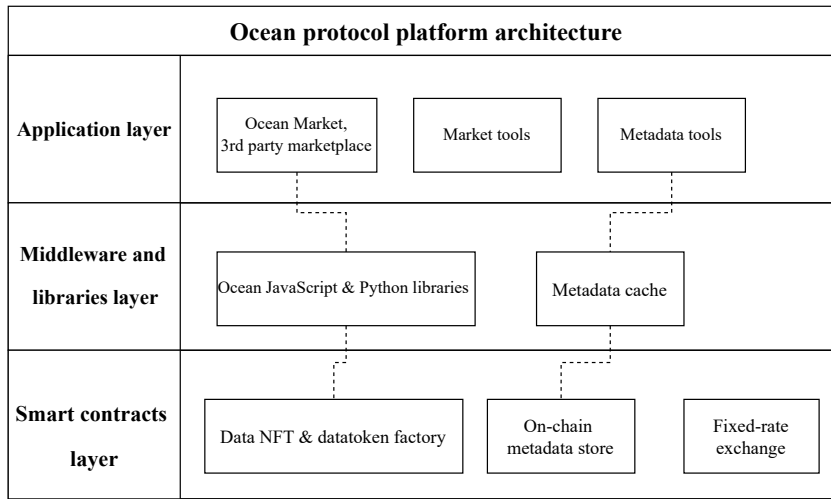


Figure 3 Overall architecture of the Ocean Protocol platform

The Ocean Protocol platform supports decentralized data circulation and transactions, and adopts a highly integrated three-layer architecture. The platform uses the Smart Contracts Layer as the underlying foundation and is deployed on EVM-compatible networks such as Ethereum. It manages data assets, NFTs generation, transactions, and access control through smart contracts. The Middleware and Libraries Layer provides JavaScript and Python libraries to simplify interaction with smart contracts and support data publishing, consumption, and metadata management. The Application Layer includes Ocean Market and 3rd Party Marketplace, as well as Market Tools and Metadata Tools, integrating functions such as data pricing, transaction matching, and privacy computing. Users use this layer to publish, discover, trade, and consume

data.

4.3.1 Multi-dimensional Comparison Between DECTP and Ocean Protocol

As shown in Table 7, the proposed DECTP framework is compared with Ocean Protocol across five dimensions: architectural hierarchy, data upload and processing, data privacy protection, data pricing mechanism, and industry-customized services. This comparison highlights the advantages of DECTP in modular scalability, automated data processing, privacy-preserving mechanisms, dynamic pricing, and industry-specific service provision.

4.3.2 Limitations of Comparison and a Framework for Future Evaluation

A comprehensive comparison would ideally include quantitative performance benchmarks. However, a direct, quantitative performance evaluation between the proposed DECTP framework and existing platforms like Ocean Protocol is not feasible at this theoretical stage. This is primarily due to the proprietary nature of operational data from commercial platforms, which are not publicly available for academic benchmarking.

To address this gap and establish a clear pathway for future empirical validation, this study proposes a multi-dimensional evaluation framework, as presented in Table 8. Should operational data become accessible, the following metrics can be utilized to conduct a rigorous comparative analysis. This framework provides a structured approach to assessing the key aspects of next-generation data trading platforms.

This framework not only guides future research but also provides a standardized methodology for stakeholders to evaluate and compare different data trading platforms in a more holistic and objective manner.

5 Suggestions for the Future Development of the Data Element Circulation Trading Platform in China

Building on the issues identified in Section 3 and the platform design principles outlined in Section 4, this section provides strategic recommendations to guide the future development of data element circulation trading platforms in China.

5.1 Enhancing Privacy Protection: Integration of Differential Privacy and Secure Multi-party Computing

To address weaknesses in current data privacy and security measures, it is essential to incorporate differential privacy into the data processing module. This approach introduces noise into data analysis processes, ensuring individual privacy even in cross-source analyses. For industries with stringent privacy requirements, such as healthcare and finance, the integration of SMPC is recommended. SMPC enables multiple parties to collaboratively analyze or train models without exposing their raw data, maintaining privacy throughout data uploading, sharing, and trading. These technologies together can establish a more robust privacy framework tailored to sector-specific demands.

Table 7 Multi-dimensional comparison between DECTP and Ocean Protocol

Comparison dimension	DECTP	Ocean Protocol
Architecture hierarchy	The multi-layer architecture includes eight layers, such as Blockchain and Smart Contracts Layer, and Federated Learning Layer. It adopts a microservice architecture and each layer can be expanded independently.	The three-layer architecture includes the Application Layer, the Middleware and Libraries Layer, and the Smart Contracts Layer. Each layer is highly integrated and the structure is relatively simple.
Data upload and processing	Built-in ETL automatically clean, deduplicate and standardize data to ensure data quality. It supports the processing of multi-source heterogeneous data and has a highly automated data conversion and standardization mechanism to improve data processing efficiency.	The platform does not provide built-in ETL data cleaning and standardization functions. Users need to pre-process the data themselves before uploading, which limits the efficiency of data management and user experience.
Data privacy protection	A high degree of privacy protection is achieved by combining differential privacy, homomorphic encryption, and secure multi-party computing. Data collaboration is carried out through the federated learning layer, allowing multiple parties to jointly train machine learning models without sharing data, ensuring data privacy.	The Ocean Protocol platform achieves privacy protection through the Compute-to-Data function, but does not provide direct support for differential privacy or homomorphic encryption, and privacy protection relies on the security of the computing environment rather than encryption or federated learning mechanisms.
Data pricing mechanism	Based on a dynamic pricing model of market supply and demand and data quality, the platform automatically adjusts data prices and takes more dimensions into consideration in pricing, making the data trading process more in line with market principles and ensuring fair and transparent prices.	Data providers set their own data prices. The platform lacks a built-in dynamic pricing model, and prices do not automatically adjust with changes in market supply and demand.
Industry customized services	DECTP supports customized services for different industries through a microservice architecture. It can quickly deploy customized functions based on specific industry standards and data requirements, and can meet the personalized needs of different industries for data circulation and processing.	Ocean Protocol serves more as a general data market platform. It does not provide customized services specifically for different industries and lacks optimized services and functions for different industries.

5.2 Clarifying Data Ownership: Leveraging Blockchain and NFTs for Ownership Confirmation

To resolve ambiguities in data ownership, the platform should adopt NFTs to uniquely identify data assets. By integrating NFTs with blockchain and smart contract systems, the platform

Table 8 Framework for comparative evaluation of data element circulation trading platforms

Metric category	Specific metric	Evaluation focus & implication
Performance metrics	Transaction	Measures the platform’s capacity to process data transactions.
	Throughput (TPS)	Higher TPS indicates better scalability for high-frequency data markets.
	Query Latency	Assesses the end-to-end time from a data request to a response. Lower latency is critical for real-time data applications.
Security and privacy metrics	Federated Learning Efficiency	Evaluates the time required for model convergence and the communication overhead during the aggregation of model updates.
	Privacy Leakage Quantification	Measures the privacy budget consumed per query, quantifying the level of privacy protection.
	Consensus Security	Evaluates the computational power (hash rate) or stake required to compromise the network, indicating resilience to 51% attacks.
Economic metrics	Transaction Success Rate	The ratio of successfully completed transactions to initiated ones, reflecting market liquidity and the efficiency of the matching algorithm.
	Price Volatility & Fairness	Analyzes the stability and responsiveness of the dynamic pricing model to market conditions, assessing its fairness and ability to reflect true data value.

can ensure transparent, traceable, and tamper-proof ownership records. For each transaction, smart contracts would automatically update NFT ownership details, minimizing disputes and enhancing accountability. Additionally, establishing a comprehensive data traceability system would ensure compliance with transaction standards while enabling buyers to evaluate data value through transparent source tracing. This dual approach strengthens trust, security, and transparency in data transactions^[34].

5.3 Expanding Industry-specific Services: Development of Customized Service Modules

To address the lack of specialized services, the platform should develop industry-specific data analysis and mining modules tailored to sectors such as healthcare, finance, and transportation. These modules would include tools designed to meet the unique data requirements and privacy constraints of each industry. For example, privacy-preserving data analysis tools could be developed for healthcare, while efficient risk analysis solutions could be offered for the financial sector. Furthermore, the platform should facilitate cross-industry data fusion by integrating data from multiple domains, enabling users to derive comprehensive insights and unlock the full potential of data.

5.4 Establishing a Dynamic Pricing Mechanism: Market-driven Pricing Models

To overcome the absence of effective pricing mechanisms, the platform should implement a dynamic pricing model that adjusts prices based on real-time supply and demand, as well as data quality metrics such as timeliness, coverage, and scarcity. Machine learning algorithms could analyze these factors to generate automated pricing adjustments, ensuring fair and transparent transactions. Integrating smart contracts into this process would enable automatic price updates in response to market fluctuations, reducing manual intervention and enhancing efficiency.

5.5 Standardizing Data Management: Development of Unified Processes and Quality Assurance

To address inconsistencies in data management, the platform should establish unified data standards and introduce data quality assessment tools within its management module. These tools would monitor data accuracy, completeness, and timeliness in real-time, ensuring that all transacted data adheres to established industry standards. Implementing standardized processes for data formatting and quality control would improve the interoperability of data across platforms and industries, fostering smoother data exchange and collaboration.

6 Conclusion

This study provides a comprehensive analysis of data element circulation trading platforms, examining their historical development, current research status, and technological foundations. By integrating advanced technologies such as blockchain, NFTs, and federated learning, the paper proposes a novel platform architecture designed to meet the evolving demands of the data market and its users. A comparative analysis with the Ocean Protocol platform highlights the scientific and innovative aspects of the proposed design. Additionally, the paper offers targeted recommendations for improving future platforms, including privacy protection measures like differential privacy and secure multi-party computation, mechanisms for confirming data ownership, customized service modules for specific industries, dynamic pricing models, and standardized data management practices.

This paper makes several theoretical contributions. Firstly, by combining cutting-edge technologies such as blockchain, NFTs, and federated learning, the study addresses the current gap in comprehensive platform design. The proposed architecture not only incorporates technical innovations but also adopts a holistic approach to platform functionality, aiming to enhance operational efficiency and user engagement. Secondly, the study delves into critical issues such as data ownership verification and privacy protection, offering practical solutions that contribute to the healthy development of the data market. These solutions bridge theoretical gaps while addressing real-world challenges. Thirdly, through an in-depth comparison with the Ocean Protocol platform, the paper demonstrates the advantages of the proposed architecture in areas such as data processing, privacy protection, and dynamic pricing, underscoring its scientific and innovative merits.

In addition to its theoretical insights, the paper carries significant practical implications

for developing data element circulation trading platforms in China. Firstly, the integration of blockchain and NFT technologies offers a transparent and reliable solution to address ambiguities in data ownership, fostering trust and accountability in transactions. Secondly, by leveraging federated learning and differential privacy, the proposed platform ensures secure collaborative data analysis while safeguarding user privacy, enhancing platform security and user confidence. Thirdly, the study introduces a market-driven pricing mechanism that dynamically adjusts prices based on data supply, demand, and quality. This innovation promotes transaction efficiency and fairness, addressing key challenges in data valuation. Finally, the development of specialized service modules tailored to sectors such as healthcare, finance, and transportation enhances data analysis capabilities and value extraction, facilitating cross-industry applications and driving broader economic benefits.

This study has several limitations that warrant further investigation. Firstly, the proposed dynamic pricing mechanisms and NFT-based ownership verification lack detailed exploration of their technical implementation and scalability. Future research should focus on developing algorithms and frameworks that can operate efficiently in high-frequency, large-scale trading environments. Secondly, the study does not delve into the efficiency of blockchain systems, particularly their computational and latency challenges, which may limit their practicality for large-scale data markets. Future work could explore optimized or hybrid blockchain models that balance decentralization with operational performance.

Acknowledgements The authors gratefully acknowledge the editor and the anonymous referee for their insightful comments and helpful suggestions that led to a marked improvement of the article.

References

- [1] Cui R, Wang Y D, Wang Y J. The impact of data element on enterprises' capital market performance: Insights from stock liquidity in China and implications for global markets. *Sustainability*, 2024, 16(9): 3585.
- [2] Yang Y, Li Y S, Liang X P. The role of data trading platforms (DTPs) in digital technology innovation: Mechanisms and evidence from China. *Journal of Policy Modeling*, 2025, 47(6): 1372–1396.
- [3] Yang S, Zhang Y L, Cui L Z, et al. A web 3.0-based trading platform for data annotation service with optimal pricing. *IEEE Transactions on Network Science and Engineering*, 2024, 11(5): 4032–4044.
- [4] Zhao Y Q. More than 80 platforms have been established across the country. What are the new trends in data transactions? (2024-06-06) [2024-07-30].
<https://tele.ofweek.com/2024-06/ART-8320505-8440-30637029.html>.
- [5] Zhang X, Cai Y T, Li Y W, et al. Impacts of information asymmetry on users' payment rates: evidence from trading data of Chinese knowledge payment platform. *Asia Pacific Journal of Marketing and Logistics*, 2024, 36(6): 1317–1336.
- [6] Ngo T T T, Dang T A, Huynh V V, et al. A systematic literature mapping on using blockchain

- technology in identity management. *IEEE Access*, 2023, 11: 26004–26032.
- [7] Wang Q, Li R J, Wang Q, et al. Non-Fungible Token (NFT): Overview, evaluation, opportunities and challenges. *arXiv preprint arXiv:2105.07447*, 2021.
 - [8] Liu W B, Chen H, Ngai E C H. BF-Meta: Secure blockchain-enhanced privacy-preserving federated learning for metaverse. *2024 IEEE International Conference on Metaverse Computing, Networking, and Applications (MetaCom)*, August 12–14, 2024. Hong Kong: IEEE, 2024, 166–172.
 - [9] Hu D H, Li Y F, Pan L X, et al. A blockchain-based trading system for big data. *Computer Networks*, 2021, 191: 107994.
 - [10] Gu B, Hu W W, Gong S M, et al. CBDTF: A distributed and trustworthy data trading framework for mobile crowdsensing. *IEEE Transactions on Vehicular Technology*, 2024, 73(3): 4207–4218.
 - [11] Gao G J, Xiao M J, Wu J, et al. DPDT: A differentially private crowd-sensed data trading mechanism. *IEEE Internet of Things Journal*, 2020, 7(1): 751–762.
 - [12] Xu G X, Zhou Z J, Dong J N, et al. A blockchain-based federated learning scheme for data sharing in industrial internet of things. *IEEE Internet of Things Journal*, 2023, 10(24): 21467–21478.
 - [13] Fernandez R C, Subramaniam P, Franklin M J. Data market platforms: Trading data assets to solve data problems. *Proceedings of the VLDB Endowment*, 2020, 13(12): 1933–1947.
 - [14] Lu Y, Wang J Y, Liu L X, et al. Get by how much you pay: A novel data pricing scheme for data trading. *Information Processing and Management*, 2024, 61(6): 103849.
 - [15] Araz O M, Choi T, Olson D L, et al. Data analytics for operational risk management. *Decision Sciences*, 2020, 51(6): 1316–1319.
 - [16] Spiekermann M. Data marketplaces: Trends and monetisation of data goods. *Intereconomics*, 2019, 54(4): 208–216.
 - [17] Jones C I, Tonetti C. Nonrivalry and the economics of data. *American Economic Review*, 2020, 110(9): 2819–2858.
 - [18] Dai W, Dai C, Choo K R, et al. SDTE: A secure blockchain-based data trading ecosystem. *IEEE Transactions on Information Forensics and Security*, 2020, 15: 725–737.
 - [19] Abbasi M, Prieto J, Shahraki A, et al. Industrial data monetization: A blockchain-based industrial IoT data trading system. *Internet of Things*, 2023, 24: 100959.
 - [20] Liu H, Tai W, Wang Y, et al. A blockchain-based spatial data trading framework. *Journal on Wireless Communications and Networking*, 2022, 2022(1): 71.
 - [21] Zhang X, Li X, Miao Y, et al. A data trading scheme with efficient data usage control for industrial IoT. *IEEE Transactions on Industrial Informatics*, 2021, 18(7): 4456–4465.
 - [22] Oh H, Park S, Lee G M, et al. Competitive data trading model with privacy valuation for multiple stakeholders in IoT data markets. *IEEE Internet of Things Journal*, 2020, 7(4): 3623–3639.
 - [23] Xiong S, Chen P, Ge S, et al. SFOM-DT: A secure and fair one-to-many data trading scheme based on blockchain. *IEEE Transactions on Information Forensics and Security*, 2024, 19: 10029–10042.
 - [24] Dosis A, Sand-Zantman W. The ownership of data. *The Journal of Law, Economics, and Organization*, 2023, 39(3): 615–641.
 - [25] Boerding A, Culik N, Doepke C, et al. Data Ownership — A property rights approach from a European perspective. *Journal of Civil Law Studies*, 2018, 11(2): 323–370.
 - [26] Zhao X Y, Ji X F, Guo Y Z. Research on the construction of a model of factors affecting the

- operation of data trading platforms under the goal of innovation incentives: A case study of Guiyang data trading platform. *Modern Intelligence*, 2023, 43(4): 101–112.
- [27] Wu J, Zhang Y. Research on the development of data element trading platform under the perspective of market-oriented allocation of factors. *Credit Reference*, 2021, 39(1): 59–66.
- [28] Wang W, Zhang M J, Wang J. Research and analysis of big data trading platforms at home and abroad. *Journal of Intelligence*, 2019, 38(2): 181–186.
- [29] Zhu H Y. The current situation and countermeasures of data trading platform development based on the perspective of data element circulation. *China Business and Market*, 2024, 38(2): 24–33.
- [30] Zhang X P. Comparative study of big data trading platforms at home and abroad. Heilongjiang: Heilongjiang University, 2020.
- [31] Zhuang J X. Analysis of three types of big data trading platform models and their advantages and disadvantages. *China Industrial Review*, 2016(10): 109–111.
- [32] Chen Z, Zheng Q, Wu Z S. Realistic dilemmas and solutions to the construction of China's data trading platform. *Reform*, 2022(2): 76–87.
- [33] Wang X M, Huang K M, Shao R J. Key technology design and practice of data trading platform supporting data element circulation. *New Industrialization*, 2023, 13(11): 40–50.
- [34] Li Y J. High-quality development of my country's data trading institutions: practical problems, theoretical logic and implementation paths. *Guangdong Social Sciences*, 2023(1): 47–56.
- [35] Moiso C, Minerva R. Towards a user-centric personal data ecosystem: The role of the bank of individuals' data. 2012 16th International Conference on Intelligence in Next Generation Networks (ICIN), October 8–11, 2012. Berlin: IEEE, 2012, 202–209.
- [36] He P Y, Wang X R. Research on the current dilemma and countermeasures of China's big data trading platform. *Modern Intelligence*, 2017, 37(8): 98–105.
- [37] Jamshidi P, Pahl C, Mendonça N C. Microservices: The journey so far and challenges ahead. *IEEE Software*, 2018, 3(35): 24–35.
- [38] Swan M. Blockchain: Blueprint for a New Economy. O'Reilly Media, Sebastopol, 2015.
- [39] Sandhu R, Coyne E, Feinstein H. Role-based access control models. *IEEE Computer*, 1996, 29(2): 38–47.
- [40] Buterin V. Ethereum: A next-generation smart contract and decentralized application platform. White Paper, 2014. [2024-07-30]. <https://caladex.io/api/files/papers/Ethereum>
- [41] Wang Y S, Kogan A. Designing confidentiality-preserving Blockchain-based transaction processing systems. *International Journal of Accounting Information Systems*, 2018, 30: 1–18.
- [42] Li J, Li J, Wang X, et al. Multi-blockchain based data trading markets with novel pricing mechanisms. *IEEE/CAA Journal of Automatica Sinica*, 2023, 10(12): 2222–2232.
- [43] Stallings W. Network and internetwork security: Principles and practice. Prentice Hall, Englewood Cliffs, 1995.
- [44] Brewer E. CAP twelve years later: How the “rules” have changed. *Computer*, 2012, 45(2): 23–29.
- [45] Nakamoto S. Bitcoin: A peer-to-peer electronic cash system. *SSRN Electronic Journal*, 2008, 3440802: 10–2139.
- [46] Luu L, Chu D H, Olickel H, et al. Making smart contracts smarter. Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, October 24–28, 2016. Vienna: ACM, 2016, 254–269.

- [47] Zhang M, Liu J, Feng K, et al. SmartAuction: A blockchain-based secure implementation of private data queries. *Future Generation Computer Systems*, 2023, 138: 198–211.
- [48] Jiang E, Qin B, Wang Q, et al. BDTS: Blockchain-based data trading system. *Information and Communications Security: 25th International Conference, ICICS 2023, November 18–20, 2023. Tianjin: Springer, 2023, 645–664.*
- [49] Giuseppi A, Manfredi S, Menegatti D, et al. Decentralised federated learning for hospital networks with application to COVID-19 detection. *IEEE Access*, 2022, 10: 92681–92691.
- [50] Bonawitz K, Ivanov V, Kreuter B, et al. Practical secure aggregation for privacy-preserving machine learning. *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, October 30–November 3, 2017. Dallas: ACM, 2017, 1175–1191.
- [51] Ocean Protocol Foundation. Ocean Protocol: Tools for the Web3 Data Economy. [2024-07-30]. <https://oceanprotocol.com/tech-whitepaper>.