



# Data Factor Market System Based on Data Component and Social Division of Labor

Yiyang Zhou

Beijing University of Posts and Telecommunications, Beijing, 100876, China

e-mail: zhouyiyang@bupt.edu.cn

**Abstract.** As a new production factor, data has become the core engine for the deepening development of the digital economy, but the current industrial scale of the data market has always failed to meet people's expectations. This paper mainly analyses the problems existing in the current data market, studies and proposes the data market system based on the construction of social division of labor. This paper mainly analyzes the problems existing in the current data market, studies and proposes the data market system based on the construction of social division of labor according to the theoretical basis of data components.

**Keywords:** Data Factor; Data Components; Social division of labor; Data market; Data Transaction.

## 1 Introduction

In today's world, as a core factor in building digital economy, data has become a brand-new factor of production and means of production. In 2022, the scale of China's digital economy reached 50.2 trillion R.M.B, accounting for 41.5% of GDP, and China's data production reached 8.1 ZB, accounting for 10.5% of the global total data production, ranking second in the world[1]. At present, a new round of technological revolution and industrial change is poised to take place, and the realization of the market-oriented allocation of data components is the key to release the value of data components. However, China's current industrial scale of the data factor market is far less than the scale of China's digital economy, and even around the data trading market are facing the dilemma of development stagnation. So, China's current data factor market and data transaction pattern have certain problems.

## 2 Problem analysis of the flow of data factor

### 2.1 The original data has congenital defects in transactions

As a brand-new factor of production, data is very different from traditional factors such as land, labor, capital, etc. It has a series of unique characteristics such as easy to copy, complex ownership, impossible to price, difficult to regulate, etc., which lead to the

inherent defects of original data compared to other factors in the transaction[2]. These characteristics lead to the inherent defects of original data compared with other factors in trading:

1. Data are replicable and non-exclusive. Unlike traditional factors of production such as land and labor, it can be replicated indefinitely without loss, multiple parties can use it simultaneously, and the marginal cost of replicating and using data is close to zero.
2. It is difficult to determine the rights and interests of data. Original data itself may have multiple data subjects involved in the process of generating it, and at the same time, the contribution of these data subjects to the value generated by the data is difficult to measure, and their relevant rights and interests cannot be clearly identified, so it is very difficult to rationally allocate the rights and interests of all parties.
3. Data is difficult to price. The value of original data is closely related to the usage scenario, and it is difficult to lock in the usage at the time of the transaction, with strong value of posteriori characteristics, thus making it difficult to determine the price at the time of the transaction.
4. Data transactions have security risks. The information contained in data is hidden, and simple data transactions may violate personal privacy, commercial secrets and national security; the risks inherent in original data transactions are incalculable.

## **2.2 Frequent data security incidents**

As part of national security, the importance and urgency of data security cannot be overstated. According to the *Q1 2023 Data Asset Breach Analysis Report* and *Q2 2023 Data Asset Breach Analysis Report* published by Threat Hunter Inc.[3], there were 987 data leakage incidents in Q1 2023, a 42% increase from Q1 2022, involving 1,204 enterprises and 38 industries including logistics, finance, and e-commerce. And the data leakage situation in the second quarter of 2023 is even more severe compared to the first quarter, with 2018 data leakage incidents and 52 industries involved. At the same time, the reasons for data leakage incidents are also varied, with internal reasons, such as the API and data is crawled by black industry attacks, employees mistakenly transmit sensitive information to the outside world, etc., and external reasons, such as Web application attacks, supply chain leaks, SMS channel leaks, malware, etc.[4]. It can be seen that data security incidents are currently showing a high incidence trend, and data security needs to be further emphasized.

## **2.3 Data trading industry is relatively small**

At present, the digital economy has become a new engine for the high-quality development of China's economy, and data has become an important production factor for building the digital economy. According to statistics, since the establishment of the first data exchange in 2014, by the end of 2022, 48 data exchanges have been set up in China, but the annual data transaction scale of all the exchanges totaled are only more than 3 billion R.M.B. compared with China's overall industrial scale of the digital

economy of 50 trillion, the industrial scale of the data factor market is too small, the business model is still in the primitive form, and the whole industry has just begun, with a great room for development.

### **3 Analysis of current mainstream data circulation models**

#### **3.1 Data black box model**

Data black box model is a traditional business model, it has established a similar "black box" intranet, all data-related computing operations are completed in a secure black box, the customer can only take away the results, not the original data. The data black box model is to protect the security of the premise as far as possible, the data collection, storage, governance, processing, the use of the whole chain of contact is reduced to a minimum, which is a typical "one-stop" service model, but the biggest disadvantage of this model is that it is unable to form a healthy industrial chain, the requirements of the practitioners are extremely high, and there are strict limitations on the number of people involved in the processing of data.

#### **3.2 Privacy computing model**

Privacy computing is a business model with high technological content, which refers to the realization of data analysis and calculation under the premise of protecting the data itself from being leaked to the outside world, to achieve the purpose of "usable and invisible" to the data, and to realize the transformation and release of the value of the data under the premise of adequately protecting privacy and security.

There are currently three literary theories in privacy computing techniques: multi-party secure computing, homomorphic encryption, and federated learning. However, either theory suffers from the following problems:

1. Privacy computing technology requires a high consumption of computing resources and a high cost of arithmetic, which is difficult for the market to afford.
2. The lack of unified standards and specifications for privacy computing technology has resulted in the inability of various privacy computing companies to work together and has made it easy to create monopolies and technical barriers at customers.
3. The industrial threshold of privacy computing is higher, the landing cost is higher, and the technical requirements for practitioners are also higher, thus encountering greater resistance in the process of universal promotion and making it difficult to form a scale.

#### **3.3 Data product model**

The data product model is a relatively new business model that allows data product developers to develop data products after licensing the original data and then list them for sale on an exchange. This service model appears on the surface to circumvent security risks in the circulation of data, but it still suffers from the following problems:

1. After data products are made, they may still contain original data, infringe on national security and personal privacy, and pose greater security risks and regulatory difficulties.
2. Data products contain property rights to multiple data resources, and ownership is unclear, making it difficult to distribute the benefits.
3. The aggregation of data, the production of data and the eventual listing and sale are highly concentrated among a few closely related companies, which is prone to the formation of new monopolies.

### **3.4 Brief summary**

To summarize, the main data circulation models in the current market all have various problems, and the scale of the data industry is always unsatisfactory, so it is necessary to adopt more advanced systems, technologies and business model design in order to push the scale of the data industry to a new level.

## **4 Design of a data factor market system for the great social division of labor based on data components**

### **4.1 Definition and value analysis of data components**

In the above, we have analyzed several existing data transaction models, all of which have drawbacks in the process of promoting the development of the data factor market on a large scale. It can be seen that the new mode of operation should not only comply with the characteristics of the data itself and be able to effectively deal with this open, complex, variable and massive supply and demand relationship, but also follow the basic laws of the market.

Data components is a data "intermediate state" circulation model proposed by Dr. Lu Zhipeng, Deputy General Manager of China Electronics Corporation and Director of CEC Data Science and Intelligent Engineering Research Institute, which is a data set formed by several related fields according to the need after data desensitization or data characteristics formed by modeling of data related fields[5].

The use of "data components" as the circulation form of data factor instead of original data for transaction can effectively realize the separation of original data and data applications, resolve the opposing contradictions between data security and data circulation, and at the same time solve the pain points of data transaction process such as right determination, valuation and pricing, and realize the circulation of marketable scale and security supervision. Therefore, data components are also the key to open the gate of data factor marketization.

#### **4.2 Linkages between the five great divisions of human society and the development of data factor markets**

Human society has experienced five major divisions of labor over the course of thousands of years, and the emergence and completion of each of these divisions of labor has had a significant impact on the social form of human civilization and the course of history[6]. The study of the construction of a data factor market system has been conducted in the past few thousand years. In the process of researching the construction of the data factor market system, we have found that the development law of the data factor market has many similarities with the formation of the five great social divisions of labor in human history. Therefore, we will use the five great divisions of labor in human society as a reference basis to explore the current market shape and future development process of data factor market.

1. The first great division of labor in human society, which was the separation of agriculture and animal husbandry, and formed a supply relationship with each other and a market.
2. The second great division of labor in human society. which was the separation of crafts and agriculture, and this separation contributed to a further increase in labor productivity; as the division of labor gradually became more refined, there was the emergence of commodity production for the purpose of direct exchange, and even the emergence of a commodity production system with multiple levels of processing.
3. The third major division of labor in human society. With the increasingly frequent exchange of commodities and the continuous expansion of the exchange area, it is necessary for some people to specialize in the exchange of commodities and to become indispensable intermediaries between commodity producers, and thus the emergence of merchants who do not engage in production but specialize in the purchase and sale of commodities, which is the third major division of labor in society. The emergence of merchants has shortened the time required for the purchase and sale of commodities, broadened the outlets for commodities, and further promoted commodity production and exchange.
4. The fourth great division of labor in human society. which was the emergence of the intellectual class, a social group that gave rise to the era of the great explosion of human civilization and the birth of countless specialists in various fields.
5. The fifth great division of labor in human society. which is the emergence of a class of professional officials engaged in the management of the State, the rule of the State, and the public governance of the State, in order to realize the unified management and governance of the State and society.

Since the beginning of human society's history, the development of material life and spiritual culture has been a process of continuous division of labor and evolution, which began with a simple and primitive division of labor and gradually developed a complex system of division of labor, and the birth of these systems has further promoted the development of human society and the economy, which is also the basis of modern industrial civilization.

As a new thing, data factor is developing from the original economic form of self-sufficiency and self-production to the form of market allocation, and the social division of labor is extremely important in this development process. The current data factor market has just formed some simple market division of labor, such as the emergence of "data merchants", "data brokers" and other roles, but due to the lack of in-depth theoretical research, and the lack of appropriate data value carrier, the market division of labor is difficult to further deepen. However, due to the lack of in-depth theoretical research and the lack of suitable data value carriers, the market division of labor is difficult to be deepened. After in-depth research, we propose a data factor-based social division of labor market system based on data components by combining the design idea of data components.

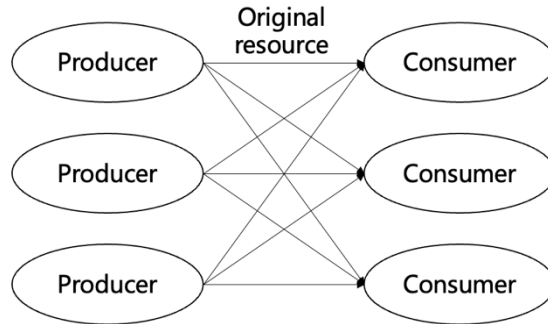
### **4.3 Design of a market system for the great social division of labor of data factor based on data components**

The five major division of labor in human society is the practice and summary of human beings for thousands of years, and also the essence of human society from primitive society to modern civilization. Therefore, we refer to the logic of the five major divisions of labor in human society and based on the core theory of "data components", we design the whole data factor market system, plan the market subject, market object, mutual relationship and operation mechanism. We will design the entire data factor market system based on the core theory of "data components", and plan the market subjects, market objects, interrelationships, and operation mechanism.

#### **1)The first great divide in the data factor market**

The first big division of labor in the data factor market is that the data generated by subject A, subject B needs, subject B produces data, subject A needs, forming a supply and demand relationship, forming a primary social division of labor. In the most primitive market form in the early period of human society, producers directly provide original resources to consumers. Similarly, in the early data factor market, data holders also directly provide their own raw data to data consumers. However, compared with traditional commodities, data is replicable and non-exclusive. At the same time, in the process of circulation transactions, there are difficulties in confirming rights, pricing, measurement, supervision, security risks and other problems[7], and it is difficult to large-scale circulation transactions, so although the division of labor has been formed, it has not formed the industrial scale.

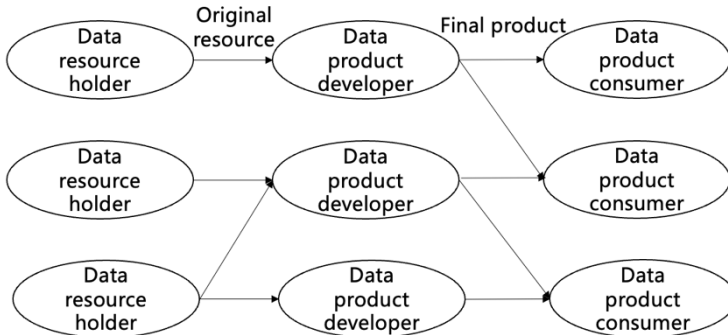
The market architecture after the first major division of the data factor market is shown in Figure 1.



**Fig. 1.** Market architecture after the first major division of the data factor market

## 2) The second great divide in the data factor market

The second great division of labor in the data factor market saw the emergence of data product developers, specializing in providing services such as data governance and data product processing for customers. Data resource holders provide and sell raw data to data product developers, who develop final data products and then sell them to data product consumers. The market architecture after the second major division of labor in the data factor market is shown in Figure 2.



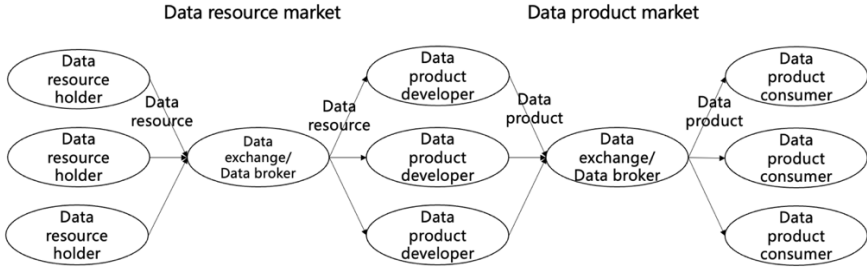
**Fig. 2.** Market architecture after the second great division of the data factor market

Since data product developers will have direct contact with raw data resources, data resource holders often require data product developers to have a certain degree of credibility, which makes it impossible for a large number of developers to participate in data product development, and ultimately creates a monopoly-style data product development service, and although a division of labor is formed, the market is still difficult to prosper.

## 3) The third great divide in the data factor market

Market players such as data exchanges and data brokers have emerged from the third major division of labor in the data factor market. The establishment of data exchanges

is intended to bring together data resource holders and data demanders for matching transactions. At the same time, after data product developers develop data products, the role of data brokers also appears to pull through the supply and demand, docking data product consumers. The market architecture after the third major division of labor in the data factor market is shown in Figure 3.



**Fig. 3.** Market architecture after the third major division of the data factor market

Since the establishment of the first data exchange in 2014, dozens of data exchanges have been set up in China, and at the same time, the role of data brokers has also appeared in the past two years in the areas where the data factor market is active, marking the completion of the third great social division of labor in China's data factor market. At the same time, we also found that, although the emergence of data exchanges and data brokers has played a certain role in promoting the activity of the data factor market, the entire industrial structure is still very simple, and the industrial scale has not taken a qualitative leap, and its core problem is still the lack of systematic design.

**4)The fourth great divide in the data factor market**

Just as the fourth great division of labor in human society has detonated human civilization, the fourth great division of labor in the data factor market will also play a crucial role in the explosion of the entire data industry. In the fourth big division of labor, there appeared the roles of intellectuals including data factor research experts, data asset evaluation experts, legal experts, education experts, etc. The emergence of these roles immediately enriched the overall data factor market system, especially the emergence of the theoretical system of data factor market allocation with data components as the core. It solves the problems of data circulation and transaction, such as difficulty in confirming rights, difficulty in pricing, difficulty in measurement, and high security risks. Meanwhile, based on data components, it designs data vault[5], data space[8] and other technical platforms, which solve the problems of data safe storage and data component processing through data vault, and solve the problems of weak, scattered and small data application scenarios through data space. To further promote the development of data resource market, data component market and data product market, the market structure of data factor market after the fourth major division of labor is shown in Figure 4.



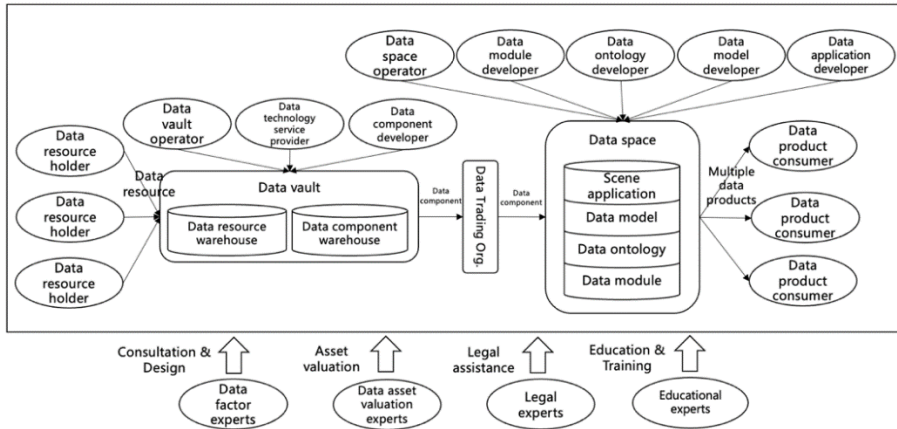


Fig. 4. Market Architecture after the Fourth Great Divide in Data Factor Markets

After the fourth major division of the data factor market, the market was further segmented and more business entities emerged.

Around the data vault, there are data resource holders, data vault operators, security operation and maintenance service providers, data component developers and other business entities.

Around the data space, there are data space operators, data component developers, data ontology developers, data model developers, data application developers, data product consumers and other business entities.

Data vaults and data spaces are connected by data exchange organizations through the transaction flow relationship of data components.

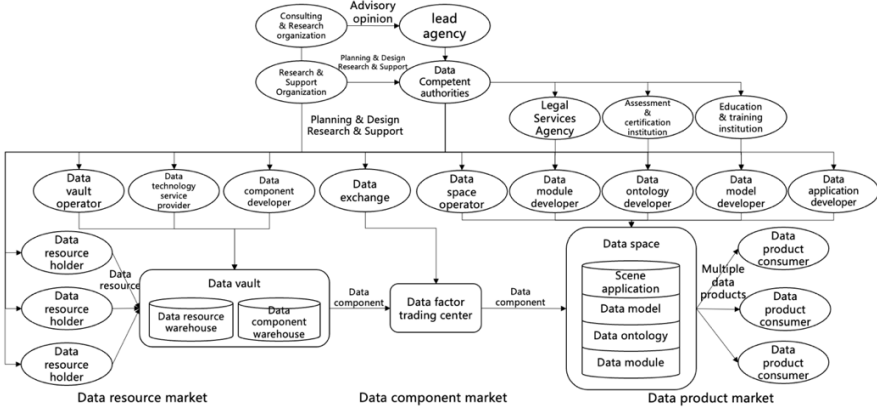
Various business carries the main body of their respective duties, specialization and division of labor, the industrial economy is beginning to take shape, the market is greatly prosperous.

### 5)The fifth great divide in data factor markets

The fifth major social division of labor was the emergence of a class of professional officials engaged in the management of the state, the rule of the state, and the public governance of the state. In the beginning stage of the data factor market, the central government arranged several ministries and commissions to manage the data industry and coordinate data security and development, but with little success.2023 On October 25, the National Data Bureau was officially inaugurated[9], marking the official debut of a professional management department representing the state to manage the data factor market.

In the data factor market before the emergence of data component technology, due to the non-standardization and value uncertainty of data, whether it is the direct transaction mode of raw data or the data circulation mode of data black box, privacy computing, data products, etc., the official class is unable to intervene in the regulation of the content of the transaction, which is also the difficulty faced by several ministries in managing the data factor market before. After the emergence of data component

technology, due to its standardization and value certainty, it provides feasibility for the regulation of the whole data circulation and trading system, and the market architecture after the fifth major division of labor in the data factor market is shown in Figure 5.



**Fig. 5.** Market architecture after the fifth major division of the data factor market

After the fifth major division of the data factor market, the three types of market, namely, data resources, data components and data products, have become more orderly under the decision-making and deployment of the coordinating and leading organizations and the unified management of the data authorities. Configuration is carried out in accordance with different security levels and regulatory requirements, and there is a reasonable division of labor among the implementing entities.

The data factor market system can be divided into four parts:

Part I: Governance subjects, including the coordinating lead agency, data authorities, etc.[10].

Part II: Market players, including data resource holders, data vault operators, data technology service providers, data component developers, data trading organizations, data space operators, data formation developers, data ontology developers, data model developers, data application developers, and data product consumers.

Part III: Supporting bodies, including consulting and research organizations, research support organizations, legal service organizations, assessment and certification organizations, education and training institutions, etc.

Part IV: Market Objects, including data resources, data components, data components, data ontologies, data models, data applications, etc.

There is a complementary relationship between governance subjects, market subjects, support subjects, and market objects, which work together to maintain the order of data factor marketization.

## 5 Market validation

Since 2021, Deyang City Government, in cooperation with China Electronics and Information Industry Group Co., LTD., has taken the lead in the pilot reform of market-

oriented allocation of data factor in prefecture-level cities in China, built a "1+4+27" data base system, and built a data factor processing and trading center, data asset registration platform, data asset evaluation platform and other technical platforms. The country's first data factor industrial park has been laid out to jointly explore the establishment of a social data factor market system based on data components[11].

On the basis of standardizing and improving the management system of the data factor market, Deyang City has set up a coordinated leading agency and data authority, established an industrial ecosystem covering data resource collection, data security storage, data component processing, data component trading, data product application, data broker, data compliance consulting, and data asset assessment, and formed three types of markets, namely, data resources, data components, and data products. By the end of 2023, Deyang City has closed the loop and landed more than 40 application scenarios such as government-enterprise-banking financial service platform, electric e-energy, marital status assessment, etc., and introduced and cultivated 65 eco-partners, including 22 data resource service providers, 2 data component developers, 29 data product developers, 2 data compliance consulting organizations, 3 data asset assessment organizations and 7 data brokerage companies, has developed more than 2,100 data components online, completed 50 million data component transactions, and driven billions of data product transactions. In 2023 alone, leaders of central ministries and commissions, leaders of provinces and cities, and experts visited Deyang 126 times for investigation and research, which caused great repercussions in the field of data factor.

The case of Deyang City's data factor market allocation reform has also been selected as one of the "2022 Digital Government Innovation Achievements and Practice Cases" by the China Information Society and the "Data Asset Evaluation Standard Pilot Test and Validation" by the Information Technology Research Center of the China Electronic Technology Standardization Institute (CETS). It was also selected as the "2023 Excellent Practice Case" by the China Institute of Electronics, the "Pilot Case of Data Factor Market Allocation Reform" by the Sichuan Provincial Data Factor Market Allocation Reform Pilot, and the "Best Example City" by the Data Factor Market Allocation Reform Comprehensive Research Institute.

## 6 Conclusions

In *The Wealth of Nations*, Adam Smith argued that "With the division of labor, the same number of laborers can do much more work than in the past, for three reasons: firstly, the skill of the laborer is improved by professional specialization; Secondly, the division of Labor eliminates the loss of time usually involved in the transfer from one type of work to another; Thirdly, the invention of many machines that simplify and reduce labor has enabled one person to do the work of many [12]".

At present, the division of labor in China's data factor market has not yet been fully formed. This is one of the fundamental reasons why the industrial scale of China's data factor market has never been able to meet expectations. Therefore, China's data factor market must complete the fifth division of labor as soon as possible in order to inject a

new and powerful impetus into the development of China's digital economy and the construction of a digital power.

We believe that after the formation of the five major divisions of labor in the data factor market, China's digital economy will flourish under the market system of the social division of labor based on data components, and the scale of China's data factor industry will surely be able to reach the same level of scale as that of the traditional four factors of production[13], and will eventually lead a new round of digital wave and industrial revolution in the world.

## References

1. National Internet Information Office of the People's Republic of China. Digital China Development Report (2022) [EB/OL]. [2023-05-23].
2. Threat Hunter Official Website [EB/OL]. <https://www.threathunter.cn>.
3. Chen Hongmin, Xiong Honglin, Xu Li, et al. Analysis of data transaction model and characteristics based on platform perspective[J]. *Big Data*,2023,9(2):56-66.
4. Guo R. Trend analysis of data leakage risk[J]. *Information Security and Technology*,2014,5(10):18-21.
5. Lu Zhipeng. Innovative Data Governance Path to Activate the Potential of Data Elements[J]. *Economy*, 2021, (06): 117-119.
6. Xie Changan, Cheng Enfu. The deepening of the division of labor: An analysis of the five great social divisions of labor and the intra-sectoral division of labor[J]. *Marxist Studies*, 2016, (12): 46-58+157.
7. Guoli, Zhou Chongyi, Yang Erlong, Hu Chengsheng. Dilemma analysis of raw data trading
8. Han Guang, Zhou Chongyi, Hao Donglin, Liu Wei. Building a Secure and Trustworthy Financial Data Space Based on Data Components and Data Vaults[J]. *Network Security and Data Governance*, 2023, 42 (09): 1-5.
9. China.gov.cn. NDA inaugurated [EB/OL]. [2023-10-25].
10. Lu Zhipeng, Meng Qingguo, Wang Yue. Data Factor Governance [M]. Tsinghua University Press, 2024: 140-142.
11. Wang Jin, He Zirui. Deyang focuses on "five problems" to empower the development of digital economy[N]. *China Reform News*, 2023-11-20(004).
12. Smith Adam. *An Inquiry into the Nature and Causes of the Wealth of Nations*[M]. University of Chicago Press: 2008-07-18.
13. Rong Ke, Lu Zhipeng. *The Theory of Data* [M]. People's Publishing House, 2022: 007-013.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

