Research on Problems and Countermeasures in the Process of Data Capitalization of China

Tongyan Shang and Haijun Zhao[✉]

School of Information, Guangdong University of Finance and Economics, Guangzhou, China
zhj88888@126.com

Abstract. Based on the clarification and general process of data capitalization, this paper discusses the infrastructure construction of data capitalization, the principle of determining the attribution of data asset rights, the dynamic value assessment in the process of data asset redistribution, the practice and supervision of data security and data capitalization in data asset circulation, which are the core problems that need to be solved urgently in the process of data capitalization in China. The purpose of this paper is to call on the whole society to pay attention to the cause of data capitalization, give strong policy support at the national level and promote the orderly development of data capitalization.

Keywords: Data Capitalization · Data Circulation · Data Rights Confirmation · Data Asset Evaluation · Coordination and Supervision

1 Introduction

With the development of the Internet and the advent of digital economic ages, the amount of global data presents an explosive growth trend. According to *the 47th China Statistical Report on Internet Development* released by *China Internet Network Information Center (CNNIC)* in February 2021, by the end of 2020, the number of Internet users in China has reached 989 million, with a penetration rate of 70.4% [2]. The sudden outbreak of the *COVID-19* epidemic in early 2020 had a huge impact in our daily life. For example, the home segregated people can only go shopping online, send instant messages, work and study from home because of the needs of epidemic prevention and control. Even the elderly group began to use smart phones to meet the need of showing the health codes and activity tracks. The popularity of mobile Internet makes it easy for collecting a lot of data.

In 2019, *the fourth Plenary Session of the 19th C.P.C Central Committee* expressed "data" as one of the production factors which is as important as labor, capital, land, knowledge, technology and management for the first time. Various industries also tended to explore risk management and business model innovation based on the massive data generated anytime and anywhere. In such an objective environment, the huge value of big data has been gradually excavated. However, data assets cannot be regarded as traditional assets because of their characteristics such as intangibility, non-exclusiveness, replicability and scale economy. Therefore, it is necessary to establish "*data asset*" subject separately as well as carry out the right confirmation and evaluation of data.

2 The Technical Path of Transforming Data Resources to Data Assets

In the current academic research, many scholars point that the data should be regarded as one of the assets because making it expensed will weaken the decision-making role of big data, distort the ratio of income and expenses as well as affect the quality of financial statements [14]. In addition, some scholars suggest that data should be classified as intangible assets for initial and follow-up measurement due to the ownership, intangibility and economic profitability of data [5]. The author believes that not all data resources can be identified as assets, and the traditional assets accounting system can not be applied to data assets directly. Instead, we should explore a new confirmation and evaluation method that adapted to the data assets. In order to promote the capitalization process, we should clarify the characteristics of data assets for distinguish them from data resources.

Data assets have the characteristics of easy circulation and replicability, diversified sources and utility, measurable cost, dynamic changes of data value and so on. In order to build the process of data assets, we need to first clarify the specific meaning of the characteristics of data assets, and then set up a separate "*data asset*" project to establish a complete closed-loop process and finally achieve the process of data asset management from processing to internal use, circulation, transaction and final disposal.

2.1 The Core Tasks and Key Problems of Data Capitalization

Data resources refer to the resources formed by information. As the original form of data, data resources are lack of standards and clear ownership to be traded or used directly. To harness the value of them, we can transform date resources into data assets after a further process. Data assets refer to the data that meets the conditions of asset recognition, that is to say, the data is owned or controlled by the enterprise, and it is expected that economic benefits will flow into the enterprise as well as the cost of data assets can be measured.

The most prominent similarity between data assets and intangible assets is intangible, while the difference is that data assets are easy to store, circulate and copy. The data stored in the memory can not only be wrote, read and modified at any time, but also can be copied and propagated infinitely without changing the original data forms, that is, the data assets are non-exclusive [13]. The easy circulation of data assets provides the feasibility for the free flow of data assets among departments, enterprises and regions.

As to the recognition conditions of assets, it has been confirmed by many practices that data assets can bring economic benefits into enterprises and play an important role in supporting business development, improving products and services, helping enterprises to find some new business opportunities and expand the customer groups [14]. Generally speaking, the economic values flow into enterprises in the form of cost saving or income increasing. However, the capitalization of data still needs to meet one of the basic accounting assumptions—monetary measurement. Obviously, monetary measurement helps us to aggregate and compare the value of assets directly. Moreover, the value of data will be reflected in the financial statements, which can improve the defect that it is

difficult to compare and understand the total amount caused by the physical measurement of data resources. To value the existing data assets comprehensively, we should consider the data assets from the perspective of value measurement.

Generally, the measurement of assets in the balance sheet can be divided into initial measurement, subsequent measurement and final disposal measurement. For example, the initial value of intangible assets should be determined by the way in which assets are acquired, such as self-development, outsourcing, capital contributed by owner and so on. At the subsequent period, they should be amortization year by year. It is necessary to recognize intangible assets depreciation reserves when the signs of impairment appear, and terminate the book value of them which are scrapped or disposed. If the service life of the intangible asset is uncertain, it will not be amortized while the impairment test will be conducted at the end of each year. As for the trading securities, the initial value and subsequent value will change with the fair value. However, the two typical traditional asset evaluation methods can not be fully applied to data assets. Although the existing asset evaluation methods have considered the impact of subsequent changes on asset value to a certain extent, the time point of evaluation is very dependent on the setting of accounting cycle, which is not in line with the ever-changing value of data assets.

The value of data assets is also affected by many factors so that it is urgent to explore the dynamic value evaluation model of data assets.

2.2 Technical Path of Data Capitalization

The transformation of data from resources to assets goes through a series of technical and economic processes. First of all, the enterprises should distinguish the data types, that is, whether the data are independent intellectual property rights information, associated intellectual property rights information or non-proprietary rights information. Among them, the non-proprietary rights information does not belong to the enterprises and should not be further processed. Second, the enterprises should collect data, verify data property rights and identify data guidelines manually or mechanically according to the degree of data structure. Third, the enterprises should preset data attributes and status guidance tables in the data attribute and status system, such as data integrity, data confidentiality, etc., and then input the collected data into the system to match the data attributes and status automatically. The fourth step are data cleaning, data desensitization and data clustering. The enterprises should rename the collected data, delete duplicates, add or delete missing values, handle or delete outliers, as well as sort the data. These processing steps will improve the value density of massive data effectively. The fifth step are data confirmation and evaluation. The enterprises should use encryption technique, access control and audit supervision to protect the data security, and then apply *Metcalfe's law* to evaluate the data assets dynamically. Besides, the ownership and value of data assets should also be audited by the third-party audit institutions. The sixth step is rights registration of data assets. It is the key premise to clarify both the ownership and rights before the circulation or transaction of data assets, so we should establish a legal registration institution to carry out data assets rights registration. The seventh step is that the data assets can be traded on the big data property rights exchange, or they can be directly traded by the data asset supplier and demander. Finally, when the two parties reach an agreement on the transaction, the registration and settlement work should be

Fig. 1. General process of data capitalization

carried out through the legal registration institution and the legal settlement institution. After the transaction, the status of data assets should still be reported to the regulatory authority regularly. The above steps can be described as the general process of data capitalization as shown in Fig. 1.

3   Infrastructure Construction of Data Capitalization

From the technical path of data capitalization, we can see that data capitalization needs the support of a series of new infrastructure constructions.

Firstly, we should build the construction of data property right indexing and right confirmation technology facilities, which are a pair of two-way and reciprocal technical work. The former is to establish a set of scientific and complete data property right indexing technology system, whose core task is to index the data resource entities in the dimension of property rights. The latter is the reverse process of the former, that is, the traceability of data property right indexing or the ownership. We have carried out some discussions [16], and comprehensively applied data citation technology, data tracing technology, reversible hiding technology, electronic forensics technology and blockchain technology to the rights confirmation. Many institutions have carried out some pioneering explorations, such as *the Certificate Cloud, Cloud Visualization, Yibaoquan.* However, the current situation is that each organization works in its own way, which inevitably leads to the confusion of ownership indexing and rights confirmation.

The second task is to building a construction of data assets trading platform. Since the promulgation of *the State Council* had been published, many data asset trading centers have been built throughout the country, such as *Global Big Data Exchange, Zhongguancun Big Data Industry Alliance, Beijing International Big Data Property Rights Exchange* and so on. Generally, the data markets dominated by enterprises belong to the third-party data trading market, which only acts as transaction intermediaries in the trading. While most of the data trading markets that led by the government are comprehensive data trading markets, which provide pure data products, data analysis technical services, data related decision-making programs, data trading intermediary services and other functions. However, there are still many problems on the operation quality of the

market, such as the lack of good trust mechanisms and benign interactive data asset trading systems. In a word, the data trading markets are still in an immature stage. We affirmed that the current data asset trading systems are far from the goal of "*promoting the circulation of data resources, establishing and improving the data resource trading and pricing mechanism, and standardizing the trading behavior*" proposed by *the State Council* [14].

The third task is to building the property right registration system of data assets. Due to the independence of data property trading institutions, there are lack of the industrial mechanism, the corresponding legal support and the scientific right confirmation mechanism which is necessary for the operation of big data trading platform. It is urgent to establish a national data asset rights registration institution on the basis of unified confirmation technical standards and rules.

4   Dynamic Confirmation of Data Assets

4.1   Principles for Determining the Ownership of Data Assets

In terms of the property rights integrity of data assets, it can be divided into independent intellectual property rights information, associated intellectual property rights information and non-proprietary rights information. In particular, non-proprietary rights information refers to the information that the enterprise does not have the ownership or control rights at all. It can not be regarded as the data assets of the enterprise, because of its legal risks. Associated intellectual property rights information is a kind of quasi-public goods. The property rights of data assets generated by independent intellectual property rights information belong to the enterprises according to the current legal system [8]. In practice, it is difficult to determine the value of data if data are regarded as the common property of the original data owners and the data gatherers [11]. Therefore, most data should belong to the enterprise after removing the privacy information.

The above data ownership is about the confirmation of static data, but the reality is that only in the rapid flow and use can data play its maximum utility value. At present, there is no such a technical specification to solve the problem in China. *The white paper on China's digital economy (2020)* also points out that there are still many gaps in "*marketization of data elements and design of circulation mechanism*". In *the National Social Science Fund Project of "Research On Data Assets Right Confirmation Under Big Data Environment (project number: 16BTQ075)"*, we have constructed data right confirmation classifier as well as created "*data right confirmation smart contract SC1 based on quality and quantity*" and "*data right confirmation smart contract SC2 based on property value and its attribution allocation*" [16], for the purpose of exploring the technology and method of data dynamic right confirmation in the context of data transaction and circulation, and giving its business process and working principle. Our conclusion is that the dynamic indexing of data responsibility based on data quality and quantity is the technical basis of dynamic data right confirmation.

4.2   Evaluation in the Process of Data Assets Redistribution

Data assets evaluation is the basis of data assets listing, circulation and transaction, and it is a quantitative index to measure the commercial value of data assets. However, the

traditional commodity pricing mechanism and the current academic methods [9] can not meet the requirements of real-time dynamic changing data asset value. There is no doubt that the circulation and transaction of data will inevitably form the redistribution of data property rights interests, and the value of data assets will also change with the data type, transaction subject, the service life, data quality and business model of data application. For different users and application scenarios, the same quality and quantity of data will produce different utility values. Because of the rapid changing, we need to identify the original data owner, user and processor by the technologies of automatic property indexing, machine learning and intelligent contract [3]. Moreover, the potential market capacity (maximum customer volume) of the same data product also needs scientific and accurate measurement. According to *the Metcalfe's law*, the network value is directly proportional to the square of the number of users linked with the network. Because data is often accessed to the customers through the network, *the Metcalfe's law* is also applicable to the evaluation of data asset spillover value.

Assumed that there is V as the value of data assets, n as the actual or permitted number of users of the data, k as the data quality coefficient and C as the production cost of data. The value of data assets is equal to the sum of the production cost of data and its external spillover value. Then, the calculation model of data assets value can be expressed as follows:

$$V = \left(1 + kn^2\right)C \tag{1}$$

In this calculation model, the data quality coefficient k can be calculated according to the six dimensions of standardization, integrity, timeliness, accuracy, consistency and accessibility proposed in the national standard *GB/T 36344-2018 "information technology: data quality evaluation indicators"*. However, the measurement indicators in this standard are extremely complicated, and it is difficult to obtain the eigenvalues of some indicators. Therefore, *the Delphi method* can be flexibly used to measure them. Specifically, on the basis of fully understanding the national standard, we can extract the connotation characteristics of each indicator in the national data quality standard, and put forward the absolute and relative quantitative indicators of data quality indicators in combination with the actual situation [10]. Moreover, the relative quantitative indicators should be recognized by the management and experts of data processing business.

It makes sense to discuss the number of users. The actual number of data users n is also the number of licensed or legal users of data. We define N as the number of potential market users *max(n)* which is equal to the critical capacity of data consumers in the whole society. In reality, N and n change with the market conditions. Then the property value of data assets based on the number of potential users in the whole society can be expressed as $max(V) = \left(1 + kN^2\right) * C$. In the specific process of data assets transaction, the theoretical transaction price should be the property value of data assets multiplied by the ratio of the number of licensed users and the number of potential users in the whole society, that is $P = \left(1 + kN^2\right)C * \frac{n}{N}$, namely $P = C * \frac{n}{N} + kCNn$. The number of licensed users can be controlled by both sides of the data trading, but the estimation of the critical capacity of data consumers in the whole society is a professional work.

5 Data Security Issues in the Circulation of Data Assets

In the process of data capitalization, five level data tags can be established according to the degree of data confidentiality. As the data can be divided into public data, semi-public data, secret data, confidential data and top-secret data, the data assets formed by different types of data should also adopt various ownership managements, security managements, evaluation methods and transaction modes. Apart from encryption technology, there are lots of methods to protect data security, such as access control, data integrity and system availability management, as well as auditing in the process of data capitalization [7]. We need to rely on the data rights management information system mentioned above to dynamically update the data rights information in real time, effectively control the data quality and safety responsibility, and prevent the malicious tampering of data ownership information under the big data circulation environment. In practical work, the enterprises should grant different degrees of access rights, operation rights and authorization rights to data asset processors, adopt the principle of separation of incompatible responsibilities, check the data assets regularly or irregularly, supervise the internal control of data assets, and establish a strict data security management system. In this regard, we can also establish and improve the data quality audit system according to *GB/T 36344-2018 "information technology: data quality evaluation indicators"*, and realize the error early warning control of ownership information and personal information according to the check relationship of data quality indicators.

Another data security issue is the protection of sensitive information. Although there is still a divergence on the ownership of personal information, scholars at home and abroad have basically reached a consensus on the point that "*it is necessary to hide the sensitive information before submitting them to the data trading platform*". At present, with the trade secret and national security information have been protected by relevant regulations, China's "*personal information protection law*" is still under consideration due to the weak sovereignty awareness for privacy protection [4]. In fact, the users have no choice but to accept the terms, although the enterprises claim that they only obtain user's privacy information after their agreement [15]. The author believes that China should enact the "*personal information protection law*" as soon as possible to clarify the rights of ownership, disposal and income of data assets, so as to provide a legal basis for the security protection of personal information.

6 Supervision of Data Capitalization

From the perspective of the whole life cycle, data assets audit should be carried out from formation, storage to the disposal. The premise of data asset management is to make clear the type, quantity, value, producer and manager and storage method of them [6]. Some scholars [12] used the data asset framework to audit the data assets of British institutions and provided solutions for data asset management and audit in China from the perspectives of institutional data assets inventory, storage location, creator and manager identity. Taking the third-party shopping platform as an example, after collecting the data of browsing, visiting, adding to shopping carts and placing orders, the user portraits are formed as well as the intelligent recommendation technologies are used so as to

improve the trading volume. This is the role transformation of the third-party shopping platform from "*credit intermediary"* to "*information intermediary"* [4]. The behavior of data collectors has a close impact on the interests of the public. Therefore, it is necessary to audit on data capitalization with the help of neutral third-party audit institutions.

In the big data environment, it is inevitable for data assets to enter the accounting system. There are issues that need to be solved such as the timely accounting information disclosure of data assets, the determination of life span and the amortization methods. Nowadays, the development of digital economy is in full swing, and the trend of data capitalization is fierce. However, the current superstructure can not fully adapt to the rapid development of digital productivity. It is urgent for the system reform of asset accounting, audit, information disclosure, mortgage financing and securitization.

To ensure the smooth progress of data capitalization, a suitable top-level design is also needed in the supervision and management system. As we all know that *Ministry of Industry and Information Technology* is responsible for the processing of big data, *Ministry of Finance* is responsible for the accounting of data capitalization, *State Administration for Market Regulation* is responsible for the data assets standardization, circulation and transaction, *State-owned Assets Supervision and Administration Commission of the State Council* is responsible for the data capitalization of state-owned enterprises, and the business management of data asset securitization is under the jurisdiction of *China Securities Regulatory Commission*. Therefore, the top-level design of data asset supervision system in China should be the specialized agency which is composed of the departments above under the coordination of *the State Council*.

7   Conclusion

To sum up, we can get the following understanding. Firstly, the core task of data capitalization is to promote the transformation from data resource to assets technically and institutionally, and the key obstacle is the dynamic evaluation of data assets. Secondly, the technical guarantee of data capitalization are the constructions of data property rights indexing technology, data asset trading platform and registration system. Thirdly, the spillover value of data assets can be measured by *the Metcalfe's law* and be limited to the critical capacity of data consumers in the whole society. Fourthly, as for the data assets security, we should not only consider the protection of personal privacy, business secret, national and other sensitive information, but also prevent the malicious tampering of data ownership information. Finally, data capitalization should be reflected in the entry of data assets into the accounting system, which needs the reform and guarantee of relevant policies, regulations and management systems.

References

1. Yu B, Zhao H (2019) Research on the construction of big dada trading platform in China. In ACM international conference proceeding series, 4th international conference on intelligent information technology, ICIIT 2019
2. China Internet Network Information Center (2021) The 47th statistical report on Internet development in China [EB/OL], 3 February 2021. http://www.cnnic.net.cn/hlwfzyj/hlwxzbg/hlwtjbg/202102/P020210203334633480104.pdf
3. Zhao H, Zhao B, Cheng S (2019) The mechanism of confirming big data property rights based on smart contract. In ACM international conference proceeding series, 4th international conference on intelligent information technology, ICIIT 2019
4. Han H, Yuan L, Li X, Qu X, Sun S (2019) Study on data capitalization in the digital economy. Credit Ref 37(04):72–78
5. Li J (2020) Research on recording data assets in national accounts. Stat Res 37(11):3–14
6. Lyon L (2016) Dealing with data: roles, rights, responsibilities and re_x0002_lationships [EB/OL], 11 January 2016. http://www.Ukoln.ac.uk/ukoln/staff/e.j.lyon/reports/dealing_with_data_report-final.pdf
7. Sheng X, Guo D (2020) Research on data security governance in open sharing of scientific data. Library Inf Serv 64(22):25–36
8. Tang L, Li X (2017) Research on data asset accounting. Chin Certified Public Accountant 02:87–89
9. Wang Y, Zhao H (2020) Data asset value assessment literature review and prospect. ICETIS 2020 J Phys Conf Ser (03):032133
10. Wang Y (2021) Research on data asset valuation under big data environment. Guangdong University of Finance and Economics
11. Wang Y, Gao F (2016) Research on property attribute of big data. Libr Inf 2016(01):29–35+43
12. Wei J, Wei H (2016) Survey and analysis of institute data asset audit from "Data Asset Framework" perspective. Libr Inf Serv 60(08):59–67+92
13. Wu Y, Jiang N, Qin Z, Lin H, Liu S, Li M (2021) Research group of Guangxi Zhuang autonomous region taxation Bureau of the State Administration of taxation. Research on the "right return" and element path of government enterprise related data–taking blockchain technology to realize the capitalization and circulation of tax data of small and medium-sized enterprises as an example. Taxation Res 2021(01):59–66
14. Yu Y (2020) Confirming big data assets to boost new economic development. Finance Account Monthly 23:52–55
15. Yuan M (2015) The debate on "the right to be forgotten": the boundary between digital memory and privacy in the era of big data. Academia Bimestrie 04:55–61
16. Zhao H (2017) Probe into the method of confirming information property right in the large data environment. J Libr Inf Sci 2(09):40–47