Research on the Definition and Construction of Data Ownership

Xinlu Wang

Dalian Ocean University, Dalian, Liaoning, China

Abstract: Accurately defining data ownership is an important premise for data circulation and sharing, this paper analyzes the existing legislation and judicial status by combing the theoretical views of data ownership with typical cases, to further clarify the current problems of data ownership in practice, which is mainly due to the complexity of the data elements and the difficulty of defining uniform standards. According to the top-level design of the "three rights" of the data twenty articles, data can be divided into personal data, enterprise data and public data according to different subjects, in different stages of the flow of the corresponding subject to the data resources to hold the right, data processing right to use, the right to operate data products, in the three rights on the basis of the construction of the right to build a hierarchical classification, with a view to realizing the flow and utilization of data, and guarantee the data ownership problems. The flow and utilization of data, to ensure the security and value realization of data.

Keywords: data ownership; data types; data rights; data flows

1. Introduction

Data is the core of the digital economy, and various emerging business models directly or indirectly rely on data to realize [3]. With the development of the data industry, there is an increasing need for a sound legal system to protect it. The protection of data rights and interests of academics currently exists two ideas: one is behavioral regulation: defining the boundaries of the behavior of others outside the data enterprise, such as improper access to and use of data to stop the behavior, through the anti-unfair competition and other laws to regulate; the second is the empowerment of the data: for example, the data property right, the data property right, and so on. In the field of private law, these two ideas have similarities and differences, as the behavioral boundaries also derive from a clear definition of rights, they are two sides of the same coin [6]. Clarifying data ownership is the starting point for building a legal system, an important prerequisite for ensuring data security, and also provides theoretical support for the trial of judicial practice related cases.

2. Literature Review

At present, there is a large amount of research and discussion on the connotation and system of data ownership in academic circles, resulting in various theories. Firstly, regarding data empowerment, some scholars believe that data itself does not have the attribute of an object, and they advocate the use of relevant legal models for protection, such as competition law, tort liability law, etc.; Other scholars believe that although data needs to rely on a certain carrier, its content and form are independent and have object attributes, and should be empowered for protection.

The existing research on data rights is controversial, and scholars have continuously proposed the theories of bundle of rights, block of rights and tree of rights to understand the structure of data rights and clarify the legal status of data property rights [15]. Those scholars who oppose the establishment of rights point out that the establishment of rights will hinder the use and sharing of data, and believe that the protection mode of behaviorism should be adopted, or the protection should be carried out by the attribution of responsibility instead of the attribution of rights. However, some scholars have responded to the issue of data rights, and have responded to the current opposition to rights, so the author recognizes the necessity of rights, on the basis of data ownership issues to sort out and put forward personal views.
3. Theoretical inquiry: typological generalization of data ownership

3.1 Public data and non-public data

Open data, as the name suggests, can be accessed by anyone through certain technical means. Based on the value orientation of openness and sharing, the public enjoys the right of free access to public data, but it does not mean that the public data can be freely copied and used, which involves the value judgment of the boundary of the right to data property. At the same time, if data subjects want to exclude others from accessing and using public data, they need to have justifiable reasons, otherwise they will be recognized as unfair competition.

Non-public data refers to data that have not been disclosed to the public, and non-public data are correlated with trade secrets. Non-public data can get better protection, but not conducive to data circulation and value matters. In addition, according to the degree of sensitivity of the data is divided into general data and sensitive data, the relevant laws on sensitive data and provisions and enumeration.

3.2 Personal data, corporate data and public data

According to the categorization of the rights of data right subjects, some scholars have proposed that the existing research has initially constructed a "three-line data right" framework based on data sovereignty, with the right to enterprise data and the right to personal information as the core [13]: the right to personal information extends the scope of the adjustment to the personal data, but does not include the data related to individuals, and does not have the right to data collected by enterprises; the right to enterprise data is a limited exclusivity right; the right to enterprise data is a limited exclusivity right. The right to personal information extends to personal data, but not to data relating to individuals, and does not extend to data collected by enterprises; the right to data of enterprises is a right of limited exclusivity. This "three-line right to data" framework is inherently consistent with the categorization of personal data, corporate data and public data.

Article 20 of the Data Act divides data into personal data, corporate data and public data based on the subject of the data, and this trichotomy is generally recognized and is important for structuring data property rights. According to Article 4 of the Personal Information Protection Law, the key to personal information is "identifiable" natural person-related information, and Article 1034 of the Civil Code has increased the scope of personal information accordingly. The provisions of the EU General Data Protection Regulation are also closely related to the "identifiability" of personal information. Thus, the standard for defining personal information is identifiability, and information that is not identifiable will be defined as non-personal information. When personal information is transformed into personal data, personal data contains both personality and property interests. Personal data is the information content of the data, as opposed to non-personal data, which focuses on the legal prior rights of the data source.

Enterprise data and public data belong to another category, which is based on the subject and distribution of data property rights, emphasizing the control of relevant subjects over a certain range of data. Enterprise data are data sets collected and processed in the course of engaging in commercial activities, and public data are data sets generated in the course of performing their duties or providing services by party and government organizations at all levels and enterprises and institutions [16]. Enterprise data and public data contain both personal data and non-personal data. The use of enterprise data for personal data should be desensitized, and the use of enterprise data should not harm the public interest. In the case of public data, which is more related to data security and national data sovereignty, public data can naturally be collected on the basis of legal authority for both personal data and corporate data.

The significance of the distinction between personal data, corporate data and public data lies in the fact that the legal purposes of the three are different, thereby providing different levels of protection. For personal data, the purpose is to protect the rights and interests of individuals from being jeopardized, while the purpose of protecting enterprise data and public data is to unleash the economic value of data and promote the development of industries. The typology of data is a prerequisite for the construction of data property rights, so the data trichotomy is of great significance.

3.3 Data sources and data processors

According to the data production theory proposed by scholars, the data production process is divided into data production and data analysis, in which data production is subdivided into the production of raw
data and the production of data sets \cite{4}. The relevant stakeholders are data producers, data holders and data utilizers.

Some scholars have proposed a scheme of separation of rights based on the right to hold, use and operate data, which proposes a dual interest structure distinguishing between the data source and the data processor \cite{12}. The data source generally enjoys the property rights and interests in the data and data products as a whole, which is mainly related to the protection of property interests, while the data processor only enjoys its prior rights and rights in the data provided by the data processor, which is related to the protection of personality interests. The distinction between data sources and data processors is useful not only for distinguishing data sources and data products, but also for distinguishing their interests and the scope of their rights.

Data sources are divided into natural persons and non-natural persons: for natural persons, it is mainly personal information, which is subject to legal obligations and statutory obligations such as consent, deletion and carrying in order to safeguard the personal personality rights and interests; for non-natural persons, it is mainly through contractual claims such as copyrights, trade secrets and format terms to safeguard their rights and interests, which mainly involves the issue of contractual fairness, and in principle, it should be explicitly stipulated by contractual interpretations that safeguard the rights and interests of non-natural persons to query and copy, but should not claim strong personality rights and interests to avoid infringement of the rights and interests of data processors. Natural person query, copy the rights and interests, but shall not claim the rights and interests of strong personality color, to avoid the infringement of the rights and interests of the data processor.

Data processors need to respect the statutory prior rights and interests of the data source, and at the same time comply with the principle of fair use and the rules of open use, exercising the right of exclusion and domination over the data within the statutory scope. The data processor's property rights in data have certain specificity, and with reference to the ownership power system, in addition to enjoying the limited right of exclusion, right of use, right of profit and right of disposal, the data processor's property rights in data should also have the negative power at the same time.

3.4 Classification under the separation of powers

Scholar Shen Satellite proposed the establishment of a hierarchical rights mechanism centered on the right to use and benefit from data, with the right to hold data resources, the right to use data processing, and the right to operate data products \cite{10}. The logical starting point for understanding the separation of the three rights lies in the ownership of data and the right to use and benefit from data (the concept adopted in Article 20 of the Data Law is the right to use data, but in reality, it also includes the right to benefit from it, and the use and benefit are the core of the right to use and benefit from data). The way to confirm the rights of the three rights to data is: firstly, in the data collection stage, after the authorization of the data source, to achieve the separation of data ownership and the enterprise's right to data utility, the enterprise obtains the right to hold the data resources; the data is assembled and processed, the enterprise obtains the right to process and use the data based on the right to data utility it enjoys, and the original data is processed to become a data product or derivative data, so that the enterprise has the data product management rights to its products. The original data is processed into data products or derivative data, so that the enterprise enjoys the right to operate data products.

The right to hold data resources is to distinguish it from data ownership, which belongs to the data source, and the data processor enjoys the right to hold data resources; the right to use data processing includes the right to process data and the right to use data; the right to process data is the right to process and handle the original data to form a new dataset, and the right to use data is the right to access, copy and so on on the collected data; the right to operate data products is the right to operate data products or data-derived products based on new data, and the ownership of data products also belongs to the data processor. The right to operate data products is the right to operate based on new data products or data-derived products. The creativity of data products reaches the scope of intellectual property protection, and intellectual property protection can be applied, and its ownership is also attributed to the data processor.

3.5 Raw, pooled and derived data

Some scholars have proposed three levels of classification of data resources based on three behaviors in the transformation of data resources: raw data resources attributed to data producers, aggregated data resources formed by data processors organizing and assembling data, and derived data formed based on
the analysis of the above data [5]. Similarly, some scholars have proposed the distinction between basic data and derived data, original data and derived data, original data and basic data are roughly the same, in which the basic data is the data initially obtained, including the personal data of the natural person, the basic data in terms of official activities and the basic data related to the production and operation activities of the enterprise. Raw data is based on some kind of fact or behavior obtained, does not depend on obtaining from other people, raw data is the premise of aggregate data and derived data, raw data is also as basic data has a key role. In general, raw data is basic data, but the two corresponding to other data concepts are different, basic data corresponds to the derivative data, raw data corresponds to the data collection and data products, the name of which there is a certain difference, but its connotation is basically the same.

The source of raw data is classified as natural information and personal information, and the datamation record of objective information is the most primitive way of producing data resources, and the data processor who wants to obtain the right to hold the data must have the consent of the subject of the raw data. Raw data is fragmented and unstructured, and only has very limited use value. The original data is processed through the pooling of data processors to form different collections of data, which realizes the interconnection of data and greatly enhances the applicable value of data resources. There is currently a certain protection gap for publicly available non-original data collections in data collections. Derived data can make data value-added, for example, the current various software by analyzing the user's preference, product recommendation to the user, which greatly enhances the user stickiness. Derived data has greater commercial value than original data and aggregate data, and its essence is also a data product. With the reanalysis of data products, data analysis can be further optimized to continuously achieve positive feedback.

3.6 Categorization of interest in data elements

Depending on the content of the data element rights and interests, i.e., the means of data protection, they can be categorized into property rights, intellectual property rights, new types of rights and interests and anti-unfair competition.

The typing of data according to different criteria helps to deeply understand the connotation of data, so as to confirm the right according to different attributes of data or the flow process. Of course, the classification of data is not limited to this, for example, according to whether the data is cross-border can be divided into cross-border data and domestic data, and according to the country of belonging to the national data and foreign data, and so on. This article is just a summary of the main types of data, and will not repeat the rest.

4. Analysis of the current status of legislation and judicial determination of data ownership issues

4.1 Status of data ownership legislation in China

Some scholars have summarized the legal system of data protection in China in both vertical and horizontal dimensions: vertically, it includes central legislation such as laws, administrative regulations and other administrative regulations, such as the Data Security Law, as well as local legislation, including local laws and regulations and local government regulations, such as the Shenzhen Special Economic Zone Data Regulations, which form a vertical legal system combining the central government and local authorities; Horizontally, the legislature has made efforts in the areas of "personal information protection system", "data circulation system", "data management system" and "data security". The legislature has made top-level designs on "personal information protection system", "data circulation system", "data management system", "data security", etc., and constructed a horizontal legal system that is compatible with China's economic development [14].

4.1.1 Data specifications need to be refined urgently

Legislation relating to the data field is mainly concentrated in the Data Security Law, as well as in departmental regulations and local legislation, and there is insufficient articulation between the top-level law and subordinate laws, such as departmental regulations. China has realized top-level design in terms of data flow and data security, but the provisions on data ownership issues are not clear, such as the attribution and content of data rights. In the Data Security Law, it is stipulated that a categorized and graded protection system be established, but the current situation is more of a policy document, which only forms the mainstream view.
For example, in the protection of personal data, China's Personal Information Protection Law for the right to personal data portability provisions are relatively general, only in the second paragraph of Article 45, a slight reference to its concept, and does not provide for the transmission of the form of data and can be used in the scope of the personal data portability of the right to the user to protect the rights of users has a vital significance, the right of portability of personal data in this area of the provisions of the right to personal data portability needs to be refined.

4.1.2 Certain gaps in data-related legislation

In terms of cross-border data, there is no relevant specialized legislation in China, and the existing legislation is too principled, there are certain legal deficiencies, for the jurisdiction of cross-border flow, extraterritorial validity of data ownership has not been clearly stipulated, to determine the ownership of cross-border data is an important prerequisite to ensure data security.

China's relevant legislation does not adequately protect the property rights of enterprise data, and the rights of enterprises over aggregated data are not clear [17]. Existing judicial practice positions data resource products as a kind of competitive property rights and interests, for example, in Taobao v. Mei Jing, the court recognized that the developer of the data product enjoys competitive property rights and interests and protected it with the anti-unfair competition law, although the anti-unfair competition law can protect the legitimate interests of the enterprise, but it may still lead damage to the interests of the operator of the data product, which cannot obtain prior relief. Although the anti-unfair competition law can protect the legitimate interests of enterprises, it may still result in data product operators' interests being jeopardized without prior relief.

4.2 Current status of judicial determinations regarding data ownership

4.2.1 Collection and use of data in violation of the principle of "triple authorization"

Weibo sued Maimai [2] The defendant exceeded the scope of the license to capture and use the information of Weibo users during the cooperation period, and after the termination of the cooperation, the defendant still used the information of users from Weibo as the information of non-Pulse users in the software, which jeopardized the security of Weibo's users' information and damaged the plaintiff's lawful competition rights and interests, and constituted unfair competition. In this case, the court judged whether the infringement was constituted according to the "triple authorization principle", i.e., "user's consent" + "platform's consent" + "user's consent", and the third party's consent was based on the open platform. "A third party utilizing user information based on the open API cooperation model should not only obtain the consent of the data provider, but also obtain the consent of the user again, fully respecting the user's will and realizing the cooperation of data economy. Microblogging v. Pulse case for data ownership is not a simple one-size-fits-all, but through unfair competition to be protected, and the rights and interests of the user's data to be confirmed, fully guarantee the safety of enterprise data and personal data, recognized the user and the enterprise in a certain degree of limited control.

Tencent sued. Dou Yin Jitterbug [11] In the case, the court held that the principle of triple authorization established in Weibo v. Pulse had become a commercial ethic for open platform network. In addition, the court recognized the data submitted by Tencent relating to user avatars and friend relationships accumulated by QQ and WeChat as commercial resources that could give Tencent a competitive advantage, affirming Tencent's rights and interests in its products.

4.2.2 Different rights for different data types

The WeChat Group Control Case [21] The Chinese court categorized the WeChat platform data interests into two data forms, one is the data resource as a whole and the other is a single data individual. According to the single data and the data resource as a whole, the network platform party enjoys different data rights and interests. In respect of single data, this part is attached to personal information rights and interests, and according to the agreement, the platform only enjoys limited use rights. In terms of the data resources as a whole, the two plaintiffs enjoy competitive rights and interests in accordance with the law, that is, the platform of the user information collected legally, and pay a certain cost of the data collection formed by the commercial value and competitive advantage of the premise of the platform to enjoy the property rights and interests. In the case of the defendant's destructive use of data resources as a whole, it constitutes unfair competition.

This case does not clarify the value of personal rights and interests of the enterprise, and only makes a preliminary division of the data rights and interests enjoyed by both: the original personal data can be reduced to specific personal information, so the individual enjoys personal information rights and
interests in the original data. Enterprises have invested certain costs in the collection of user information and should enjoy certain property rights and interests. For the original data, the data control subject can only enjoy the limited right of use according to the agreement. According to the characteristics of network resource sharing, the data rights and interests of a single user is not "who controls, who enjoys", the use of data controlled by others as long as it does not violate the principle of "lawfulness, necessity, legitimacy, consent, and not excessive", generally will not be recognized as an infringement.

In the case of unfair competition dispute between a (China) software company and an Anhui technology company, the court found that the user did not have property rights in the personal information provided by the user, and that the enterprise had a limited right to use the original data, and a competing property right in the data products resulting from its processing.

In the case of Unfair Competition Dispute between Zhejiang Sou Dao Network Technology Co., Ltd. and Hangzhou Ju Ke Tong Technology Co., Ltd., the court similarly held that a single raw data enterprise enjoys a limited right of use, and that the enterprise enjoys a competing property right and interest in the data resource products formed by aggregation.

4.2.3 Substantial substitution exists for the data product in question

The court in the case of Wei Meng sued Yi Fang categorized Internet data into public data and non-public data from the perspective of balancing interests, where public data is data for which no access privileges have been set, and non-public data is data for which access privileges have been set by logging in or other measures. The company has captured the public and non-public data of Wei Meng through illegal means and processed them to form data products to be sold to users. This behavior leads to the leakage of personal information on the one hand, and damages Wei Meng's right to the relevant data on the other hand, which constitutes unfair competition. The means by which Ant Square captured the public data was not legitimate, so the data products formed by Ant Square did not have a legitimate basis and harmed the rights and interests of others.

Alibaba sued Code Note Company, in which the defendant Code Note Company used the data published by Alibaba directly on its platform, applied the data in question beyond the necessary limits, harmed the interests of Alibaba, violated recognized business ethics, and constituted unfair competition.

Taobao sued Mei Jing company. The data content provided by the "Business Counselor" data product in the case was derived from the original user information data, but was developed by Taobao to form a new derivative product. The court's decision recognized Taobao's interest in the data and denied any property ownership of the data. Mei Jing's "Gu Gu Mutual Support Platform" substantially replaced the "Business Counselor" data product, undermining Taobao's competitive advantage and constituting unfair competition.

By analyzing the above typical cases, judicial practice tends to protect data as property rights and interests in the sense of competition law. In addition to the anti-unfair competition law, which provides the legal basis for enterprise data protection, the copyright law and patent law also provide effective ways of protection. As to what kind of data is protected by property rights and interests, the derivative data (also including data products) formed through in-depth development and processing is generally not controversial, and the enterprise enjoys competitive and property interests; as to the original data, the enterprise enjoys the right to use it in a limited way, and the data collected by the platform is generally taken into account in terms of its commercial value and the importance of the platform's operation; as to the collection of data formed from the user's information and its published data or as a whole, the platform enjoys a property interest. In determining unfair competition, whether it conforms to the "principle of triple authorization" and whether the excessive use of other people's data constitutes substantial similarity are important considerations.

The same data in different scenarios and at different stages of circulation may lead to differences in data ownership. The key to determining data ownership lies in balancing the rights and interests of data sources, data holders and data users, so as to maximize the value of data. For example, when there is a conflict between enterprise data and public data, only after maximizing the public interest can the interests of the enterprise be effectively safeguarded. The protection of data rights and interests has different needs in different scenarios, and we need to encourage openness and sharing, as well as fully protect the rights and interests of enterprises and individuals, so as to enhance the flow of data and maximize the value of data.
5. System building: data classification and rights

From the perspective of data circulation and shared use, the non-exclusive character of data should be guaranteed and the creation of property rights in data should be avoided. EU academics have already reached a consensus on the rejection of data property rights legislation and the promotion of a data access regime, which is leading data legislation [6]. EU academics on "who owns the data" extensive discussion, the results are not known, proved that "who owns the data" question cannot be answered, then the best way to solve the flow of data is to answer the question of "who uses the data?" This is the basic factual judgment of data rights. The experience of the European Union provides certain reference for the formulation of data protection legislation in China. We should not only provide the basis for adjudication, but also form an incentive mechanism through the protection of rights and interests, and realize the flow and value of data through the legislation on data rights, so as to provide a solid guarantee for the development of China's digital economy.

Article 20 of China's data puts forward the theory of three rights, although it is not defined at the level of national legislation, there exists a large amount of research in the academic community, which establishes a basic framework for the construction of a data base system, and lays the foundation for the clear establishment of a hierarchical and classified authorization system on the basis of confirmed rights. The right design of the three rights should aim at the circulation of data: first, the right of data resource holders contains two aspects of rights: namely, the right to consent to the access or transfer of data held by others, and the right of independent management of the data for lawful holding, management and prevention of infringement; second, the right of data processing and use contains the right to innovative labor or substantial processing of data and the right to use the collected data in business activities; third, the right of data product operation contains the right to use data in business activities; and third, the right of data product operation contains the right to use data in business activities. Third, the right to operate data products includes the right to operate data products lawfully and to impose restrictions on third parties [9]. This right division dilutes data ownership, establishes a new path according to the needs of different subjects, protects the legitimate rights and interests of all subjects, and lays a good foundation for data governance.

Data elements have multiple attributes, both property attributes and personality attributes; at the same time, data also have multiple characteristics, such as sharing, multi-attribute and dynamics. Therefore, the realization of the rights and interests of differentiated data lies in the hierarchical classification to be guaranteed, which is mainly reflected in the hierarchical classification of data, subjects and scenes to confirm the rights. First, according to the different data rights subject is divided into individuals, enterprises and the government, also both personal data, enterprise data and public data; second, according to the different data flow stage is divided into data collection stage, transmission stage, storage and processing stage, application and destruction stage, including data resource holder rights, data processing right of use, the right to operate data products; third, combined with the judicial practice, the common application of field contains personal information, civil, competition law and new types of fields, in addition to cross-border data rights. China's "Practice Guidelines for Network Security Standards - Guidelines for Classification and Rating of Network Data" stipulates the specific methods for data classification and provides examples of classification and rating, classifying data into three levels, i.e., general data, important data, and core data, and introduces scenario-based governance methods based on which, such as in the face of the data magnitude, relevance, precision, and whether or not to disclose the data, the data are handled accordingly. Make the corresponding upgrade and downgrade processing[7]. In general, data will conform to various classification and grading methods, be given different labels, and generate corresponding rights and obligations, so as to realize the dynamic management of classification and grading on the basis of safeguarding the interests of all parties.

6. Conclusion

In the era of the digital economy, data has become a new type of production factor posing a serious test for the current legal system, and although the current legal system provides for data security and protection of the rights and interests of personal information, the protection of data rights is still inadequate. Article 20 on data is only a policy document and needs to be transformed into legislation related to data rights and protection. The path of data governance should be aimed at data circulation, rationally allocating the dynamic data rights and interests of various types of data subjects in different scenarios, seeking the maximum common denominator between the protection of data rights and interests and the circulation of data, and constructing a reasonable data tenure system is an important prerequisite.
for the circulation of data elements in the market. In this paper, by combing the views of scholars and judicial practice, the author agrees that under the premise of the separation of the three rights, according to the data, the subject, and the scene of the different hierarchical classification of rights. Currently, there are still many difficulties in the protection of data rights and interests, and it is necessary to follow the two-way dimension of legislation and justice to achieve the dual purpose of protection and utilization, and systematically build a mechanism for defining the ownership of data, to meet the efficient use of data in different occasions, and to realize the circulation and utilization of data among different subjects, so as to continuously create value.

References

[18] Zhejiang (2019) 0108 Civil Preliminary Trial No. 5049
[19] Zhejiang (2018) 01 Civil Final Trial No. 7312
[20] Zhejiang (2020) 01 Civil Final Trial No. 5889