On the logical rules of enterprise data validation

--Observations based on one hundred and fifty adjudication documents

Zhou Wen Office of Cooperation and Exchange, Suzhou Medical College, Soochow University

Abstract
Enterprises are essential participants in the data market. Properly defining the ownership of enterprise data can stimulate better the vitality of data and the enthusiasm of enterprise subjects. At present, the classification and grading of enterprise data are in a wide range of views, and it is difficult to unify the standard of confirmation of rights, so it is necessary to seek a proper rule of proof of ownership. Based on the observation of judicial practice, data rights should be confirmed in two steps: the first step is whether there are conditions for the confirmation of ownership, i.e., the necessity of data rights, including three specific aspects: the data has a particular scale and value; there is a data processing relationship between the user and the enterprise; and the purpose, scope, and way of using the data are legal. The second step is the specific steps to be taken to establish the right: in turn, five specific criteria are identified: whether the enterprise’s access to data is authorized, whether the data is desensitized, whether it is costly or intellectually demanding, whether the data can be recovered, and the type of data product.

Keywords: enterprise data, validation, logic rules, two-step process, judicial perspective

I. Presentation of the problem
Data, as a new type of property, take a form different from the usual perception of forms of property in the past. With the rise of the digital economy, data has become a vital productivity, and according to Coase’s theorem, clarifying the ownership of data is a prerequisite for transactions. Applying property law would ignore the protection of private interests if the data is public and violate the principle of property rights if it is personal. The application of contract law leads to a complete neglect of the object of the data transaction. Applying intellectual property law is very restrictive, as there is a wide variation in the specific forms of data to which intellectual rights are directed. For example, copyright is limited to compiled data; trade secrets are limited to personal data and data rights are limited to structured data. The Civil Code of the People’s Republic of China (from now on referred to as “the Civil Code”) only contains provisions of a leading nature, and there are no supporting laws to address the judicial practice of data rights. As a result, there needs to be more clarity in applying the law. The big data industry is booming, and data trading platforms have been set up, while the lack of rules on data rights has, to a certain extent, affected the data trading process.

On 30 March 2020, the Central Committee of the Communist Party of China (CPC) and the State Council issued the Opinions on Building a Perfect Institutional Mechanism for Market-based Allocation of Factors, in which they proposed to “study improving the nature of property rights according to the nature of data” and explicitly mentioned “data rights.” However, from the perspective of practice, there are currently two problems in defining data ownership. Firstly, there are different dimensions of observation and inconsistent standards. For example, some scholars have discussed from three levels: state, enterprise, and individual. Secondly, the current research has paid more attention to data with vital personal attributes and data with solid public attributes but needs more discussion on the issue of the determination of the rights of enterprise data with compound attributes. In contrast, enterprises, as essential subjects of data transactions, should be more concerned and studied for their rights determination rules. To this end, this paper selects enterprise data as the object of analysis, observes and analyses it from the perspective of judicial practice, and attempts to establish a set of logical rules for the confirmation of enterprise data in order to provide a theoretical analysis tool for actively promoting data transactions.

II. Status of data classification and grading standards and the focus of controversy

(i) Different perspectives on enterprise data validation
From the standpoint of practical operation, data
classification and grading are closely related to data validation and have a significant influence on data validation, and there are several views on the issue of classification and grading.

Viewpoint I. The ‘identity classification approach.’ This view is that the criteria for classifying data are primarily based on its characteristics and that there needs to be uniformity in understanding factors. Some scholars consider the attributes of data to be: high fixed initial cost\(^{(3)}\); others consider the characteristics of data to be: bi-directionality of use, indirectness of benefits realization, and uncertainty\(^{(4)}\), still, others consider the characteristics of data to be: companionship (corresponding to object, time, space, state, and attribute), independence, non-homogeneity, exclusivity (non-public data), and non-competitiveness.\(^{(5)}\)

Viewpoint two is the ‘big data taxonomy.’ In this view, data are classified into two categories based on whether they have been technically processed by a data collection: readers of individual data and data that a data collection has technically processed.

Viewpoint 3: “Application Scenario Taxonomy.” According to this view, data should be classified according to the application scenarios of the data and divided into four categories: customer data, business data, management data, and derived data.

Viewpoint 4: The “security level classification approach.” This view is that data should be classified according to its security level. Data should be classified into four categories: Top Secret, Confidential, Internal, and Public.

Viewpoint 5: “Platform Taxonomy.” This view is that platform data belongs to the enterprise and that the data has public attributes.\(^{(6)}\)

(ii) Review of data classification standards and their core controversial issues

It is easy to see that the academic community has a diverse understanding of data classification and grading; some have little to do with data ownership, such as viewpoint one, while some directly determine data ownership, such as viewpoint five, while some do not have much to do with the issue of data ownership, such as viewpoint four, and some even with the expression of customer data, in fact, it is impossible to judge the state of data ownership from this, it is not difficult to find that the relationship between data classification and grading and it is easy to see that the relationship between data classification and grading is not easy to operate; Viewpoint 2 is relatively simple and easy to observe, but it is not easy to discover the inherent rights of data; Viewpoint 3 classifies data according to the general rules of data ownership from a judicial perspective.

III. General rules for enterprise data validation

What makes data unique as a kind of property is that “data carries personality rights, property rights, and the relationship between the transfer of property rights arising from personality rights and property rights”.\(^{(8)}\) This, coupled with the fact that Data itself is reproducible, transferable, and adaptable, leads to a much more complex path and method of data rights determination than ordinary property. From the actual situation of data flow, data flow usually involves three problems: user to user, user to platform, and platform to platform. User-to-platform is relatively widespread of these three modes, while media will play an essential role in the future digital economy with their data advantages. In view of this, in order to better discuss the issue of data ownership, the observation dimension of this article is mainly limited to user-to-platform.

Since the prerequisite for data rights is that the data has a corresponding value to others or to society, this paper takes a two-step approach to data rights. The first step...
examines the prerequisites for data rights, i.e., what data is necessary for requests to be established; based on this, the second step examines how to prove ownership, i.e., the specific steps to be taken to establish rights.

(i) Prerequisites for data validation

This paper argues that the prerequisites for data validation are threefold.

1. Data must have a specific scale or value.

In combination with judicial practice, the source of individual data is usually the user, which is relatively single and has vital personal attributes, so the protection of private rights and interests should take precedence at this time, and the Personal Information Protection Law or the Civil Code is sufficient to solve the problem of ownership of such data. In this article, the so-called specific scale or value is mainly reflected in two aspects: the data content has value, and the data can bring certain economic benefits or competitive advantages to the data holder.

On data content having value. Only when data content has value does it generate demand for use and the possibility of data flow? In the case of Taobao v. Fairview Business Consulting,[9] the platform integrated in-depth analysis of raw data on a large scale to form an intuitive, predictive, and statistical big data product. The value of the data content is reflected in its ability to provide guidance, increase the ease of operation of the industry and improve the efficiency of the economy and society. In the case of Zhang and the Municipal Bureau of Finance Government Information Disclosure,[10] one of the reasons the court allowed the platform not to disclose the data involved in the case was that the data contained company financial, production, sales, and operational statistics, and the value of its data content itself had commercial value. Therefore, the competing property interests in the data did have a need for protection.

About the fact that data can bring a specific economic benefit or competitive advantage to the data holder, the term competitive advantage refers to core competencies, i.e., the degree to which the goods or services offered by a market player are attractive to consumers. High-value data can lead to potential trading opportunities and enable a company to stand out in a highly competitive market, thus continuously enhancing the economic benefits or competitive advantage. Therefore, any party with such data can gain a head start and advantage in the market. In the case of Taobao v. Fairview “Business Counselor,” the big data product was able to win business benefits for the platform and accumulate a competitive advantage in the industry. In the case of Coolcom and Zheng, infringement of trade secrets,[11] the business data in the case had reference value, i.e., commercial value, to different users’ preferences and business status. In the tort liability dispute between Minister Gong Co. v. Hongwan Co.,[12] the destroyed settlement data was a possible property interest, and as an object of electronic data, it was the object of protection under tort law.

As can be seen, the scale of the data formed by aggregation and the value it provides in terms of convenience or usefulness, this value becomes a profit-seeking target for businesses, and the role of the data is realized, which ultimately promotes the frequency of data flow even more. Therefore, data of a certain scale or value is the primary prerequisite for data rights.

2. Presence of data collection and processing relationships

It is also vital that valuable data is effectively linked to the enterprise platform, i.e., that there is a data collection and processing relationship between the user and the platform. The data collection and processing relationship is an essential way for the data to form scale and for the ownership of the data to face reallocation, which is also a prerequisite for establishing a mechanism for identifying rights between users and platforms. Data processing itself is, in most cases, a civil relationship, and if the data collection and processing process involves both the user and the platform agreeing on the ownership of the data, the division of ownership is clear and should be handled following the agreement between the two parties, provided that it does not violate the law. For example, in the case of Wanfu Jinan Company v. Real Estate Technology Development Service Department,[13] the allocation of data rights and interests in real estate transaction information was ultimately based on an agreement signed by both parties.

3. The purpose, manner, and scope of use are legal

The object of the right must be legitimate data, not only for collection but also for subsequent processing. The existence of a data collection and processing relationship does not mean that the subsequent use of the information is legitimate. Therefore, it is necessary to regulate the legality of acquiring data.

In the case of Akiyip Inc. v. Feilu Inc.,[14] Feilu Inc. caused the false data of video playback on Akiyip’s website to increase through virtual user clicks. Although there was no direct competition between the parties, such conduct harmed the other party’s commercial interests, and the purpose, manner, and scope of the use of the data was unlawful. The court noted that “unrestrained use of information would be detrimental to encouraging commercial input, industrial innovation, and honest business, and ultimately undermine a healthy competitive mechanism.”[15] In the case of Baidu Inc. and Hantao Inc., other unfair competition dispute,[16] the search and crawling of information from other platforms, although not in breach of Internet technology agreements, the scope
and manner of displaying data was unreasonable and resulted in substantial substitution for other platforms. In short, examining the legality of the purpose, manner, and scope of data use is about allocating data ownership. The purpose of the legitimate use of data is to bring about a virtuous cycle of social and economic development. Firstly, if platforms plunder honest data from other platforms to provide homogenized services and disguise themselves as substitutes for other platforms, it is equivalent to taking a shortcut to innovation. In aggregate, there is no real impetus for social creation. Secondly, due to the economies of scale of data, large-scale data can bring a more significant proportion of value\(^{[17]}\), and with it, the rapid growth of the platform. Even though in the short term there is significant growth of individual platforms and increased frequency of data flows, in the long time, this has the detrimental consequence of cribbing and is very harmful to the economic health of the industry. This type of scale of development can quickly gather industry strengths and resources, making it easy to form oligopolies. Not only does this not contribute to social innovation, but it can also hinder innovation on other platforms, undermine the industry ecology, and reduce the incentive for social innovation. Therefore, the lawfulness of the purpose, manner, and scope of using data is another prerequisite for data rights.

It is also important to note that where there is a need to protect data that does not have the prerequisites for data rights, protecting the rights and interests of individuals should take precedence. In the case of data that is not subject to economic activity but which falls within the scope of personal data, protecting the rights and interests of individuals shall also take precedence.

Having discussed the prerequisites for data validation above, the question of how data with validated value should be validated between the user and the data processor is the next step in data validation to be discussed.

(ii) Steps in the identification of rights

![Data Classification and Identification Roadmap](image)

**Figure 1 Data Classification and Identification Roadmap**
The main idea behind the steps to set up data validation is to build a framework to layer the data and facilitate the verification of each layer. Following this idea and considering judicial practice, data validation can be divided into five steps: authorization, desensitization, cost or intellectual effort, recovery of the situation, and examination of the type of product.

1. Is it authorized

Article 13 of the Personal Data Protection Act states that information processors must obtain the consent of individuals before processing information. Data of a certain scale consists of individual data, which originates from individuals, and therefore data of a specific scale also inevitably has personal attributes. Moreover, data of a certain scale also condenses the labor and financial effort of the enterprise platform. Therefore the acquisition and use of data require the authorized consent of the user and the venue, both from the perspective of personal privacy and the perspective of platform revenue. Authorization implies the disposition of data by the user, and the user’s right to dispose of it should be fully respected, provided that it does not violate public order and morality, so authorization is used as the first step in confirming data rights.

In *Sina Weibo v. Pulse*[^18^], the Pulse platform obtained and used the correspondence between non-Pulse users’ contacts and Sina Weibo users (user information such as avatar information, hashtag information, education information, occupation information, etc.) in the Pulse user’s mobile phone address book without the user’s consent and authorization from the Weibo platform, undermining the non-Pulse users’ right to informed choice and violating the principles of good faith and accepted business ethics. The data was obtained without the user’s consent, and the platform’s permission, and no transfer of data ownership occurred.

On the contrary, if the user is not authorized, it will lead to infringement. For example, in the case of *copyright infringement and unfair competition dispute between Tencent and Byte Jump*[^19^], Byte Jump agreed with the user that the user was responsible for uploading videos of Tencent’s games and could share the proceeds obtained; such an unauthorized act would constitute copyright infringement.

Determining whether or not data is authorized before it is used can go a long way toward avoiding disputes over the validity of the data. Assuming that the data has been authorized and the scope of use is within the range of the authorization, it is logical that the ownership of the data should be allocated following the contractual requirements and that a contractual claim may be brought when one of the parties is affected by a breach of contract from the other party to the contract.[^20^]

Conversely, in cases where the data is not authorized or is used beyond the scope of the authorization, which amounts to a lack of knowledge or consent from the data provider, further determination of data ownership is required in conjunction with step two.

2. Whether desensitized or not

“sensitive” in “desensitization” refers to sensitive personal information. According to Article 28 of the Personal Information Protection Law, sensitive personal information is personal information that, if leaked or used unlawfully, could easily violate a natural person’s human dignity or endanger the safety of their person or property. The proper handling of sensitive information is directly related to the question of whether the right to privacy is violated, as stipulated in Article 1032 of the Civil Code, which states that “no organization or individual shall violate the right to privacy of others using spying, intrusion, leakage or disclosure.” The right to privacy is a personal right that is dominant, absolute, and exclusive. Therefore, the processing of sensitive data is a key step in the confirmation of data rights.

Individuals have the right to dispose of the data property. Where authorization exists, this indicates that the user is willing to cede an interest in personal information, and the ownership of the data should be determined following the approval. For example, in the case of subjects who enter into clinical trial protocols or consent forms for issues, the ownership of medical information relating to their subjects should be determined following the scope of the authorization.

Protecting personal privacy information in the absence of authorization should be given a high priority, aiming to protect the absolute right from infringement. For example, in the case of *Luo v. China Pacific Property and Casualty Insurance Co. Ltd. privacy dispute*[^21^], Pacific Property and Casualty Insurance Company illegally collected personal information (mobile phone number, name, vehicle insurance expiry date) and called the plaintiff to promote vehicle insurance, resulting in moral damage to the user. This is direct evidence that the flow of un-desensitized data can lead to significant challenges to personal privacy.

In the case of *a privacy dispute between a user and Fung*[^22^], the judicial decision to uphold the rights of the parties in relation to the leaked name, phone number, and itinerary information was primarily from a privacy perspective.

As can be seen from the previous section, the protection of the rights of individuals should take precedence over data containing sensitive data that has yet to be desensitized. For data that has been desensitized, it is necessary to determine the ownership of the data in conjunction with step 3.

In the other case, the data itself does not contain sensitive
data. Such data also needs to be combined with the third step to determine the ownership of the data. For example, in the case of *Wanfu Jinan Company v. Real Estate Technology Development Service Department in a contract dispute*, the publication of information on the sale of commercial properties by real estate enterprises was a mandatory requirement of China’s system of clear price indication for the sale of commercial properties. It did not contain personal information, so other decisive factors were assigned to the data.

3. whether costly or intellectual work is involved

In the case of data that has been desensitized, the subject who has paid for the cost of academic work is entitled to the value added by their labor.

The cost paid deserves the ownership of the data. In the case of the *dispute of unfair competition between Deacon and Leaders v. China Service Company*, although the right to use the member account and password belonged to the user, the database corresponding to the account contained the cost paid by the data processing platform, which had a high reference value and direct economic interest. The data processed is exclusive and has long been different from ordinary customer information, and its property rights and goods belong to the data processing platform. Other platforms can view, obtain and use the dealer’s database information by “crashing the database,” which is an act of “seeking business opportunities and gaining competitive advantages for oneself without paying for the work or improperly using the fruits of others’ work, which is tantamount to “gaining something for nothing.” This is tantamount to “gaining something for nothing” and “eating people’s money.”

For example, in the case of *Flush Inc. v. Lighthouse Inc.* and the case of *Baidu Inc. and Hantao Inc.*, the data processing platform invested in costs and produced data products. It had legally protected rights and interests in the products, including but not limited to the right to dispose of and In the case, the data processing platform invested costs and produced data products, and had legally protectable interests in the products, including but not limited to the right to dispose of them and the right to receive benefits.

In practice, the courts will look not only at whether costs are incurred but also prudently at what costs are incurred. Costs must be more than equated with data processing, as defined in this article. Such costs must be able to transform the data and not interfere with the realization of the original data owner’s interest.

In the case of the *dispute between AbbVie and Liang Mou, harmed trade secrets*, the medical insurance system on the industry’s impact analysis and sales forecast report of the table without specific content or calculation methods, it is difficult to determine that it can bring competitive advantage to the operator, and thus difficult to carry out data confirmation. In the case of the *unfair competition dispute between Tencent and Sodao*, the group control software directly accessed WeChat information without authorization. Although the WeChat function was added, it did not change the data. Ultimately, the original data platform enjoyed a competitive interest in the data.

Whether or not intellectual work is involved is another form of input. Academic work, also known as mental work, can be divided into knowledge creation, knowledge transfer, knowledge management, and knowledge utilization, and is infinitely broad and creative. The intellectual work is what gives the Data its originality and is what sets the processing platform apart from other competitors.

The sale of the data processing platform’s salary and employment industry analysis data without any payment of research and development, operating costs, or consideration constituted unfair competition in the case of *Aipling v. Xuezhi*. The data processing platform poured intellectual labor into the data, which resulted in data that could predict and analyze employment in a particular industry with a reference value. This process transforms the potential productivity of intelligence into direct productivity, giving the data a new look and a different function than before, creating a whole new dimension of value.

In the case of *Wanfu Jinan Company v. Department of Real Estate Technology Development Services*, because Wanfu Jinan Company distinguished the massive amount of comprehensive real estate information into current status files, preview files, historical files, commodity house filings, stock house filings, etc., and made it public on the platform, it invested a large amount of intellectual labor in the development and promotion of the information in question, and it was entitled to the revenue corresponding to its investment by the Cooperation Agreement. It is entitled to the proceeds corresponding to its investment following the Cooperation Agreement.

As can be seen, a certain level of cost or intellectual effort is essential to determine whether the transfer of data rights has occurred. Where no cost or intellectual effort has been incurred, ownership of the data has not been transferred, and the interest of the original data platform is protected. Data that has been subject to cost or intellectual effort requires further discussion.

4. Can it be recovered

“‘The anonymization of data is not irreversible with newer iterations of accessing big data, advances in crawling and analysis techniques and artificial intelligence technologies.’ [33] The potential for infringement of the rights of others exists if the data is recovered so that it
can point to an individual or is essentially the product of someone else’s labor. Therefore, while data that has been subject to cost or intellectual labor is altered to some extent, the transfer of ownership of the data is subject to the condition that it does not infringe on the rights of others.

In practice, there is no shortage of similar situations: in the case of copyright infringement and unfair competition disputes between Tencent and Byte Jump, the game video made by the game player is only a new way of showing Tencent’s game, and the basis for its picture generation is the operation of Tencent’s game. It is easy to see that, removing the player innovation part, the game video points to the game itself, so the player cannot replace Tencent as the copyright owner of the game video. In the Taobao v. Fairview “business counselor” unfair competition case, if the original network data is processed by the platform not out of the scope of the original network user information, the data should still be subject to the control of the network user. However, in the case of derivative data that does not directly correspond to the information of the network users or the original network data, the network operator has its own independent property rights and interests in the data. Therefore, the allocation of ownership of the processed data is still subject to the post-recovery state. If the data remains within the scope of the original network user information after the cost or intellectual effort, the protection of personal information shall take precedence. The check on the recoverability of data is to protect the rights of data processors better. The current state of technology can achieve a certain level of data recovery, and data that can be recovered after recovery to reveal personal information should take precedence over the protection of the rights of individuals. If the personal information cannot be displayed after restoration or if it cannot be restored, you will need to proceed to step 5 for identification.

5. Product types

The final form in which the data is presented can affect the application of the law. Data products can be divided into derived and non-derived depending on the extent to which data processing changes the data. Derived data products are based on some of the original data and are deeply processed by specific algorithms as if the atomic bonds were broken and the scattered atoms were reconstituted into molecules with completely different characteristics. This type of data product is different from the original data and emphasizes the originality of the platform. After the above four steps of screening and excluding the possibility of infringing the legal interests of others, ownership of the data is vested in the platform, which has the right to possess, use, benefit from, and dispose of such data as a matter of course.

In the case of Tencent v. Changyou Cloud Company and Hero Mutual Entertainment Company in dispute over copyright ownership and infringement, the spatial layout structure of the game scene map belonged to the developer’s concretized creation of the virtual scene, or even the virtual world it envisaged, and the copyright law applied. The court specifically stated that “any person may only claim copyright protection and compensation belonging to the profits contributed by his copyright in respect of his original creation.” Data is turned into a special derivative product through original use. Its platform enjoys exclusive and dispositive rights over it. There are also cases in other areas, such as the case of Zhongcai Zhilian v. Dongfang Fortune Company in an unfair competition dispute and the case of Liu and Sangyou Mother and Baby Company in a dispute over infringement of the right to disseminate works on the information network, where the data involved were protected by copyright law due to the originality of the author’s views and value judgments. The WeChat public number, its data, and its right to use it all have property attributes in the case of the dispute between Palm Technology Company and Ai’s shareholder’s contribution. The value of the WeChat public number and the value of its operation was shared between PDA and Ai.

Non-derivative data products are mostly splicing arrangements or compilations of the original data, which is also presented in whole or in part in the final product, except that the data moves from disorder to order and directly expresses the data processor’s creativity. Its owner also falls within the category of protected subjects, who generally enjoy the right to property, personal rights, and rights of use.

In the unfair competition dispute of Guomi v. Yuanguang, Yuanguang crawled Guomi’s data (real-time bus information data) without Guomi’s permission. Although the real-time bus information data is an objective fact, after data restructuring, even without changing its original content, the big data product can achieve competitive advantage and economic benefits and has the property of intangible property. The platform also has the right to gain and dispose of the products formed by using objectively factual data.

This type of data product is more like data that requires human intervention to function. In the case of the labor contract dispute between Liu and Zhenghua, the data in question was data from the detection of pipelines in urban areas. The court held that a contract for contracting is a contract in which the contractor completes the work following the requirements of the fixer and delivers the work product. In contrast, the fixer pays the remuneration, i.e. it affirms the right to revenue of the custom-made
person for the data. In White Rabbit v. Ding Rong, Inc.\(^{[43]}\), White Rabbit’s original compilation and collation of the trademark data compilation were protected by copyright law.

In summary, although a number of rights and interests are concomitantly related, it is easier to start with the type of data product to anticipate the applicable legal path and the corresponding rights and interests.

IV. Remainder: Prospects for data validation rules

With the development of the digital economy, data, as a new type of production factor, is destined to become more active and play a more critical role. However, this mode of application of data, which “gathers sand to form a tower, and gathers the artifacts to form a fur,” has become more prominent in the mutual clustering of its scale value advantage, making people pay more attention to it. In this flow, each participating subject, as a data contractor or owner, will put forward corresponding power claims in the face of the data, destined to cause continuous arguments due to the issue of data rights, thus showing more and more the complexity of data rights. The purpose of this article is to simplify the operation of data rights through the identification of the necessary elements of the process of data rights and the inclusion of these elements to realize better the establishment of a better interface mechanism between data and the use of relevant laws at this time, and thus facilitate the sorting out of the related subjects of rights, to promote the continuous development and progress of data rights.

Reference

[34] (2021) Yu 01 Min Zhong No. 3805.
[38] (2021) Jing 0105 Min Chu No. 48603.
[40] (2020) Yu Min Zai 463.
[41] (2017) Yue 03 Min Chu No. 822.