

# Research on the Ownership of Data Elements under the Trend of Data Sharing

Yiming Gao

Central University of Finance and Economics, Beijing

835636239@qq.com

**Keywords:** property rights theory, ownership issues, income rights, quantification of property rights, transaction models

**Abstract:** First of all, this essay introduces the research background based on the development of big data at home and abroad, and demonstrates and analyzes the great practical significance of the definition of data property rights to the economic development of my country. Secondly, based on the theory of property rights, public product theory, and principal-agent theory, analyze the ownership of data property rights. According to the form and value of different links in the transaction chain, the data is divided into metadata, structured data, and visualization data. On this basis, discuss the ownership of the data source, seller, and buyer, as well as the attribution of ownership, use rights and income rights. Among them, the source party has absolute rights to the metadata, and the buyer enjoys the power of structured and visualized data to varying degrees. In order to further explore the distribution of income rights among different subjects, this paper establishes a quantitative model for defining data property rights based on the Bazel property rights theory. It believes that the three variables of contribution, competition intentions of other subjects, and policies (external factors) are used to obtain the data of subjects. The degree of ownership of property rights further quantifies the ownership of property rights. Finally, this essay uses the Guiyang Big Data Exchange as an example to point out the problems in the profit model and feedback mechanism of the existing data trading model. It is recommended that the big data platform further improve the feedback mechanism; in addition, the state and government should further increase data processing and business analysis The cultivation of comprehensive talents, the improvement and implementation of the big data security assurance system, and the stable talent input and legal guarantee for the vigorous development of the big data industry.

## 1. Introduction

### 1.1. Research background

With the rapid development of the Internet, e-commerce, and social networks, the global reserves of big data are growing rapidly, becoming the foundation for the development of the big data industry. According to the monitoring data of the International Data Corporation (IDC), the global big data reserves in 2013 were 4.3ZB (equivalent to 4.724 billion mobile hard drives with a capacity of 1TB). In 2016, big data reserves increased the fastest, with a year-on-year increase of 87.2%. As of 2018, the global big data reserves reached 33.0ZB, an increase of 52.8% year-on-year.

According to Wikibon's research data, the global big data market has grown rapidly from 2012 to 2019. As of 2019, it has reached 49 billion US dollars, and the market size is equivalent to four times that of 2012. It is predicted that from 2020 to 2024, the scale of the big data market will continue to grow. The growth rate will slow down and the growth rate will remain at about 12%. By 2024, it will reach 84 billion US dollars, with a compound annual growth rate of 12.3%.

According to CCID data, the scale of China's big data industry in 2018 was 438.45 billion yuan, a year-on-year increase of 23.5%; by 2021, it is predicted that the scale of China's big data industry will exceed 800 billion yuan, and the annual growth rate will slow down, but still maintain Above 20%.

## 1.2. Research significance

In the era of big data, the issue of property rights and ownership of data elements has not been effectively resolved, which has largely led to data fragmentation and hindered the smooth realization of data sharing. Based on the existing data classification methods and the definition of property rights, this paper gives a new perspective on the definition of property rights from two aspects of original data and information. By studying the issue of data property rights, different types of data are defined. On the one hand, it standardizes the data transaction process and breaks the cross-externality of the use of big data, thereby promoting data sharing; on the other hand, it is conducive to the digital transformation of enterprises, thereby greatly improving the internal efficiency of the enterprise, Enhance its competitiveness, increase market vitality, and promote high-quality and efficient economic development.

Also, the existing data pricing and transaction models are analyzed, and opinions and improvement measures are put forward to promote the emergence of a more scientific big data pricing theory.

This article mainly adopts qualitative and quantitative analysis methods, case analysis methods, and literature research methods.

## 1.3. Research ideas

This article is mainly divided into four parts.

In the first part, a comprehensive analysis of domestic and foreign researchers in the field of data property ownership and transaction mode found that the question of how and to what extent data property rights need to be resolved urgently, which leads to the thinking of this article on the data property rights.

The second part analyzes the ownership of data property rights based on property rights theory, public product theory, and principal-agent theory. According to the form and value of different links in the transaction chain, the data is divided into metadata, structured data, and visualized data, and on this basis, the data source, seller, and buyer are discussed, and the ownership, use rights, and income rights are the attributable problems.

In the third part, when researching the property right definition of the three omnipotent in the previous article, it is found that the income as a quantifiable power cannot be analyzed only by qualitative methods. Therefore, based on the Bazel property rights theory, a quantitative model of data property rights definition is established.

In the fourth part, thinking about the data transaction model is induced by the ownership of data property rights. Taking the Guiyang Big Data Exchange as an example, it points out the problems of the existing data transaction model. Finally, it puts forward feasible policies, laws, and industry recommendations.

## 2. Literature review

Regarding the property rights of big data, European Union [1] passed the General Data Protection Act, stipulating the rights of multiple data subjects, and specifically mentioned the "right to forget data", that is, "the data subject has the right to request the data controller Delete the relevant personal data, the administrator must not be in arrears without reason." [1] Nadezhda Purtova (2017) provided the latest information on the debate on personal data property rights from a European perspective. He believes that "the latest developments in data processing technologies and practices, especially AI-driven big data analysis, make personal data a difficult to implement personal property rights. The main reasons are: firstly, the data processing caused by one person's decision will inevitably have spillover effects on others; Secondly, the establishment and management of transparent property rights in terms of property objects, and the establishment of rights holders also have Challenged." [2]

In terms of big data transaction models, Zhuang Jinxin [3] believes that: "There are three types of big data transaction platform models: big data analysis result transaction model, data product

transaction model, and transaction intermediary model." Li Jiyu [4] believes: "There are three trading modes: ownership, right to use, and right to profit." Zhang Min, Zhu Xueyan [5] propose: "Analyzing the two typical transaction modes of China's big data transaction and their legal relationship based on this, the main content of the subject, object, data quality, and big data transaction supervision of big data transactions should be used to build a legal system for big data transactions in China. Cui Hongmei [6] explores the platform big data transaction model from data quality issues, functional balance contradictions, and data supply and demand issues in data transactions, and propose a multi-relational network structure. A comprehensive model of data transaction. In summary, there are still deficiencies in the research on big data right confirmation and transaction at home and abroad. The process of data circulation may involve the sharing of rights by multiple parties, and the question of how and to what extent is still unresolved; in addition, the data cannot be static, and the analysis and processing process will bring a lot of added value to the data. These added values are different. The issue of ownership between subjects needs to be resolved urgently.

This article will address the above problems and solve the "quantity" and "degree" in the ownership of data property rights from a new perspective of "raw data and information", and analyze the existing big data transaction models to promote a more mature big data right confirmation theory as soon as possible Appeared, and promoted the progress and development of my country's big data transactions.

### **3. Classification of data elements**

#### **(1) Commercial Big Data**

Commercial big data has many different classification standards according to multiple industries. Business big data does not only include narrowly defined enterprise-related data, but also data with potential commercial value such as related society and government departments. Business big data is still mainly driven by enterprises. The ultimate goal of an enterprise is profit. In order to achieve this goal, an enterprise needs to analyze its customers in-depth, use commercial big data to establish a complete user profile, and conduct precision marketing, thereby increasing customer acquisition and expanding profits. While enterprises gain benefits through big data, they are also constantly promoting the progress of big data technology and promoting the continuous improvement of commercial big data.

#### **(2) Government big data**

Government big data involves relevant data generated by various government departments in the process of performing their functions. The main driving force of government big data is government functional departments and institutions. Government big data is related to people's livelihood and public utilities. The collection and analysis of government big data can effectively promote the development of public utilities and improve government efficiency. In recent years, as the use of commercial big data has gradually deepened, government big data is attracting attention. To a certain extent, public big data can be disclosed, and the value of data can be maximized by analyzing and processing public big data with the help of commercial forces.

#### **(3) Individual data**

In real life, many individual data have not been discovered, which does not mean that it has no value, but has not been discovered by existing business models. Therefore, this article classifies individual data separately. This part of the data has huge potential and will be continuously converted to commercial data and government data in the future.

#### **(4) The relationship between data elements**

Commercial big data, government big data, and individual data are intertwined. Some public data that can be used by enterprises can be classified as commercial big data or government big data; the part of individual data with commercial value and public value is classified into commercial big data or government big data.

## 4. Quantitative model of data property rights definition based on Bazel's property rights theory

### 4.1. Bazel's theory of property rights

Bazel's property rights theory points out: "The definition of asset rights is a function of the income stream determined by the asset service flow, the proportion of personal contribution value, and the various transaction costs formed in the transaction"[7]. The higher the income, the higher the 100 million yuan that the transaction subject clearly defines the property rights; the higher the cost, The lower the willingness of transaction entities to clearly define property rights. Therefore, the clarity of the definition of property rights is positively correlated with the transaction income of the transaction subject and negatively correlated with the transaction cost. At the same time, the higher the proportion of the value contributed by one party to the transaction, the greater its impact on the variability of assets, and the more rights will tend to be allocated to the party with greater variability.

Another core concept of the Bazel property rights model is the public domain. In the above analysis, when the cost of defining property rights exceeds the benefits, the transaction subject will give up this part of the property rights, and the undefined part of the property will fall into the public domain. The public domain mentioned here is not a public good, but a collection of unclearly defined properties that property rights voluntarily give up for individuals seeking to maximize benefits. Most of the early studies of economists such as Knight (1924), Gordon (1954), and Coase (1960) believed that property rights are absolute rights, which are either completely defined or non-existent. Bazel believes that precisely because of transaction costs, it is impossible to completely define property rights.

### 4.2. Model

In the previous study on the ownership of data property rights, the categories of ownership and use rights can be divided intuitively and qualitatively, that is, whether they have ownership and whether they have use rights, but the income rights are different and cannot be analyzed only qualitatively. As a concept that can be quantified, revenue is the main concern of all entities in data transactions. This paper establishes a quantitative model for defining data property rights based on the Bazel property rights theory, formally to quantitatively analyze the degree of ownership of income rights by each transaction entity, and complement and combine the model and theoretical analysis to provide a new approach to the issue of data property rights. Perspective.

### 4.3. Quantitative model for defining data property rights

#### 4.3.1 Model assumptions

- 1) Assuming that each data transaction subject enterprise follows the principle of maximizing benefits;
- 2) Assuming that from the perspective of one transaction subject, the other subjects are all property right competitors;
- 3) Assuming that data property rights can bring profits, the more data property rights you own, the greater the corporate profit.

#### 4.3.2 Model establishment

$$P_i(e_i, o_i, g) \quad P_i \in (0,1)$$

It represents the degree of ownership of the data by the transaction entity (the degree of property right definition).

Among them, the data item  $i$  indicates the contribution of the  $i$ -th item to the data,  $o_i$  is the intention of other subjects to compete for the data property rights, and external factors  $g$  is the government, laws, and regulations.

It is defined  $C(P_i)$  as the cost function of the transaction entity owning data property rights, and  $R(P_i)$  is the revenue function of the transaction entity owning data property rights. According to

Bazel's theoretical model of property rights, the clearer the definition of property rights, the higher the cost, and the more difficult the definition of property rights. Therefore,  $C'(P_i) > 0$ ,  $C''(P_i) > 0$ , the benefit of owning more property rights subjects will increase. Increasing its marginal benefit will also decrease, therefore.  $R'(P_i) > 0$ ,  $R''(P_i) < 0$ . This is also in line with the law of increasing marginal costs and decreasing marginal benefits in Confucian classics. The cost and benefit functions are shown in the figure 1.

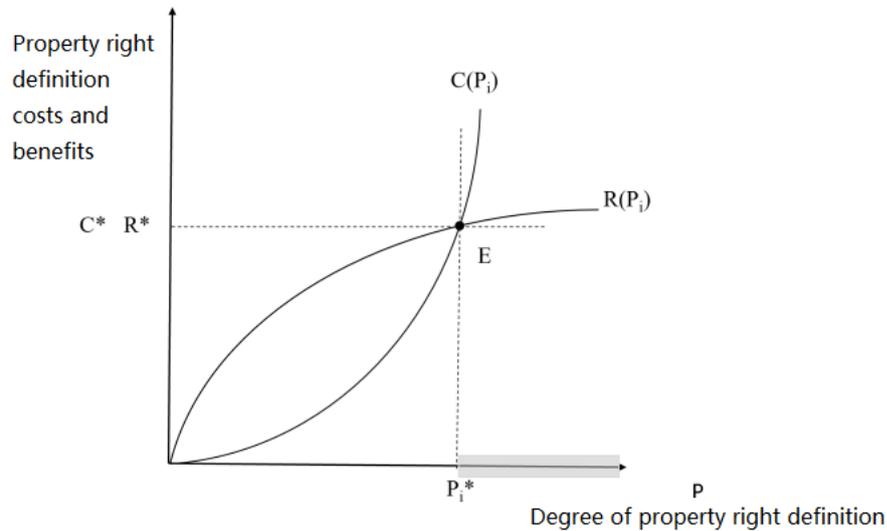


Figure 1 The cost and benefit curve of property right definition

The cost function and income function reach equilibrium at a point  $P_i^*$ , which is the boundary defined by data property rights. At that time  $P_i < P_i^*$ , the transaction subject tended to further increase the degree of data property ownership. At that time  $P_i > P_i^*$ , the transaction subject followed the principle of maximizing benefits and gave up this part of the data property rights. Data property rights fall into the public domain. The gray part in the picture in the public domain.

#### 4.4. Model application for defining data income rights

A complete data transaction chain includes "generation-collection-processing-analysis-final use", in which the processing party and the analysis party both process the data and generate added value, so the problem of the definition of the right to data income is particularly prominent. Next, apply the data property right definition model to the problem of defining the income rights of the processor and the analyst.

From the perspective of the analyst,

$$P(e, o, g) \quad P \in (0,1)$$

It is the degree of the analysis party's ownership of the data income right,  $e$  indicates the degree of contribution of the analysis party to the data,  $o$  is the processor's intention to compete for the data income right, and  $g$  is the government, laws and regulations and other external factors.

Specific indicators for measuring variables  $e$  is shown table 1.

Table 1 Variable primary and secondary index table

First-level index	Second-level index	First-level index	Second-level index
Analyze the quality of data, the spread of data in and out of the market, the level of visualization, etc.	Analyze the quality of data, the spread of data in and out of the market, the level of visualization, etc.	Analyze the quality of data, the spread of data in and out of the market, the level of visualization, etc.	Analyze the quality of data, the spread of data in and out of the market, the level of visualization, etc.
Analyst input cycle of data analysis, labor input, etc.	Analyst input cycle of data analysis, labor input, etc.	Analyst input cycle of data analysis, labor input, etc.	Analyst input cycle of data analysis, labor input, etc.
Other analysts' technical maturity, customer satisfaction, etc.			

Specific indicators for measuring variables  $\theta$  is shown table 2.

Table 2 Variable primary and secondary index table

First-level index	Second-level index	First-level index	Second-level index
Processing data quality, data inflow and outflow market spread, processing level, etc.		Processing data quality, data inflow and outflow market spread, processing level, etc.	
Processing party input	Data processing cycle, labor input, etc.	Processing party input	Data processing cycle, labor input, etc.
Potential benefits	Current value of data, estimated value of data, etc.	Potential benefits	Current value of data, estimated value of data, etc.

Specific indicators for measuring variables  $g$  can include: the degree of inclination of existing policies, laws and regulations, and industry market rules to protect upstream and downstream companies in the data chain.

If the data analyst provides more added value to the data and invests a lot of labor costs, time costs, etc., it will increase its degree of data revenue rights, that is,  $\frac{\partial P}{\partial e} \geq 0$ . If the data processor also invests labor costs and time Cost, and has a high estimated value of the data, driven by interests, will compete for the analysis party's right to use the data, that is,  $\frac{\partial P}{\partial o} \leq 0$ . Domestic policies and regulations related to data property rights have not been perfected, and the influence of exogenous variable  $g$  on  $P$  will be uncertain with the trend of future government, laws and regulations. The specific operation and calculation ideas of the model are as follows:

(1) Collect data. Divide the indicators into qualitative indicators (such as visualization level, technological maturity, customer satisfaction, etc.), quantitative indicators (market spread, cyclical market, cost input, etc.), survey and collect real data on quantitative indicators, and score quantitative indicators through expert scoring methods (0-10 points).

(2) Standard data. Through the vector normalization method, the various data are normalized at the same magnitude.

(3) Determine the weight of each indicator. Based on the entropy weight method, according to the degree of variation of each indicator value, different weights are assigned to the indicators.

(4) Standardize the subsequent substitution function and determine the degree of ownership of the income right by the processor (the degree of property right definition).

## 5. Model limitations

### 5.1. Promotion of variables

The selection of variables in the data property right definition model in this article follows the Bazel property rights theory, mainly considering the three perspectives of the transaction party itself, the competitor, and external factors (government, industry), but in the actual data transaction process, more participation may be involved square. When faced with more participants, according to Bazel's property rights theory, several variables can be substituted into the function of the degree of ownership of the subject's property rights to extend the model.

### 5.2. The subjectivity of indicator selection

In the equation for measuring the degree of ownership of property rights, different data transaction entities may be inclined to different indicators when facing different data transaction objects, so it is difficult to comprehensively measure it with a unified standard. At this time, the selection of indicators is subjective and requires the coordination of various data transaction parties.

### 5.3. The definition equation of data property rights is difficult to specify

This paper proposes a data property right delimitation model, gives the abstract form of the parametric equation and the direction of the influence of each variable on the dependent variable,

but has not yet concretized it. The specific reasons are:

(1) The regression equation requires a large amount of real sample data. Due to the reproducibility of the data, its privacy and security issues are particularly sensitive, so it is difficult to obtain a large amount of real data as sample support;

(2) There is a lack of research on the definition of domestic data property rights, and it is difficult to form a situation of mutual learning and mutual integration.

(3) When faced with different data subjects and data property rights, the definition of their property rights cannot be generalized, and it is also difficult to use one equation as the standard for the definition.

In the face of the above problems, the application of this model still has practical and feasible solutions:

Some innovative data trading platforms have emerged in China, and these platforms, as intermediaries, will play a main role in the definition of data property rights. As an intermediary agency, the data trading platform holds a lot of first-hand data such as transaction volume, transaction price, data analysis level, gold content, etc. The data sample size is large and the coverage is wide, and a high-fit regression equation can be established based on the data property right definition model. Also, the data trading platform can give full play to its carrier role, promote the continuous enrichment and improvement of the model index system, and make the automation and electronic definition of data property rights a reality.

## References

- [1] European Union. General Data Protection Regulation[EB/OL].<https://gdpr-info.eu/art-17-gdpr/>, 2018—01-22.
- [2] Nadezhda Purtova. Do Property Rights in Personal Data Make Sense After the Big Data Turn?. 2017, 10(2):208-222.
- [3] Zhuang Jinxin. Comparison and Strategy Analysis of Three Big Data Trading Platforms. China Computer
- [4] Li Jiyu. Discussion on big data transaction mode .Mobile Communications, 2016, 40(05):41-44.
- [5] Zhang Min,Zhu Xueyan. Legislative Thinking on Big Data Transaction in my country. Learning and Practice, 2018(07):60-70.
- [6] Cui Hongmei. Construction of a comprehensive model of big data transaction with multi-relational network structure. Science and Technology Innovation, 2020(24):70-71.
- [7] Ding Zhifan. Bazer's Economic Organization Research and Its Contributions——Based on the Analysis Framework of Property Rights. Journal of Hebei University of Economics and Trade, 2017, 38(02):29-35.