

# The Mechanism of Confirming Big Data Property Rights Based on Smart Contract

Haijun Zhao  
Guangdong University of Finance  
and Economics, China  
Zhj8888@126.com

Bang Zhao  
Guangzhou JesseHealth  
Rehabilitation Co., Ltd, China  
ChinaLtd262318077@qq.co  
m

Susu Cheng  
China Merchants Life Insurance  
Co., Ltd. Guangdong Branch.  
China  
Chenss002@cmrh.com

## ABSTRACT

The confirmation of data property rights is one of the most important functions of big data trading institutions. This paper builds a specialized classifier which aims to the normalization process of data property confirmation in the process of big data transaction and proposes the mechanism of confirming big data property rights based on Smart Contract.

## CCS Concepts

• **Applied computing**→**Computer forensics**→**Evidence collection, storage and analysis.**

## Keywords

Smart contract; Big Data Exchange; Confirmation of Data Property; Mechanism.

## 1. INTRODUCTION

After comparing several important big data trading platforms in China, we found that many trading institutions run in chaos. As shown in table 1. The trading patterns, the trading rules, data property confirmation and the pricing mode of data products in these trading institutions are lack of legal basis, in addition to 24-hour market opening. When it comes to responsibility of "data property confirmation", many big data trading institutions directly dumped it to the sellers, such as Zhongguancun Shuhai data trading platform. Some of them adopt the "submitting title certificate + expert evaluation", such as Global Big Data Exchange. None of these is in line with the market rules of the Platform Economics that platform operators should ensure that goods traded on their platforms are legitimate products instead of "counterfeit". So this paper aims to glean the legal basis of confirming big data property rights and discusses the feasibility of using technology to confirm data property rights and classify them. Finally, we may build the mechanism of confirming big data property rights in big data trading markets by using Smart contract technology of block chain.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Request permissions from [Permissions@acm.org](mailto:Permissions@acm.org).

ICIT '19, February 20–23, 2019, Da Nang, Viet Nam

© 2019 Copyright is held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-6633-5/19/02...\$15.00

DOI: <https://doi.org/10.1145/3321454.3321461>

## 2. THE PREMISE OF TRADING BIG DATA IS THAT DATA PROPERTY CAN BE CONFIRMED

The essence of transactions in data assets is the value exchange of big data property rights. In market, the "data asset package" usually flows from the seller to the buyer in a certain period of time and content, and the buyer pays the seller a certain reward and obtains the possession right and the right of use of the "data asset package". But in this case, how much the buyer should pay? As we know, the value determines the price. So, the price of this package is based on the property value owned by the owner. In terms of the legitimacy of the transaction, the owner identity, the nature and structure of this package should be confirmed before this transaction starts. And then, we can move to the second step: evaluate its value. Only by proving the legal identity of the owner of this package, can this transaction may legal. Only by evaluating this package scientifically and ensuring the share that the owner occupied, can it be priced accurately. This is so called data property confirmation. The premise of trading big data is that data property can be confirmed. First, we should identity the owner of big data. Second, we should evaluate its value. There are many researches aimed at confirming big data property in a static environment. But in current circulation environment of big data, data resources are flowing and variable. So it has great significance to develop and utilize the big data assets of building a dynamic model which aims to evaluate it.

## 3. BUSINESS CONTENT OF CONFIRMING BIG DATA PROPERTY RIGHTS

Information resources can be divided into property information, non-property information and accompanying property information [1]. Similarly, big data property rights can be divided into property data, non-property data and accompanying property data. Property data are absolutely produced by data producer. He or she has the complete ownership and the disposition right to it. Non-proprietary data refers to the external data collected by the data owner from the outside world. The data owner has no ownership and disposal right to this part of data. He or she only can use it. Accompanying property data are new data that were processed by data processor, which condenses the general labor value of the data processor, and made the processed data gain value. So, he or she has the certain value claim for this part of the data. In general, the process of data validation can be divided into three steps. First, every type of property data should be confirmed in "data asset package". We must figure out the proportion of it. Second, we should evaluate its value separately. There are differences in the way of calculating value between property data and accompanying property data.

**Table 1. The major big data trading platforms in China**

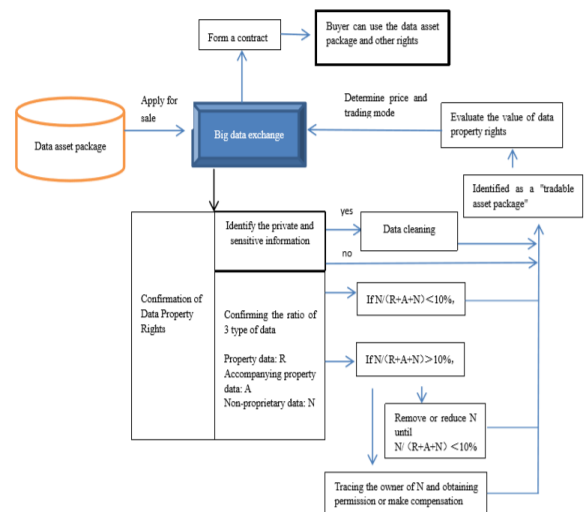
Platform item	Global Big Data Exchange (www.gbde.com)	Central China Big Data Exchange (www.ccbde.com)	Shanghai Data Exchange (www.chinadep.com)	Datang Technology Co. LTD (www.datatang.com)	Zhongguancun Shuhai Data Trading Platform (www.datarong.com)
Trading mode	Big data analysis result mode	Transaction intermediary mode	Data product transaction mode	Data product transaction mode	Transaction intermediary mode
Trading rules	<p>①Transaction content: not initial data, but the results after data cleaning, modeling and analysis.</p> <p>②Trading qualification: Club member after approved by platform</p> <p>③Trading varieties: nearly 4000 types, financial big data, medical big data</p> <p>④Transaction format: API data interface, data terminal, online.</p>	<p>①Transaction content: original or processed digital information,</p> <p>②Trading qualification: registered and approved by platform</p> <p>③Transaction varieties: Internet big data, medical big data, energy big data, financial big data, transportation and tourism big data.</p> <p>④Transaction format: API data interface</p>	<p>①Transaction content: marketing application data products and credit application data products.</p> <p>②Trading qualification: registered and approved by platform.</p> <p>③Trading variety: face marketing application and credit application, 30 data single goods listed trading.</p> <p>④Transaction format: Deploy Configuration file to a dedicated server</p>	<p>①Transaction content: data customization, data mall, mobile application data service.</p> <p>②Trade qualification: registered as a website user</p> <p>③Transaction varieties: speech recognition, image recognition video processing, social network/e-commerce.</p> <p>④Transaction format : the website gives the buyer a link to purchase data for download.</p>	<p>①Transaction content: the platform itself does not store and analyze data, but ACTS as a transaction channel.</p> <p>②Trade qualification: registered as a website user</p> <p>③Trade variety: trade variety mainly depends on seller, such as mobile communication, social communication, public transportation</p> <p>④Transaction format: The platform and data provider offer usage key of API after payment</p>
Confirmation of data Property	The seller submit of ownership certificate + expert review	The ownership of the data belongs to the data provider	The property rights of the data belongs to the registered member	Can ensure a clear ownership relationship of data products.	Ensure that the seller owns the property rights and it is legitimate and reliable.
Data value assessment	Assess the value of the data according to its timeliness, capacity and type.	The value of the data shall be determined by the data seller or evaluated by the platform professionals.	The registered members independently evaluate the value of the transaction data to conduct data pricing.	The platform independently evaluates the value of the data according to the characteristics of data.	The seller assesses the value of the data and then prices the data autonomously.
Pricing	agreement pricing, fixed pricing and set pricing.	Depends on seller or specialists from platform	After the supply side sealed price bidding, the demand side selected the best	The quotation is mainly provided by the website, and the buyer purchases on demand.	The price quoted by the seller will be charged on a monthly basis or by the number of calls.

Last, we should ensure that this transaction is legal. The identity and distribution of owners in this non- property data should be identified. Only the ratio of non-property did not exceed the statutory criteria <sup>1</sup>, can this transaction is legal. Otherwise, this data must be removed from data product<sup>2</sup>.

From the perspective of privacy protection and national information security, personal privacy, commercial secrets and sensitive information that endangers national security are not allowed to be found in this data product. Otherwise, this product need to be cleaned based on the "right of to be forgotten". Therefore, another important work of confirming big data property is to confirm whether there are sensitive and private information components in this product.

<sup>1</sup> According to the regulations on the Protection of Books and periodicals, the limit of "appropriate citation" shall not exceed 2500 words or 1/10 of the cited works in non-poetic category. "when a work is cited, the total amount cited shall not exceed 1/10 of the work". Among them, "must not exceed 1/10 of the total amount" is the legal standard.

<sup>2</sup> This regulation has two advantages, one is to prevent data infringement, the other is to prevent excessive redundancy of data resources.



**Figure 1. The process of confirming big data property rights.**

To sum up, process of confirming big data property rights can be described as figure 1.

#### 4. TECHNIQUES FOR CONFIRMING BIG DATA PROPERTY RIGHTS AND THEIR LIMITATIONS

**Table 2. Techniques for confirming big data property rights**

Techniques	Key technology	Appropriate object	Legal basis	limitations
Data Citation	Data indexing; Data fingerprint; Data tracing	Fine-grained structured data sets and databases	Rules of data citation, copyright Law, WIPO copyright	Only for data from literature to its cited or the other way around
Data provenance	Meta data; The model and management of Data provenance ("7W")	From single data to data set, from database to data warehouse.	data tracing Standard, copyright Law, WIPO copyright	Only for data set, database and data warehouse
Data reversible hiding technology	Data compression technology without loss.(data fingerprint; digital watermarking)	Text, 2D Image, 3D Image, Video and Audio data	Technical Standard, copyright Law, WIPO copyright, Electronic signature Law	Only for single file or single database
Computer Forensics	Data recovery; Data fixation; Evidence identification.	Text, calligraphy, Picture, recording, Video, Web, Mail	data Security Standard, Real right Law, copyright Law, WIPO copyright, Electronic signature Law	Only for those data which electronic evidence are fixed
Block Chain	Distributed storage; common maintenance; Consensus mechanism; security	Real-time data flow of data Resources based on Block chain Technology	Block chain and Technical Standards of Intelligent contracts, copyright Law, WIPO copyright, Electronic signatures Law, contract Law	Only for data was recorded as block chain

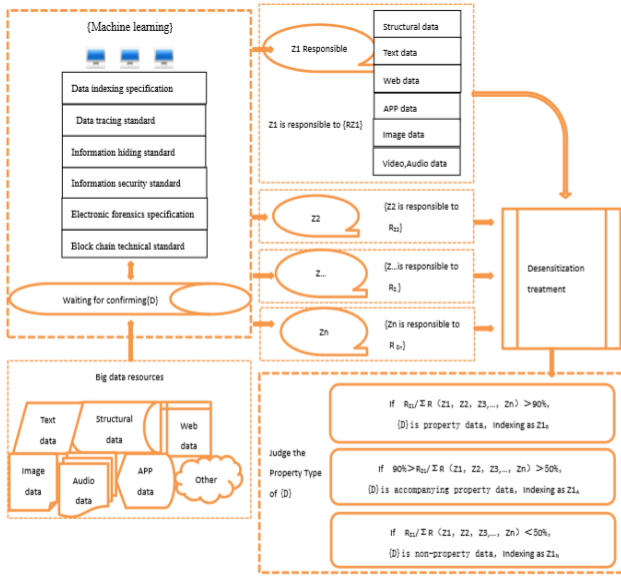
Strictly speaking, confirming big data property is a judicial identification process which is related to the attribution of data property rights. On one hand, we should use certain techniques to trace data. On the other hand, the tracing results must have legal effect. In big data, the "data asset package" which circulating in the market often contains many data form different platforms with

different granularity, different formats, different media, different structures. Different data needs to be processed with different techniques [2]. We have already summed some feasible techniques and methods for confirming big data property which are data citation technology, data provenance technology, data reversible hiding technology, computer forensic technology and block chain technology. But they have their limitations. As shown in table 2. For example, data citation technology only can be used to trace data from literature to its cited or the other way around. Data provenance technology only access to data set, database and data warehouse. Data reversible hiding technology is limited applied to single file or single database. Computer forensic technology only can be used to those data resources which electronic evidence is fixed. As to block chain technology, the data resources needed to be recorded as block chain. These techniques have something in common that they all related to "The right of management information" and standardized technology [3]. Study on the standard of data provenance technology is of some significance to promote the development of data provenance technology. E.g. PROV is approved by W3C. OAIS which is put forward by CCSD [4] and released by ISO [5] and so on. In today's world, more and more data confirming company emerged in china. Such as Cunnar, Atlasdata and Ebaoquan and so on. They all provide data confirming services. By comparing these companies, it is found that there is no uniform technical standard for indexing data. It may lead to confusion and error while indexing data. Therefore, there is an urgent need to build up the standard of management which related to data property. Only by using these techniques which are shown in Table 2 and Smart contract technology comprehensively, can we found a scientific and complete solution of confirming the whole category data resources and covering their whole life.

#### 5. SOLUTION OF CONFIRMING BIG DATA PROPERTY RIGHTS BY SMART CONTRACT

According to above study, we propose that confirming big data property rights can be divided into two parts: one is based on quality and quantity of data and another is based on property value and its distribution. The former is about auditing data contents, analysis data format and calculating the capacity of big data. The latter is aimed at evaluating the value of the data asset and indexing it by scientific model which based on the former work. People in this market are trade on the Internet automatically. So, this market is changing every minute. The traditional way of confirming big data property no longer meets the practical needs of the increasingly developed big data trading market. We must use modern intelligent technology to cope with dynamic data, which requires comprehensive application of machine learning or deep learning technology and Smart contract technology.

Smart contract is a calculable transaction agreement to execute the terms of the contract by computer language instead of legal language. Essentially, it is a computer program that can automatically execute the terms of a contract. When a set of conditions is coded in advance are triggered, the smart contract will execute the corresponding contract terms by the protocol and user interface [6]. This gives us the idea that we can use smart contracts to implement the data confirmation process as shown in figure 1. We designed a specialized classifier for confirming big data property rights to implement this process that works as shown in figure 2.



**Figure 2. The specialized classifier for confirming big data property rights.**

According to the working principle of this classifier in Fig.2, we put forward the smart contract SC1 which is aimed at confirming quality and quantity of big data property rights as shown in Table 3. Table 4 is the design philosophy of smart contract SC2 which is aimed at confirming value of big data property and its distribution which is based on the business process and working principle as shown in figure 1 and figure 2.

**Table 3. The design logic of Smart Contract SC1**

Conditions	Trigger conditions and execution results
1: Responsible person has been identified and the data has been desensitized	A: If both condition 1 and condition 2 are satisfied, then Z1 is responsible to {data set D}. Automatically identify as Z1 <sub>R</sub> and write this block in block chain
2: $RZ1/\Sigma R (Z1, Z2, Z3, \dots, Zn) > 90\%$	B: If both condition 1 and condition 3 are satisfied, then Z1 is accompanying responsible to {data set D}. Automatically identify as Z1 <sub>A</sub> and write this block in block chain
3: $90\% > RZ1/\Sigma R (Z1, Z2, Z3, \dots, Zn) > 50\%$	C: If condition 1 and condition 4 are satisfied, then Z1 is non-responsible to {data set D}. Automatically identify as Z1 <sub>N</sub> and write this block in block chain
4: $RZ1/\Sigma R (Z1, Z2, Z3, \dots, Zn) < 50\%$	

The sample code snippet of smart contract SC1 which described in Table 3 is showed as follow. Similarly, we can design the programme codes of Smart contract SC2.

**Table 4. The design logic of Smart Contract SC2**

Conditions	Primary trigger condition and execution result	Secondary trigger condition and execution result
1: {D} has been indexed as Z1 <sub>R</sub> automatically	A: If both conditions 1 and 4 are satisfied, {D} is tradable data for Z1. Automatically identify as Z1KD and write this block including information of rights in block chain.	a: In case A, the model Mi in condition 6 is invoked to calculate the property value of Z1KD and write this block in block chain.
2: {D} has been indexed as Z1 <sub>A</sub> automatically	B: If conditions 1 and 5 are met at the same time, or conditions 2 and 5 are met at the same time, or conditions 3 and 4 are met at the same time, then {D} is the illegally traded asset of Z1, marked as Z1FD and written into the block chain	b: In case B, terminate the Z1FD value evaluation
3: {D} has been indexed as Z1 <sub>N</sub> automatically	C: If conditions 2 and 4 are met at the same time, {D} is the permitted transaction data asset of Z1 after recognized by (Z2,Z3...Zn) and can be identified as Z1XD. Otherwise, the ownership status of {D} is transformed from Z1A to Z1R and remove the data owned by (Z2,Z3...Zn).	c: In case C, either recognized by (Z2,Z3...Zn) or delete the the data owned by (Z2,Z3...Zn) so that he ownership status of {D} can be transformed from Z1A to Z1R. Then, we can use the model Mi in condition 6 to calculate the property value of Z1XD or Z1ZD and write this process in the block chain.
4: The provider of {D} is Z1		
5: The provider of {D} is not Z1		
6: Big data property rights valuation model set {Mi,(i=1,2,3,...,n)}	Meanwhile,,this data set can be marked as Z1ZD and written into the block chain according to the block chain record rule.	

```

1 function getSituation(desensitized, rightRatio){
2   if (desensitized && rightRatio >= 0.9){
3     return "ZR";
4   }
5   else if (desensitized && rightRatio < 0.9 && rightRatio >= 0.5){
6     return "ZA";
7   }
8   else if (desensitized && rightRatio < 0.5){
9     return "ZN";
10  }
11  else {
12    return "";
13  }
14 };
15 var desensitized = checkDesensitization(data);
16 var rightRatio = getRightRatio(data, author);
17 var situation = getSituation(desensitized, rightRatio);
18 switch (situation)
19 {
20   case "ZR":
21     ZR(data, author); //claim "author" owns the property right of "data", and broadcast.
22     break;
23   case "ZA":
24     ZA(data, author); //claim "author" partially owns the property right of "data", and broadcast.
25     break;
26   case "ZN":
27     ZN(data, author); //claim "author" does NOT own the property right of "data", and broadcast.
28     break;
29   default:
30     break;
31 }

```

**Figure 3. The sample code snippet of smart contract SC1.**

## 6. THE MECHANISM OF CONFIRMING BIG DATA PROPERTY RIGHTS BASED ON SMART CONTRACT

Smart contract technology can not only be applied to confirming big data property rights, but also in big data trading process. In this regard, scholars had put forward many valuable solutions. Hua Jie did some research about applying block chain smart contract to intellectual property transactions and their legal obstacles [7]. Sheng Nianzu put forward an evaluated model of data asset which is conducted by third party based on block chain smart contract [8]. Cao Yang did some research about solving the problem about online use of music works by smart contract [9]. Quan Lixin proposed a dual-block chain model to solve the market circulation problem of digital educational resources [10]. Du Zhenhua put forward the idea of the construction of data property right system form data market [11]. Yan Shu discussed the application of data circulation based on block chain smart contract form overseas [12]. These studies indicate that the application of block chain smart contract in big data trading is getting more and more mature. However, there are many problems left here. Such as problems about the construction of the unified big data chain, smart contract library and its online platform, the legal effect of smart contract, the disclosure mechanism for this market and the standard system of right big data property management, and so on. These are all important components of big data property confirming system in big data Exchange. They are all important practical problems that need to be solved urgently in the development of big data industrial. In order to figure out this, we should spare no effort to accelerate the pace of smart contract research.

## 7. CONCLUSION

To conclude, there is a great deal of random disorder in our data exchange market. This disorder shows in trading patterns, trading rules, data property confirmation, data value evaluation model and pricing methods, and so on. Our research attempts to build an integrated confirming solution that access to all categories of data with entire data life-circle based on smart contract technology and other techniques. Therefore, we proposed a logical design scheme of big data smart contract for confirming big data property rights, which is expected to be of some significance to the construction of the big data exchange market in our country.

## 8. REFERENCES

- [1] Haijun Zhao, J. 2012 . Classification of Information Resources based on property right and Construction of Information confirmation right Theory. *Library & Information*, (04):89-97.
- [2] Susu Cheng, Haijun Zhao, A. 2018. An overview of techniques for Confirming Big Data Property Rights. In *Proceedings of ICIT2018[C]*. New York, ACM, 59-64.
- [3] Yu Yang, J. 2010. Research On the significance of right Management Information Standardization. *Modern economic information*, (08) :47.
- [4] [4] Moreau L, Groth P, J. 2013. PROV-Overview: An overview of the PROV family of documents. *PW3C Note*, (4): 1-9.
- [5] CCSDS Secretariat, Program Integration Division (Code M-), National Aeronautics and Space Administration. 2002. Reference Model for an Open Archival Information System (OAIS) [M]. Washington, DC, USA. CCSDS Secretariat.

- [6] Hu Kai, Bai Xiaomin, J. 2016. Formal Verification method of Intelligent contract [J]. *Research on Information Security*, (12):1080-1089.
- [7] Hua Jie, J. 2018. Application of Block chain Technology and Intelligent contract in intellectual property right confirmation and transaction and its legal Regulation. *intellectual property Rights*, (02):13-19.
- [8] Sheng Nianzu, Li Fang, J. 2018. A method of data Asset based on Block chain Intelligent contract in Internet of things, *Journal of Zhejiang University*, (11):1-9.
- [9] Cao Yang, Bo Jun,J. 2018. Block chain Technology and copyright Protection of Internet Music works. *South China Sea Jurisprudence*, (3):77-87.
- [10] Quan lixin, Xiong Qian, J. 2018. Application of Block chain Technology in the Circulation of Digital Education Resources. *Audio-visual Education*, (08):78-84.
- [11] Du Zhenhua, Cha Hongwang, J. 2016. Realistic consideration of data property right system. *Chongqing Social Sciences*, (08):19-25.
- [12] Yan Shu, Qing Sude, J. 2018Application of Block chain in data Circulation. *big data*, (01): 3-12.