

## Research on Scientific Data Management and Sharing in the Environment of Big Data

Ningning Liu

Department of Economics and Management, Xidian University, Xi'an, China

15771906603@163.com

---

*Abstract: In today's environment of big data, data has become an indispensable resource, and scientific data are important resources for the sustainable development of science in the information age. This paper provides a reference for the sharing of scientific data in China by discussing the necessity, the challenge and the related problems of scientific data sharing in big data environment.*

*Keywords: big data, scientific data sharing, challenge, Countermeasure*

---

### 1. INTRODUCTION

With the rapid development of computer information technology, we have ushered in the data explosion growth of big data, data has become one of the essential elements of various fields, and scientific data has become an important scientific and sustainable development of the important social resources. The development of a country depends largely on technology, and scientific data management and innovation is an important way to accelerate technological development and technological innovation. Therefore, in the background of big data, how to achieve scientific data collation, storage, sharing and service has become an important social issue of the times. This paper studies the necessity and challenge of scientific data sharing in the big data environment, and puts forward some countermeasures to promote the sharing of scientific data in order to promote the sharing of scientific data and its value – added in China.

### 2. OVERVIEW

#### 2.1 The concept of big data

With the development of e-commerce and mobile Internet, a comprehensive access to a large-scale production, sharing and application, data information age, big data age has come, the importance of big data has been recognized in various fields, but there is no uniform definition. McKinsey defines big data in the Global Research Report "Big Data: Innovation,

Competition and Productivity in the Next Front": big data is a data group that exceeds the size of traditional database software tools for crawling, storing, managing, and analyzing capabilities <sup>[1]</sup>. The Big data has a "4V" feature, Volume, Variety, Velocity, and the most important Value. Volume is a huge amount of data and data integrity, the amount of data has been developed from TB to PB and ZB, can be described as massive, huge and even excessive. Variety means that the data type is numerous, more and more performance for the web, pictures, video, images and other semi-structured and unstructured data information. Velocity refers to the high speed of data transmission and fast, faster to meet real-time needs. Value is the ultimate meaning of big data - to meet the needs of mankind, in the shortest possible time to tap its value.

"Big data" of the "big", not only lies in the big data capacity, but also through the massive data exchange, integration and analysis, discover new knowledge, create new value, bring "big knowledge" "big technology" "Big development"<sup>[2]</sup>.

## **2.2 The concept of scientific data**

Scientific data is the basic data of the nature, characteristics and changes of the objective world of the human society science and technology activities, and the data products and related information systematically processed according to the different needs <sup>[3]</sup>. Scientific data covers a number of disciplines, is the information age science and sustainable development of important resources in the government decision-making, technological innovation, social and economic development and many other aspects of a significant role. Scientific data is not only the pre-scientific research results, but also to further develop the basis of scientific research, scientific data sharing in the knowledge innovation system in the strategic role has become a consensus.

China's scientific data sharing started late, the Ministry of Science and Technology in 2002 to open the scientific data sharing project, in the full policy support and financial protection under development so far, basically formed a wide coverage, reasonable structure of the national scientific data sharing platform system, into a good The operation of the service phase <sup>[4]</sup>.

## **3. THE NECESSITY OF SCIENTIFIC DATA SHARING IN THE BACKGROUND OF BIG DATA**

### **3.1 National development strategy needs**

The arrival of big data makes the data become an important asset of the information age, and it is becoming an important factor of production compared with material assets and human capital. The strategic resource of scientific data is becoming an important asset of society. The development of a country and enterprise depends largely on its level of technological

innovation and technological innovation, and the management of scientific data is an important way to realize scientific and technological progress and innovation. It is an important way to support the long-term sustainable development of national innovation technology. Guarantee. Whether it is resources, energy development, or high-tech industrialization, are in the accumulation of scientific data and support to achieve the results of theoretical and technological innovation. In the era of knowledge economy today, more and more knowledge products in a variety of ways to promote the rapid economic development, knowledge and information for the purpose of information management, processing and publishing to become a rapidly developing industry, in the era of big data to promote scientific and technological progress And economic development, that is, the so-called "big data economy".

For a long time, China has accumulated a wealth of scientific data resources, but most do not have effective management and utilization, digital level is relatively low, many databases are often confined to the department or the use of the unit, or even personal use, resulting in scientific and technological resources A huge waste. Therefore, breaking the scientific data barriers, the implementation of scientific data sharing, is the inevitable requirement of national development strategy, to enhance China's overall level of science and technology, improve scientific and technological innovation, promote social and economic development is important.

### **3.2 Scientific research and academic exchange needs**

Scientific research process with the emergence of scientific data, both the results and accumulation, but also the late scientific data research indispensable resource base. Especially in the era of big data, the amount of scientific data surge, scientific research more and more dependent on the system, high credibility of the scientific data analysis, disciplines of the major breakthroughs and innovations are mostly multidisciplinary cross, the integration of the results [5]. With the development of information technology and science, the current discipline is no longer a single subject, interdisciplinary is the inevitable development, so cross-regional, interdisciplinary cooperation and resource exchange become the inevitable trend of scientific research.

A series of major research projects and the rise of complex scientific issues raised, ushered in an unprecedented scientific research cooperation and academic exchanges. To achieve scientific data management and sharing, scientific research can be free from regional and academic constraints to obtain large amounts of scientific data resources, and in the massive data to explore the potential of innovation, making scientific data in the sharing and application of value-added, to enhance scientific research efficiency, Promote scientific and technological innovation.

### **3.3 Enterprise and public demand**

Under the vigorous development of big data, the demand for scientific data is not only the demand for scientific data, such as government, higher education institutions, scientific research institutes, enterprises and the public. Scientific data are no longer the field of knowledge of professional data personnel. The development of innovative enterprises depends on scientific and technological innovation, especially for small and medium-sized innovative enterprises, to achieve scientific data sharing and use can effectively reduce the cost of scientific and technological innovation, promote small and medium enterprises of technological innovation and enterprise development, so as to effectively improve social and economic development.

Scientific data for the public, not only can improve their own knowledge and scientific literacy, but also indispensable daily life important information resources. With the development of mobile Internet and the popularity of personal computers, smart phones and other intelligent devices, the public demand for scientific data is becoming stronger and stronger after the Internet goes to millions of households. For example, in the shopping site, the platform based on the user's buying habits and browsing records, predict the user needs for the user personalized recommendation. The driver through the mobile phone loading and positioning system to determine the traffic conditions and driving routes, but also can use big data to predict congestion warning, congestion and parking and distribution and other information. In the field of health care, data analysis of large-scale complex data helps to identify disease causes more effectively and provide targeted prevention, diagnosis and treatment. It can be seen that in the era of big data, the demand of scientific data from non-professionals is becoming more and more important, and the requirements of scientific data release, data form and data quality will be higher and higher.

## **4. CHALLENGES OF SCIENTIFIC DATA SHARING IN THE ENVIRONMENT OF BIG DATA**

### **4.1 Scientific data sharing laws and regulations system is not perfect**

Europe and the United States and other developed countries, government departments have developed a corresponding scientific data laws and regulations, the preservation and management of scientific data to establish a clear specification. The Data Protection Act 1998 and the Freedom of Information Act 2000 and the Environmental Information Regulations also provide for public access to the scientific data generated by the public institutions of the United Kingdom shall be responded within 20 working days unless the exemption or exception is made <sup>[6]</sup>. The basic laws of the People's Republic of China in the field of science and technology, the People's Republic of China, promote the transformation of scientific and technological achievements Law "does not make specific provisions on the sharing of

scientific data, they establish the legal principles can only play a guiding role in scientific data sharing activities from the administrative level, the " People's Republic of China Government Information Disclosure Ordinance "on the scientific data Management has produced a normative role, but its provisions do not explicitly mention scientific data.

Therefore, it is an inevitable requirement for scientific research and social development to formulate laws and regulations that regulate the management and sharing of scientific data or revise the relevant provisions in the original relevant laws and regulations.

#### **4.2 Challenges for the integration of scientific data**

In the era of big data, massive amounts of data are stored in a variety of forms. Scientific research needs to use the scientific data need to find a variety of ways to access the data, the data may also be pictures, video and other unstructured data types, these scientific data for a number of research institutions, universities or enterprises, is internal data, even if Some of the data is open, it may be difficult because of the data format and other issues. Therefore, the integration of scientific data is imminent, through the feasible integration mechanism, the valuable scientific data unified display and use, build a scientific data sharing platform, the data sharing platform as a carrier to achieve data integration.

Big data background in the integration of scientific data is now facing the main challenges are: (1) To build a scientific data sharing platform network protocol, remote implementation of the program and so on, heterogeneous platform system how to communicate effectively and contact. (2) The diversity of scientific data storage, text, video, pictures and relational data, how to use a unified data syntax for data description. (3) How to establish effective metadata standards. (4) Scientific research process may need different data sources, how will the heterogeneous data semantic association. (5) The current need for big data algorithm for virtual integration, in order to improve the effectiveness of integration.

#### **4.3 Challenges faced by scientific data sharing**

Scientific data sharing is the primary consideration of intellectual property issues, scientific data is produced in the research process of creative labor, is the wisdom of scientific research personnel crystallization of scientific data access to the need for scientific research personnel to pay in many ways, but also from the unit And related platform help, so scientific research data is an important asset of scientific research personnel and units, and its intellectual property rights should be protected.

The establishment of data sharing incentive mechanism to achieve the scientific data of big data integration, scientific and extensive data, long-term sharing is the basis of intellectual property system. Therefore, how to achieve scientific data sharing on the basis of government intervention in the protection of intellectual property rights is another challenge for scientific data sharing. This requires government intervention, through the encouragement and support

of laws and regulations, clear scientific data and metadata property rights issues, to fully protect the data provider's intellectual property, to stimulate and promote data owners to provide data spontaneously, can't damage the interests of researchers, Nor does it make intellectual property issues an obstacle to the sharing of scientific data.

## **5. COUNTERMEASURE OF PROMOTING SCIENTIFIC DATA SHARING IN BIG ENVIRONMENT**

### **5.1 To strengthen policy support and promote government to open and share scientific data**

Scientific data management in China's scientific and technological development in the basic and guiding strategic position can't be ignored, only the scientific data management work to rise and integration into the national strategy, capital investment, policies and regulations, technical standards, the establishment of professional personnel, Organizational security and other aspects in order to obtain the national level system arrangements, overall planning. Therefore, the scientific data management and sharing has yet to be the highest level of national decision-making from the national strategic level of development planning and organization, collaborative multi-sectoral, multi-disciplinary, multi-unit to promote.

Government scientific data to master and control a large number of scientific data, should take the initiative to share these scientific data. Developed countries usually put the scientific data held by the government into the scope of information disclosure, the development of sound laws and regulations to promote the disclosure of government information, the United States and Britain and other countries have opened the government open data portal, and provide "one-stop" data services. China should learn from the experience of government information disclosure in developed countries such as Europe and the United States, and gradually improve the relevant laws and regulations of government information disclosure, make specific provisions on the disclosure and sharing of scientific data, so that the scientific data can be transparent; "One-stop" data integration services website, encourage enterprises, individuals to use the government to open up the data to create new application value, and can become the government departments to open and share the power of scientific data.

### **5.2 To explore China's scientific data sharing development mode**

Scientific data has great value-added potential, and its value-added use has long been explored in the international community, then we need to explore suitable for China's scientific data sharing model. Scientific data are divided into strategic scientific data, public welfare scientific data and commercial scientific data according to basic social attributes. The corresponding data model is: the state constraints of the free share of the model and the market constraints of the industrial mode of operation <sup>[7]</sup>.

At present, scientific data such as government departments, data centers and other public departments are mainly concentrated in the fields of land, geography, meteorology, transportation, public health and agriculture. These data are generally produced by national investment. Therefore, while meeting their own business needs, we should vigorously develop value-added use services, and we must adhere to the principle of public welfare as the primary principle, to protect the public sector to lead the welfare of all citizens of the scientific data of value-added use. In order to ensure the sustainability of the value-added use of scientific data in the public sector, it is necessary to explore the service mechanism of market-oriented operation. China can learn from Europe and the United States and other developed countries, the successful use of scientific data value-added experience to encourage the private sector or individuals involved in value-added development to meet the diversity of social needs. For profit-making institutions or personal investment generated by the scientific data, in its value-added development can be implemented in commercial operation, to explore the paid service model, cultivate the relevant scientific data products market and industry.

### **5.3 Establish the operational mechanism of scientific data sharing**

Scientific data sharing embodies the concept of cooperation, openness and sharing in the information age. It is an interdisciplinary, cross-cutting and large amount of data sharing and cooperation. Under the environment of big data, we should establish the main function of innovation and management, and adhere to the government as the guide, the technology intermediary service as the data manager, play the core role of the data center shared service operation mechanism.

Scientific data resources scattered to colleges and universities, enterprises, research institutions as the main elements of innovation as the core elements, the data resources are not balanced, mutual data sharing is not smooth, there are serious "data barriers." China's universities and research institutions produce a lot of data and knowledge each year, but because of the lack of attention to the integration of knowledge information, transfer, sharing and commercialization, the conversion rate is very low. Therefore, it is necessary to use the scientific data sharing platform as a link to the information platform of the production and research, to build collaborative information network, to establish long-term, stable and lasting cooperative relations, to strengthen the sharing of resources, information communication, effectively break the data barriers, , The division of scientific data resources to achieve sharing, communication and integration, enhance the results of the transformation of vitality, making the scientific data in the process of sharing and commercialization to value.

The establishment of data sharing incentive mechanism, a clear intellectual property issues, to achieve regional data integration of big data. Only the government involved in the identification of scientific data and metadata property rights issues, to fully protect the data provider's intellectual property, in order to stimulate and promote data owners to provide data spontaneously. Under the initiative of the government, the data platform as the main carrier of

data service should give full play to its core role, while protecting the data owner's intellectual property rights, to ensure that users can obtain more useful scientific data. So the data center needs to participate in the data service process to help data providers and data users more efficiently and more quickly to achieve scientific data sharing, but also by tracking the data service process to analyze and evaluate the sharing of results, and ultimately better to achieve big data and Large scientific integration, and comprehensively promote the sharing and application of scientific data.

#### **5.4 Training of relevant personnel in scientific data management**

As a basic, strategic, long-term scientific data sharing work, the need to build a stable professional and technical personnel. Foreign countries attach great importance to the cultivation of such talents, has successfully implemented a number of training projects, such as the University of Illinois Library and Information Science Graduate School of data monitoring education program, the United States Syracuse University Library "scientific data literacy" project. And China's current library and information science curriculum has not yet seen with the scientific data management-related courses, scientific data management expertise is the urgent task of training<sup>[8]</sup>.

To carry out scientific data sharing personnel training. Training methods are: (1) Colleges cultivate professional talents. Through the National Science and Technology and professional institutions of higher learning to establish long-term cooperation mechanism in the curriculum, training methods and employment direction and other aspects of the agreement, so that the combination of theoretical training and quality training for scientific data sharing work to provide human resources protection.(2)To conduct on-the-job training. Constantly improve the skills of science and technology resource management personnel, standardize the industry qualification management, regular scientific data sharing management practitioners internal training, the training effect and title assessment, incentive mechanism, personnel selection and assessment linked to comprehensively improve the quality of practitioners.(3) International Talent Cooperation and Training Exchange Mechanism. By the relevant departments of science and technology cooperation and other relevant departments, and other countries to establish scientific and technological resources management personnel training mechanism, through cooperation in training, exchange students and participate in international science and technology projects, etc., the establishment of China's scientific and technological resources management personnel international cooperation and exchange mechanism.

#### **REFERENCES**

- [1] Wu Hequan. Opportunities and Challenges in the Big Data Age [J]. Information Security and Communications Privacy, 2013 (7): 9-10.
- [2] Tu Zipei. Big data: the coming data revolution [M]. Guangxi: Guangxi Normal University

- Press, 2012.
- [3] Huang Dingcheng, Guo Zengyan. Scientific data sharing management research [M] Beijing: China Science and Technology Press, 2002.
- [4] Zhang Xinxing. Construction and service of scientific data sharing platform [J]. Modern Information, 2016, (11): 109-113.
- [5] WANG Liangxu, WU Lizong, NAN Zhuojian. A comprehensive scientific data sharing system based on B2C architecture [J]. Remote Sensing Technology and Application, 2013, 28 (3): 355-361.
- [6] Liu Runda, Lin Haiqing, Zhao Yanping. Study on the status quo of geo-shared data resources based on metadata analysis [J]. 2009, 34 (4): 80-85.
- [7] Zuo Jian'an, Chen Ya. Study on Scientific Data Sharing Model Based on Large Data Environment [J]. Journal of Information Technology, 2013 (12): 151-154.
- [8] Man Rui, Wang Jian. Opening and sharing of scientific data metadata in large data age [J]. Modern Information, 2016, 36 (3): 38-41.