Exploration and Analysis of University Scientific Research Data Management Strategies under Big Data Environment

Wei Wang*, Xinhua Zhang, Yabin Shi, Feng Gao and Wenpeng Xu
Education and Research Support Center, Dalian Naval Academy of the Navy, Dalian, China

*Corresponding author e-mail: xiyang@dlust.edu.cn

Abstract. From the perspective of the current big data environment, briefly describe the urgency of university scientific research data management, and use this as a basis to discuss the university scientific research data management strategy under the big data environment, and propose to conduct sufficient research on scientific research data management and pay attention to scientific research data Strategies such as dynamic management, strengthening the storage and security management of scientific research data, and enhancing the value of scientific research data results through publication and citation.

Keywords: Big Data, Scientific Research Data, Scientific Research Data Management

With the development of information technologies such as big data, Internet of Things, mobile Internet, and cloud computing, the ways of academic use and dissemination are constantly changing. A large number of native digital scientific research materials are constantly being produced. Research Data Management (RDM) and Research Data Services (RDS) has become an important task faced by scientific research institutions. This article is mainly based on the urgency of universities for scientific research data management under the big data environment, and proposes strategies and methods for universities to manage scientific research data, in order to provide reference and reference for universities to build scientific research data management service innovation systems, thereby promoting scientific research data management and Development of shared services.

1. University Urgently Needs Scientific Research Data Management in the Era of Big Data
In the big data environment, scientific research data management services based on massive data make full use of big data technology and big data thinking to collect, manage and understand scientific research data, and provide a full range of management services for scientific research activities. The Chinese government attaches great importance to the management and sharing of scientific research data, and has successively formulated a series of policies and plans. In January 2018, the second meeting of the Central Leading Group for Comprehensively Deepening Reform deliberated and approved the "Scientific Data Management Measures", which was formally issued by the State Council in April 2018 [1], which provided guidance for the formulation of my country’s scientific data
management and sharing policies. Macro guidance.

Compared with other social institutions, colleges and universities have a relatively stable organizational structure and operation methods. However, because colleges and universities have both educational and scientific research tasks, they often have scientific research projects and discipline development faults due to the flow of graduate students and other scientific and technological talents. At the same time, there are certain problems in the transfer of scientific research work and data due to graduate students leaving school and the work transfer of scientific research backbones. Therefore, in the face of heterogeneously distributed massive amounts of information in the big data environment, and at the same time to ensure the sustainable and stable development of scientific research projects, universities should establish a scientific and effective scientific research management mechanism to ensure the orderly development of scientific research activities and at the same time attach importance to scientific research Data storage and management.

The development of scientific research activities not only requires a lot of manpower costs, but also requires adequate financial support. At present, most scientific research projects in colleges and universities are inseparable from funding from schools, social organizations and the state finance. Therefore, the management of scientific research data can not only allow the process of scientific research activities to be fully recorded, but also reflect the effectiveness of scientific research workers to a certain extent. Secondly, scientific research data management will display the researchers’ scientific research plans, timetables, data standards, sharing conditions, etc., which will help funding agencies to effectively carry out preliminary evaluation and process control, and at the same time facilitate the formulation of acceptance criteria, Promote the development of scientific research activities.

2. University Research Data Management Strategy

2.1. Carry out Adequate Research

The data platform is the basic guarantee and support for the development of scientific research data management services. In the early stage of construction, it can use the literature research method, questionnaire method, Delphi method, life cycle method, etc. to conduct research, and guide the specific implementation and management. In the initial stage of construction, the first choice is to use the Internet, library database resources and other channels to retrieve relevant literature and academic results through the literature research method, repeat research, clarify the research progress and methods at home and abroad, and build a research framework for thematic research data integration and platform construction.

In the early stage of construction, two methods, questionnaire survey method and Delphi method, can be used to determine specific issues such as resource integration requirements, implementation plans, and features. At the same time, the readers who use the data platform are surveyed and their opinions are collected. Experts in the industry are invited to evaluate and test the database, and make appropriate adjustments in the middle and later stages of the project to ensure that the project meets the needs of readers and users.

2.2. Pay Attention to the Dynamic Management of Scientific Research Data

The life cycle of scientific research data is the management of data as a red line throughout the entire life process of the scientific research process, providing clear planning and important decisions for data users and data managers. Using the scientific research data life cycle method to manage the scientific research data platform can help platform managers and users to obtain the maximum value at the lowest cost at all stages of the scientific research data life cycle. Some scholars divide the life cycle of scientific research data management into five stages, namely data access, data organization, data storage, data sharing, and data security.

Serial number stage content
1 Data access online and offline
2 Data organization, Data application, data standard, metadata standard, data indexing, etc.
3 Data storage, Data ownership, data archiving, data deletion, data destruction, data custody, data backup, data review, data restoration, etc.
4 Data sharing, data transfer, data transfer, etc.
5 Data security, intellectual property rights, data ownership, etc.

2.3. Strengthen the Collection and Processing Of Scientific Research Data and Product Services
During the construction of the scientific research data sharing and management platform, it is first necessary to plan and construct a scientific and reasonable plan and implement it in accordance with the standardized process. The digitization of various paper documents and the collection of electronic documents are the main tasks of platform construction. In addition, it also includes in-depth processing of various types of information data for a subject subject, providing more accurate and scientific knowledge services, and various databases with unique content jointly developed for the same goal. The purpose of the construction of the scientific research data platform is to provide resources for users to provide services, in order to help users make better use of resources to provide in-depth knowledge services for learning and scientific research. Efforts are made to establish associations between various knowledge points [7] to help users discover knowledge more quickly, dig deeper into the information of famous scholars or research hotspots in the field, etc., and refer to the professional thesaurus to build the thesaurus and subject the metadata Index and cluster by subject.

2.4. Pay Attention to the Storage and Safety Management of Scientific Research Data
Due to the dynamic nature of scientific data, from the beginning of scientific research to the submission of final scientific data, a large amount of intermediate data will be generated during this process. This part of the data is temporary and dynamic, and it is possible for scientific research. As a guarantee of continuous progress, some data may have long-term preservation value after evaluation after the end of scientific research. Therefore, this part of the data needs to be maintained and archived temporarily. After the entire scientific research process is over, part of the scientific data needs to be stored for a long time as the final scientific research results. Moreover, these data need to be maintained by the corresponding subject during the archiving and long-term preservation process and to ensure that they can be discovered and obtained [4].

According to my country’s "Scientific Data Management Measures" issued by the General Office of the State Council in 2018, the hierarchical construction of scientific data centers will become an important carrier for promoting the open sharing of scientific research data [8]. It should be equipped with data storage, management, service and Necessary measures such as security, establish and improve the preservation system of scientific data, so as to ensure the safety and integrity of scientific research data storage [1]. The sharing and management of scientific research data can be achieved by setting different resource access permissions for different types of users. For a data topic, category, and page, the administrator can access the user group of the resource library, and all users of the user group have the authority to use the resource library, category, and page; you can also set the IP address to access the resource library Segment, all users in the address range have the right to use the resource library, category, and page. During the trial operation and maintenance phase of the platform, the teachers and students in the school can be conditionally open to use through registration, login, identity authentication, etc., and the staff is responsible for system security management. Problems were found through trial operation, and the contents and functions of the database were adjusted and improved. Focus on strengthening the various cross-database retrieval and faceted retrieval functions of the database, and establish a unified long-term mechanism for metadata harvesting and object data acquisition, in order to achieve a long-term sustainable development mechanism based on self-learning and self-organizing platforms.

2.5. Enhance the Value of Scientific Research Data Results through Publication and Citation
The publication and citation of scientific research data is an important way to realize the value of data.
The publication of scientific research data means that the data generated in scientific research activities has a relatively fixed form and minimizes its possible deviations, and finally realizes the sharing of scientific research labor results. The published scientific research data will also be further verified and quoted by researchers from all parties, so as to maximize the value of the data [9], and can effectively reduce false data and erroneous data.

(1) Pay attention to the scientific research data management plan, and require scientific research institutions and researchers to submit the data management plan of this scientific research project when applying for publicly funded scientific research projects to ensure the effective development of subsequent scientific research data management;

(2) In accordance with the relevant national laws, regulations and policies and combining with the actual situation of the university, it is the basic requirement of the university's scientific research data management policy to establish as detailed as possible, safe and operable data retention, data disposal and data storage systems;

(3) Data sharing shall comply with the requirements of laws, regulations and ethics, and establish a confidentiality and sharing system hierarchically and categorically, respecting the labor results of scientific research institutions and researchers, keeping scientific research secrets and maintaining national security, and effectively using scientific research data. The life cycle of the company guarantees the right to know and fair use of other researchers and the public. Hierarchical classification not only needs to classify data according to discipline and confidentiality level, but also needs to classify and manage the shared objects and the degree of sharing. Sharing objects include universities, project team members, off-campus collaborators or cooperative institutions, all researchers of the school, corporate institutions or the public in need, the public, etc.; the degree of sharing such as restricted access, mediated access, and open access;

(4) Clearly define the specific responsibilities of various departments and staff of universities in scientific research data management, give full play to the roles of core leaders and project scientific researchers, so that the various systems of scientific research data management are truly implemented;

(5) Provide complete technical standards, business specifications and process guidelines and other technical conditions, such as data management plan templates, the recommended list of data warehousing or off-campus data warehousing by the institution, metadata standards, data sharing license agreement templates, data acquisition declaration templates, etc. [10];

(6) Paying attention to the cost of scientific research data management, respecting the labor contribution of scientific researchers in scientific research data management, and providing corresponding funding support are necessary conditions for the effective operation of scientific research data management

3. Conclusion
The collection, storage and management of scientific research data has become an important task of scientific research management in universities. The educational level of colleges and universities reflects the ability of academic "reproduction", while the level of scientific research can show the academic height of the instructor and the educated. The formulation of scientific research data management policies is an urgent requirement for implementing the National Scientific Data Management Measures and strengthening the management of scientific research data in universities. At the same time, universities formulating research data management policies can accumulate valuable experience for future research data management legislation. The sharing and management of scientific research data is a long-term and continuous work in the scientific research management of universities. Constantly enriching resources and improving the system will provide sufficient motivation for universities to carry out long-term scientific research activities.

References
[1] The Central People’s Government of the People’s Republic of China! Notice of the General Office of the State Council on Issuing the Measures for the Administration of Scientific Data.
[2] Ding Pei. Research on the Management Policy of Scientific Research Data in Foreign Universities. Library Forum, 2014(5): 103-110.

[3] Sun Jianjun, Ke Qing. On the construction of the national digital information resource strategic system. Journal of the Library Science in China, 2007(5):81-86.

[4] Fan Junhao. Research on the role of libraries in scientific data management. Library and Information Service, 2014(6): 38-42.

[5] Juanle Wang, Kun Bu, Yanjie Wang, Yating Shao. Progress in Activities of WDS-China Data Centers. Data Science Journal, 2020, 19(1).

[6] Ren Yazhong. The Feasibility Study of University Library Participating in Data Governance. Journal of Library Science, 2020, 42(07): 14-17.

[7] Yang Yang, Yang Ye. Research on Knowledge Management and Collaborative Innovation Platform Construction in Scientific Research Institutions. Information Science, 2020, 38(09): 101-106.

[8] Liu Wei, Liang Bing, Qu Baoqiang. Research on service model based on technology management data [J/OL]. Information theory and practice: 1-10 [2020-10-25]. http://kns.cnki.net/kcms/detail/11.1762.G3.20200728.1553.006.html.

[9] Yin Huaiqiong, Xiong Yongjun, Liu Haixia. Analysis and Enlightenment of Scientific Research Data Management Service of University Libraries at Home and Abroad. Research in Library Science, 2020(11): 33-41.

[10] Liu Guifeng, Qian Jinlin, Zhang Jiyong. Research on the Model Construction of Scientific Research Data Governance in Chinese Universities. Information Science, 2020, 38(04): 28-36.