

# Research on Data Monitoring of University Library Driven by Big Data

Xin Huang

Weifang University Library, Weifang 261061, China

723607079@QQ.com

**Abstract.** Data guardianship and big data will inevitably have an impact on University libraries; foreign scholars in the field of Library and information actively carry out research and achieve great results, while our country does not attach enough importance to it. Big data and data guardianship are mutually reinforcing. Big data promotes the data guardianship of University libraries, mainly in the following aspects: university libraries are the important subjects of data guardianship in the era of big data; big data promotes the research of data guardianship in University libraries; big data promotes the data supervision of University Libraries Protect talents.

**Keywords:** Data; management of big data; University Libraries.

## 1. Introduction

With the rapid development of modern information technology, communication technology and E-Science, a large number of scientific data have been generated. The role of scientific data is gradually being valued, especially in the long-term preservation, maintenance, management and secondary development and utilization of scientific data. Therefore, data curation and Big Data have emerged as new research fields in recent two years. Foreign libraries (especially American University libraries) have been fully aware of the opportunities and challenges brought by data monitoring and big data, and have tried to expand the service boundary of libraries to a new stage of scientific research and academic exchanges - the data stage, while Chinese libraries seem to lag behind in this respect. As for the meaning of data guardianship, the author quotes the definition of the Joint Information Systems Committee (JISC): "Data guardianship is an activity that manages and improves data from the beginning of data generation for the purpose of ensuring the current use of data and for future rediscovery and reuse. For dynamic data sets, data monitoring means that data needs to be continuously supplemented and updated to meet user needs"[1]. As for big data, there is no accepted and accurate definition. The typical description is put forward by the National Science Foundation (NCF). Big data refers to "large-scale, diversified data generated by scientific instruments, sensor devices, Internet transactions, e-mail, audio and video software, network click stream and other data sources." Complex and long-term distributed data sets"[2] Chinese library scholars pay close attention to the hotspots of data monitoring and large data and actively carry out research. In order to explore the current academic research situation in this field, the author searched the full-text database of academic journals of China HowNet, and found that domestic scholars are more interested in the research of big data, which is far more popular than the research of data monitoring.

The author classifies and collates the literature published at present, and concludes that the research on data monitoring in academic circles mainly focuses on the following topics: data monitoring in American University Libraries and Its Enlightenment to China [3-5]; data monitoring education and its development abroad [6-7]; research and practice of scientific data monitoring abroad [8-9]; The establishment and demand of data guardianship posts in libraries [10]; the necessity of data guardianship in University Libraries in China and suggestions on strategies [11]. And the research on big data mainly involves: big numbers.

According to the concept definition, characteristics, etc. [12]; measures for libraries to cope with big data [13]; benefits brought by big data for libraries and changes in library service mode [14]; impact of big data on library scientific research and academic environment for the construction of Library big data system [15].

## **2. Current Situation of Data Monitoring and Large Data Research in Libraries in China**

University libraries devote themselves to the development of information services and the improvement of readers' satisfaction, but often neglect the study of scientific data.

Knowledge service and long-term development in libraries are not conducive. Data guardianship and big data, as the products of new information technology, will inevitably have a certain impact and impact on libraries. Scholars in the field of Library and information will also pay close attention to the research progress of data guardianship and big data. From the relevant research literature published at present, it can be seen that the scholars in the library circle have studied data monitoring and large data from different perspectives, and have achieved certain results; in contrast, the enthusiasm for large data research is more than that of data monitoring research, and it is also continuing to "warm up". This is of great theoretical and practical significance for improving the knowledge level of subject librarians participating in scientific data service in University libraries, improving the ability of data analysis and maintenance in University libraries, and further enhancing the social value of libraries. Needless to say, there are still some shortcomings in the current research on data monitoring and big data, such as: most of them focus on the general introduction of foreign development and application, but less on the summary of domestic experience; most scholars focus on theoretical research and pay less attention to the relevant practice, some even the same people. Some scholars have put forward some strategies and suggestions to solve the problem, but they are relatively general and lack of necessary feasibility analysis, so it is still difficult to implement them in reality. The main regret is that the data monitoring and large data of university libraries have not been combined for research. Maybe someone wants to ask, "Is there a connection between data monitoring and big data in libraries?" The author can answer this question with certainty: there is an inevitable link between the two. It is important and urgent for University Libraries in the era of big data to carry out research on them, which is often ignored by researchers. In order to verify this conclusion, we may refer back to the search records in Table 1. It is not difficult to see that up to now, there is no literature in China that combines data monitoring with large data for comprehensive research (the search record is 0). Is there an inevitable connection between the current big data and data guardianship? How will University Libraries view it? In this regard, the author has seriously considered, and based on an innovative perspective, the university library data monitoring and large data are combined to study, hoping to attract the attention and attention of the library community, which is the reason for this study.

## **3. Big Data and Data Monitoring: They Complement Each Other**

Data in data monitoring refers to the original data generated through scientific experiments and research, which is different from the digital coding stored in computer. At present, there is no consensus on the exact meaning of data guardianship in academic circles, but this does not affect the research of data guardianship. The importance of data is obvious to all: it is an important part of project research results, and also a necessary basis for scientific and technological innovation. If scientific data are not managed scientifically and efficiently, the project results will not be scientifically validated and the social value of sharing and utilizing scientific data will not be brought into play. Data guardianship is responsible for the selection, evaluation, organization and maintenance of scientific data. Therefore, data guardianship is of great significance to the maximum value of scientific data. The object of data monitoring is scientific data, and the management of massive scientific data is a very complex project, so it is necessary to build a scientific and effective data monitoring model. Some scholars believe that data monitoring model should include three parts: data generated by scientific research, data publication and data monitoring; data monitoring needs the participation of data guardians and scientific researchers, the support of national policies and education, and university library is the best Department of data monitoring [8]. There are still many problems and obstacles in the monitoring of scientific data in China's scientific research institutions, especially in small and medium-sized scientific research institutions.

Big data is one of the frontier technologies in the field of data analysis. It can be easily understood that big data technology is the ability to quickly obtain valuable information from various types of data. It is generally believed that large data has four basic characteristics, or "4V" for short: huge Volume, diversity, fast processing speed and great value [16]. The application of big data has become the focus of public attention. The business circles in developed countries (especially in the United States) have seen the business opportunities of applying big data, and set off an upsurge of research and application of big data. For the library community, big data technology can also help the library community to improve the basic service system and expand value-added additional services. Big data is different from the concept of massive data. At the same time, there are differences between big data and existing database technology, such as table 2 [17].

Although data guardianship and big data are two different concepts, they are inevitably linked. The author summarizes them as complementary relationships (as shown in Figure 1). First, data monitoring will effectively challenge large data. University library is the center of knowledge service and academic information exchange. With the change of information technology and communication mode, the data types of libraries are increasing gradually. Large-scale data will become an important part of future digital libraries. In the past, University Libraries lacked effective management of these big data, especially scientific research data, which led to the failure of data to play its due value. The challenges brought by big data to university libraries are various, such as data storage ability, analysis and processing ability of complex data and innovation of service mode; data monitoring can effectively maintain and manage data, improve the ability structure and skill requirements of subject librarians, and help university libraries cope with big data ring. Challenges in the environment. Secondly, big data stimulates the demand for data guardianship talents. Big data collection, organization, storage and so on are driving the demand for data monitoring talents, while university libraries just lack data consulting librarians, data processors and analysts. In order to expand the scope of service and turn the target to data monitoring and management services, university libraries should timely carry out data monitoring education and training programs, rationally set up data monitoring posts and train qualified data monitoring personnel. Finally, both big data and data guardianship promote the data management of University libraries, put forward new knowledge and skills requirements for Librarians in the big data era, bring new opportunities for the transformation of librarians' responsibilities and the service reform of University libraries, and seize the position of data management and service.

## **4. Data Monitoring of University Library Driven by Big Data**

### **4.1 University Libraries are the Important Subjects of Data Monitoring in the Era of Big Data**

McKinsey, a world-renowned consulting firm, first proposed the concept of "big data"; 2012 became the first year of big data, and the era of big data has come [18]. However, how to manage and maintain these vast data scientifically and how to train professional data guardians to deal with the challenges brought by the big data environment is a headache for many organizations. University libraries are one of the important "positions" of big data, which mainly includes digital resources of libraries, unstructured data formed by readers'browsing history, and a large number of unstructured data in the era of social network. Big data can help university libraries to achieve personalized services for different readers, and provide resource evaluation opinions for the library's database procurement department. With the increasing of large scientific research data, university libraries will gradually realize the transformation of service objectives to a new form of service - data services. Therefore, it is necessary to monitor the scientific data. In the era of big data, university libraries will become an important subject of data monitoring, because university libraries have good conditions for data monitoring: first, they have hardware facilities for data monitoring. After long-term construction of digital libraries, university libraries have special data servers and storage facilities, which can also be used for data monitoring. Second, university libraries have gradually enhanced the awareness of data monitoring. The traditional service modes of university libraries are affected and challenged. In order not to be marginalized, we must actively innovate and explore new service modes.

Librarians have gradually realized that university libraries can try to provide users with various information services of scientific data in order to better serve scientific research and give full play to the value of knowledge preservation and service center of University libraries. Traditional institutional repositories or academic exchange posts are transformed into data monitoring posts. University libraries can provide a unified data monitoring platform with stable operation and high efficiency by virtue of their rich knowledge resources.

#### **4.2 Big Data Promotes the Research of Data Monitoring in University Libraries**

Foreign colleges and libraries have made great achievements in the theoretical research and practical application of data monitoring, especially in the United States. Some scholars summarized the research and practice characteristics of data monitoring in American University libraries: formulating clear data monitoring planning and development strategies in order to promote the research process; pursuing good collaboration of data monitoring; carrying out data monitoring education and cultivating outstanding talents, etc. [4]. In addition, the Cornell University Library of the United States launched a data monitoring research and service project, Data Staging Repository (DataStaR), which is based on institutional databases and mainly aimed at scientific researchers. It has been approved by many fund organizations and has made special propaganda on data monitoring at the declaration meeting [19]. The emergence and development of big data stimulate new job demands for data guardianship talents, and urge foreign scholars to carry out research on data guardianship education and skills training. Some scholars point out that there are three main types of research contents in data guardianship education by foreign scholars: one is to analyze the roles and responsibilities of relevant stakeholders in data guardianship as a whole; the other is to specialize in the role orientation and professional skills of Librarians in data guardianship; and the third is to trace the data management skills education and training institutions. Understand its training plan for data management talents [6]. In recent years, foreign scholars have gradually shifted their research focus to the field research done by the College of Library and Information in order to meet the needs of the society by setting up data monitoring courses and training talents who meet the actual needs. Typical examples are: J. Kim, School of Information, University of North Texas, etc. to design the College's "Information: Guardianship, Archiving, Management, Preservation". (Information: Curation, Archive, Manage, Preserve) project curriculum system and system research [20]; University of Illinois Graduate School of Library and Information M. H. Cragin and other research for the design of the Institute's "Data Curation Education Program" curriculum [21]; C. Lee, School of Library and Information, Chapel Hill, University of North Carolina, conducted a survey on building a core knowledge competency matrix for data monitoring [22]. Foreign data monitoring research and application practice provide new enlightenment and reference for the construction of curriculum system of graphic information education and the training of talents in University libraries; however, unfortunately, our country has not paid attention to the research and application of data monitoring, and the current research literature only briefly mentions it, and there is no systematic specialization. Research. Big data will provide new research trends and research hotspots for data monitoring of University Libraries in China. By collecting and summarizing scientific research data and reader data, and then conducting large data analysis, the library can analyze users'needs and trends of research hotspots in related fields, and scientifically manage and maintain the analyzed data. Nursing, providing decision-making reference services for scientific research departments and scholars; through data monitoring, letting University Libraries integrate into the environment of scientific research field, timely understanding of scientific research trends, and enhancing the social value of University libraries.

## **5. Conclusion**

With the increasing of scientific data, its role has been paid more and more attention. Foreign libraries (especially in the United States as a typical representative) have realized that they have to deal with the opportunities and challenges of big data, and have begun to set up special data

monitoring posts, such as Data Scientist and Data Research Scientist. Arch Scientist, Data Services Librarian and Data Humanist expand the service boundary of libraries. Collection, collection, organization and preservation of big data stimulate the demand for data guardianship posts in University libraries, and give birth to data guardianship talents. These data guardianship posts are set up to adapt to the new data service roles and business of University Libraries in the big data environment. They belong to a new type of posts for libraries in China. University libraries are the main explorers and leaders of data monitoring. In order to keep up with the trend of big data era, it is necessary to re-examine the role orientation in scientific data management, carry out data monitoring education and training programs, and train existing subject service members (such as subject librarians, teaching librarians, liaison librarians, etc.). It is the sacred mission of librarians and informationists to be high-quality data librarians or data processors. In the future, librarians will go out and participate in the scientific research activities of scientific researchers, and continue to realize the management of big data outside libraries.

## References

- [1]. Lord P, Macdonald A. Data Curation for e-Science in the UK: An Audit to Establish Requirements for Future Curation and Provision [EB/OL]. [2013-12-22]. [http:// www. jisc.ac. uk/uploaded\\_ documents/e-science report final.pdf](http://www.jisc.ac.uk/uploaded_documents/e-science_report_final.pdf).
- [2]. What's Big Data? [EB/OL]. [2013-12-22]. [http://www.nsf.gov/funding/pgm\\_summ.jsp? Pims\\_id = 504767](http://www.nsf.gov/funding/pgm_summ.jsp?Pims_id=504767).
- [3]. Yang Helin. Data Guardianship: A New Exploration of American University Libraries [J]. *Journal of University Library*, 2011, 29 (2): 18-21, 41.
- [4]. Cheng Lianjuan. Practice and Enlightenment of Data Guardianship in American University Libraries [J]. *Library Journal*, 2012, 31 (1): 76-78.
- [5]. Yang Helin. A New Thought on the Construction of Institutional Repositories in American University Libraries from the Perspective of Data Guardianship: Enlightenment from DataStaR [J]. *Journal of University Library*, 2012, 30 (2): 23-28.
- [6]. Yelan. Research on Data Guardianship Education and Career Development Abroad [J]. *Journal of University Library*, 2013, 31 (3): 22-28, 37.
- [7]. Xia Yaojuan. Practice and Enlightenment of Data Guardianship Education and Training Abroad [J]. *New Century Library*, 2013 (6): 35-37.
- [8]. Zhang Qiuyan. Research on monitoring of scientific data in Colleges and universities [J]. *Information Science*, 2013, 31 (5): 42-45.
- [9]. Yin Shenqin, Zhang Jilong, Dou Fang. Standards and Practice of European Scientific Data Guardianship - UKDA Case Study [J]. *Library Journal*, 2013, 32 (6): 76-80.
- [10]. Yelan. Setting up and demand analysis of data guardianship posts in foreign libraries [J]. *Journal of University Library*, 2013, 31 (5): 5-12.
- [11]. Pei Yuxiang. Preliminary study on data monitoring in University Libraries [J]. *Library theory and practice*, 2013 (8): 79-81.
- [12]. Jiangshan, Wang Gang. Revelation of Big Data to Library [J]. *Library Work and Research*, 2013 (4): 52-54, 79.
- [13]. Rong Chunlin. Strategic Research on the Application of Big Data in Public Libraries [J]. *Library Construction*, 2013 (7): 91-95.

- [14]. Han Cuifeng. Innovation and development of library services in the era of big data [J]. *Library*, 2013 (1): 121-122.
- [15]. Zhang Xingwang. Academic Environment and Strategic Thinking on the Construction of Library Big Data System [J]. *Information and Information Work*, 2013 (2): 12-17.
- [16]. The White House. Big Data Across the Federal Government [EB/OL]. [2013-12-22]. [http://www.whitehouse.gov/sites/default/files/microsites/ostp/big\\_data\\_fact\\_sheet.pdf](http://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_fact_sheet.pdf).
- [17]. Big Data: The Next Frontier for Innovation, Competition and Productivity [EB/OL]. [2013-11-22]. <http://www.mckinsey.com/Features/Big Data>.
- [18]. The Wall Street Journal. Big-Data Success Stories: Splunk [EB/OL]. [2013-12-22]. <http://blogs.wsj.com/venturecapital/2011/10/21/big-data-success-stories-splunk/>.
- [19]. DataStaR [EB/OL]. [2013-12-22]. <http://DataStaR.mannlib.cornell.edu/>.
- [20]. Kim J, Warge E, Moen W. Digital Curation in the Academic Library Job Market [EB/OL]. [2013-12-22]. <http://www.asis.org/asist2012/proceedings/submissions/283.pdf>.
- [21]. Cragin M H, Palmer C L, Varvel Jr V E, et al. Analyzing Data Curation Job Descriptions [EB/OL]. [2013-12-22]. <http://www.ideals.illinois.edu/handle/2142/14544>.
- [22]. Lee C. What do Job Postings Indicate about Digital Curation Competencies [EB/OL]. [2013-12-22]. <http://ils.unc.edu/digccurr/digccurr-saa-research-forum-2008.pdf>.
- [23]. School of Information Management, Wuhan University. Ronald Larsen, Dean of School of Information Management, University of Pittsburgh, USA, visited the College for exchange: "I-school Education and New Generation of Data Processors" [EB/OL]. [2013-12-22]. [http://www.sim.whu.edu.cn/board/show\\_board\\_news.php?Board\\_news\\_id=2531](http://www.sim.whu.edu.cn/board/show_board_news.php?Board_news_id=2531).