

# Study on Experience of Scientific Research Data Service in American University Libraries

Dong Ning

Shandong Technology & Business University, Yantai, China

673077003@qq.com

**Keywords:** Scientific data, Research data service, Data librarian, University library

**Abstract:** Scientific research data is an important data and information source in the field of scientific research in the era of big data. Academic libraries in various countries have brought scientific research data management services into their service areas and participated in the whole process of scientific research. By using literature research methods, this paper exemplifies the practice of scientific research data services provided by American university libraries, to provide reference for the development of scientific research data services in China's libraries.

## 1. Introduction

Scientific research data is the data, data products and related information generated and processed by human beings in the whole process of scientific research [1]. Scientific research data is the product of scientific research, and has important research and predictive development effects on the development of scientific research projects. The service and management of scientific research data has become a hot spot in the development of the scientific and library circles. Library-related organizations such as IFLA and OCLC regard scientific data management as an important conference topic and research topic [2,3]. The UK established the Digital Curation Center (DCC) to specialize in the research and practice of scientific data management.

Domestic scholars' research on scientific research data services is still in the process of exploration and preliminary structure. Huang Ruhua and Li Nan conducted a study on licenses types of open data [4]. Zhang Shanshang, Liu Jingjing, Gu Liping, Cui Wenjian, Zhang Zhuoran analyzed a research on rights and interests' issues of research data reuse [5]. Ku Liping (also translated to Gu Liping) elaborated on research data life cycle, data management service policies [6], and data property and use boundary [7]. This paper uses the research methods of literature research to track the practice of foreign university libraries participating in scientific research data services in recent years, to provide experience for domestic university libraries.

## 2. University Libraries gradually participate in RDS

Research data services (RDS) is a series of management of artificial or intelligent discovery, correlation, aggregation, reuse, etc. of data scattered in the literature in the context of data-intensive production, so that users can directly solve problems. Service, which can improve the use of existing knowledge and data, and promote the generation and dissemination of new knowledge. In July 2013, the Primary Research Group (a survey company) released its survey of scientific data services for well-known university libraries around the world. Among the respondents, about 72.22% of university libraries in the United States provide scientific data services, while in other countries only about 50% [8].

The high participation rate of scientific data services in American university libraries is inseparable from the mandatory policy requirements of the US research funding agencies for scientific data management programs proposed by research funding projects. In the Guidelines for Project Funding and Management published by National Science Foundation (NSF) in January 2011, applicants were required to submit additional data management plans to maximize the dissemination

of their funded research results [9]. The National Institutes of Health (NIH) clearly stated in the Data Sharing and Implementation Guidelines that projects with an annual budget of more than \$500,000 must publicly share data results [10].

### **3. Practice of RDS in American University Libraries**

In 2013, the White House Office of Science and Technology Policy (OSTP) issued "Office of Science and Technology Policy: Promoting open data, open government, and open science", requiring that all government-funded research results, whether research data or final publications, must be accessible to the public and users of the scientific community. With the introduction of this policy, American university libraries have incorporated RDS into their regular service system. Its scientific research data service mainly includes the following contents:

#### **3.1 Research Guidelines**

American research libraries provide research/subject guides services and science librarians (with relevant knowledge background under various subject portals) service. Research guidelines usually collect knowledge and data in related fields under certain topics. Take the University of Chicago Research Guidelines Service as an example, it has four browsing paths: disciplines, courses, data types and phonology. Under each item, the following are gathered: (1) Web data sets, including online research documents, experimental records, observation data and research aids; (2) Numeric and factual databases; (3) Bibliographic index and periodical e-books; (4) APP applications include comprehensive applications, such as BrowZine, a citation management tool, professional applications, such as ChemSpider, Access Medicine, AgileMD, etc.

#### **3.2 Academic Reference Management**

Academic references can help research users grasp the main developments of research projects in the research field. Therefore, academic reference management is of great significance for users to fully understand the research situation, rationally plan the research route, and clarify the relationship between citations. Academic reference management tools that are currently widely used include: Zotero, RefWorks, EndNote, Citation Linker, and Mendeley.

#### **3.3 Academic Institutional Repository**

In 2013, OCLC "advancing the research mission" concluded that social media has become a new generation of academic communication tools. Scientific research activities are increasingly showing a tendency to become scattered and networked. Institutional repository has achieved sustainable development as an early prototype of academic exchange reform in the network environment. The Cornell University Library's Ar Xiv is a super-research database with more than a million visits per week, providing 735 000 full-text documents (statistics as of October 2016). Ar Xiv stores content in the fields of physics, mathematics, computer science, statistics, biology and economics [12]. The DASH Institutional Library is an open science community created by Harvard University that regularly sends creators a statistical report of aggregated document utilization and supports web applications such as Google and Twitter to retrieve, reference, and evaluate documents [13].

#### **3.4 Data Management and Analysis System**

As mentioned in Part 1, since January 2011, in the Guidelines for Project Funding and Management, NSF has required applicants to submit additional Data Management Plan (DMP) to maximize the dissemination and sharing of their funded research results [14]. Data management and analysis has become a new skill that scientists and technicians need to master in the data age. University libraries provide corresponding services to scientific researchers in data management analysis.

The Scientific Research Data Service Group of the University of Illinois Urbana Champaign Library concluded that the scientific research data management plan should include data storage, data naming, data sharing, time allocation of scientific research activities, personnel allocation of

scientific research activities, reference materials and policies of data, ownership and rights of achievements, etc. The library's scientific research data service group uses DMP Tool tools to provide detailed requirements for scientific research data management plan of various institutions and provide corresponding consultation and guidance services to promote scientific research workers to successfully complete the scientific research data management plan.

### **3.5 Knowledge Discovery System**

Cornell University's VIVO research discovery system [15] is a collaborative research data discovery system based on the Semantic Web. It provides multiple search paths for researchers, research institutions, research sites, experimental equipment, research topics, etc., and presents knowledge nodes such as abstracts, concepts, chapters, events, etc. in fragments, so that users can discover more academic associates and related institutions from the research roadmap of researchers, thus realizing knowledge discovery and deep mining from point to area.

### **3.6 Research Gateway**

Research gateway, also known as work space, is an omni-directional management service for promoting the sustainable development of research and facing the scientific research process. Taking the comprehensive research gateway of the Arizona State University Library as an example, its services include: (1) external funding opportunities. Every Wednesday, the project sponsorship will be announced in the form of "project announcement"; (2) strategic proposals for project bidding; (3) organization of seminars and training; (4) opportunities for peer review experts; (5) recommendation of expert partners. Use Elsevier's Scopus database to query the expert users who are engaged in similar research within the organization, or to set up a research team with interdisciplinary background; (6) human resource management scheme; (7) financial management. Including financial report, project income, project deficit and cash management; (8) guiding the preparation of project plan; (9) document and data management; (10) project management [16].

## **4. The Role and Vocational Training of Data Librarians**

### **4.1 New Role Orientation as Data Manager**

Based on the task of undertaking scientific research data services, the role and responsibility of librarians will change accordingly in the big data environment. Winston Tabb, director of the Hopkins University Library, describes libraries in the e-Science environment: "Libraries are part of a distributed network; data becomes a collection; data services are provided; librarians are data scientists."

### **4.2 Diversified Professional Education and Training**

With the increasing role of Libraries in scientific data management and the demand for professionals in scientific data management, the United States attaches great importance to the training of professionals in scientific data management and services. In response to this trend, LIS College provides a wealth of courses and educational projects. Currently, there are seven courses and training programs in scientific data management in the United States (Table 1), which are certified by the British Center for Scientific Data Management.

In addition to professional education in the field of information and graphics, American universities, research institutes, libraries and other organizations also provide education and training programs related to scientific data management. In addition to professional curricula for postgraduate students, University of North Carolina at Chapel Hill started its annual training program GigCCurr in 2009 to improve the digital management skills of digital collection managers. Library of Congress has launched a project called Digital Preservation Outreach & Education (DPOE), which aims to train Librarians in digital preservation skills.

Table 1. DCC-certified American Science Data Management Professionals and Projects.

University	Major	Academic Degree	Course Format
Arizona State University	Mathematical information management	Master 's degree	Remote lecture
University of California, Berkeley	Information management and information system	Master 's degree	Face-to-face
University of Illinois Urbana Champaign	Master of Library and Information, Scientific data management direction	Master 's degree	Face-to-face
University of North Carolina at Chapel Hill	GigCCurrl	Master, Doctor 's degree	Face-to-face
San Jose State University	Archives and records management	Master 's degree	Remote lecture
Simons College, Library and Information College	Digital supervision	Master 's degree	Remote lecture
The Johns Hopkins University	Digital management	Master 's degree	Online lecture

Source: British Data Center DCC [17]

## 5. Conclusion

Scientific research data service, which attracts many research institutions and scholars in the world to conduct in-depth research, is a broad and complex topic. As a supporting organization for educational and scientific research, university libraries should actively learn from foreign excellent experience, train data librarians according to our own policies and environment, and gradually develop scientific research data services.

## References

- [1] Wei Dongyuan, Zhu Zhaoyu.: How to Share the Scientific Data among the Professional Libraries. *Library Tribune* 37(12), 253-256 (2007).
- [2] The Role of Libraries in Data Curation, Access, and Preservation: An International Perspective, <http://www.ifla.org/news/the-role-of-libraries-in-data-curation-access-and-preservation-an-international-perspective>, 2013/10/03.
- [3] Role of Libraries in Data Curation, <http://www.oclc.org/research/activities/datacuration.html>. 2013/10/03.
- [4] Huang Ruhua, Li Nan.: A Study of License Types of Open Data. *Library* (8),16-21(2016).
- [5] Zhang Shanshan, Liu Jingjing, Gu Liping, Cui Wenjian, Zhang Zhuoran.: A Research on the Rights and Interests Issues of Research Data Reuse. *Documentation, Information & Knowledge* (1), 105-113(2018).
- [6] Ku Liping.: Data Management Services in the Transition of Research Model: An Approach of Implementing Open Access, Open Data and Open Science. *Journal of Library Science in China* (11),43-58(2018).
- [7] Gu Liping, Shu Fan.: Research on Data Property Rights and Use Boundary. *e-Science Technology & Application* (3), 34-39(2018).
- [8] Gary Price. Highlights From: International Survey of Academic Library Data Curation Practices, <http://www.infodocket.com/2013/07/11/highlights-from-international-survey-of-academic-library-data-curation-practices/> ,2013/10/09.
- [9] National Science Foundation Dissemination and sharing of research results. <http://www.nsf.gov/bfa/dias/policy/dmp.jsp>.2016/10/22.
- [10] National Institutes of Health. Policy for sharing of data obtained in NIH supported or conducted genome-wide association studies. <http://grants.nih.gov/grants/guide/no-tice-files/NOT-OD-07-088.html>.2016/10/22.
- [11] Office of Science and Technology Policy. Promoting open data, open government, and open science, <https://www.whitehouse.gov/administration/eop/ostp/initiatives# Openness>.2016/10/20.

- [12] Institutional repository of Cornell University. Ar Xiv, <https://www.library.cornell.edu/>.2019/01/18.
- [13] Institutional repository of Harvard University. DASH. <https://dash.harvard.edu/>. 2019/01/18.
- [14] National Science Foundation. Dissemination and sharing of research results, <http://www.nsf.gov/bfa/dias/policy/dmp.jsp>.2016/10/22.
- [15] Cornell University Library.VIVO, <http://vivo.cornell.edu/>. 2018/10/22.
- [16] Arizona State University Library, <https://www.asu.edu/>. 2019/01/18.
- [17] Data management, curation education and training, <http://www.dcc.ac.uk/training/data-management-courses-andtraining>. 2013/12/19