International Education Forum, 2024, Volume 2, Issue 6
http://ojs.bbwpublisher.com/index.php/IEF
Online ISSN: 2981-8605

Research and Reflection on Data Librarian Services in University Libraries

Wenting Wang*

Inner Mongolia University of Technology Library, Hohhot 010051, China

***Corresponding author:** Wenting Wang, 20110000001@imut.edu.cn

Abstract: Utilizing online surveys and literature research, this paper examines the roles and data services provided by data librarians in university libraries both domestically and internationally. Focusing on six aspects—construction and management of data management platforms, data support services, data reference consulting services, data analysis and visualization services, data literacy education and training, and data storage and security management. The study summarizes and analyzes the similarities and differences in the data services offered by university libraries worldwide. The paper advocates for universities to actively engage in the construction of data management platforms, optimize data mining, storage, and visualization technologies, and enhance librarians' data literacy levels to provide references for training data librarians and improving data services in domestic universities.

Keywords: University library; Data librarian; Data services; Data literacy

Online publication: August 9, 2024

1. Introduction

With the normalization of data-intensive scientific research paradigms, scientific data has become the most important knowledge resource in research, and data management services have become a major direction for university libraries in the era of big data to expand and enhance their services. Data librarians are tasked with meeting the multi-dimensional capability enhancement needs of researchers in data acquisition, statistical analysis, management, storage, and sharing. In the contemporary digital research environment, the data literacy of data librarians and the services they provide have become a key competitive advantage for university libraries.

This paper employs online surveys and literature research to investigate and analyze the roles and data management services of data librarians in university libraries, introducing their responsibilities and services in six areas including construction and management of data platforms, data analysis and visualization, and data literacy education and training, to provide references for the establishment of data librarian positions, skill training, and service enhancement.

2. Overview of data librarians

The concept of data librarians was proposed by J. Liscouski in 1997, defining them as librarians capable of providing storage, retrieval, search, and record access services for laboratory data [1–2]. In 2007, J. Read defined data librarians as librarians who provide data services [3]. The 2008 special issue of Nature on Big Data marked the emergence of the "big data" concept. In the same year, A. Swan and S. Brown introduced roles related to data, including data scientists, data administrators, data creators, and data librarians, defining data librarians as those professionally trained in data management, preservation, and storage with industry qualifications. Chinese scholar Meng Xiangbao believes that data librarians are trained to acquire, organize, process, analyze, store, and reuse data, and have obtained industry qualifications [4].

Overall, scholars agree that data librarians are specially trained in data literacy and possess professional skills in data management, storage, and supervision. They are expected to provide specialized data services based on user needs, supporting users throughout the scientific research lifecycle.

3. Data librarian services in universities

3.1. Construction and management of data management platforms

Higher education institutions are responsible for scientific data management, which involves collecting, processing, and preserving data according to standards to ensure data quality; maintaining data confidentiality and security; and establishing scientific data management systems to promote data sharing services [5]. Data librarians manage data resources generated throughout the scientific research lifecycle by effectively collecting, storing, and organizing data, creating data navigations, and maintaining data management platforms to facilitate data retrieval, sharing, and reuse.

University libraries often use institutional repositories or data management platforms for data resource construction and management. For example, MIT's DSpace and Cornell University's Fedora can perform general processes of data collection, preservation, and publication, commonly found in institutional repositories [6]. Beijing University of Aeronautics and Astronautics faculty achievements database provides archival, management, retrieval, and open access services for the school's research outputs [7]. Wuhan University's research data management service platform offers data preservation, management, and sharing services [8]. All these data management platforms are led by university libraries with data librarians responsible for processing and managing collected scientific data, conducting data navigation, and offering related data services to facilitate the sharing and reuse of scientific data.

3.2. Data support services

Data support services involve data librarians assisting relevant departments in establishing and perfecting scientific data management systems within their institutions to ensure standard-compliant data collection, processing, and long-term preservation while maintaining data quality. Additionally, they help researchers or research teams to formulate data management plans.

3.2.1. Establishment of comprehensive scientific data management and sharing systems

The perfection of scientific data management and sharing policies is fundamental for providing data services. University policies on scientific data management should encompass data management plans, data standards, data management services, data preservation, data access, data sharing, and data ownership, among others. Data librarians should participate in crafting these policies, clarifying storage standards, and addressing issues related to data storage, quality assurance, and stakeholder rights protection during data service provision.

In practice, many universities fall short of implementing comprehensive data management systems, with most researchers being unaware of their institution's policies on data management and sharing. Data librarians must take on the responsibility of actively promoting these policies, engaging all stakeholders involved in the data lifecycle, and addressing the issues of limited data sharing due to researchers' lack of policy awareness.

3.2.2. Assistance in formulating scientific data management plans

Research funding agencies in the UK and the US require applicants to provide a scientific data management plan when applying for project funding. These plans should detail data overview, organization, metadata, data preservation, data sharing and usage methods, privacy policies, confidentiality, intellectual property, and ethics, to ensure long-term data safety and utility. Funded scientific data from various technology plans should be submitted to relevant data centers for management.

Therefore, providing support or guidance in formulating data management plans is one of the crucial responsibilities of data librarians. For instance, at Purdue University, data librarians assist researchers in creating data management plans as required by funding bodies, aiding in the successful application for project funding. Similarly, at the University of Michigan and the University of Oxford, data librarians assess the data management needs of researchers or teams, helping to create internal scientific data management plans to ensure standardized and rational data management [9]. The Chinese Academy of Sciences explicitly states that a data management plan is a prerequisite for project approval and acceptance, necessitating data librarians' cooperation in project data management planning, defining internal roles, and data collection and processing requirements to ensure the authenticity, integrity, and usability of scientific data.

In China, only a few universities provide such data service support, with most data management plans being self-developed and completed by researchers themselves, indicating a low involvement of data librarians.

3.3. Data reference consultation services

Data reference consultation services allow data librarians to assist users with retrieving and accessing various forms of scientific data, such as numbers, charts, images, and sounds, primarily through data management platforms. Some users unfamiliar with platform operations require navigation or consultation on how to effectively use these data management systems.

Stanford University, through its library-affiliated social science data and software center, offers convenient data retrieval and analysis services. Purdue University's data librarians provide consultations on scientific data management platforms and processes. The University of California, Berkeley, offers guidance on data collection and access, while Peking University provides open data resource retrieval and data management platform usage consultations.

3.4. Data analysis and visualization services

Data analysis services involve data librarians using computer and digital technologies to provide specialized data analysis services based on user needs. They may collaborate with IT departments to clean, filter, merge, categorize, and reformat various data sources, and use technical tools to provide statistical analysis, qualitative analysis, and visualization services, along with corresponding statistical analysis tools and technical consultations.

The University of Edinburgh, using its data management platform, offers data cleaning and filtering services based on user needs, generating various data analysis visualizations and reports [10]. Harvard University enhances data interconnectivity through its platform, linking data with relevant literature to aid users in integrating diverse resources and using visualization technologies to enhance graphical, informational, and web

interaction displays for visual effects and support [11]. Peking University and Fudan University also rely on their data management platforms to offer online statistical analysis and visualization displays.

These data analysis and visualization services are crucially dependent on data management platforms and involve specialized fields like computer technology. Most universities independently appoint data librarians with explicit professional requirements in statistics and computer science, while some collaborate with IT departments to provide these services. Beyond platform-based services, some Chinese universities focus on institutional publication and patent data statistical analysis and offer technical consultations or training on data analysis tools [14].

3.5. Data literacy education and training

Data literacy education and training involve data librarians customizing training content around different stages of the data lifecycle according to researchers' needs. This includes general data literacy education, subject-specific data literacy training aligned with academic disciplines, systematic instruction on scientific data theories, and the use of common data management and statistical analysis tools. The goal is to enhance researchers' understanding of writing scientific data management plans, utilizing data management platforms, and knowledge of data governance processes, thereby improving their skills in organizing, describing, managing, and reusing data.

At Georgia State University, job postings for data literacy and learning specialists require candidates to teach users about data management tools and skills and collaborate with other instructors to integrate data literacy education into relevant courses. They also provide consultation services or embedded course guidance on data resource usage. Harvard University, in addition to offering offline data management training, also provides MOOC resources on data knowledge and tool usage and customizes lectures and seminars on scientific data for specific audiences [12]. Wuhan University and Beijing Normal University libraries utilize "micro-classes" to teach data resource retrieval and usage techniques through graphics and videos [13–14]. Peking University has explicit requirements in its data librarian job responsibilities to provide data literacy training or course services, including lectures on creating Excel charts and basic Python as part of their data literacy offerings and specifically developing study materials on data literacy and statistical data resources for big data environments [15].

Most universities conduct data literacy education mainly focused on acquiring data resources and using data statistical analysis software, often in the form of general courses or specialized lectures. However, there is less emphasis on aspects of the data lifecycle like data publishing and storage, resulting in data literacy education that is not as systematic and comprehensive as it could be.

3.6. Data storage and security management

Data storage and security management involve data librarians using data management platforms to ensure the confidentiality and security of scientific data while actively fulfilling responsibilities related to scientific data-sharing services. Data librarians leverage their expertise in data resource management to work with researchers to develop long-term data preservation strategies and standards. They organize and collect scientific data according to metadata standards in data management platforms, ensuring data is stored in a uniform and standardized format and preserved through institutional knowledge bases or data management platforms for future sharing and reuse.

At Simon Fraser University, data management and preservation specialists are required to establish data preservation standards and management rules, assist researchers in uploading and submitting scientific data, and provide IT service support [16]. Wuhan University, Fudan University, and Peking University rely on established

scientific data management platforms where data librarians assist researchers with data upload and submission tasks, providing long-term data preservation, security management, and sharing services.

University library data librarians should fully utilize their strengths in data resource preservation and organization, actively participate in setting data management norms and standards, and through institutional repositories or scientific data management platforms, ensure the long-term storage and security management of scientific data to guarantee its subsequent sharing and reuse.

4. Conclusion

Data services represent a new area in the transformation and upgrading of library services. Universities with high service levels and a rational structure may establish dedicated data governance positions, actively engage in the construction of data management platforms, and enhance data mining, storage, and visualization analysis technologies to broaden the scope and depth of data services from a technical perspective. For universities that are not equipped to set up separate data librarian positions, it is encouraged for librarians to enhance their personal data literacy through professional skill training and other means. By embedding such services, librarians can better fulfill their responsibilities in supporting researchers to create data management plans, acquire data resources, ensure secure storage, and provide data literacy education.

Funding

Inner Mongolia Autonomous Region Directly Affiliated University's Basic Research Business Fee Project (No. JY20220133)

Disclosure statement

The author declares no conflict of interest.

References

[1] Liscouski J, 1997, The Data Librarian: Introducing the Data Librarian. Journal of Automatic Chemistry, 19(6): 199–204.

[2] Read EJ, 2007, Data Services in Academic Libraries: Assessing Needs and Promoting Service. Reference & User Services Quarterly, 46(3): 61–75.

[3] Swan A, Brown S, 2021, The Skills, Role and Career Structure of Data Scientists and Curators: An Assessment of Current Practice and Future Needs, http://www.jisc.ac.uk/whatwedo/programmes/digitalrepositories2007/dataskillscareers.aspx

[4] Meng XB, Qian P, 2013, Position Setup and Management Mechanisms of Data Librarians in Foreign University Libraries. Library & Information, 2013(4): 12–17.

[5] General Office of the State Council, PRC, 2022, Notice of The General Office of the State Council on Printing and Distributing Measures for the Management of Scientific Data, https://www.gov.cn/zhengce/content/2018-04/02/content_5279272.htm

[6] Zhang JL, Yin SQ, Zhang Y, et al., 2015, Sharing and Services of Social Science Data: A Case Study of Fudan University's Social Science Data Sharing Platform. Journal of Academic Libraries, 2015(1): 74–78.

[7] Zhang GD, Song S, 2017, Construction of a Faculty Achievement Database at Beihang University. Proceedings of

the 2017 Beijing University Digital Library Annual Conference, 1–26.

[8] Wuhan University Library, 2022, Research Data Management Platform, http://www.lib.whu.edu.cn/webfile/category/Scientific_research_data_management_platform/317.html

[9] Chinese Academy of Sciences, 2022, Regulations on Scientific Data Management and Open Sharing, https://www.cas.cn/tz/201902/t20190220_4679797.shtml

[10] Yu DF, 2021, Practices and Insights from Harvard University Library's Research Data Management Services. Library, 2021(10): 47–53.

[11] Harvard Library, 2022, Visualization Support, https://library.harvard.edu/services-tools/visualization-support

[12] Wang SM, 2019, Analysis of Content Elements in Data Management Plans of Chinese University Libraries and Their Application, thesis, Northeast Normal University.

[13] Wuhan University Library, 2021, Xiao Bu Micro-Courses, http://www.lib.whu.edu.cn/web/index.asp?obj_id=805

[14] Beijing Normal University Library, 2021, Xiao Tu Micro-Courses, http://www.lib.bnu.edu.cn/content/wei-ke-cheng

[15] Peking University Library, 2022, Data Literacy and Introduction to Statistical Data Resources, http://www.lib.pku.edu.cn/portal/fw/yixiaoshijiangzuo/ziyuanchazhao

[16] The International Association for Social Science Information Service and Technology, 2021, Data Curation and Digital Preservation Specialist, http://www.iassistdata.org/resources/jobs/3147