

Research Data Management Implementation at Peking University Library: Foster and Promote Open Science and Open Data postprint

Authors: Nie, Hua, Luo, Pengcheng, Fu, Ping, Nie, Hua

Date: 2022-11-27T00:00:00+00:00

Abstract

Research Data Management (RDM) has become increasingly important for more and more academic institutions. Using the Peking University Open Research Data Repository (PKU-ORDR) project as an example, this paper will review a library-based university-wide open research data repository project and related RDM services implementation process including project kickoff, needs assessment, partnerships establishment, software investigation and selection, software customization, as well as data curation services and training. Through the review, some issues revealed during the stages of the implementation process are also discussed and addressed in the paper such as awareness of research data, demands from data providers and users, data policies and requirements from home institution, requirements from funding agencies and publishers, the collaboration between administrative units and libraries, and concerns from data providers and users. The significance of the study is that the paper shows an example of creating an Open Data repository and RDM services for other Chinese academic libraries planning to implement their RDM services for their home institutions. The authors of the paper have also observed since the PKU-ORDR and RDM services implemented in 2015, the Peking University Library (PKUL) has helped numerous researchers to support the entire research life cycle and enhanced Open Science (OS) practices on campus, as well as impacted the national OS movement in China through various national events and activities hosted by the PKUL.

Full Text

Preamble

Research Data Management Implementation at Peking University Library: Foster and Promote Open Science and Open Data

Hua Nie^{1†}, Pengcheng Luo¹ & Ping Fu²

¹Peking University Library, Beijing 100871, China

²Central Washington University Library, 400 E. University Way, Ellensburg WA 98926, USA

Keywords: Research data management implementation; Open Science; Open Data; Research life cycle; Data sharing and reuse

Citation: Nie, H., Luo, P.C., Fu, P.: Research data management implementation at Peking University Library: Foster and promote Open Science and Open Data. *Data Intelligence* 3(1), 189-204 (2021). doi: 10.1162/dint_a_{00088}

Received: September 29, 2020; **Revised:** December 14, 2020; **Accepted:** December 22, 2020

ABSTRACT

Research Data Management (RDM) has become increasingly important for academic institutions worldwide. Using the Peking University Open Research Data Repository (PKU-ORDR) project as a case study, this paper reviews a library-based, university-wide open research data repository initiative and its associated RDM services implementation process, encompassing project kickoff, needs assessment, partnership establishment, software investigation and selection, software customization, data curation services, and training.

Through this review, the paper also discusses and addresses issues revealed during various implementation stages, including researcher awareness of research data management, demands from data providers and users, data policies and requirements from the home institution, funding agencies, and publishers, collaboration between administrative units and libraries, and concerns from data providers and users. The significance of this study lies in providing an exemplary model for creating an Open Data repository and RDM services that other Chinese academic libraries can reference when planning to implement their own RDM services. Since the PKU-ORDR and RDM services were launched in 2015, the authors have observed that Peking University Library (PKUL) has supported numerous researchers throughout the entire research lifecycle, enhanced Open Science (OS) practices on campus, and influenced the national OS movement in China through various events and activities hosted by PKUL.

Corresponding author: Hua Nie (Email: hnie@lib.pku.edu.cn; ORCID: 0000-0002-4522-5049).

1. INTRODUCTION

Open Science (OS) has made science more efficient, reliable, and responsive to societal challenges while reshaping scholarly communication landscapes [1]. Open access to research data is regarded as a key driver for OS [2], and the scientific research community believes that data should be open, accessible, and reusable.

Data sharing and reuse help strengthen researchers' and institutional data stewardship [3,4,5]. To better foster OS, Open Data (OD), and the transformation of scholarly communication, an increasing number of academic libraries have provided or plan to provide library-based research data management (RDM) services for their institutions. Tenopir et al.'s study [6] suggested that academic libraries are ideal centers for research data service activities on campuses, providing unique opportunities for libraries to become even more active participants in the knowledge creation cycle. More recent studies agree that academic libraries and librarians, as active stakeholders, have played a significant role in fostering the open access movement, transforming scholarly communication landscapes, and facilitating RDM services [6,7,8,9,10,11,12].

Although RDM has rapidly developed in research universities worldwide, RDM in Chinese academic libraries remains at an early stage. For example, re3data.org (the Registry of Research Data Repositories) listed more than 2,000 research data repositories in 2019, among which only 42 were related to China. Re3data.org is an OS tool that offers researchers, funding agencies, libraries, and publishers an overview of existing international repositories for research data [13].

As a top research university library in China, Peking University Library (PKUL) has long actively sought opportunities to raise awareness, foster collaboration, and initiate projects both within and outside the University. PKUL has also been actively involved in numerous national and international endeavors to foster the open access movement and transform scholarly communication over the past decade. Since 2010, PKUL has implemented several initiatives to support the dynamically changing scholarly communication environment at the University: the Peking University (PKU) Institutional Repository in 2010, PKU Open Journals in 2015, Scholars @ PKU in 2015, and the PKU Open Research Data Repository (PKU-ORDR) in 2015. Specifically, the PKU-ORDR was created in 2015 to facilitate more effective and efficient data preservation, sharing, and reuse, while providing incentives for making data readily accessible to researchers and the general public.

This paper reviews how PKUL implemented the PKU-ORDR project and RDM services to foster and support OS and OD on campus and influence Chinese OS communities. The implementation phases reviewed include project kickoff and needs assessment, partnership establishment, software investigation and selection, software localization and customization, and the implementation of RDM policies and services. The paper also discusses and addresses issues revealed during the implementation process, such as awareness of research data, demands from data providers and users, data policies and requirements from the home institution, funding agencies, and publishers, collaboration between administrative units and libraries, and concerns from data providers and users.

The significance of this study is that PKU-ORDR demonstrates a successful example of creating an OD repository and RDM services for other Chinese academic libraries planning to implement their own RDM services. The authors

have observed that since the PKU-ORDR and RDM services were implemented in 2015, PKUL has helped numerous researchers support the entire research lifecycle, enhanced OS practices on campus, and impacted the national OS movement in China through hosting various national events and activities.

2. RELATED WORKS

This review focuses on aspects addressed in this paper, including RDM and academic libraries & librarians, RDM and open research data repositories and systems, collaborations between research units and libraries, service support and promotion, repository implementation, data curation, and skills training for research support staff and librarians.

Tenopir et al. [6] noted that as science becomes more collaborative, data-intensive, and computational, academic researchers face a series of data management needs. Meanwhile, Moon's study [14] shows that research funding agencies require researchers to provide Data Management Plans (DMPs) when applying for grants, and publishers also require researchers to provide data when publishing research results. Curdt's study [15] indicated that science conducted in cross-institutional, interdisciplinary, and long-term research projects requires active sharing of data, documents, and further information. Thus, RDM services should be established to support all researchers throughout their individual research studies.

Tenopir et al. [6] also claimed that academic libraries may be ideal centers for RDM service activities on campuses. Cox et al. [10] reported an international study of RDM activities, services, and capabilities in higher education libraries, finding that libraries have provided leadership in RDM, particularly in advocacy and policy development. However, services provided by libraries are still limited, focusing especially on advisory and consultancy services. Tripathi et al. [16] studied RDM services implemented by different university libraries in India for managing, organizing, curating, and preserving research data generated at their universities' departments and laboratories for data reuse and sharing, and suggested a model for university libraries to follow when deploying RDM services.

Johnston et al. [17] compared six institutions' RDM support levels within the Data Curation Network project and developed a shared staffing model for data curation across multiple institutions to support researchers in meeting their data-sharing goals through library-based data repository and curation services. Lee et al. [18] interviewed American university institutional repository (IR) staff and provided a rich, qualitative description of research data curation and use practices in IRs. In particular, Lee et al. identified data curation and use activities in IRs, as well as IR structures, roles played, skills needed, contradictions and problems exposed, solutions sought, and workarounds applied.

Curdt and Hoffmeister [19] shared their design and implementation of RDM services for a multidisciplinary and collaborative research project. McKinney et

al. [20] described how Harvard University established a diffraction data publication system, the Structural Biology Data Grid (SBDG), to preserve primary experimental data sets supporting scientific publications. All data sets published through the SBDG are freely available to the research community under a public domain dedication license, with metadata compliant with the DataCite Schema. They also shared practices where the SBDG collaborated with the Institute for Quantitative Social Science at Harvard University to extend the Dataverse open-source data repository system to structural biology data sets.

Mannheimer et al. [21] described how data repositories and academic libraries can partner with researchers to address challenges associated with qualitative data sharing, suggesting that data repositories and academic libraries could help researchers address some challenges related to ethical and lawful qualitative data sharing. Dovidonyte's study [22] described the Lithuanian landscape of OS policies and institutional involvement in OS practices, discussing prerequisites for sustainable and consistent OS implementation such as OS infrastructure, incentives for researchers, research assessment, and repositories' compliance with European Council requirements at a national level.

Pontika [23] analyzed and found that academic libraries have created new academic librarian positions to support OS, OD, scholarly communication, and RDM on their campuses. However, researchers are still unfamiliar with RDM best practices, and research support staff including librarians face difficulty providing support to researchers across different disciplines and career stages [24].

Alonso-Arévalo [25] agreed that the management of research data is one of the major challenges facing scientific and research libraries in the coming years. Already half of American universities have a work plan on this issue, and all trend reports agree that RDM will be one of the priorities and future issues for research libraries. Söderholm et al. [26] found that a network-based collaboration model fostering individuals' interconnectedness is crucial for surviving the built-in dynamism of RDM. Tang and Hu emphasized in their study [27] that for growing RDM services, institutional commitment to resources and training opportunities is crucial. As an emergent profession, data librarians need to be nurtured, mentored, and further trained.

All these studies have provided theoretical, useful, and practical insights and examples, while also revealing challenges and issues in the RDM implementation process faced by researchers, academic libraries, and librarians.

3.1 Kick-Off and Needs Assessment

To kick off the PKU-ORDR project, PKUL conducted a campus-wide survey in 2013 to better understand RDM needs and requirements from researchers and research teams. The purpose was to identify real needs and collect data to create a strategic roadmap and framework to meet RDM requirements. The analysis results were summarized and published in the *Journal of Library and Information Service* [28]. The survey focused on: awareness and current practices of

RDM including data preservation, sharing, and reuse; description and features of research data; the current state of RDM; and expectations of RDM services. Results showed that 87.5% of respondents were willing to share research data under certain conditions. The primary motivation was recognition of data sharing's value, the positive relationship between data use and citations, increased data visibility, and credits awarded to data providers. However, researchers' biggest concern was plagiarism.

PKUL also interviewed 23 research teams from multiple disciplines on campus. Face-to-face communication helped discover valuable information about the current state of RDM, including long-term preservation, data sharing, and data reuse. Zhu et al. [28] summarized three major findings: (1) Research data sharing behavior is significantly influenced by disciplines. For example, biology is a data-driven and data-intensive discipline where open access is already common practice, with well-established data sharing standards and norms. (2) An embargo period for data sharing is generally expected and required. Almost all interviewed researchers emphasized that data should be shared only after results are formally published, addressing concerns about possible plagiarism. (3) Data sharing behavior is more spontaneous and passive than active, lacking proper incentives, necessary maintenance, and a well-established mechanism for data citation, recognition, credits, and feedback from data users.

Interviews also revealed that data management needs vary greatly across disciplines. Bioinformatics researchers need very large data storage to preserve massive amounts of process data from experiments. Computer Science researchers are willing to share data, but their datasets are often very large, making sharing more expensive. For example, the Chinese Web dataset collected by the Institute of Network Computing and Information Systems over ten years exceeds 100TB. Business researchers hope to obtain more valuable enterprise and government data for use in classes and research. Researchers from the Institute of Social Science Survey (ISSS) hope to maximize their survey data's value through sharing, though ISSS has established relatively strict user application procedures for data access. Faced with such diverse data management needs, PKUL analyzed priorities and decided to build initial service infrastructure to meet the highest-priority needs, focusing on data closer to its final state and easier to share. PKUL collaborated with institutions inside and outside campus to build PKU-ORDR, making data more accessible for PKU researchers to reuse and share.

3.2 The Establishment of the Collaborative Model

As one of the most active advocates of RDM at the University, PKUL made numerous efforts to convince various university research administrative units to invest in and support creation of a library-based RDM framework. PKUL also sought potential partners within the University, as cooperation and collaboration with administrative units and other campus units are vital to project success and critical for sustainability.

PKUL selected the Institute of Social Science Survey (ISSS) as a working partner to collaborate on RDM service development. ISSS was created to act as a social science data survey coordinator and interdisciplinary empirical research platform, enabling Peking University and other research institutions worldwide to study China's social problems and conduct social science research, primarily through undertaking large-scale social survey projects and openly sharing survey data. Thus, ISSS was an ideal collaborative candidate for the Library. ISSS also plays a leading role on campus in providing workshops and training classes in data access, curation, and analysis methods for the social science research community.

In 2014, Peking University received a grant from the National Natural Science Foundation of China for the China Survey Data Archive (CSDA) project, which aimed to develop a data repository administered by the University Management Science Data Center (MSDC), a department within ISSS. This grant provided PKUL an opportunity to build a more collaborative relationship with ISSS. With assistance from research administrative units such as the Office of Science Research and the Office of Social Science Research, PKUL and ISSS decided to work together on this project. Initially, responsibilities were split as follows: MSDC, supervised by ISSS, was responsible for research data collection and cleaning, standardization and analysis, data repository platform testing, and feedback. PKUL was responsible for requirements analysis, functional design, software selection, development and maintenance of the data repository, data storage, classification and metadata, systems administration, and associated technical and technological services.

However, ISSS and PKUL soon discovered through analysis of survey and interview data that this could be an opportunity to build a strong showcase for OD for the nation, as there were only a very limited number of subject-specific and/or research team-oriented data storages and services available at either the institutional or national level. The initiative was named the PKU-ORDR (PKU Open Research Data Repository) project as a sub-project of CSDA, with the goal of developing infrastructure to support PKU researchers in managing their data more effectively and efficiently and providing RDM services ranging from storage to consultation.

The strategic objectives of PKU-ORDR are summarized as follows: - To publish high-quality research data and disseminate academic outputs through an open platform; - To promote OS, facilitate data sharing and reuse, and encourage research reproducibility; - To enable and track data citations and usage metrics; - To explore data publishing and long-term preservation solutions; - To foster innovation and cross-disciplinary integration.

In addition to ISSS, PKUL cooperated with other internal units and external organizations to enrich PKU-ORDR's data content. Through collaboration with the Center for Bioinformatics of Peking University, PKUL created linked data in PKU-ORDR connecting to the Bioinformatics database. Through cooperation with the Beijing Information Resources Management Center, PKU-ORDR

interoperated with the Beijing Government Data Resource System (BGDRS), enabling registered PKU-ORDR users to download data directly from BGDRS. Through cooperation with the National Information Center, PKU-ORDR collected valuable enterprise datasets across the country. All these collaborations have greatly enriched data content and expanded the disciplinary scope of PKU-ORDR.

3.3 The Establishment of the Open Research Data Repository

The first step of PKU-ORDR was creating an open research data repository to meet data storage and sharing needs. This establishment included software selection as a framework and subsequent customization.

3.3.1 Software Selection

Several types of RDM software were available at that time, including various institutional repositories (IRs) supporting RDM, with either open-source or proprietary solutions. PKUL evaluated and assessed existing software including Dataverse, Data Conservancy, CKAN, Dryad, ICPSR, Genbank, Figshare, and Nesstar. The implementation team also deployed and tested open-source solutions such as Dataverse, Data Conservancy, CKAN, and DSpace.

The team adopted a software metrics tool and created criteria to evaluate these solutions. General criteria considered included business and industry expertise, market knowledge, program/project management capabilities, methodology, communications, and independence and objectivity. Additionally, as shown in Table 1, four specific criteria were particularly important: Metadata standard and interoperability; Permissions management and access control; DOI identifier and version management; Online analysis and visualization. Notably, the Dataverse metadata schema consists of a compulsory citation metadata block and multiple optional discipline metadata blocks that can be easily customized. The default discipline metadata block is DDI for Social Sciences, and Dataverse also provides several other discipline metadata blocks such as Biomedical, Geospatial, Astronomy, and Astrophysics. Therefore, the Dataverse metadata schema is flexible enough to theoretically adapt to any discipline.

After systematic comparison and assessment, the Dataverse solution was chosen as the development tool. Dataverse was originally developed by Harvard's Institute for Quantitative Social Science (IQSS), along with many collaborators and contributors worldwide. As of August 7, 2020, it had 59 installations worldwide.

3.3.2 Software Customization

Although Dataverse was chosen as the framework, customization presented challenges. The software development and version release phases are shown in Figure

1 [Figure 1: see original paper]. Project milestones are summarized as follows: (1) By the end of May 2015, PKUL completed testing and function building with the Chinese version of Dataverse v3.3 adopted from Fudan University. (2) Starting in June 2015, PKUL continued working on system architecture, localization and customization, and functionality refinement based on Dataverse 4.0, which was then released and PKU-ORDR was formally launched in December 2015. Version 4.0 was used between June 2015 and July 2019. During that period, Harvard University released more than ten minor versions with multiple functions added to the framework such as metadata harvesting, private URLs, and cloud storage support. (3) In July 2019, PKUL decided to upgrade the platform to v4.14 by adopting changes made by Harvard University and fixing numerous bugs in early v4.0. PKUL also completed function customization and data migration during this upgrade.

Highlights of local customization include: (1) user management, (2) bilingual interface, (3) usage statistics, (4) data contests, (5) other functions such as DataCite DOI registration, dataset-related publications, and (6) custom homepage.

To enhance user management, PKUL implemented the PKU-IAAA single sign-on system to enable users to quickly and securely authenticate permissions and gain instant access to the OD repository, with relevant patron information carried into the data repository through PKU-IAAA. Furthermore, PKUL enabled a group download function so users can download multiple files within one dataset with a single request, while original Dataverse only allows one file per request. PKUL also created two types of user accounts: regular and advanced. A regular account can be upgraded to advanced when a user submits an application and provides additional required information to obtain more privileges.

A bilingual interface is essential since the repository is open to anyone worldwide. While researchers typically publish outputs in English to increase visibility, the majority of users come from China and are more comfortable with Chinese. Therefore, PKUL made the Dataverse repository interface and metadata support both English and Chinese. The user interface can switch between languages, search results display in both languages, and notifications to users are customized in bilingual format.

Regarding usage statistics, original Dataverse only tracks download numbers, which fails to satisfy PKU data providers' statistical requirements. Therefore, PKUL enabled logging of user applications, administrator verification, user browsing, and downloads. Elasticsearch indexes these logs so data providers can query and download real-time data. Meanwhile, Baidu Analytics was implemented to analyze user sources, devices used, search keywords, and pages visited. Furthermore, PKUL hosted two national data contests to promote open research data repositories. A contest module was added to facilitate user enrollment and data use, allowing participants to form teams, submit papers, and access research data directly using their repository accounts. The module also provides functions such as a contest homepage and paper display gallery.

Additionally, PKUL added many other functions to Dataverse. Dataverse 4.0 only provided Handle identifier registration, while PKU-ORDR adopted DataCite DOI registration—a module later adopted by Harvard University and other Dataverse institutions. Since some PKU-ORDR datasets are high-quality (e.g., the China Family Panel Studies dataset has been cited by numerous research papers), PKUL used APIs to interoperate with PKU-IR, retrieving papers from PKU-IR and displaying associated papers on the PKU-ORDR platform. Also, Dataverse 4.0 did not support homepage customization, but PKUL developed a custom homepage technology now adopted by Harvard University. As shown in Figure 2 [Figure 2: see original paper], numerous efforts were made between 2014 and 2019, with key milestones highlighted.

3.4.1 Usage

PKU-ORDR enhanced PKUL's infrastructure for data storage and sharing. Through collaboration with academic departments, PKU-ORDR has collected numerous high-quality datasets, including China Family Panel Studies, China Health and Retirement Longitudinal Study, Beijing Area Study, Comprehensive Language Knowledge Base, and AutismKB, an Evidence-based Knowledge Base of Autism.

As of August 2020, PKU-ORDR has released 66 Dataverses, 305 datasets, and 2,036 data files. Total downloads have exceeded 620,000, with an average of about 500 daily visitors and 2,700 page views. In recent years, visitors from more than 89 countries have accessed the repository, with the top five being China, the United States, the United Kingdom, Japan, and South Korea. Registered users have reached 32,000. Figures 3 and 4 show the top 10 institutions in terms of registered users in China and abroad, respectively, all from prestigious research universities either in China or internationally.

3.4.2 Data Curation and Skills Training

To cultivate past research data for future use, PKUL offered several data curation services. Collaborating with ISSS, PKUL hosted an RDM Seminar in 2015, inviting experts from the Inter-University Consortium for Political and Social Research (ICPSR), USA, and the Data Archive, UK, to deliver data management training. Trainees included teachers and students from PKU and other peer universities in China. PKUL also sent librarians to participate in relevant training activities hosted by other universities to improve their RDM curation skills. To improve students' data search skills, PKUL offered one-hour workshops teaching students to identify data sources and use scientific methods to acquire relevant research data, statistical data, and Internet data. PKUL also provided a series of lectures teaching students how to use data analytics tools.

3.4.3 Service Promotion

To improve visibility of PKU's data services, PKU promoted PKU-ORDR through various channels, including marketing on the PKU homepage, social media public accounts of student groups, annual ISSS conferences, and various domestic and international RDM-related conferences. To improve data accessibility, PKU-ORDR provided metadata to re3data.org (an international data repository registry), DataCite Search, Data Citation Index (data discovery systems), and search engines such as Baidu and Google. Additionally, collaborating with other campus units, the National Information Center, PKU hosted two national contests titled "National Data-Driven Research Contests for Colleges and Universities" in 2018 and 2019 to promote PKU-ORDR use and RDM services and train students' data searching and acquisition skills.

The contest included six stages: training workshops and lectures, enrollment, paper submission, paper evaluation, and oral defense. During the first stage, organizers provided training on contest rules, data analysis and mining, data management and sharing, data resources, and acquisition. During enrollment, contestants registered in groups, submitted selected topics, and applied for research data from PKU-ORDR. During paper submission, contestants conducted research using PKU-ORDR data or original data they collected, wrote essays, and submitted papers with data to organizers. During evaluation, organizers first conducted formal assessment and plagiarism checks; qualified papers were then evaluated by invited experts, with each essay reviewed and graded by two experts. Papers were ranked accordingly. During oral defense, top-ranked teams delivered statements and reports evaluated by more than 10 experts to determine final ranking. After the contest, winning teams shared their research data and presented at the Jing Ling Big Data Summits, with excellent essays published in Chinese core journals in special topic issues.

The contests attracted numerous students from major research universities nationwide. The first contest enrolled nearly 600 teams with about 2,000 contestants from more than 160 universities and colleges, from 28 provinces, majoring in 59 disciplines including Computer Science, Information Technology, Management Science & Engineering, Applied Economics, Statistics, Public Health & Preventive Medicine, Library and Information Science, and Chinese History. Ultimately, 289 teams with about 1,000 participants submitted research papers. The second contest enrolled 600 teams with 1,704 participants from 29 provinces, covering Applied Economics, Computer Science, Information Technology, Statistics, Sociology, Library and Information Science, Management Science & Engineering, and Public Health & Preventive Medicine.

Through these contests, students from different disciplines gained experience on the same OD platform. The competition greatly promoted data-driven research paradigms and PKU-ORDR visibility. Between December 2017 and May 2018 during the two contests, online visitors to PKU-ORDR increased 10-fold, registered users increased 5-fold, and data downloads increased 7-fold compared to

before. More external websites linked to PKU-ORDR, and the repository's ranking and exposure in search engines greatly improved. Simultaneously, original data submitted by contestants greatly enriched the repository's content.

4. CONCLUSIONS

This paper reviewed the PKU-ORDR implementation process and RDM service creation at PKUL. Through this review, the authors found that needs assessment and collaboration are vital to the success of library-based, university-wide RDM projects. Raising researchers' awareness of OS and OD is critical. Software identification and selection is a complicated and time-consuming process; the software must meet essential criteria such as stability and sustainability. Communication is critical throughout the process, particularly with administrative and academic units. Data curation programs such as workshops, lectures, and contests can improve researchers', students', and librarians' data searching and acquisition skills while promoting services on campus and in broader research communities. RDM policies must be created and implemented. In summary, this represents a learning curve and cumulative process in both theory and practice. The authors hope this practical paper offers insights to academic libraries planning to implement OD repositories and/or RDM services for their institutions. Although PKUL has made great efforts in RDM construction and contributed to OS and OD communities on campus and across China, PKUL recognizes it still has a long way to go, with many challenges and opportunities ahead for libraries and librarians.

AUTHOR CONTRIBUTIONS

All authors wrote and revised the manuscript and made substantial contributions to the paper's design. H. Nie (hnie@lib.pku.edu.cn) directed overall planning and contributed primarily to the article's framework design, Introduction, RDM Implementation (especially "Kick-Off and Needs Assessment" and "The Establishment of the Collaborative Model"), and Conclusions. P.C. Luo (luopc@lib.pku.edu.cn) contributed primarily to RDM Implementation (especially "The Establishment of the Open Research Data Repository" and "Usage, Data Curation, Skills Training and Services Promotion"). P. Fu (Ping.Fu@cwu.edu) contributed primarily to Related Works, partially to Introduction and Conclusions, and co-contributed to layout, composition, writing, and source utilization.

REFERENCES

- [1] Burgelman, J.-C., et al.: Open science, open data, and open scholarship: European policies to make science fit for the twenty-first century. *Frontiers in Big Data* 2, Article 43 (2019)
- [2] Giglia, E.: Open Access to research data as a driver for Open Science. *JLIS.it: Italian Journal of Library and Information Science* 6(2), 225-247 (2015)
- [3] Tenopir, C., et al.: Changes in data sharing and data reuse practices and

- perceptions among scientists worldwide. PLoS ONE 10(8), e0134826 (2015)
- [4] Shen, Y. (2016). Research data sharing and reuse practices of academic faculty researchers: A study of the Virginia Tech data landscape. International Journal of Digital Curation 10(2), 157-175 (2016)
- [5] Schober, D., et al.: NmrML: A community supported open data standard for the description, storage, and exchange of NMR data. Analytical Chemistry 90(1), 649-656 (2018)
- [6] Tenopir, C., et al.: Research data management services in academic research libraries and perceptions of librarians. Library & Information Science Research 36(2), 84-90 (2014)
- [7] Cox, A. M., Pinfield, S.: Research data management and libraries: Current activities and future priorities. Journal of Librarianship and Information Science 46(4), 299-316 (2014)
- [8] Kruse, F., & Thestrup, J. B.: Research libraries' new role in research data management, current trends and visions in Denmark. LIBER Quarterly 23(4), 310-335 (2014)
- [9] Chiware, E., Mathe, Z.: Academic libraries' role in research data management services: A South African perspective. South African Journal of Library and Information Science 81(2), 1-10 (2016)
- [10] Cox, A.M., et al.: Developments in research data management in academic libraries: Towards an understanding of research data service maturity. Journal of the Association for Information Science and Technology 68(9), 2182-2200 (2017)
- [11] Ayris, P., Ignat, T.: Defining the role of libraries in the Open Science landscape: A reflection on current European practice. Open Information Science 2(1), 1-22 (2018)
- [12] Hamad, F., Al-Fadel, M., Al-Soub, A.: Awareness of research data management services at academic libraries in Jordan: Roles, responsibilities and challenges. The New Review of Academic Librarianship, 1-21 (2019). Available at: <https://doi.org/10.1080/13614533.2019.1691027>. Accessed 22 December, 2020
- [13] Pampel, H., et al.: Making research data repositories visible: The re3data.org registry. PLoS ONE 8(11), e78080 (2013)
- [14] Moon, J.: Developing a research data management service -a case study. Partnership 9(1), 1-14 (2014)
- [15] Curdt, C.: Supporting the interdisciplinary, long-term research project 'patterns in soil-vegetation-atmosphere-systems' by data management services. Data Science Journal 18(5), 1-9 (2019)
- [16] Tripathi, M., Shukla, A., Sonkar, S. K.: Research data management practices in university libraries: A study. DESIDOC Journal of Library & Information Technology 37(6), 417 (2017)
- [17] Johnston, L., et al.: Data curation network: How do we compare? A snapshot of six academic library institutions' data repository and curation services. Journal of Escience Librarianship 6(1), e1102 (2017)
- [18] Lee, D.J., Stvilia, B.: Practices of research data curation in institutional repositories: A qualitative view from repository staff. PLoS ONE 12(3),

e0173987 (2017)

- [19] Curdt, C., Hoffmeister, D.: Research data management services for a multidisciplinary, collaborative research project. *Program: Electronic Library and Information Systems* 49(4), 494-512 (2015)
- [20] McKinney, B., et al.: Extension of research data repository system to support direct compute access to biomedical datasets: Enhancing Dataverse to support large datasets. *Annals of the New York Academy of Sciences* 1387(1), 95-104 (2017)
- [21] Mannheimer, S., et al.: Qualitative data sharing: Data repositories and academic libraries as key partners in addressing challenges. *The American Behavioral Scientist (Beverly Hills)* 63(5), 643-664 (2018)
- [22] Dovidonyte, R.: Implementation of Open Science in Lithuania. *Nordic perspectives on Open Science*, 1-12 (2019). Available at: <https://doi.org/10.7557/11.4828>. Accessed 22 December, 2020
- [23] Pontika, N.: Roles and jobs in the open research scholarly communications environment: Analysing job descriptions to predict future trends. *LIBER Quarterly* 29(1), 1-20 (2019)
- [24] Clare, C., et al.: Engaging researchers with data management: The cookbook. Open Book Publishers, Cambridge (2019). Available at <https://doi.org/10.11647/OBP.0185>.
- [25] Alonso-Arévalo, J.: Research data management (RDM) on the horizon of academics and research libraries. *Cuadernos De Documentación Multimedia* 30, 75-88 (2019)
- [26] Söderholm, M., Sunikka, A.: Collaboration in RDM activities -Practices and development at Aalto University. In: *The 12th Munin Conference on Scholarly Publishing*, pp. 1-10 (2017). Available at: <https://doi.org/10.7557/5.4247>. Accessed 22 December, 2020
- [27] Tang, R., Hu, Z.: Providing research data management (RDM) services in libraries: Preparedness, roles, challenges, and training for RDM practice. *Data and Information Management* 3(2), 84-101 (2019)
- [28] Zhu, L., et al.: (2016). The construction of Peking University open research data platform: Exploration and practice. *Library and Information Service* 60(4), 44-51 (2016) (in Chinese)

AUTHOR BIOGRAPHY

Hua Nie is Professor of Library Services at Peking University Library, China. Her research interests include open access, scholarly communication, digital humanities, and library marketing. ORCID: 0000-0002-4522-5049

Pengcheng Luo is Data Librarian at Peking University Library, China. His research interests include data management and open data. ORCID: 0000-0001-9598-0715

Ping Fu is Professor of Library Services and Science & Data Management Librarian at Central Washington University, USA. His research interests include integrated library systems, scholarly communication, and research data man-

agement. ORCID: 0000-0002-5386-748X

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv –Machine translation. Verify with original.