

Research on Library Research Data Management Services for First-Class University Construction: From the Perspective of DLC-SH Integration (Postprint)

Authors: Qin Shun

Date: 2023-04-01T00:00:00+00:00

Abstract

[Purpose/Significance] The development of research data management services in first-class university libraries is both a practical necessity for open scientific exchange and collaboration, and an objective requirement for research data sharing and utilization, while also representing an inevitable trend in adapting to the construction of a first-class higher education system. [Methodology/Process] Selecting representative first-class university libraries both domestically and internationally as samples, this study investigates the core supporting elements, primary cyclical processes and content, key stakeholders and their rights and responsibilities in the construction of their research data management services. It deeply explores the experiences and challenges in their service construction from an integrated Data Lifecycle-Stakeholder (DLC-SH) perspective. [Results/Conclusions] Based on the DLC-SH perspective, this paper proposes an innovative framework for library research data management services in China, oriented towards future first-class university construction. This framework is governed by interconnected dimensions including macro-level top-level design and policy planning, meso-level service pathway envisioning and foresight, and micro-level service system construction and implementation, thereby providing reference and guidance for subsequent research and practice.

Full Text

Abstract

[Purpose/Significance] The development of research data management services in first-class university libraries is not only a practical necessity for open scientific exchange and cooperation and an objective requirement for research data sharing and utilization, but also an inevitable trend in adapting to the

construction of a first-class higher education system. **[Method/Process]** This study selected representative first-class university libraries at home and abroad as samples, investigating their core supporting elements, main circulation processes and content, and key stakeholders and their rights and responsibilities in the construction of research data management services. From the perspective of integrating the data life cycle and stakeholders (DLC-SH), the paper explores in depth the experiences and problems in their service construction. **[Result/Conclusion]** Based on the DLC-SH perspective, this paper proposes a renewal framework for China's future library research data management services oriented toward first-class university construction, comprising interconnected aspects of macro-level top-level design and policy planning, meso-level service path conception and prospect, and micro-level service system construction and implementation, providing reference for subsequent research and practice.

Keywords: first-class university; research data management service; data life cycle; stakeholder

Classification Number: G250

DOI: 10.13266/j.issn.0252-3116.2021.04.004

Data-intensive research is emerging, and data-intensive scientific discovery has gradually developed into the fourth paradigm of scientific research [1-2]. As an important component of scientific research, research data serves as a crucial resource, evidence, and driving force supporting research discoveries in digital humanities, evidence-based library science, and other fields. At present, the scientific value of making research data findable, accessible, interoperable, and reusable (FAIR), along with data science and research data management services, has become a new theme and focus of international competition and cooperation. For instance, in November 2019, the Committee on Data for Science and Technology (CODATA) of the International Science Council released the *Beijing Declaration on Research Data* [3], affirming the progress of research data policies worldwide and articulating ten principles for advancing multilateral cooperation in relevant fields. Countries and regions such as the United States and the United Kingdom have established research data policies and supporting management and service systems at national, funding organization, research institution, library, and data center levels. In China, the *Measures for the Management of Scientific Data* issued by the State Council in March 2018 [4] elevated scientific data management and sharing to a national-level systematic layout, establishing a “graded management, division of responsibilities” system. The measures clearly stipulate that “universities and other legal entities are the responsible subjects for scientific data management,” bearing responsibilities for optimizing scientific data management services and accepting social evaluation and supervision. This has fostered a two-dimensional architecture and system combining the data life cycle (DLC) and stakeholders (SH), which this paper terms the “DLC-SH perspective.”

In the era of data science, the industry must recognize the driving forces behind the innovation of library research data management services oriented toward

first-class university construction: it is both the practical need for open scientific exchange and cooperation and the objective requirement for research data sharing and utilization, as well as the inevitable trend of adapting to first-class higher education system construction. University libraries have long been natural institutions for storing, disseminating, and utilizing data, information, and knowledge resources. They possess certain advantages in research data resource construction, management optimization, and service provision, and shoulder the mission and responsibility of research data management services. Using them as the core subject for research data management service construction offers a favorable policy environment, theoretical support, and practical foundation, making it feasible to use them as the starting point, breakthrough point, and foothold for promoting first-class university construction. Based on this, this paper takes university libraries—the core stakeholders of research data management services—as the main subject, examining and revealing the landscape and paths that need exploration in the practice of research data management services in first-class university libraries at home and abroad from the DLC-SH perspective, with the aim of providing reference for their service innovation and high-quality development.

1. Literature Review and Brief Commentary

Research data management services represent a pathway for scientific research to transform toward open access, open data, open science, and other new research models [6]. With the continuous improvement of the data science and open science-driven environment, and with the increasingly prominent core position and enabling role of data, the importance of research data management services has become more pronounced, and data-enabled library science research and practice have continued to gain momentum. Using multi-source data from Google Dataset, Harvard Dataverse, and Figshare as evidence to analyze current international library science research characteristics and frontier hotspots, the results show that research data management and sharing will further become a focus of attention for society and the library community, as well as a persistent hotspot in library science research and practice [7].

Literature review reveals that current domestic and international research is abundant but lacks deep integration of perspectives, concentrating mainly on two viewpoints: First, the data life cycle perspective, such as Zhou Lihong and H.A. Piracha, who summarized practical experiences of university libraries in research data management service processes and proposed targeted recommendations based on case analysis and surveys [8-10]. Second, the stakeholder perspective, focusing on exploring the role positioning of university libraries in research data management services [11-12], data librarian team building and training [13-14], and data literacy education and user training [15-16]. Research in the library and information science field combining data life cycle and stakeholder perspectives mainly involves analysis of influencing factors in research data sharing and interpretation of their roles, as well as the construction of

research data management policy frameworks [17-18], with weak connections to research and practice on research data management services in first-class university libraries. Notably, adopting the integrated DLC-SH perspective as a theoretical foundation to explore university library research data management service construction is beneficial for promoting their development toward first-class, high-quality directions from an integrated viewpoint.

In summary, scholars at home and abroad have conducted detailed analyses of cognitive issues in university library research data management services from different perspectives and dimensions, but several problems remain: First, the conceptual confusion among scientific data, research data, and research data needs to be clarified. The scopes of these three concepts have certain overlaps; scientific data is strongly correlated with scientific and technological development, while research data and research data are more strongly correlated with research discoveries. The latter two can both be translated as “Research Data.” The business scope of university libraries is to develop embedded research data management services oriented toward research subjects and covering the full research life cycle. Therefore, this paper collectively refers to them as research data, while retaining original expressions when referencing peer scholars. Second, research on university library research data management services is still in its infancy, with few studies exploring the overall patterns of library research data management services for first-class university construction from the DLC-SH perspective. Literature review shows that research on research data management services oriented toward first-class university construction has established a certain foundation. For example, Chu Jiewang et al. consider scientific data services as an important pathway for university libraries to achieve the “Double First-Class” construction strategy [19]; Cui Haiyuan et al., using Peking University Library as an example, explored first-class research data management services supporting user needs for data-driven research paradigms, including practical aspects such as research outcomes, framework design, construction experience, and service effectiveness [20]; Guo Jiaxin et al., by analyzing typical Data Curation (DC) projects and practical activities abroad, proposed new ideas for scientific data management in China’s “Double First-Class” university libraries from three aspects: management objectives, platform construction, and team building [21]. The above research points the way for this paper to expand to multi-sample, multi-dimensional research covering diverse stakeholders and the full life cycle process of first-class university library research data management service construction.

2. Investigation and Analysis of Current Status of Research Data Management Services in First-Class University Libraries

2.1 Survey Design

The survey samples for this study were representative first-class university libraries at home and abroad. For the international sample, we selected the libraries of the top 30 world-class universities from the 2019 Academic Ranking of World Universities (ARWU) published by Shanghai Ranking [22]. For China, we selected the libraries of the 42 “Double First-Class” construction universities from the list published by the Ministry of Education, Ministry of Finance, and National Development and Reform Commission [23]. Both groups are collectively referred to as first-class university libraries. The rationale for sample selection is that the construction of research data management services in these libraries is inseparable from first-class university construction. On the one hand, both lists use indices such as scientific research, technology R&D, and cultural leadership as meta-indicators with relatively high weights [22-23]. On the other hand, preliminary research shows that first-class university libraries tend to have relatively developed research data management services with rich practical experience. For instance, libraries at first-class universities such as Harvard University, Stanford University, Peking University, and Wuhan University have all formed relatively scientific and reasonable service processes and systems.

The focus of the integrated DLC-SH perspective is to clarify the division of responsibilities among stakeholders in each stage of the full data life cycle. According to statistics from the “Data Life Cycle Models and Concepts” published by the Committee on Earth Observation Satellites (CEOS) in April 2012 [24], there are already more than 55 different types of data life cycle models globally. Based on preliminary research and drawing on main components of internationally renowned data life cycle models and Sheng Xiaoping et al.’s classification of stakeholders in scientific data open sharing [25], this paper constructs an integrated DLC-SH perspective-based university library research data management service model (see Figure 1 [Figure 1: see original paper]).

As shown in Figure 1, the integrated DLC-SH perspective-based university library research data management service model involves 4 stages, 8 processes, and several stakeholders. The research approach is as follows: Survey indicators cover core supporting elements, mainly including relevant policies, technical tools, and talent teams; main service circulation processes and content from the data life cycle perspective, namely 4 stages and 8 processes of planning, management, sharing, and utilization; and key stakeholders and their rights and responsibilities from the stakeholder perspective, such as sorting out stakeholders oriented to user needs and their service provision forms. From the integrated DLC-SH perspective, using methods such as web surveys and literature research, and obtaining relevant information through search engines, site

search boxes, emails, and telephone consultations, this paper summarizes and concludes various indicator data and content to provide evidence support for the research.

2.2 Survey Results Analysis

2.2.1 Analysis of Core Supporting Elements Core supporting elements such as research data policies, technical tool resources, and talent teams are the basic conditions for the development of research data management services in first-class university libraries. Optimizing these core supporting elements from the DLC-SH perspective is conducive to achieving the goal of high-quality development of research data management services linking multiple stakeholders and covering the full data life cycle process.

- (1) The formulation and implementation of research data policies provide a legal environment guarantee for the construction of research data management services in first-class university libraries. The survey shows that a total of 18 research data policies were formulated by 16 domestic and foreign university libraries in the samples, with only Wuhan University in China formulating the *Standards and Specifications for Scientific Data Management in Universities* [26]. The core stakeholders in policy formulation are universities and their libraries, with common naming conventions including service guidelines and management, sharing, and publishing policies. The published policy frameworks and content have the following characteristics: They tend to organize policies from the data life cycle perspective, involving business processes such as data creation and organization, preservation and archiving, publishing and sharing, analysis and citation, security and supervision, which basically match the integrated DLC-SH perspective-based university library research data management service model; Formulating research data policies from the DLC-SH perspective has become a trend, with conclusions similar to Xing Wenming et al.'s research findings [18], namely that in addition to work content within data life cycle stages, they also involve core stakeholders and their division of rights and responsibilities. For example, the *Imperial College Research Data Management Policy* [27] clearly specifies the library's responsibilities and roles in providing research data management literacy training, general DMP templates, and data repositories; To some extent, they neglect the rights and responsibilities statements and policy norms for libraries as the most important stakeholders, with only the Oxford University Library formulating the *Oxford University Library Research Data, Records Management and Preservation Policy*, which specifies specific responsibilities such as data storage and data literacy training (e.g., ORA-Data service training) [28].
- (2) Technical tools provide software support for the construction of research data management services in first-class university libraries under digital transformation. Based on the 10 major research stages divided by the

Center for Open Science (COS) “Open Science Framework” [29] and the research data management service model constructed in this paper, the author sorted out the provision of more than 520 types of technical platforms and over 320 types of tool resources by domestic and foreign first-class university libraries. Representative applications are shown in Table 1 and Table 2 . The applications have the following characteristics: In terms of technical platforms, they mainly include four forms: disciplinary data repositories, general storage platforms, cloud backup platforms, and self-developed platforms. Platform development and application in foreign sample libraries are more extensive, with advantages such as good application environments, diverse types, and coexistence of personalization and standardization. In China, platform construction and application are relatively lagging, characterized by “few platforms that have not yet formed scale,” with deficiencies such as single funding sources, lack of open source and cooperation concepts, and incomplete platform service functions [30]. In terms of tool resources, foreign supply focuses on embedding the full research life cycle and offers rich types. Based on data life cycle stages and main functions, they can be categorized into tools for basic environments, DMP and data collection, description, naming, transfer, security, publishing, cleaning, mining, analysis, visualization, citation, and verification. Domestic tool resource development and application are limited to data analysis, visualization, and citation, with problems such as weak open source awareness, partial application that deviates from the data life cycle process, and an urgent need to clarify the relationship between “more and comprehensive” versus “less and refined.”

- (3) Talent team building is also an important foundation for the development of research data management services in first-class university libraries. Among them, data librarians, platform staff, and peer reviewers are the most critical stakeholders. The May 2017 report *SPEC Kit 354: Data Curation* released by the Association of Research Libraries (ARL) pointed out: “The staff quality level of 124 ARL member libraries has some mismatch with their interest in and demand for data management services, with most libraries not having dedicated data management positions” [31]. The survey results on data management position setting, structure, and responsibilities in the samples basically align with the ARL report conclusions. Data analysis reveals that a total of 24 libraries in the sample clearly established positions in full-time (13 libraries), part-time (15 libraries), or parallel forms, with 7 adopting the method of linking subject librarians, mainly providing services such as DMP writing assistance, data storage and organization, and data literacy cultivation. In Chinese samples, library data librarian allocation is even more scarce, with relatively simple position responsibilities. For example, only three first-class university libraries—Peking University, Wuhan University, and Fudan University—explicitly employ data librarians responsible for information literacy and data consultation services [8, 20]. Additionally, domestic and foreign

technical platforms mainly link to disciplinary data repositories and general repositories, and many university libraries self-develop or secondary-develop technical platforms based on open source or commercial software, usually equipped with professional staff and peer reviewers responsible for full life cycle research data management services. For example, Harvard University' s "Dataverse" project, which is favored by the international data management field and has become a platform for many university libraries, is developed based on an open source architecture, in collaboration with Simon Fraser University and Stanford University to build the Public Knowledge Project, and works with its developers and data librarians to develop and maintain the "OJS Dataverse" plugin to support double-blind peer review, editing, publishing, and dissemination of journal articles and accompanying research data [32]; platform staff in China relatively lack the ability to assist in DMP formulation and data review.

2.2.2 Analysis of Main Circulation Processes and Content The main circulation processes of library research data management services oriented toward first-class university construction include four stages and eight processes: planning, management, sharing, and utilization. The following analysis focuses on the core stakeholder—university libraries—and their service circulation processes and main content.

- (1) **Planning Stage.** The main implementation method in the planning stage is formulating a Data Management Plan (DMP) to guide all or part of the data life cycle management work. A DMP describes the plan for handling data during a research project from implementation to completion, including stakeholders such as libraries and researchers; a DMP submitted for peer review can serve as a primary indicator and basis for evaluating research projects [33]. The survey shows (see Table 3) that all 28 foreign sample libraries clearly provide DMP services, while domestic exploration of DMP content elements is almost nonexistent [34]. Foreign sample libraries mainly provide DMP services through tools and templates, with main components including data collection, organization and description, storage and backup, sharing and reuse, rights and security, and DMP roles and responsibilities.
- (2) **Management Stage.** The core task of the management stage is data creation and storage. The survey results show (see Table 4) that related services in foreign libraries have richer connotations, while China mainly focuses on providing data storage services. The core points of research data creation include naming and organization, metadata description, format specification, and version control. The focus of research data storage lies in classification level specification, security review, data repositories, data backup, and data transfer. For example, Harvard University Library' s Research Data Management @ Harvard project focuses on data collection, storage, and security, and provides best practice guidance and tool support

[35].

- (3) **Sharing Stage.** In the sharing stage, clarifying the open publishing model for university library research data sharing is a prerequisite for data analysis and reuse. Based on the current practice of first-class university libraries at home and abroad, as well as A. Sarretta' s *Research Data Life Cycle* [36] and A. Pablo et al.' s *LERU Roadmap for Research Data: Research Workflow & Data Publishing Life Cycle Model* [37], this paper optimizes a research workflow-oriented research data publishing and sharing life cycle model for first-class university libraries (see Figure 2 [Figure 2: see original paper]). The survey found that university library research data publishing and sharing services mainly involve processes such as data collection, description and organization, publishing and distribution, and reuse and evaluation. Under optimal conditions, these should run through the three processes of research, publishing, and supervision. However, currently few sample libraries at home and abroad comprehensively involve this entire process. University library research data publishing and sharing services mainly have three models: data storage and sharing, publishing and distribution, and archiving and sharing.
- (4) **Utilization Stage.** The utilization stage mainly provides data analysis and reuse, evaluation and verification services for research users. The former aims to maximize data value, while the latter helps purify the academic environment and feeds back into research output [8]. The survey shows (see Table 5) that domestic and foreign sample libraries' provision of research data analysis and reuse services concentrates on specialized services such as analysis and statistics, visualization, retrieval and discovery, and citation. Among them, analysis and statistics and visualization services emphasize tool resource provision and service training and consultation; all first-class university libraries with technical platforms provide retrieval and discovery services; data citation services place greater emphasis on establishing relevant guidelines, standards, and norms, as well as developing and applying data citation tools. For example, MIT Libraries' *Citing Sources: Data* includes relevant reference standards for citing data resources [39].

Research data evaluation and verification services include the following approaches: collaborating with Institutional Review Boards (IRB), Data Trust Council (DTC), and other organizations to evaluate and review whether research data complies with ethical norms, such as Johns Hopkins University Library' s establishment of the JHM Data Trust Council and issuance of data sharing policies [40]; constructing quality assurance and control frameworks based on factors such as completeness, validity, and accuracy, and providing relevant guidelines and standards, such as Yale University Library' s *Framework for Assessing Data Quality in Clinical Research Datasets* [41]; using DMPs to evaluate and test data to identify quality issues during data collection and preservation, such as using DMP Tool and RSAM systems to assess risks and compliance in data

management and sharing processes for risk identification and appropriate regulatory strategies; and emphasizing training in research data verification, such as Cambridge University and Yale University libraries' attention to data archiving review and belief that libraries should participate in verifying the integrity of research data [8].

2.2.3 Analysis of Key Stakeholders and Their Rights and Responsibilities As mentioned earlier, the core stakeholders in first-class university library research data management service construction include government, research and funding agencies, universities and their research management departments, libraries and archives, data centers, publishers and platforms, researchers, and service personnel, who bear different rights and responsibilities in policy formulation, data life cycle process management, technical tool development and application, and talent team building (see Table 6).

From the perspective of service provision and utilization, data librarians and researchers are the most important stakeholders. The survey shows that the main forms of service provision by first-class university libraries oriented to research user needs include reference consultation, data literacy cultivation, and service quality control, which are guaranteed through the data librarian system. Typical cases of reference consultation services include Columbia University Libraries' implementation of *Research Data Services Consultation Guidelines* [42]; the Digital Scholarship Section (DSS) of the Association of College and Research Libraries (ACRL) established collaborative research data reference consultation services through the "Research Data Q" community Q&A platform [43]. In terms of data literacy cultivation, foreign service patterns show characteristics of "demand-oriented and focused," mainly including course training and workshops (25 libraries) and online training and tutorials (9 libraries). For example, some libraries link to the Research Data Management Librarian Academy (RDMLA), providing training on research data management fundamentals, research data culture navigation, data analysis and visualization tool overviews, and coding tool introductions [44]. In contrast, although Chinese sample libraries attach high importance to data literacy cultivation, they suffer from problems such as complicated training forms but insufficient depth, failure to form integrated training systems and regular training mechanisms, and relatively neglecting data librarian talent cultivation and literacy training [15]. Additionally, introducing Data Quality Control Plans and Checklist Management Systems are major innovative measures by foreign first-class university libraries to promote research data management service quality control, focusing on data quality control and management quality control respectively, while exploration in this area is relatively rare in China.

3. Innovation Framework for First-Class University Library Research Data Management Services

In the new open society, data has become an important production factor driving innovation, research, and social governance [45]. Through investigation, this paper summarizes some achievements in the construction of research data management services in some first-class university libraries at home and abroad from the DLC-SH perspective: the legal environment is continuously improving, and service provision is gradually diversifying; technology leads data enablement, with novel and diverse service forms; personnel team allocation is relatively scientific, and service rights and responsibilities division is basically reasonable; multiple measures are taken for literacy cultivation, and multi-pronged approaches are implemented for service supervision.

Meanwhile, some global and non-negligible problems also exist, manifested in the partial disconnect between policies and services, with unclear core positioning of stakeholders; one-sided technology application, with the relationship between personalization and standardization needing clarification; personnel quantity and quality still not meeting standards, with mismatches between team allocation and service demand; backward training and supervision mechanisms, with an urgent need to improve service quality evaluation systems. Overall, the construction of research data management services in Chinese university libraries is relatively weak, with problems such as unclear “first-class” goal positioning, failure to embed into the research data life cycle and associate with multiple stakeholders, few relevant practical cases, scarce technical platforms and tool resources, and imperfect standards, norms, and talent teams [21].

Library renewal must consider environmental changes, fundamental development, and innovative strategies [46]. The realization of university libraries’ contribution to the “Double First-Class” university construction strategy cannot be separated from first-class university library services [19]. In view of this, Chinese university library research data management service construction should seize the opportunity. Based on the constructed research data management service model, the division of key stakeholders and their rights and responsibilities (see Figure 1 and Table 6), and survey analysis, this paper argues that the integrated DLC-SH perspective should serve as the theoretical and practical basis, benchmarking the “Double First-Class” construction goals, and actively planning from three interconnected aspects: macro-level policy formulation and improvement, meso-level service path conception and prospect, and micro-level service system construction and implementation, to boost the renewal, transformation, and development of first-class university library research data management services.

3.1 Macro Level: Research Data Policy Formulation and Improvement

3.1.1 Formulating Reference Frameworks for University and Library Research Data Policies

Formulating and implementing standardized and

normalized policies helps ensure controllable quality of university library research data management services. It should be a priority for governments, universities, and their libraries to lead the formulation of reference frameworks for research data policies to guide the formulation and improvement of implementation rules and provide established reference standards. This is not only a clear requirement of the *Measures* but also the correct development direction guided by practical experience in Europe, America, and other regions. For example, Article 9(2) of the *Measures* stipulates: “Universities and other legal entities have the responsibility to establish and improve their own scientific data-related management systems” [4]; the League of European Research Universities (LERU) released policies such as the *LERU Statement on Open Research Data* [47] and *LERU Roadmap for Research Data* [48] to guide and standardize participation by research, education, and information institutions in research data management service construction. Drawing on representative policy outlines/templates such as the *Outline of Research Data Management Policy for Australian Universities/Institutions* [49], the UK’ s *Five Steps to Developing a Research Data Policy* [50], and Switzerland’ s *Research Data Management Policy Template* [51], as well as the research data policy frameworks of sample university libraries, this paper proposes a reference framework for Chinese university and library research data policies from the DLC-SH perspective (see Table 7).

3.1.2 Ensuring Smooth Implementation of University and Library Research Data Policies

On the one hand, reference frameworks for university and library research data policies should be used to construct standardized and normalized implementation policies and mechanisms from the DLC-SH perspective. In research data management and sharing, legal and policy frameworks for the full data life cycle process can be strengthened at the macro level, with specialized implementation groups or committees established to guide, evaluate, and supervise university library research data management and sharing work. Policies should clarify the core position, role, and specific rights and responsibilities of university libraries in the research data management service construction process to meet first-class university construction needs. On the other hand, research data management and service-driven policies should be formulated and implemented, focusing on identifying, breaking through, and regulating obstacles and risks in research data management and sharing from the DLC-SH perspective, supporting first-class university library research data management service construction. The policy environment for collaborative driving of personalized and standardized services should be improved to promote scientific openness and technological open source, empowering the construction of more open, diverse, inclusive, and competitive first-class university library research data management services.

3.2 Meso Level: Service Path Conception and Prospect

3.2.1 “Motivation and Rights/Responsibilities” Support: Adjusting and Improving Service Division

First-class university libraries at home

and abroad usually clarify work priorities at each data life cycle stage and standardize the division of rights and responsibilities among stakeholders in research data policies, management systems, and data librarian job descriptions. The motivation for research data management service construction comes from people, with clear rights and responsibilities as the basic guarantee. Therefore, future innovation in Chinese library research data management services oriented toward first-class university construction should be supported by “motivation and rights/responsibilities” to adjust and improve the division of research data management services. Explorable paths include:

First, actively clarifying the rights, responsibilities, and division of labor among stakeholders in the research data life cycle process, especially the core positioning and rights/responsibilities of university libraries. On the one hand, specific responsibilities of stakeholders centered on university libraries can be standardized in Research Data Management Service Guides (LibGuides). On the other hand, the specific responsibilities and division of labor for service personnel such as data librarians, platform staff, and peer reviewers at each data life cycle stage should be standardized.

Second, adjusting and improving missing or weak research data management service processes and stakeholder division based on the data life cycle model to help achieve seamless connection of the full research data management service life cycle process, overcoming the one-sidedness of Chinese university library research data management service construction oriented toward the full data life cycle process.

3.2.2 “Humanities and Technology” Drive: Integrating and Reconstructing Service Processes First-class university library research data management service construction is far from being a single technology “wisdom.” Falling into the narrow domain of “technology supremacy” will lead to unsustainable research data management and service construction. Therefore, future innovation in Chinese library research data management services should be driven by “humanities and technology” to integrate and reconstruct service processes. The design concept of “service-oriented, technology-enabled” should be adhered to, continuously integrating and reconstructing service processes covering the DLC-SH perspective.

First, uphold the value pursuit of “goodness and excellence” in concept. Emphasize the cultivation of data, information, and humanistic literacy for both research users and service personnel, permeating the concept of “people-oriented, service-first” throughout the data life cycle process to supply data-driven services and continuously enrich service content. Simultaneously, “technology as a tool” should not be neglected—attention should be paid to the development and application of technical platforms and tool resources covering the full data life cycle process.

Second, emphasize business process reengineering for university library research

data management. University libraries and other stakeholders should actively learn from foreign experiences, follow principles such as embedding into the full spectrum of research processes and full data life cycle flow, placing users at the center of research data management business processes, and prioritizing the reengineering of key research data management processes. They should transform from “document flow” to “data flow” and “service flow” to optimally configure and link business processes such as university library research data management, sharing, and reuse. Meanwhile, university libraries and other stakeholders should update service processes and architectural design according to the purposes of open scientific exchange and cooperation, research data sharing and utilization, and adaptation to first-class higher education system construction.

3.2.3 “Demand and Innovation” Leadership: Organizing and Reshaping Service Patterns Facing the development goals of first-class university construction, “demand and innovation” will remain the core driving force for continuously leading, organizing, and reshaping the service patterns of Chinese university library research data management service construction. Chinese university libraries should continuously organize and reshape the “trinity” service elements of data resources, stakeholders, and data services, driving continuous innovation and research communication through demand discovery.

First, elevate the supply and utilization of research data resources to a strategic level. Avoid falling into the shallow “pushing resources” trap; explore using artificial intelligence to generate metadata, and employ ontology and semantic web technologies for metadata creation, linked data, and knowledge discovery analysis to achieve fine-grained data and knowledge unlocking and organization, deeply revealing and mining the knowledge and data mechanisms of data, literature, and other information resources [7], and effectively improving the quality of data resource supply.

Second, conduct demand cognition surveys and analysis of research data management services for multiple stakeholders and users from different disciplines to deeply grasp the data needs and behavioral manifestations of various stakeholders, such as referencing the ARL’ s detailed analysis of the match between data librarians’ quality levels and their interest in and demand for data management services [31].

Finally, innovate forms of research data management services. Innovation in Chinese library research data management services oriented toward first-class university construction should enhance the development level of core supporting elements from multiple aspects including legal policies, resources, management, technology, and talent, reshape service mechanisms from the integrated DLC-SH perspective, and drive services toward high-quality innovative development, demonstrating the value of libraries as “data centers” and “data spaces” for co-located and cross-domain collaboration.

3.3 Micro Level: Service System Overall Construction and Implementation

3.3.1 Enriching Service Content: Emphasizing Literacy Training and Cognitive Promotion Micro-level institutional research data management service content should at least involve the 4 stages and 8 processes of the integrated DLC-SH perspective-based university library research data management service model constructed in this paper.

In the planning stage, DMP formulation should be elevated to strategic deployment and the work agenda. Explorations can be made to formulate and implement universal DMP guidelines, introduce and develop DMP tools adapted to the specific needs of Chinese university library research data management services. Researchers can also follow W. K. Michener's *Ten Simple Rules for Creating a Good Data Management Plan* for DMP writing [33].

In the management stage, scientific basic work should be done well for research data creation and storage, including naming and organization, metadata description, format specification, version control, and data classification level specification, security review, data repositories, data backup, and data transfer. For example, multiple stakeholders including university libraries can refer to E. M. Hart et al.'s *Ten Simple Rules for Digital Data Storage* for data creation and storage [52]. Given the dynamic and composite nature of data, university libraries can consider matching it with dataset semantic vectors, using timestamps to mark the stage of the data in the life cycle, and analyzing logical relationships between datasets and within different data elements according to dataset granularity standards [8].

In the sharing stage, key issues such as data integration, identification, citation, and review in the process of university library research data OA publishing and sharing should be emphasized to build a data publishing ecosystem and quality control system [38, 53]. One-stop data integration should be achieved to realize data sequential organization and promote data reuse and scientific research innovation. Reference should be made to the national standard *Information Technology—Scientific Data Citation* (GB/T 35294-2017) [54] to assign unique permanent identifiers to data, adopting a strategy that emphasizes both data identification and author identification. The formulation of university library research data citation guidelines and standards and the development and application of related citation tools should be strengthened to ensure clear intellectual property rights. Intelligent review after data publication should be conducted using automated means [55] to ensure data quality is controllable.

In the utilization stage, professional data analysis and mining, evaluation and verification services should be provided. A unified data integration and retrieval platform can be established jointly with the university library community, such as learning from Google Dataset's practice of building an open data ecosystem to achieve intelligent connection, knowledge graph coordination, and intelligent ranking and retrieval of data [56]. Alternatively, technologies such as the OAI-

PMH protocol, Multi-Agent cross-database retrieval, and semantic association can be used to achieve data integration organization, harvesting and exchange, and unified retrieval. Based on data librarians, precise research data analysis and reuse services embedded in research/teaching teams and workflows should be provided. Data quality evaluation and verification work should be done well to ensure that research data and services provided by university libraries have enabling, empowering, and evidentiary functions.

In addition to enriching service content, innovation in library research data management services oriented toward first-class university construction also needs to emphasize literacy training and cognitive promotion. University libraries should strengthen training on specific business within the above processes to keep pace with global development, enabling business work to adapt to user needs and allowing researchers and other stakeholders to understand and support service work. Cognitive promotion and library image building should be emphasized through news media, social networks, and other forms to promote university library research data management services through combinations of reality and virtuality, intra-library and inter-library, and domestic and international channels.

3.3.2 Expanding Service Approaches: Strengthening Technical Tool Development and Application

- (1) **Establish a Rational Open Source Concept and Strengthen Multi-party Incentive Mechanisms.** The Chinese university library community should uphold a rational open mindset, adopt an “open, low-barrier” cooperation model while strengthening independent research and development capabilities for scientific openness and technological open source, and independently build a new open source ecosystem. Technology should be used wisely and guided toward goodness, emphasizing instrumental rationality and value return to achieve autonomous control of data openness and technological open source [57]. Multi-party incentive mechanisms for technical tool development and application should be strengthened, with funding, software and hardware facilities, performance evaluation support, and policy incentives driving the maintenance and updating of technical platforms and tool resources. Review and supervision mechanisms should be established to control technical tool development and application from multiple angles.
- (2) **Focus on Data Life Cycle and Functional Integrity in Technical Platform and Tool Development and Application.** Based on clarifying the relationships between “personalization” and “standardization” and between “more and comprehensive” and “less and refined,” focus should be placed on driving technical tool development and application from the perspective of data life cycle and functional integrity. First, emphasize functional integrity in self-developed or secondary-developed platforms, and conduct system selection based on system characteristics and

first-class university and discipline construction needs. Second, conduct scientific positioning, set limited goals, and implement gradually. Third, develop and apply technical tools based on the data life cycle process, pursuing “specialization and refinement” in functions on the basis of ensuring “completeness” in types, and encouraging multiple stakeholders such as university libraries to develop and apply tool resources for DMPs and data collection (including metadata harvesting and exchange), publishing, citation, etc. Finally, do well in the selection of technical tools, selectively providing technical platforms and tool resources adapted to multi-stage and multi-disciplinary needs, along with relevant usage training tutorials. In short, traditional reference consultation and subject service methods should be combined with technical platform and tool development and application to expand service approaches.

3.3.3 Integrating Service Teams: Broadening the Breadth and Depth of Embedded Services

- (1) **Properly Handle the Relationship Between “Quantity” and “Quality” in Service Team Building.** To adapt to the development needs of universities’ “large disciplines” and “small disciplines,” an appropriate number of high-quality subject data librarians should be allocated, and an embedded data librarian service system in research workflows and research data life cycles should be explored. A high-level technical tool development and service team should be built with funding and software and hardware facility support. Attention should be paid to peer reviewers’ scientific evaluation, knowledge reserves in data structure and metadata standards, and their time and energy costs. The core competitiveness of data librarians, platform staff, and peer reviewers should be enhanced through education, training, and promoting the self-development of professional groups to improve user satisfaction [58].
- (2) **Achieve Scientific Position Allocation and Clear Rights/Responsibilities for Research Data Management Positions.** Specifically, the specialization of data librarian positions needs to be promoted, with flexible data management position settings tailored to each library’s circumstances. The “embedded,” “fluid,” and “collaborative” characteristics of research data management service team building should be grasped to promote data librarians becoming high-level academic partners. Attention should be paid to the diverse disciplinary backgrounds of service personnel and their scientific allocation and rights/responsibilities definition, with professional skills and competency training conducted according to job responsibilities, emphasizing talent cultivation and absorption oriented to service demand [59-60].

3.3.4 Strengthening Service Supervision: Promoting Standardized Evaluation and Service Effectiveness Improvement

Service supervision

and evaluation can identify weak links, optimize research data management service construction paths, and promote virtuous cycles in research data management services. The supervision and evaluation of first-class university library research data management services in China is a relatively weak link in the data life cycle stage. An explorable path is combining self-evaluation with multi-stakeholder supervision and evaluation. Internationally, existing research data management service self-evaluation models mainly include the CMM model, DMM model, and RISE-DE evaluation model. In China, Dang Hongli et al. referenced the DMM model to measure library data management capabilities from five aspects: basic capability and data service, security, technology, and environment capabilities [61-62]. China's first-class university construction-oriented research data management service evaluation can learn from the more advanced RISE-DE self-evaluation model to build a maturity evaluation model and evaluation indicator system suitable for national and institutional conditions, identifying gaps, addressing weaknesses, and promoting innovation according to first-class university construction needs. Additionally, on the basis of introducing Data Quality Control Plans and Checklist Management Systems, attention should be paid to combining them with multi-stakeholder supervision and evaluation methods to promote standardized evaluation and service effectiveness improvement.

References

- [1] CODATA China Committee. *Big Data Era Research Activities* [M]. Beijing: Science Press, 2014: 2-18.
- [2] Pan Jiaofeng, Zhang Xiaolin. *The Fourth Paradigm: Data-Intensive Scientific Discovery* [M]. Beijing: Science Press, 2012: 1-5.
- [3] CODATA. *The Beijing Declaration on Research Data* [EB/OL]. [2020-06-16]. <http://www.codata.org/uploads/Beijing%20Declaration-19-11-07-FINAL.pdf>.
- [4] General Office of the State Council. *Notice on Issuing the Measures for the Management of Scientific Data* [EB/OL]. [2020-06-25]. http://www.gov.cn/zhengce/content/2018-04/02/content_{5279272}.htm.
- [5] Xing Wenming, Hong Fanglin, Li Xiaoyan. A Two-Dimensional Perspective on Scientific Data Management Systems—Interpretation of the Measures for the Management of Scientific Data [J]. *Library and Information Service*, 2019, 63(23): 30-37.
- [6] Gu Liping. Data Management Services in the Transformation of Research Models: A Pathway to Achieving Open Access, Open Data, and Open Science [J]. *Journal of Library Science in China*, 2018, 44(6): 43-55.
- [7] Qin Shun, Liu Kunxiong, Zhao Wei. Characteristics and Frontiers of International Library Science Research Enabled by Data—Evidence from Google Dataset [J]. *Library and Information*, 2019(4): 74-83.
- [8] *Journal of Library and Information Service Magazine*. *Research and Practice of Data Management* [M]. Beijing: Ocean Press, 2018: 3-356.
- [9] Zhou Lihong, Duan Xinyu, Song Yaqian. Investigation and Analysis of Research Data Management Services in Chinese University Libraries [J].

Library and Information Service, 2017, 61(20): 77-86.

[10] Piracha HA, Ameen K. Policy and Planning of Research Data Management in University Libraries of Pakistan [J]. *Collection and Curation*, 2019, 8(2): 39-44.

[11] Perrier L, Blondal E, Macdonald H. Exploring the Experiences of Academic Libraries with Research Data Management: A Meta-Ethnographic Analysis of Qualitative Studies [J]. *Library & Information Science Research*, 2018, 40(3/4): 173-183.

[12] Cox AM, Kennan MA, Lyon L, et al. Maturing Research Data Services and the Transformation of Academic Libraries [J]. *Journal of Documentation*, 2019, 75(6): 1432-1462.

[13] Zuo Simin, Zhu Jiaqi, Liang Yiming. Role Transformation of University Librarians Driven by Open Scientific Data [J]. *Library Development*, 2018(11): 23-27, 35.

[14] Nitecki DA, Davis MEK. Expanding Academic Librarians' Roles in the Research Life Cycle [J]. *Libri*, 2019, 69(2): 117-125.

[15] Song Jiali, Song Chunjian, Cheng Jiejing. Investigation and Analysis of University Library Users' Scientific Data Literacy Capabilities and Needs [J]. *Library Science Research*, 2019(5): 56-70.

[16] Oliver JC, Kollen C, Hickson B, et al. Data Science Support at the Academic Library [J]. *Journal of Library Administration*, 2019, 59(3): 241-257.

[17] Liu Guifeng, Pu Jingrong, Qian Jinlin. Analysis and Interpretation of Influencing Factors in Research Data Sharing [J]. *Library Tribune*, 2018, 38(11): 10-17, 26.

[18] Xing Wenming, Tang Yajing, Qin Shun. Interpretation and Enlightenment of Foreign Educational Institution Research Data Management Policy Outlines [J]. *Digital Library Forum*, 2019(5): 9-16.

[19] Chu Jiewang, Zhang Yu, Liu Qingqing. Strategic Thinking on University Libraries' "Double First-Class" Construction [J]. *Journal of Academic Libraries*, 2019, 37(1): 6-16.

[20] Cui Haiyuan, Luo Pengcheng, Li Guojun, et al. First-Class University Research Data Management Service System Research and Construction—Taking Peking University Library as an Example [J]. *Journal of Academic Libraries*, 2019, 37(2): 42-48.

[21] Guo Jiaxin, Fan Xin. Foreign Scientific Data Management Experience and Its Enlightenment for China's "Double First-Class" University Libraries [J]. *Journal of Documentation and Data Studies*, 2019, 1(3): 26-37.

[22] Shanghai Ranking Consultancy. Academic Ranking of World Universities 2019 [EB/OL]. [2020-06-10]. <http://www.shanghairanking.com/ARWU2019.html>.

[23] Ministry of Education, Ministry of Finance, National Development and Reform Commission. Notice on Publishing the List of World-Class Universities and First-Class Discipline Construction Universities and Disciplines [EB/OL]. [2020-06-10]. http://www.moe.gov.cn/srcsite/A22/moe_{843}/201709/t20170921_{314942}.html.

[24] CEOS, WGISS, DSIG. Data Life Cycle Models and Concepts [EB/OL].

[2020-06-16]. <https://my.usgs.gov/confluence/download/attachments/82935852/Data%20Life%20Cycle%20Model.pdf>.

[25] Sheng Xiaoping, Wang Yi. Responsibilities and Roles of Stakeholders in

- Scientific Data Open Sharing—Analysis Based on International Organization Scientific Data Open Sharing Policies [J]. *Library and Information Service*, 2019, 63(9): 21-30.
- [26] Liu Ziheng, Zeng Liying. Investigation and Comparison of Scientific Data Management and Sharing Platforms in Chinese Universities [J]. *Information and Documentation Services*, 2017(6): 90-95.
- [27] Imperial College London. Imperial RDM Policy [EB/OL]. [2020-06-18]. <https://www.imperial.ac.uk/media/imperial-college/research-and-innovation/research-office/public/Imperial-College-RDM-Policy.pdf>.
- [28] Bodleian Libraries. Bodleian Libraries' Policy on the Management and Preservation of Research Data and Records [EB/OL]. [2020-06-18]. https://www.bodleian.ox.ac.uk/__data/assets/pdf_file/0007/168640/Bodleian-policy-on-the-Management-of-Research-Data-and-Records-FINAL-April2014.pdf.
- [29] Center for Open Science. Data Management Planning: Open Science Framework [EB/OL]. [2020-01-10]. https://guides.nyu.edu/data_management/osf.
- [30] Cui Xu, Zhao Ximei, Wang Zheng, et al. Analysis of Achievements, Deficiencies, Countermeasures, and Trends of Scientific Data Management Platform Construction in China—Based on Domestic and International Comparison [J]. *Library and Information Service*, 2019, 63(17): 31-39.
- [31] Hudson-Vitale C, Imker H, Johnston LR, et al. SPEC Kit 354: Data Curation (May 2017) [R]. Washington: ARL, 2017.
- [32] Altman M, Castro E, Crosas M, et al. Open Journal Systems and Data-verse Integration—Helping Journals to Upgrade Data Publication for Reusable Research [EB/OL]. [2020-10-10]. <https://journal.code4lib.org/articles/10989>.
- [33] Michener WK. Ten Simple Rules for Creating a Good Data Management Plan [J]. *PLOS Computational Biology*, 2015, 11(10): e1005097.
- [34] Wang Siming. Analysis and Applied Research on Content Elements of Data Management Plans in Chinese University Libraries [D]. Changchun: Northeast Normal University, 2019: 1-79.
- [35] Harvard University. Research Data Management @ Harvard [EB/OL]. [2020-10-10]. <https://researchdatamanagement.harvard.edu/>.
- [36] Sarretta A. Research Data Life Cycle [EB/OL]. [2020-06-27]. <http://doi.org/10.5281/zenodo.1149049>.
- [37] Pablo A, Paula A, Serge F, et al. LERU Roadmap for Research Data—Model of the Research Workflow & Data Publishing Life Cycle [R]. Leuven: LERU Publications, 2013.
- [38] Qin Shun, Wang Quanli, Xing Wenming. Investigation and Enlightenment of European and American Scientific Data Open Access Publishing Platform Services [J]. *Library and Information Service*, 2019, 63(13): 129-136.
- [39] MIT Libraries. Citing Sources: Cite Data [EB/OL]. [2020-10-12]. <https://libguides.mit.edu/c.php?g=176032&p=1159520>.
- [40] The Sheridan Libraries. Data Trust Council Data Sharing Policies [EB/OL]. [2020-10-11]. https://guides.library.jhu.edu/protecting_identifiers/datatrust.
- [41] Yale University Library. Research Data Management for the Health Sciences: Data Quality [EB/OL]. [2020-01-23]. https://guides.library.yale.edu/rdm_healthsci/quality.
- [42] Columbia University Libraries. Research Data Services Consultation Guidelines [EB/OL]. [2020-01-16]. <https://library.columbia.edu/services/research->

data-services/guidelines.html.

[43] ACRL. DSS Research Data Q Editorial Board [EB/OL]. [2020-01-16]. <http://www.ala.org/acrl/dss/acr-dssdq>.

[44] Shipman JP, Tang R. The Collaborative Creation of a Research Data Management Librarian Academy (RDMLA) [J]. *Information Services & Use*, 2019(3): 1-5.

[45] Chen Chuanfu, Li Qiushi. Open Society and Library Development [J]. *Journal of Library Science in China*, 2020, 46(1): 16-37.

[46] Wang Shiwei. On Artificial Intelligence and Library Renewal [J]. *Documentation, Information & Knowledge*, 2019(4): 35-42.

[47] LERU. Statement on Open Research Data [EB/OL]. [2020-01-30]. <https://www.leru.org/publications/leru-statement-on-open-research-data>.

[48] LERU. Roadmap for Research Data [EB/OL]. [2020-01-30]. <https://www.leru.org/publications/leru-roadmap-for-research-data>.

[49] ANDS. Outline of a Research Data Management Policy for Australian Universities/Institutions [EB/OL]. [2020-10-10]. <https://rdc-drc.ca/wp-content/uploads/Institutional-Research-Data-Management-Policies.pdf>.

[50] DCC. Five Steps to Developing a Research Data Policy [EB/OL]. [2020-10-10]. <http://www.dcc.ac.uk/sites/default/files/documents/publications/DCC-FiveStepsToDevelopingAnRDMpolicy.pdf>.

[51] DLCM. Research Data Management Policy Template [EB/OL]. [2020-10-10]. https://www.dlcm.ch/download_{file}/force/68/372.

[52] Hart EM, Barmby P, LeBauer D, et al. Ten Simple Rules for Digital Data Storage [J]. *PLOS Computational Biology*, 2016, 12(10): e1005097.

[53] Tu Zhifang. Construction of Scientific Data Publishing Ecosystem and Quality Control System [J]. *Library and Information*, 2019(1): 125-134.

[54] General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China, Standardization Administration of China. *Information Technology—Scientific Data Citation* [EB/OL]. [2020-10-11]. <http://c.gb688.cn/bzgk/gb/showGb?type=online&hcno=A495CA355BAF00D962AA8DD84C3B2C16>.

[55] Qu Baoqiang, Wang Kai. Scientific Data Peer Review from the Perspective of Data Publishing [J]. *Library Journal*, 2017, 36(10): 71-77.

[56] Matthew B, Natasha N. Building Google Dataset Search and Fostering an Open Data Ecosystem [EB/OL]. [2020-02-06]. <https://ai.googleblog.com/2018/09/building-google-dataset-search-and.html>.

[57] Zeng Jianxun. Data Openness and Technology Open Source Should Be Autonomously Controllable [J]. *Digital Library Forum*, 2019(6): 1.

[58] Gu Liping, Zhang Xiaoyue. Practical Analysis of Data Librarians in the Open Science Environment [J]. *Library and Information Knowledge*, 2020(2): 60-74, 112.

[59] Li Mei. Research Data Service Innovation and New Roles of Data Librarians in University Libraries—A Case Study of the University of Edinburgh [J]. *Library and Information*, 2019(3): 122-128.

[60] Xie Jiahui, Xu Junhua. Investigation and Analysis of Research Data Management Positions in American Research Libraries [J]. *Library Research and Work*, 2020(1): 74-79.

- [61] Zhou Lei, Liu Liyong. German RISE-DE Research Data Management Service Self-Evaluation Model [J]. *Library Science Research*, 2019(21): 38-47, 9.
- [62] Dang Hongli, Tan Haibing. Application of DMM-Based Data Management Maturity Model in Service Evaluation [J]. *Modern Information*, 2017(9): 118-121.

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

Source: ChinaXiv –Machine translation. Verify with original.