Scientific Data Services in Chinese University Libraries: Research Progress, Hotspots, and Prospects (Postprint)

Authors: Qi Ning, Yuan Yumo

Date: 2024-06-13T00:00:00+00:00

Abstract

Integrating bibliometric methods with visual analysis approaches, this study conducts a bibliometric analysis of research on scientific data services in Chinese university libraries from perspectives including publication trends and core research forces. The article utilizes CiteSpace software to perform knowledge mapping analysis of hotspots in this field, encompassing keyword co-occurrence, burst detection, evolution, and clustering. From the functional perspective of university libraries, it comprehensively reviews the research hotspots and thematic distribution in areas such as scientific data lifecycle management, scientific data service models, scientific data literacy education, and scientific data infrastructure support. Finally, from five dimensions—national policy, university development, scientific research innovation, library services, and librarian professional competence—it proposes research prospects for the field of scientific data services in Chinese university libraries.

Full Text

Preamble

Scientific Data Services in Chinese University Libraries: Research Progress, Hotspots, and Prospects

This study employs bibliometric and visualization analysis methods to conduct a comprehensive bibliometric analysis of research on scientific data services in Chinese university libraries, examining publication trends and core research strengths. Using CiteSpace software, we analyze knowledge mapping features including keyword co-occurrence, burst detection, evolution, and clustering to systematically review research hotspots and thematic distributions across scientific data lifecycle management, service models, literacy education, and foundational support from the perspective of university library functions. Finally, we

propose research prospects for scientific data services in Chinese university libraries across five dimensions: national policy, university development, scientific research innovation, library services, and librarian professional competencies.

Keywords: university library; scientific data service; bibliometrics; knowledge graph; research topics

---

1. Data Sources and Research Methods

We conducted a professional search in the China National Knowledge Infrastructure (CNKI) database, with the search date extending to December 2022. The search strategy combined the terms "高校图书馆" (university library) OR "大学图书馆" (college library) with "科学数据" (scientific data). After manual screening to exclude irrelevant documents, we obtained 427 relevant publications. This study integrates bibliometric methods and utilizes CiteSpace, SPSS, and Excel software to perform bibliometric analysis of publication trends, core research authors, core research teams, and core research institutions in this field. We employ CiteSpace 6.1.R6 for visual knowledge mapping analysis of research progress, thematic distribution, and future trends in the domain.

---

2. Research Progress on Scientific Data Services in Chinese University Libraries

2.1 Publication Trend Analysis

Annual publication trends can indicate the developmental status and research trajectory of a field during specific time periods, providing auxiliary insights into the overall landscape of scientific data services in university libraries. Based on the annual publication volume and cumulative publications, we created a trend chart with a fitted curve for the field of scientific data services in Chinese university libraries.

The annual publication curve reveals that research in this field first emerged in 2011, with no more than 10 publications annually during 2011–2013, indicating relatively low overall output. The period from 2014 to 2020 witnessed rapid growth in publications. In 2015 and 2018, the State Council issued the "Big Data Development Action Outline" and the "Measures for the Management of Scientific Data," respectively, which further promoted standardized management and development of scientific data, elevating scientific data to a strategic resource supporting national scientific innovation and socio-economic development. Since 2021, related research has entered a stable development phase.

[Figure 1: see original paper] Publication Trend and Fitted Curve of Scientific Data Services in Chinese University Libraries

The cumulative publication curve shows a more regular pattern, with its numerical values displaying systematic changes that can be represented by functional or parametric equations, enabling predictive analysis based on curve patterns. This study establishes a relationship between publication year and cumulative volume, employing quadratic function fitting from regression analysis to forecast future publication trends in this field. The dashed line in the figure represents the fitted curve with the equation:

$$) = 3.072 + 12.073t + 34.659$$

where t represents the publication year. The fitting effect is satisfactory with $R^2 = 0.990$. The dashed line indicates that related research will continue to show growth trends in the coming years.

Table: Quadratic Function Fitting Curve Summary for Scientific Data Services in Chinese University Libraries

Parameter	Estimate	F-value	Degrees of Freedom	Significance
Constant	34.659			
t	12.073	1,876.957	2	0.000
t^2	3.072			

The fitted equation is: $) = 3.072 + 12.073t + 34.659$

2.2 Core Research Authors Analysis

Based on comprehensive index calculations, the author with the maximum publications ranked first with 9 papers. According to Price's Law, the formula for determining the minimum publication threshold for core authors is $M = 0.749 \times \sqrt{N_{\max}}$, where $N_{\max}$ represents the maximum number of publications by any author. Substituting the values yields $M \approx 3.317$, meaning authors with 4 or more publications can be considered core author candidates.

Citation frequency is another important criterion for identifying core authors. The formula for minimum citation threshold derived from Price's Law is $c = 0.749 \times \sqrt{c_{\max}}$, where $c_{\max}$ represents the highest single-paper citation count in the field. Meng Xiangbao and Li Aiguo's paper "Research on Scientific Data Literacy Education in Foreign University Libraries" achieved 220 citations, the highest in the field. Substituting $c_{\max} = 220$ yields $c \approx 11.08$, meaning authors with cumulative citations of 12 or more can also be listed as core author candidates.

A total of 16 authors meet both criteria. We employ the comprehensive index evaluation method for precise identification of core research authors. The process involves: (1) identifying important indicators, (2) calculating indicator values, (3) determining indicator weights through expert interviews, and

(4) setting thresholds and calculating comprehensive indices. Four evaluation indicators were identified: total publications in the field (weight: 30%), first-author publications (weight: 30%), total citation frequency (weight: 20%), and first-author citation frequency (weight: 20%).

According to Price's Law, candidates must simultaneously meet the publication threshold ($M \approx 3.317$) and citation threshold ($c \approx 11.08$). This yields 16 qualified core author candidates. The comprehensive index W_i for core research authors is calculated as:

$$W_i = \left(\frac{A_i}{A_{\text{avg}}} \times 30 + \frac{B_i}{B_{\text{avg}}} \times 30 + \frac{C_i}{C_{\text{avg}}} \times 20 + \frac{D_i}{D_{\text{avg}}} \times 20 \right) \times 100$$

where $i = 1, 2, 3, 4, \ldots, 16$ represents each candidate. The average indices for core author candidates are: $A_{\text{avg}} = 5.31$, $B_{\text{avg}} = 3.81$, $C_{\text{avg}} = 173$, $D_{\text{avg}} = 121$. Substituting these averages yields a threshold W_i value of 100. Candidates with scores above 100 are recognized as core research authors in this field.

Table: Comprehensive Index Statistics of Core Research Authors

Six authors are identified as core research authors, with Meng Xiangbao ranking highest due to leading both publication volume and citation frequency, particularly with citations far exceeding other researchers. Zhang Qun leads in publication volume, with over half as first-author, demonstrating high contribution to the field. Both authors' comprehensive indices exceed the average by more than twofold, reflecting their significant influence. Liu Guifeng and Shen Tingting also show high publication volume and citation frequency, representing frontier developments. Chen Yuanyuan and Ye Lan have fewer publications, but Ye Lan's high citation frequency indicates strong representativeness. These authors collectively drive the field's deepening development.

2.3 Core Research Teams Analysis

We visualize author collaboration using CiteSpace to generate a co-authorship network map. In the map, node size correlates with publication volume, and connecting lines indicate collaborative relationships. The network density is 0.116, suggesting relatively loose cooperation that requires strengthened academic exchange.

Four stable collaborative groups emerge: (1) A cross-institutional team led by Meng Xiangbao focusing on scientific data consortium construction and literacy cultivation; (2) An intra-institutional team comprising Liu Guifeng, Lu Zhangping, and Liu Qiong focusing on research data management and governance; (3) A mentor-student team including Wu Ming, Hu Hui, and Chen Xiujuan focusing on data literacy education and library scientific data services; and (4) A department-level team of Li Guojun and Zhu Ling focusing on open data platforms and research data management.

2.4 Core Research Institutions Analysis

Core research institutions possess competitive advantages in talent support, resource guarantee, and technological backing. According to Price's Law, the formula for determining core institutions is $M_i = 0.749 \times \sqrt{I_{\max}}$, where $I_{\max}$ is the maximum institutional publication volume. With $I_{\max} = 16$ (Chinese Academy of Sciences Documentation and Information Center), we obtain $M_i \approx 2.996$, meaning institutions with 3+ publications qualify as core candidates.

The comprehensive index Z_i for core institutions is calculated as:

$$Z_i = \left(\frac{X_i}{X_{\text{avg}}} \times 50 + \frac{Y_i}{Y_{\text{avg}}} \times 50 \right) \times 100$$

where X_i represents total institutional publications and Y_i represents total citations. The average indices are $X_{\text{avg}} = 5.31$ and $Y_{\text{avg}} = 121$. Substituting yields a threshold Z_i value of 100.

Thirteen core research institutions are identified, led by Southeast University Library with the highest publication volume and citations exceeding the average by 1.5 times, demonstrating significant attention from scholars. Heilongjiang University Information Management College, though not a core institution, shows the highest comprehensive index among non-core institutions, suggesting potential to become core with continued research.

[Figure 2: see original paper] Author Co-authorship Network

[Figure 3: see original paper] Institutional Co-occurrence Network

The institutional collaboration network reveals four main networks formed by core institutions: (1) Chinese Academy of Sciences Documentation and Information Center with Chinese Academy of Sciences University and Computer Network Information Center; (2) Shanghai University Library and Information Science & Archives Department with Wuhan University Information Management School; (3) Southeast University Library, Nanjing University Library, and Sun Yat-sen University Information Management School; and (4) Peking University Library, Peking University Information Management Department, and Institute of Scientific and Technical Information of China. These networks promote diversified research directions and produce valuable reference 成果 for university library scientific data services.

---

3. Research Hotspots in Scientific Data Services of Chinese University Libraries

Based on the above analysis, we examine research hotspots, evolutionary processes, and thematic distributions in this field. Keywords provide condensed

summaries of paper themes and core content. Visual analysis of keywords enables rapid understanding of research hotspots, perspectives, and developmental trajectories across different periods.

3.1 Keyword Co-occurrence Analysis

In keyword co-occurrence networks, each node represents a key term. Node size corresponds to frequency, with higher frequencies producing larger nodes. Connecting lines indicate keywords appearing together in the same paper. A node's betweenness centrality represents its importance in the network and correlation with other nodes. Keywords with betweenness centrality $ $0.1 occupy relatively important positions.

Keywords such as "scientific data" (centrality: 0.31), "university library" (0.25), "data management" (0.21), "data literacy" (0.13), "data service" (0.11), and "data curator" (0.11) demonstrate high frequency and centrality, indicating they are core research hotspots. Although terms like "subject service" (0.09) and "E-science" (0.06) have relatively lower frequency, their centrality values show they are also important research directions closely related to the field's development.

[Figure 4: see original paper] Keyword Co-occurrence Network

3.2 Burst Keyword Analysis

Burst keyword analysis identifies high-frequency terms in specific periods, helping recognize research trends and frontiers. The terms "open data" and "scientific support" emerged as strongest burst keywords in 2011, lasting nearly three years. The introduction of the "open data" concept into Chinese university libraries promoted integration between disciplinary services and data services, highlighting the functional role of scientific data sharing.

Keywords like "E-research," "data curator," and "service system" emerged in 2012 with burst strength exceeding 3.0, indicating scholars' widespread attention to data curation services development in foreign university libraries. "Data curator" began receiving focused academic attention in 2017 with a duration exceeding three years, demonstrating that data curators and data services have become key development and research priorities in recent years.

[Figure 5: see original paper] Keyword Burst Analysis

3.3 Keyword Timeline Analysis

The keyword timeline map intuitively displays thematic evolution and development trends. Setting the time span to 2011–2022 and using CiteSpace's Timeview function, we observe that research publications increased from few to many while keywords gradually multiplied.

Phase 1 (2011–2014): Initial Stage
Publications were limited, with keywords and clusters beginning preliminary

evolution. Research direction expanded from library data curation to data curators and data services. Literature primarily focused on basic concepts of data management, problems in domestic university library scientific data services, and enlightenment from foreign practices, laying a solid foundation for subsequent research.

Phase 2 (2015–2020): Ascending Stage
Publication volume increased significantly compared to the previous stage. While research content and paradigms showed new changes based on earlier work, the number of keywords and clusters increased substantially, presenting a multi-perspective research landscape. This phase's achievements guided future development.

Phase 3 (2021–2022): Deepening Stage
Publication volume decreased slightly, but research quality improved. With rapid digital technology development, new paradigms such as smart environments emerged, expanding research scope and potentially attracting more scholars to related themes.

[Figure 6: see original paper] Keyword Evolution Timeline

3.4 Keyword Clustering Analysis

Clustering analysis groups keywords with strong co-occurrence relationships, revealing research hotspots and thematic structures. Using the log-likelihood ratio (LLR) method for keyword clustering, we obtain clustering results. The clustering modularity value is $Q = 0.973$ and the average silhouette value is $S = 0.872$, indicating significant clustering structure and reliable results.

The analysis yields 11 clusters. To further organize hotspot research themes, we categorize these clusters into four functional areas of university libraries: scientific data lifecycle management, scientific data service models, scientific data literacy education, and scientific data foundational support.

[Figure 7: see original paper] Keyword Clustering Map

3.4.1 Scientific Data Lifecycle Management University library management functions encompass scientific data collection, organization, preservation, and sharing. Research deeply explores each stage of scientific data management, including development characteristics of collection, functional roles of scientific data sharing, open sharing service models, and regulatory models for scientific data curation. Studies investigate implementation processes and practical operational experiences, enriching lifecycle management frameworks in Chinese university libraries.

3.4.2 Scientific Data Service Models Drawing on advanced experiences from foreign university libraries, domestic scholars construct localized scientific data service models. Some build models based on user needs, policy support,

and staffing; others develop localized service frameworks from process, service, and support perspectives; some create data management plan service frameworks based on information ecology theory. Service strategies include both consulting services (management planning, discovery, citation, protection, storage, sharing) and technical services (data identification, metadata creation and updating, blockchain technology for security, and constructing reliable scientific data management systems throughout the research cycle).

3.4.3 Scientific Data Literacy Education Scientific data literacy education represents a crucial educational function of university libraries. Theoretical research initially summarized data literacy concepts across four dimensions: data awareness, data knowledge, data skills, and data ethics. Further refinement expanded the definition to emphasize systematic, standardized lifecycle management of scientific data to maximize value. Research explores teaching models, content, and implementation methods, addressing key issues in researcher data literacy education from perspectives of educational content, models, stakeholder concerns, and participation dimensions. Studies have constructed data literacy frameworks for librarians and evaluation systems for graduate students, enriching teaching content and practice forms with researchers, librarians, and graduate students as target audiences.

3.4.4 Scientific Data Foundational Support University libraries rely on policies at national and institutional levels and the professional competencies of data curators as foundational support for scientific data services. Policy research compares Chinese and U.S. scientific data policies, recommending two-tier management models referencing Dutch university libraries. Despite progress since the 2018 "Measures for Scientific Data Management" and the Chinese Academy of Sciences' "Management and Open Sharing Measures," China's scientific data open sharing remains in its early stages. Future research should integrate advanced digital and intelligent technologies (metaverse, AI, etc.) to broaden innovation and practice.

Research on data curators examines role positioning, competency frameworks, and development strategies from multidimensional perspectives. As digital society evolves, cultivating librarians' diverse data literacy capabilities becomes a new growth point for service effectiveness. Studies propose building a tripartite capability guarantee system linking role positioning, competency development, and collaborative construction, which helps university libraries clarify their roles and accelerate innovative development of scientific data services.

---

4. Research Prospects for Scientific Data Services in Chinese University Libraries

From four dimensions—national policy orientation, university development premise, research innovation core, and librarian capability focus—we propose research prospects and development directions.

4.1 National Policy Orientation: Promoting Refined Scientific Data Management

The 2018 State Council's "Measures for Scientific Data Management" strengthened national-level institutional frameworks, imposing new requirements on responsible entities including research institutes, universities, and enterprises. University libraries should actively conduct standardized scientific data curation services to assist researchers in lifecycle data management and improve data quality. Future research should refine data curation implementation standards and policies for collection, sharing, and disposal, and explore process management and security supervision from perspectives of personal privacy and data ecosystem safety.

4.2 University Development Premise: Promoting Scientific Data Openness and Sharing

As China implements innovation-driven development and science/technology power strategies, data elements have become central to the digital economy. University libraries should leverage their scale and scenario advantages in scientific data to empower national economic and social digital development. Future research should combine university disciplinary strengths, exploring incentive mechanisms and security review mechanisms for data exchange from multi-stakeholder perspectives, while incorporating advanced technologies to deepen innovation breadth and depth.

4.3 Research Innovation Core: Enabling Multi-modal Scientific Data Services

The ACRL's "2021 Environmental Scan" identified scientific data management services as a key future direction for academic libraries. China's future development should enable multi-modal services through both virtual and physical channels: (1) From a resource integration perspective, build one-stop scientific data resource platforms promoting multi-department collaboration and online learning; (2) From a space reconstruction perspective, use AI, AR, and VR technologies to create intelligent data service spaces enabling interdisciplinary collaboration. Libraries should design hierarchical service systems based on user characteristics: cultivating data awareness and practical abilities for undergraduates, providing deep academic support on technology policies and data intellectual property for researchers, and constructing refined services integrated throughout the research cycle.

4.4 Librarian Capability Focus: Optimizing Data Curator Team Leadership

University library data services have attracted industry attention, and cultivating librarians' diverse data literacy capabilities will become a new growth point. The ACRL's "2023 Environmental Scan" indicates that academic librarians need to master data analysis skills while shaping data ethics awareness. Future research should focus on domestic data curator needs, training systems, promotion mechanisms, and career development prospects. Offering professional data science and skills training to all librarians will continuously strengthen their capabilities in data services, enabling them to lead future development in research data management.

---

5. Conclusion

Chinese university libraries have launched research on scientific data lifecycle management, literacy education, and foundational support with preliminary results achieved. However, improvements are still needed in policy construction, service development, and librarian capability enhancement. As noted by Jiang Yunzhong, Party Secretary of Tsinghua University Library, at the 2021 China University Research Data Management Working Group Meeting: "Universities are crucial bases for national scientific and technological innovation, producing vast amounts of research data vital to national innovation and socio-economic development. Managing and using this data to maximize effectiveness is the mission of university libraries." Developing scientific data services can strengthen collaboration among stakeholders, forming synergies in data management and services to provide robust support for university scientific data sharing and reuse.

References

[1] Executive Office of the President. Increasing access to the results of federally funded scientific research[EB/OL]. [2023-01-10]. http://www.gov.cn/zhengce/2018-04/05/content_{5279957}.htm.

[2] China University Research Data Management Working Group. Introduction to the Working Group[EB/OL]. [2023-06-13]. http://society.library.sh.cn/node/2611.

[3] Xiao X, et al. Research on scientific data services in foreign libraries under the E-science environment[J]. Library and Information Service, 2011(2): 18-21, 41.

[4] Yang W, et al. Progress in foreign university library scientific research data management services[J]. Journal of the National Library of China, 2017(5): 88-97.

[5] Xia W, et al. Review and prospects of foreign university library scientific data services[J]. Journal of Sichuan Library Science, 2019(3): 67-70.

[6] Cai S. Review and enlightenment of foreign university library scientific data literacy education[J]. Library Tribune, 2020(10): 122-129.

[7] Chen Y, et al. Overview of domestic and foreign university library scientific data services from theoretical, policy, and educational perspectives[J]. Library Work and Study, 2021(5): 8-11, 47.

[8] Yan S. Analysis of research status and trends in university library scientific data services[J]. Library Work and Study, 2017(10): 17-23, 30.

[9] Chen W. Review and trend analysis of scientific data management research in Chinese libraries[J]. Library Work and Study, 2019(12): 86-91.

[10] Li J. Review of scientific research data management services in China (2010–2020)[J]. Library Work and Study, 2020(12): 17-23.

[11] Jin M. Research status and trend analysis of scientific data literacy in Chinese libraries[J]. Library Forum, 2019(11): 78-84.

[12] Min J. Overview of scientific data management research in Chinese libraries[J]. Library Science Research, 2020(12): 39-47.

[13] Chen W. Altmetrics research in China[J]. Library Science Research, 2020(8): 43-53.

[14] Based on bibliometric analysis of core research strength in data literacy in China[J]. Journal of Library Science in China, 2020(4): 66-72.

[15] Research on constructing scientific data management system embedded with lifecycle theory[J]. Library and Information Service, 2020(10): 34-42.

[16] Practice of research data management services at Monash University Library[J]. Digital Library Forum, 2022(5): 33-38.

[17] Research on open sharing services of scientific data in U.S. research libraries based on data lifecycle[J]. Library and Information, 2019(1): 135-144.

[18] Research on collaborative supervision model for scientific research data in university libraries[J]. Theory and Practice, 2017(3): 14-19.

[19] Research on data curation elements and optimization strategies for university libraries in the big data era[J]. Library Work and Study, 2019(1): 58-64.

[20] Research on embedded research services for data curation in university libraries[J]. Library Science Research, 2020(3): 54-60.

[21] Exploration of embedded service models for scientific data management in universities[J]. Information Studies: Theory & Application, 2017(2): 77-82.

[22] Research on scientific data management plan service framework for university libraries[J]. National Library Journal of China, 2019(4): 21-31.

[23] Current status and enlightenment of scientific data services in foreign university libraries: Case studies of five libraries[J]. Library Work and Study, 2017(10): 31-36.

[24] Approaches and contents of scientific data services provided by university libraries[J]. Library Science Research, 2018(4): 80-85.

[25] Research on data curator positions in U.S. university libraries[J]. Library Construction, 2019(1): 135-140, 146.

[26] Thoughts on data service models and new roles of data curators[J]. Library Work and Study, 2014(3): 97-99.

[27] Research on scientific data literacy education in foreign university libraries[J]. Journal of Academic Libraries, 2014(3): 11-16.

[28] Research on scientific data management from user data literacy education perspective[J]. Library and Information, 2015(4): 139-141, 109.

[29] Construction of scientific data literacy education system model for university libraries[J]. Journal of Academic Libraries, 2016(1): 96-102.

[30] Review and enlightenment of data literacy education for foreign researchers[J]. Information and Documentation Services, 2017(3): 102-106.

[31] Research on librarians' scientific data literacy competency framework and development strategies[J]. Library Theory and Practice, 2018(9): 90-95.

[32] Research on data literacy status of Chinese graduate students in research activities: From dual lifecycle theory perspective[J]. Library and Information Service, 2020(7): 84-93.

[33] Comparison of Chinese and U.S. scientific data policies[J]. Library Science Research, 2022(11): 113-121.

[34] Practice and enlightenment of research data management services in Dutch university libraries[J]. Library Science Research, 2020(15): 91-101.

[35] Core roles and training strategies of data curators[J]. Library and Information Service, 2022(7): 144-152.

[36] Current status and enlightenment of data curator management services in foreign university libraries[J]. Library Work and Study, 2022(1): 45-53.

[37] Exploration of data curator role transformation[J]. Library Work and Study, 2022(1): 54-60.

[38] Data curator training based on position functions and demand characteristics[J]. Library Work and Study, 2022(8): 53-62.

[39] Notice of the General Office of the State Council on issuing Measures for Scientific Data Management[EB/OL]. (2018-04-02). http://www.gov.cn/zhengce/content/2018-04/02/content_{5279272}.htm.

[40] Chinese Academy of Sciences. Measures for scientific data management and open sharing of Chinese Academy of Sciences[EB/OL]. [2023-02-10]. https://www.cas.cn/sygz/201902/t20190221_{4679910}.shtml.

[41] Association of College & Research Libraries. 2020 top trends in academic libraries[EB/OL]. [2023-02-10]. https://crln.acrl.org/index.php/crlnews/article/view/24478.

[42] China University Research Data Management Working Group Meeting held[EB/OL]. [2023-06-13]. https://learning.sohu.com/a/514054666_{121124303}.

[43] New Media Consortium. NMC horizon report: 2017 higher education edition[R]. Austin: The New Media Consortium, 2017.

[44] ACRL Insider. 2021 environmental scan[EB/OL]. [2023-02-05]. https://acrl.ala.org/acrlinsider/2021-acrl-environmental-scan/.

[45] Steering Committee for Academic Libraries of Higher Education Institutions, Ministry of Education. Compass report on modernization of university libraries[EB/OL]. [2023-02-10]. http://scal.edu.cn/zxdt/202201140221.

[46] ACRL Insider. 2023 environmental scan[EB/OL]. [2023-03-06]. https://acrl.ala.org/acrlinsider/2023-acrl-environmental-scan/.

[47] Research on research support services and librarian role transformation in foreign university libraries[J]. Library Work and Study, 2023(5): 99-106.

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