Achievements, Shortcomings, Strategies, and Trends in the Development of China' s Scientific Data Management Platforms: A Postprint Based on Domestic and International Comparative Perspectives

Authors: Cui Xu, Zhao Ximei, Wang Zheng, Du Fengrui

Date: 2023-07-26T00:00:00+00:00

Abstract

[Purpose/Significance] China' s scientific data management platforms have achieved certain accomplishments, yet gaps remain compared with international counterparts. This study aims to diagnose and position the current development status of China' s scientific data management platforms through vertical temporal trajectory analysis and horizontal domestic-international comparison, while providing international experience and countermeasures for the optimization of domestic platforms. [Method/Process] Employing web-based survey methods, the research is conducted from two perspectives: vertical temporal trajectory and horizontal domestic-international comparison. The achievements of China' s scientific data management platforms are analyzed through vertical temporal trajectory, while existing problems are examined through horizontal comparison. [Results/Conclusions] Achievements: The relevant policy system is continuously improving; the disciplinary fields covered by datasets are increasingly expanding; and the internationalization development of data management platform construction has begun to show initial results. Deficiencies: Single-source funding; incomplete platform service functions; lack of collaborative construction concepts in some platforms; minimal participation from university libraries, among others. Countermeasures: Establish a diversified funding mechanism; construct a data management value chain; strengthen cooperation among heterogeneous institutions; expand platform service modalities; university libraries should become an important force in data management. Development trends: The construction of scientific data management platforms will become a crucial undertaking for scientific research information service institutions; the number of scientific data management

organizations and personnel will continue to expand; and the scale and international competitiveness of scientific management will be enhanced through the establishment of scientific data management platform alliances and powerful scientific data centers.

Full Text

Preamble

Volume 63, Issue 9, May 2019
Achievements, Deficiencies, Countermeasures, and Trends in Chinese Scientific Data Management Platform Construction: An International Comparative Perspective

Cui Xu, Zhao Ximei, Wang Zheng, Du Fengrui
School of Public Management, Northwest University, Xi' an 710127

Abstract

[**Purpose/Significance**] China' s scientific data management platforms have achieved certain accomplishments, yet gaps remain when compared internationally. This study aims to assess and position the current development status of Chinese scientific data management platforms through both longitudinal temporal analysis and horizontal international comparison, while providing international experience and countermeasures for domestic platform optimization. [**Method/Process**] Using web-based surveys, this research examines platforms from two perspectives: longitudinal temporal trajectory and horizontal domestic-international comparison. Achievements are analyzed through the longitudinal dimension, while deficiencies are identified through horizontal comparison. [**Results/Conclusions**] (1) Achievements: The policy framework continues to improve; dataset coverage across disciplinary fields is expanding; and international development of platform construction has yielded initial results. (2) Deficiencies: Single-source funding; incomplete platform service functions; lack of collaborative construction concepts in some platforms; and limited university library participation. (3) Countermeasures: Establish diversified funding mechanisms; create data management value chains; strengthen cooperation between heterogeneous institutions; expand platform service methods; and position university libraries as important forces in data management. (4) Trends: Scientific data management platform construction will become a critical function of research information service institutions; the number of data management institutions and personnel will continue to grow; and scale and international competitiveness will be enhanced through establishing platform alliances and robust scientific data centers.

Classification Number: G203
Keywords: scientific data, scientific data management platform, platform achievements, platform comparison

1. Literature Review

International research on data management platforms primarily focuses on three areas: (1) Platform surveys. M. Halbert [2] investigated data management services on library websites at four universities (Iowa State, Kansas State, Oklahoma State, and University of Nebraska-Lincoln). (2) Introductions to existing platforms. A. L. Vaccarino et al. [3] introduced Brain-CODE, a Canadian data management platform for collecting, storing, integrating, sharing, and analyzing diverse neurological data. T. Nind et al. [4] presented an open-source platform for managing clinical medical data that supports seamless collaboration between research groups and data quality assessment. (3) Platform technology development. A. D' Anca et al. [5] proposed an efficient, secure, and interoperable solution for oceanographic data storage and navigation. L. Persoon et al. [6] designed a preclinical research data management platform comprising workflow management, data management, and storage modules. B. Wang et al. [7] introduced a semantic web-based platform for computational chemistry data management.

Domestic research primarily follows two directions: (1) Platform construction studies, including introductions to international platforms [8-9], domain-specific platforms [10], construction standards and norms [11-12], and performance evaluation systems [13]. (2) Platform technology development and application, covering system software practices [14], operational mechanisms and system models [15-18], and system design and implementation across environmental science [19-22], economics [23-24], seismology [25], and basic research [26-27], as well as regional and industry-specific platforms [28-30]. However, existing domestic research emphasizes technology, applications, and problems, lacking longitudinal summaries of achievements and trend analysis. This paper provides more comprehensive international comparisons, offering a holistic analysis of Chinese scientific data management platform construction.

2. Comparative Analysis of Domestic and International Data Management Platforms

2.1 Funding Sources

Domestic platforms rely almost exclusively on government funding, including: (1) The National Basic Science Data Sharing Service Platform, funded by Ministry of Science and Technology special funds; (2) Independently constructed institutional platforms funded by public finance through government special funds, "211 Project" funds, Ministry of Education special funds, National Natural Science Foundation funds, or university self-investment. In contrast, international platforms benefit from more diverse funding channels, including professional associations, private sector contributions, individual donations, and membership fees, in addition to government support.

2.2 Re3data Registration Statistics

Re3data serves as an internationally recognized large-scale scientific database sharing platform reflecting national construction scale and capability. As of July 2018, 70 countries, 2 regions (Hong Kong and Taiwan), and international organizations had registered 2,960 databases. Mainland China ranked 11th globally with 38 registered institutions.

[Figure 1: see original paper] shows registration statistics for 17 countries, which can be divided into three tiers: Tier 1 (100+ databases): United States, Germany, United Kingdom, Canada; Tier 2 (50-100 databases): France, Australia, Switzerland, Japan, Netherlands; Tier 3 (<50 databases): India, China, Austria, Belgium, Italy, Spain, Norway, Russia. The U.S. has approximately 25 times China' s registration count; Germany has about 8 times; and Tier 1 countries average 439.75 databases—11 times China' s count. This substantial gap indicates significant room for growth in China' s platform construction quantity.

2.3 Representative Software Development Models

Platform construction approaches include independent development (including collaborative), commercial software, open-source, and third-party commissioning. Table 1 compares influential platforms with typical software development models.

Table 1 Comparison of Representative Software Development Models for Domestic and International Scientific Data Management Platforms

Domestic Platforms	International Platforms	Development Model
National Basic Science Data Sharing Service Platform (http://www.nsdata.cn/)	University of Michigan Inter-university Consortium for Political and Social Research (ICPSR) (https://www.icpsr.umich.edu/icpsrweb/)	Open-source (Dataverse)
Peking University Open Research Data Platform (http://opendata.pku.edu.cn/)	Harvard-MIT Data Center (Dataverse) (https://dataverse.harvard.edu/)	Open-source (Dataverse)
BGI Genomics (http://www.genomics.cn/)	National Snow and Ice Data Center (NSIDC) (https://nsidc.org/data)	Independent development
Tsinghua University China Economic and Social Data Research Center	National Center for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov/)	Dataverse + Entrez search software

Domestic Platforms	International Platforms	Development Model
Sun Yat-sen University Academic Research Database Sharing Program	DataONE (https://www.dataone.org/)	Open-source (Dspace)
Renmin University China National Survey Database (http://cnsda.ruc.edu.cn/)	Cornell University Roper Center (RoperExpress)	Self-built SDA
Wuhan University University Scientific Data Sharing Platform	University of Chicago General Social Survey (GSS) DataExplorer	Open-source (Dspace)
Fudan University Social Science Data Platform	U.S. Bureau of Labor Statistics National Longitudinal Surveys (NLS)	Open-source (Dataverse)
Hunan University Economic Data Research Center	Penn State University Libraries (ScholarSphere)	-
Shanghai Foreign Studies University Data Academic Service Platform	UK Data Archive/Norwegian Centre for Research Data (Nesstar)	Open-source (Dataverse)
National Rural Fixed Observation Point Survey Management and Data Analysis Platform	University of Virginia Library Data Services	-
National Health Commission Floating Population Data Platform	MIT Libraries (Dspace)	-
Land Survey Results Sharing and Application Service Platform	-	Independent development

Notes: Re3data URL: https://www.re3data.org/browse/by-country/; The National Basic Science Data Sharing Service Platform includes 15 disciplinary sub-platforms registered on Re3data (see Table 2); External access unavailable.

Domestic software development patterns include: (1) Independent development, such as the Chinese Academy of Sciences' integrated data management and collaboration system (TeamDR, VisualDB, RSR, DVIZ, etc.); (2) Open-source-

based secondary development (e.g., Peking University, Wuhan University platforms). However, domestic platforms lack open-source philosophy—most are for internal use only, unlike influential international open-source systems like Dataverse and Dspace, which increases national development costs.

2.4 Service Function Comparison

Platform service functions can be categorized as core (data management planning, creation, storage, acquisition, analysis, sharing) and supplementary (user guides, training). As shown in [Figure 2: see original paper], international platforms offer comprehensive services, including data management planning modules that guide users through legal, ethical, and procedural considerations before research begins. Training methods are diverse, including workshops, lectures, online learning, and one-on-one consultations, with detailed contact information for support staff.

Domestic platforms exhibit several shortcomings: (1) Most lack data management planning modules; (2) Training methods are monolithic, primarily lecture-based, with few offering multiple formats or consultation services; (3) Some platforms restrict downloads or only serve institutional users; (4) Some platforms provide only literature services rather than datasets; (5) Many have small, discontinuous datasets that compromise analytical quality.

2.5 Open Sharing Comparison

China' s platforms employ several sharing models: (1) **Scope-based**: international sharing (Re3data-registered platforms), national sharing (e.g., rural observation systems), and institutional sharing (e.g., Wuhan University platform); (2) **Organization-based**: membership models (common internationally, with China having limited participation), registration models (widely used domestically), qualification review models (e.g., health commission platforms), and proactive public disclosure models (e.g., land survey platforms).

Overall, membership-based sharing dominates internationally, with major countries like the U.S., Germany, UK, and Canada actively participating in and even initiating international data sharing organizations. China has limited participation and has not yet established an internationally led data sharing organization.

3. Achievements and Deficiencies in Chinese Platform Construction

3.1 Achievements

Chinese scientific data management platform development can be divided into three stages: (1) Spontaneous mode (1970s-2001), characterized by autonomous institute and research group construction with limited sharing; (2) Data integration (late 2001-2012), marked by the launch of the Scientific Data Sharing

Project; (3) International sharing (2013-present), signified by Re3data registration.

Policy System Improvement: From national guidelines like the "Big Data Development Action Outline" and "Scientific Data Management Measures" to organizational standards (e.g., 38 technical specifications for agricultural science data, 23 for materials science), the policy framework has become increasingly comprehensive.

Expanding Disciplinary Coverage: Initially focused on basic science, platforms now cover agriculture, forestry, oceanography, meteorology, seismology, earth systems science, and population/health. The National Basic Science Data Sharing Service Platform has integrated 254 specialized databases across physics, chemistry, astronomy, space, and biology, sharing 431TB of data with 17.12 million visits by end-2017.

Initial Internationalization Progress: China participates in World Data Centers (WDC) and disciplinary international organizations across multiple fields. Some platforms, like the Cotton Functional Genomics Database, incorporate data from both domestic and international sources.

3.2 Deficiencies

3.2.1 Single Funding Sources Limit Balanced Development: Platform construction and operation require sustained, substantial funding. Over-reliance on government special funds (Ministry of Science and Technology, Ministry of Education, NSFC, "211 Project") creates financial burdens and uneven development, with only major platforms receiving adequate support. Donations are rare, except for Tsinghua' s 20 million RMB donation in 2018.

3.2.2 Lack of Open-Source Philosophy Increases Development Costs: While university platforms use open-source software, Chinese localization efforts suffer from poor inter-team communication, delaying problem resolution. Independently developed government and commercial platforms remain closed systems, unlike internationally influential open-source solutions like Dataverse and Dspace, raising national development costs.

3.2.3 Limited Collaboration Hinders Scale: International platforms like ICPSR demonstrate extensive collaboration (776 member institutions). While Chinese Academy of Sciences platforms involve multiple partners, most university and institutional platforms operate independently, lacking cross-institutional cooperation between universities, research institutes, and enterprises, thereby limiting data scale and national scientific competitiveness.

3.2.4 Incomplete Service Functions Reduce User Engagement: Deficiencies in data management planning, monotonous training formats, download restrictions, limited service scope, and small/discontinuous datasets weaken user experience and platform utility for research support.

3.2.5 Limited University Library Participation: Unlike the widespread and collaborative involvement of Western university libraries, fewer than 10 Chinese university libraries participate in research data management initiatives, indicating they have not yet become major forces in this domain.

4. Countermeasures

4.1 Diversify Funding Sources: Establish multi-channel funding mechanisms by learning from international models like the Harvard-MIT Data Center, which secured multi-million dollar grants from NSF and other agencies plus congressional library support. Encourage local governments, universities, and foundations to support platforms while attracting social organizations, enterprises, and individual donations through preferential policies and membership models.

4.2 Promote Domestic Software Development: Encourage Chinese IT companies and non-profit organizations to develop data management software (including open-source solutions) to reduce costs, avoid redundant development, and promote widespread data service adoption across research institutions.

4.3 Strengthen Inter-Institutional Collaboration: Following the DataONE model, establish coordination and member institutions, partnering with libraries, database vendors, and sponsors to create collaborative chains among research institutions, universities, and enterprises, enabling researchers to access data beyond institutional limitations.

4.4 Enhance Service Functions: Learn from international platforms by adding data management planning modules, expanding consultation staff, and diversifying training methods (workshops, lectures, online learning, one-on-one consultations) to improve user experience and create value chains for research services.

4.5 Position University Libraries as Key Players: University libraries should actively collaborate with IT enterprises for technical support and with academic departments, research institutes, and teams for dataset construction, becoming central forces in data management platform development.

5. Development Trends

Trend 1: Platform Construction as Core Institutional Function: As research becomes increasingly data-dependent, centralized data management will reduce costs, prevent data loss, and improve research quality, making it a critical function for university libraries, research libraries, and information service institutions.

Trend 2: Expanding Data Management Institutions and Personnel: Growing demand for data management platforms will require more specialized professionals. With current talent training still weak and few universities offering relevant programs, professional education will become a key development

area.

Trend 3: Platform Alliances for Enhanced Scale and Competitiveness: Following international experience, establishing scientific data management platform alliances among university libraries, research institute libraries, and enterprise data centers will address dataset scale and quality issues, providing stronger support for national scientific innovation.

Trend 4: Robust Data Centers for International Influence: To become a scientific powerhouse, China must become a data powerhouse. Building high-quality, large-scale scientific data centers will increase China' s role as host organizations for international data centers, enhancing global influence.

References

[1] re3data.org[EB/OL].[2018-06-04].https://www.re3data.org/browse/by-country/.

[2] HALBERT M. Prospects for research data management. Research data management: principles, practices, and prospects[R/OL].[2018-03-12].http://libres.uncg.edu/ir/uncg/f/M_{{{Halbert}}*{{Prospects}}*}{2013}.pdf.

[3] VACCARINO A L, DHARSEE M, STROTHER S, et al. Brain-CODE: a secure neuroinformatics platform for management, federation, sharing and analysis of multi-dimensional neuroscience data[J/OL]. Frontiers in neuroinformatics[2018-06-10].https://doi.org/10.3389/fninf.2018.00028.

[4] NIND T, GALLOWAY J, MCALLISTER G, et al. The research data management platform (RDMP): a novel, process-driven, open-source tool for the management of longitudinal cohorts of clinical data[J]. GigaScience, 2018, 7(7):1-12.

[5] D' ANCA A, CONTE L, NASSISI P, et al. A multi-service data management platform to improve image-guided precision preclinical biological research[C/OL]. Paper for special issue for 4th Conference on Image guided precision radiotherapy[2018-10-30].https://www.birpublications.org/doi/10.1259/bjr.20180455.

[6] PERSOON L, HOOFS V, KRUIJSSE F V D, et al. A novel data management platform for image-guided precision radiotherapy[J]. Natural hazards and earth system sciences, 2017(17):171-184.

[7] WANG B, DOBOSHA P A, CHALK S, et al. Computational chemistry data management platform based on the semantic web[J]. The journal of physical chemistry, 2017, 121(1):298-307.

[8] ZHANG Shasha, HUANG Guobin, GENG Qian. Analysis of UK scientific data publishing platforms based on re3data[J]. Library Science Journal, 2016(5):133-136.

[9] ZHAO Weili, CHEN Xiaoyi, JIN Hong. Scientific data sharing platforms supporting advantageous industrial development[J]. Science Technology and Economy, 2008(2):53-55.

[10] CHEN Xiujuan, WU Ming, HU Hui. Embedding library data management services into research workflows: a chemistry case study[J]. Library Tribune, 2016(3):49-55, 74-78.

[11] HE Yi, WANG Shuguang, LIU Wenhao. Application of Infobutton in national population and health science data sharing platform[J]. China Digital Medicine, 2016(1):80-83.

[12] GU Shuangshuang. Construction of a nursing science data service platform[J]. Journal of Medical Informatics, 2014(7):33-36, 49.

[13] SI Li, LI Yueting, XING Wenming, et al. Empirical research on performance evaluation of China' s scientific data sharing platforms[J]. Library Theory and Practice, 2014(9):30-35.

[14] ZHANG Jilong, YIN Shenqin, ZHANG Yong, et al. Social science data sharing and services: a case study of Fudan University Social Science Data Platform[J]. Journal of Academic Libraries, 2015(1):74-79.

[15] SUN Xiange. Construction of scientific data integration and sharing service platform for university libraries in cloud environment[J]. Library Science Journal, 2016(5):133-136.

[16] GAO Qin, ZHONG Xiaoli. Research on construction of scientific data curation platform in university libraries[J]. Library Science Journal, 2015(8):113-116.

[17] DAI Qiongjie. Research on operation mechanism of Shaanxi scientific data sharing platform[D]. Xi' an: Xidian University, 2011.

[18] LI Huaan, CHANG Chaowen. Design and implementation of Web service middleware for scientific data sharing platform[J]. Microcomputer Information, 2006(12):247-249.

[19] FANG Li, WANG Wenjie, GAO Zhenji, et al. Design and practice of environmental science data sharing platform based on SOA[J]. Journal of Environmental Engineering Technology, 2014(4):333-340.

[20] AN Jiwen, ZHUANG Dafang, YUAN Wen. Design of resource and environmental science data sharing platform for geoscience computing[J]. Geoinformation Science, 2007(3):34-39.

[21] MA Hongwang. Research and implementation of environmental science data sharing platform based on Geoportal[D]. Xiangtan: Hunan University of Science and Technology, 2012.

[22] CHENG Wei. Research and implementation of space environment science data sharing platform[D]. Wuhan: China University of Geosciences, 2010.

[23] LI Huaan. Research and implementation of scientific data sharing platform based on global database[D]. Zhengzhou: PLA Information Engineering University, 2006.

[24] FU Yan. Design and implementation of East China Sea scientific data node platform based on WebGIS[J]. Sci-tech Information Development & Economy, 2008(14):148-150.

[25] YU Chao, CAI Weihua, HONG Zhenzhou. Design and implementation of earthquake scientific data sharing platform based on GIS[J]. Computer Applications and Software, 2011(7):86-88, 158.

[26] ZHU Ling, NIE Hua, CUI Haiyuan, et al. Construction of Peking University open research data platform: exploration and practice[J]. Library and Information Service, 2016, 60(4):44-51.

[27] LUO Pengcheng, ZHU Ling, CUI Haiyuan, et al. Construction of Peking University open research data platform based on Dataverse[J]. Library and Information Service, 2016, 60(3):52-58.

[28] WANG Zhi, WANG Li. Design and implementation of Datong scientific data sharing platform[J]. Software Guide, 2015, 14(12):81-83.

[29] XU Shujuan, JIANG Yanxia. Research on constructing industry scientific data sharing platform for coal university library consortium[J]. Information Research, 2016(9):67-70, 75.

[30] ZHANG Tong, LI Mingjia. Design of scientific research data management platform for higher vocational colleges: a case study of Liaoning Financial Vocational College[J]. Liaoning Higher Vocational Technical Institute Journal, 2016(9):92-94.

[31] WANG Tingting. Research on scientific data security supervision for military academy libraries based on trusted cloud computing[J]. Digital Library Forum, 2017(6):16-24.

[32] WU Ningbo. Research on construction of university open scientific research data scope and application platform[D]. Guiyang: Guizhou University of Finance and Economics, 2016.

[33] LIU Runda, PENG Jie, TU Yong. A classification and navigation platform construction scheme for scientific data resources based on multi-dimensional keywords and classification association[J]. New Technology of Library and Information Service, 2010(9):67-70, 75.

[34] LI Jianhui, SHEN Zhihong, MENG Xiaofeng. Scientific big data management: concepts, technologies, and systems[J]. Journal of Computer Research and Development, 2017(2):235-247.

[35] National Basic Science Data Sharing Service Platform[EB/OL].[2018-06-05].http://nsdata.cn/collection/about.

[36] ZHANG Jilong, ZHU Qin, YIN Shenqin. Social science data sharing and services in the United States[J]. Journal of Academic Libraries, 2013(5):13-17.

[37] CUI Yuhong. Analysis of institutional research data management practice: models, core services, and priority strategies[J]. Information Studies: Theory & Application, 2017(8):19-22, 29.

[38] CHOURASIA A, WONG M, MISHIN D, et al. SeedMe: a scientific data sharing and collaboration platform[EB/OL].[2017-11-14].https://www.seedme.org/sites/seedme.org/files/publi xsede2016-preprint.pdf.

[39] KRAFT A, RAZUM M, POTTHOFF J, et al. The radar project: a service for research data archival and publication[EB/OL].[2017-12-16].https://www.mdpi.com/2220-9964/5/3/28/htm.

[40] HERRICK R, HORTON W, OLSEN T, et al. XNAT central: open sourcing imaging research data[J]. NeuroImage, 2016(1):1093-1096.

[41] HERZINGER S, GU W, SATAGOPAM V, et al. SmartR: an open-source platform for interactive visual analytics for translational research data[J/OL].[2017-12-21].https://www.ncbi.nlm.nih.gov/pubmed/28334291. Doi:10.1093/bioinformatics/btx137.

[42] HUANG J, ZHANG X C, EISENHAUER G, et al. Scibox: online sharing of scientific data via the cloud[C/OL]. IEEE 28th International Parallel and Distributed Processing Symposium. Phoenix: AZ, 2014:145-154[2018-03-04].https://www.cc.gatech.edu/~jhuang95/papers/scibox-ipdps14-slides.pdf.

Author Contributions

Cui Xu: Designed research framework, guided writing, drafted sections, revised manuscript;
Zhao Ximei: Conducted data surveys, drafted sections;
Wang Zheng: Conducted data surveys, revised manuscript;
Du Fengrui: Data compilation, chart preparation.

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