Open Access Publishing Platforms for Scientific Data in Europe and America: A Survey of Services and Implications (Postprint)

Authors: Qin Shun, Wang Quanli, Xing Wenming

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

Abstract

[Purpose/Significance] The value of data and the importance of scientific data open access publishing have become widely recognized, and the development experience of European and American scientific data open access publishing platforms can serve as a valuable reference. [Method/Process] This study selects 14 typical scientific data open access publishing platforms from European and American regions as samples, and based on the "collection-distribution-reuse" lifecycle of scientific data publishing, investigates five aspects: scientific data publishing policies or visions, scientific data integration, identification, and interaction, scientific data publishing and distribution, scientific data citation, and data lifecycle management and publishing quality control, thereby summarizing the characteristics and experience of their service construction. [Results/Conclusion] This study derives beneficial insights for the service construction of China' s scientific data open access publishing platforms: formulating policies covering the entire lifecycle of scientific data publishing, and emphasizing key issues in the construction of scientific data publishing services such as data integration, data identification, data citation, and data review.

Full Text

Preamble

Research and Enlightenment on the Services of Open Access Publishing Platforms for Scientific Data in Europe and America

Qin Shun, Wang Quanli, Xing Wenming

School of Public Administration, Xiangtan University, Xiangtan 411105

Abstract

[**Purpose/Significance**] The value of data and the importance of open access publishing for scientific data have become widely recognized, and the construction experience of European and American scientific data open access publishing platforms holds significant reference value. [**Method/Process**] This study selected 14 typical scientific data open access publishing platforms in Europe and America as samples. Based on the "collect-distribute-reuse" lifecycle of scientific data publishing, we investigated five aspects: scientific data publishing policies or visions; scientific data integration, identification, and interaction; scientific data publishing and distribution; scientific data citation; and data lifecycle management and publishing quality control. We then summarized the characteristics and experiences of their service construction. [**Result/Conclusion**] We distilled beneficial insights for the service construction of China' s scientific data open access publishing platforms: formulate policies covering the entire lifecycle of scientific data publishing, and emphasize key issues in scientific data publishing service construction such as data integration, data identification, data citation, and data review.

Keywords: scientific data; open access; data publishing; data identification; data citation

Introduction

Scientific data serves as a prerequisite for supporting scientific and technological innovation, economic development, and national security, and holds high reusable value. Consequently, scientific data openness and publishing have received significant attention from international organizations and nations. The EU Horizon 2020 program proposes ensuring open access to scientific publications and promoting open access to scientific data for publicly funded research in Europe. The FOSTER Plus project, funded under this program, aims to facilitate the practical implementation of open science for Horizon 2020 and beyond, identifying "open access" and "open data" as the two main forces driving the open science movement [1]. The OECD' s 2004 *Declaration on Access to Research Data from Public Funding* [2] and the European Commission' s 2016 *OA2020 Initiative Action Plan* [3] have both extended open science to the field of open scientific data publishing. The Chinese government also attaches great importance to scientific data sharing and publishing. Article 22 of the *Measures for the Management of Scientific Data*, issued by the General Office of the State Council in March 2018, states: "Competent departments and legal entities should actively promote the publishing and dissemination of scientific data, and support researchers in organizing and publishing scientific data with clear property rights, accuracy, completeness, and high sharing value" [4], establishing an open publishing model for scientific data sharing.

Previous research on scientific data publishing has primarily focused on four aspects: (1) Motivations for scientific data publishing. For instance, D.S. Sayogo

et al. [5], through survey results from the DataONE project working group, identified two key factors motivating data sharing and publishing: data management skills and organizational support, and recognition mechanisms for dataset creators formed through legal policies. (2) Scientific data publishing models. There are various classification methods, including two, three, four, and five-model approaches. Representative works include Huang Guobin et al. [6], who summarized integrated publishing and independent publishing models for scientific data, and Tu Zhifang [7], who identified three models: independent data publishing, data publishing as paper appendices, and data paper publishing. (3) Key issues in scientific data publishing. Scholars have emphasized research on individual stages or processes of scientific data publishing, with some examining it from a data lifecycle perspective. Tu Zhifang [7] argues that data identification, data citation, and data review are key issues distinguishing data publishing from general data sharing. (4) Scientific data publishing platforms. Most scholars have focused on introducing platform functions in detail, such as D. Roman et al. [8] describing DataGraff' s data conversion, publishing, and hosting functions; Wang Dandan [9] analyzing user experience requirements and basic functional requirements for scientific data publishing platforms through contextual interviews with researchers at Nanyang Technological University and testing of the Dataverse platform; and Ma Jianling [10] reviewing research data management tools from a data lifecycle perspective, including distribution and publishing tools (platforms). Tu Zhifang et al. [11] also reviewed some scientific data open access publishing platforms and their functions. While these scholars have addressed cognitive issues in scientific data publishing from different perspectives and dimensions, few have conducted in-depth analysis of service construction for scientific data open access publishing platforms or explored them holistically. Therefore, through investigation and comparative analysis of typical European and American scientific data open access publishing platforms (hereinafter referred to as "publishing platforms"), this study aims to clarify their overall service status and characteristics, and discuss their implications for the service construction of China' s publishing platforms.

2. Survey Objects and Methods

2.1 Survey Objects

As scientific data open access publishing is a relatively new service, the number of international publishing platforms offering this service remains limited. To ensure representativeness, we established two sample selection criteria: (1) Platforms must meet the openness characteristics of scientific data open access (OA). Publishing platforms should possess fundamental open source and open features, specifically following the Creative Commons Attribution 4.0 International License (CC BY 4.0) and GitHub open source support, which are highly relevant to our survey indicators. (2) Platform organizers must be regionally representative. Europe and America have extensive practical experience in scientific data open access publishing and have long been leaders in this field, so

we selected platforms organized in these regions.

Based on these criteria, and through extensive review of scientific data publishing literature and web surveys, we selected 14 well-known publishing platforms as survey objects according to their representativeness, novelty, and data availability, with 7 platforms each from Europe and America (see Table 1). These are: Dataverse, developed by the Institute for Quantitative Social Science (IQSS) and Harvard University Library and Harvard University Information Technology organization; DSpace, jointly developed by MIT Libraries and Hewlett-Packard Labs; DataONE, developed under the NSF-funded DataNet program; DataConservancy, also initiated under the NSF-funded DataNet program; DataStaR, a data staging repository operated by Cornell University' s Albert R. Mann Library; Dryad, supported technically by vagrant-dryad with NSF funding; Samvera (formerly known as the Hydra project before May 2017), developed through collaboration between Stanford University and DuraSpace; CKAN, built by the non-profit Open Knowledge Foundation (OKF); CERN Open Data, supported technically by CERN Data Centre; DataGraft, operated by the EW-Shopp, proDataMarket, and euBusinessGraph projects; Figshare, launched by Mark Hahnel and supported by Digital Science; PANGAEA, hosted by the Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research (AWI) and the Center for Marine Environmental Sciences (MARUM) at the University of Bremen; Pure, an open data solution developed by Elsevier; and Zenodo, supported by CERN Data Centre and OpenAIRE.

2.2 Survey Methods

Based on the "collect-distribute-reuse" lifecycle of scientific data publishing, we developed five survey indicators for open access platform services: (1) Scientific data publishing policies or visions, analyzing platforms' relevant policies for scientific data open access; (2) Scientific data integration, identification, and interaction, examining platforms' tools and methods for data integration, identification, and interaction; (3) Scientific data publishing and distribution, exploring platforms' publishing and distribution pathways and models; (4) Scientific data citation, reviewing platforms' data citation tools and methods; and (5) Data lifecycle management and publishing quality control, investigating the completeness of scientific data publishing lifecycles and quality control mechanisms on publishing platforms. We primarily employed web surveys, supplemented by literature research, using in-depth platform functionality testing to systematically summarize each indicator. The survey period was from October 16, 2018 to December 2, 2018.

3. Survey Results Analysis

3.1 Scientific Data Publishing Policies or Visions

Open access to scientific data represents an inevitable trend. The UK Engineering and Physical Sciences Research Council (EPSRC) notes: "Unrestricted ac-

cess to scientific data is essential for accelerating research progress. The volume of scientific and scholarly data grows exponentially each year, yet infrastructure, policy, and technical safeguards to utilize this important resource remain lacking" [12]. Establishing scientific data open access publishing platforms is an important practice for promoting data openness both domestically and internationally, and developing policies covering the entire lifecycle of scientific data open access publishing serves as the prerequisite for their development. Our survey (see Table 2) found that all publishing platforms have visions for promoting scientific data sharing, publishing, and reuse, with approximately 35.7% having formal policies. Platforms such as Dataverse, Dryad, CERN Open Data, and PANGAEA have formulated detailed specifications, guidelines, and terms for scientific data publishing. For example, Dryad promotes scientific data publishing policies including Dryad data publishing content standards, prohibited data descriptions, data withdrawal and deletion procedures, rights and obligations of submitters and users, and privacy policies [13]. Dataverse provides community norms, Harvard Dataverse general terms of use, data preservation policies, data privacy policies, Dataverse API terms of use, sample data use agreements, and dataset replication guidelines [14].

3.2 Scientific Data Integration, Identification, and Interaction

3.2.1 Scientific Data Integration Using Semantic Web, XML, and other technologies for data encapsulation, integration, and semantic association represents a currently widespread approach. Our survey found that all platforms use XML for data encapsulation and integration. Notable personalized approaches include DataONE using the EML standard (Ecological Metadata Language) for metadata editing and integration, DataStaR employing Semantic Web for scientific data semantic linking, DataConservancy utilizing RMap (linked data graph) and GUI for data association and dataset packaging, DataGraft explicitly using RDF resource description framework for data description, and PANGAEA and Zenodo implementing OAI-PMH interfaces for data harvesting, collection, and integration to broaden scientific data sharing and publishing scope. Evidently, semantic linking represents a major trend in scientific data integration.

3.2.2 Scientific Data Identification Data identification is the prerequisite for data publishing, distribution, and citation, and also serves as the hub for data encapsulation, integration, and interaction. The massive volume of data increases the difficulty of locating and identifying scientific data, necessitating data integration and identification. Digital Object Identifier (DOI) features unique and permanent identification, making it suitable for and beneficial to data open access publishing, with the assigned DOI accompanying data throughout the entire publishing process [24]. As shown in Table 3 , 10 of the 14 publishing platforms use DataCite, EZID, CrossRef, and other tools for DOI registration, while 12 platforms collaborate with ORCID Inc. to assign unique author IDs—Open Researcher and Contributor IDs. Both DOI and ORCID help resolve digital copyright management and intellectual property issues, reducing

data interaction disputes.

3.2.3 Scientific Data Interaction Table 3 shows that 11 publishing platforms enable scientific data interaction through APIs, with RESTful APIs (Representational State Transfer) being the primary approach. RESTful APIs feature unified URI interfaces and can implement multi-format data calls based on HTTP protocols, greatly expanding the coverage of scientific data open access publishing. Additionally, platforms like DataStaR have innovated data interaction methods through self-developed frameworks. However, using APIs for scientific data identification, association, and interaction represents a unified and standardized pathway.

In summary, scientific data integration, identification, and interaction are closely linked components in the data lifecycle perspective of scientific data publishing, forming the core business workflow for achieving orderly data publishing. Our analysis reveals that while European and American publishing platforms have developed and applied personalized technical tools for these functions, they also lack standardized and normalized guidance to some extent. System heterogeneity combined with diversified publishing tools significantly increases the difficulty of scientific data integration, identification, and interaction work.

3.3 Scientific Data Publishing and Distribution

European and American publishing platforms have achieved remarkable results in scientific data publishing and distribution. For example, Dataverse has collected 52,449 datasets and generated 4,102,900 downloads [25], and open source frameworks like Dataverse and DSpace have been adopted or cooperatively developed by Chinese institutions such as Peking University and Wuhan University for scientific data management and publishing platforms. Our survey (see Table 4) found that most platforms support formats such as zip, xlsx, and csv, demonstrating diversified support for scientific data publishing formats. Dryad provides the most detailed specifications, establishing preferred formats and format support levels for text, image, audio, video, compressed documents, and other data types, categorizing them into three support levels: full support, limited support, and raw bitstream access [13].

Knowledge sharing and intellectual property protection constitute the main contradiction facing scientific data publishing and content distribution, and establishing binding sharing agreements serves as the primary solution. Most European and American publishing platforms follow agreed sharing agreements for scientific data publishing and distribution. Table 4 shows that 11 platforms explicitly follow sharing agreements such as CC BY 4.0 and CC0, emphasizing intellectual property protection during scientific data publishing and distribution. Different sharing agreements have their respective advantages. Following Creative Commons (CC) and other knowledge sharing agreements provides exemptions and norms that protect researchers who use or redistribute data

authors' works from copyright infringement concerns while safeguarding data owners' rights [26], balancing copyright protection with sharing and utilization to clarify rights and responsibilities.

3.4 Scientific Data Citation

Data citation is a critical link in data publishing and an effective way to protect the rights and interests of data authors and managers [7]. Six platforms including Dataverse endorse the Joint Declaration of Data Citation Principles (FORCE11) and the FAIR data sharing principles of "Findable, Accessible, Interoperable, and Reusable" [27]. Most platforms use DOI and Universal Numeric Fingerprints (UNF) technology for data citation, following citation principles from the International Council for Science Committee on Data (CODATA) and DataCite. Through DataCite and Crossref, they assign DOIs to scientific data and generate DOI Citation Formatter programs that support different citation languages and styles [28]. Currently, the main tools supporting DOI-based scientific data citation are DataCite, Mendeley, EZID, and Zotero, which offer excellent support. For instance, Zotero is suitable for different data formats, with plugins supporting Firefox, Chrome, and Safari browsers, as well as word processing software including Word, LibreOffice, BibTeX, and LaTeX [29]. Figure 1 [Figure 1: see original paper] illustrates commonly used scientific data citation tools on publishing platforms.

3.5 Data Lifecycle Management and Publishing Quality Control

Data lifecycle management is a necessary condition for data publishing quality control. Publishing platforms with complete data publishing lifecycle management also have relatively comprehensive data review mechanisms and quality control. DataONE divides the data lifecycle into eight components: planning, collection, assurance, description, preservation, discovery, integration, and analysis [30]. After investigating the data lifecycle management of the 14 publishing platforms and referencing DataONE' s data lifecycle model and A. Sarretta' s *Research Data Lifecycle* [31], we developed a scientific data publishing lifecycle model (DPLC) comprising six stages: Collect, Identify, Publish, Distribute, Reuse, and Evaluate (see Figure 2 [Figure 2: see original paper]). Currently, few publishing platforms fully cover the scientific data publishing lifecycle, particularly in the evaluation stage, where publishing quality control urgently needs optimization. The balance between review and sharing remains under further research and practice. Introducing data quality control plans and peer review systems would benefit data publishing quality control. For example, Dryad commits that its staff and peer reviewers will examine data security, academic integrity, and technical correctness before data release [32].

4. Implications for China' s Scientific Data Open Access Publishing Platform Service Construction

China' s scientific data open access publishing platform construction started relatively late, with related services also lagging behind. However, these platforms have received significant attention from national and research education institutions, which have conducted beneficial explorations in data integration and processing, resource discovery and retrieval, data download (reuse), and sharing, indicating substantial room for development. To date, the Ministry of Science and Technology and the Ministry of Finance have established national science and technology resource sharing service platforms in eight fields including basic sciences, agriculture, and forestry. The Chinese Academy of Sciences has led the establishment of the National Basic Science Data Sharing Service Platform, covering a distributed, scalable storage system of "one primary, one backup + 12 sub-centers." Peking University, Wuhan University, and others have built scientific data sharing platforms within university systems. However, overall, China' s scientific data open access publishing practice remains weak, manifested by fewer publishing platforms, scarce related tools, and incomplete standards, technical conditions, and talent teams.

Improving China' s publishing platform scientific data open access services requires absorbing accumulated experience from European and American platforms: (1) Strengths of European and American platforms include well-developed policy systems, strong service awareness, diverse service content, high service sharing levels, robust standard systems, and more mature technical conditions. Notably, these platforms particularly emphasize the development and application of scientific data publishing-related tools from a data lifecycle perspective, effectively improving scientific data open access publishing quality and efficiency. (2) Weaknesses of European and American platforms include deviation between personalized and standardized construction. Personalized and diverse technical standards somewhat limit scientific data open access publishing, as key issues such as data integration, identification, citation, and review are constrained by disparate processes, making heterogeneous integration difficult. Therefore, data publishing standards, norms, and their technical implementation will become a major focus of future research and practice [7]. In light of this, China' s publishing platform service construction should adopt active optimization strategies.

4.1 Formulate Policies Covering the Full Lifecycle of Scientific Data Publishing

4.1.1 Standardization and Normalization Policies Achieving open access, open data, and open science requires policy research on all aspects of scientific data sharing [33]. Scientific data publishing services face complex intellectual property issues and constitute complicated business operations. To ensure worry-free implementation, developing and implementing standardization and normalization policies is particularly urgent. The issuance of the *Mea-*

sures for the Management of Scientific Data provides guidance for scientific data sharing and publishing, but the field remains exploratory [34]. Publishing platform service construction should emphasize standardization and normalization while exploring diverse and complementary open access models of integrated publishing and independent publishing. At the macro level of scientific data publishing, we should strengthen the construction of data integration, publishing services, and platform exchange and sharing, formulate standardized and normalized policies covering the entire scientific data publishing lifecycle, and establish specialized implementation groups or committees to guide and supervise publishing platform data publishing work. Relevant publishing platforms also need to formulate implementation details (such as Dryad' s data publishing content standards) based on the above, covering six layers: data collection, identification, publishing, distribution, reuse, and evaluation. Additionally, we must explore multi-party cooperation mechanisms for standardization and normalization to resolve conflicts in service concepts and technical methods during scientific data open access publishing processes.

4.1.2 Open Source and Independent R&D-driven Policies A close examination of European and American platform construction confirms the powerful advantages of open source and independent R&D, with 13 platforms supporting GitHub for code hosting and review, project management, and software building [35]. In an open source environment, open access policy-driven and standardization-constrained approaches offer solutions to China' s weak foundation in scientific data publishing. First, formulating open access policies and encouraging open source and independent R&D should be important starting points. The open access model for scientific data operates as: data authors create scientific data → store it in scientific data open access publishing platforms (data repositories) → scholars freely utilize scientific data to create new academic achievements. This model requires policy, financial, and technical support; otherwise, it cannot be sustained. Second, we must properly handle the relationship between personalized and standardized R&D. While encouraging the adoption of advanced open source platforms, we should emphasize standardized integration and the enhancement of independent R&D capabilities. For example, Peking University Library developed a disciplinary open data navigation using the Dataverse platform architecture [36], supporting scientific data open access publishing, while the Lanzhou Documentation and Information Center of the Chinese Academy of Sciences independently developed the Global Research Project Database (ProjectGate), providing data submission and review services [37]. Furthermore, we need to promote open source and independent R&D-driven policies and publishing platform usage methods to enhance the actual effectiveness of scientific data open access.

4.2 Emphasize Key Issues in Scientific Data Publishing Service Construction

4.2.1 Data Integration: Promoting Organized Data Organized integration of scientific data is the prerequisite for open access publishing. Current major challenges include system heterogeneity, inconsistent scientific data description syntax, non-uniform scientific data metadata formats, and lack of semantic association between scientific data [38]. We can learn from DataConservancy' s four key components: DC Packaging Specification, Package Ingest Service, RMap, and API-X architecture [17] to form a standardized and normalized workflow from data integration to data publishing. Through one-stop data integration, data authors can organize scientific data systematically, promoting data reuse and scientific research innovation.

4.2.2 Data Identification: Assigning Unique Permanent Identifiers China' s scientific data publishing mainly involves simple data release and sharing, which still differs from open access publishing characterized by "reliable sources, trustworthy quality, public release, public utilization, unique identification, clear intellectual property, and formal citation" [7]. Regarding data identification, we need to assign unique permanent identifiers that facilitate data interaction to promote scientific data open access publishing. The specific strategy should emphasize both data identification and author identification. DOI and ORCID both feature unique permanent identification, ensuring permanent association between scientific data and data authors—a hot practice in foreign publishing platforms that also has deep applications in China' s scientific paper identification field. Differently, Object Identifier (OID) is the unique permanent identifier for scientific data formally adopted by China' s scientific systems, obtained by scientific data authorities from the national OID registration center [39]. Therefore, we must deepen the application of OID, DOI, and ORCID identification tools to pave the way for standardized scientific data citation.

4.2.3 Data Citation: Ensuring Clear Intellectual Property Rights The open access service practices of the 14 publishing platforms demonstrate that data citation can ensure clear intellectual property rights and guarantee proper scientific data use. Therefore, we must strengthen the formulation of scientific data citation standards and the development and application of scientific data citation tools. In December 2017, China issued the national standard *GB/T 35294-2017 Information Technology–Scientific Data Citation*, which standardizes scientific data citation through "general scientific data citation formats" and "OID-based scientific data citation methods" [39]. The citation formats are: (1) General scientific data citation format: Author. Title (Version). Creating Institution [Creating Institution], Creation Time. Distributing Institution [Distributing Institution], Distribution Time. Unique Identifier; Resolution Address. (2) OID-based scientific data citation method: Scientific Data OID Identifier Prefix.Publisher Code.Scientific Data Unique Code. Both citation standards can clarify data sources, ensure scientific data uniqueness, and facilitate dec-

laration of scientific data dissemination paths. China' s publishing platforms should, under the guidance of these standards, fully consider researchers' data needs and develop automated generation tools for data identification and citation applicable to multiple data formats and complex ports, safeguarding data authors' intellectual property rights while improving the quality and efficiency of scientific data citation and dissemination.

4.2.4 Data Review: Ensuring Reliable Data Quality Publishing platforms conduct data review through two main entities: platform staff and peer reviewers. Staff focus on technical quality review, while peer reviewers emphasize scientific quality assessment. Technical quality review covers data integrity and description adequacy, while scientific quality review primarily includes data integrity, description detail, and data usefulness [7]. Scientific data peer review is one important means to ensure data quality and is significant for producing correct scientific results; some tools and processes may facilitate rapid and convenient data peer review [40]. Peer review remains weakly applied in publishing platform data review. We can learn from Dataverse' s practices by introducing data quality control plans, such as using Data Management Plans (DMP) and DMPTool to assess and test data quality, preparing for simplified peer review [41]. Additionally, publishing platform data review processes should consider intelligent review, staff quality control, and parallel peer review, balancing supervision and service to address semantic and pragmatic level data evaluation and achieve multidisciplinary, multi-domain data quality control—these issues require further research and practice.

Conclusion

This study, based on the "collect-distribute-reuse" lifecycle of scientific data publishing and employing literature research, web surveys, and comparative analysis, examined the service content of 14 scientific data open access publishing platforms regarding scientific data publishing policies or visions; scientific data integration, identification, and interaction; scientific data publishing and distribution; scientific data citation; and data lifecycle management and publishing quality control. We believe the accumulated experience can provide reference for China' s publishing platform service construction in scientific data publishing policy formulation and key business development, hoping to boost the sustained and healthy development of China' s scientific data open access publishing work. This study has limitations: we only analyzed 14 platforms from Europe and America, resulting in a slightly small sample size. Future research should further expand survey objects and scope, refine survey indicators, and enhance the accuracy and application value of research results.

References

[1] FOSTER Plus. Open science [EB/OL]. [2018-11-02]. https://www.fosteropenscience.eu/foster#taxonomy.
[2] OECD. Declaration on access to research data from public funding

[EB/OL]. [2018-11-02]. https://legalinstruments.oecd.org/en/instruments/157. [3] European Commission Directorate-General for Research/Innovation. Guidelines to the rules on open access to scientific publications and open access to research data in Horizon 2020 [EB/OL]. [2018-11-02]. https://legalinstruments.oecd.org/en/instruments/157. [4] General Office of the State Council. Notice on issuing the Measures for the Management of Scientific Data [EB/OL]. [2018-11-03]. http://www.gov.cn/zhengce/content/2018-04/02/content_{5279272}.htm. [5] SAYOGO D S, PARDOTO T A. Exploring the determinants of scientific data sharing: understanding the motivation to publish research data [J]. Government information quarterly, 2013, 30(S1): 19-31. [6] HUANG Guobin, WANG Shu, QU Yajie. A comparative study of scientific data publishing models [J]. Journal of academic libraries, 2018, 36(1): 34-40, 33. [7] TU Zhifang. A review of basic issues and identification of key problems in scientific data publishing [J]. Library, 2018(6): 86-92, 100. [8] ROMAN D, DIMITROV M, NIKOLOV N, et al. Datagraft: simplifying open data publishing [C]//European Semantic Web Conference: The Semantic Web. Berlin: Springer, 2016: 101-106. [9] WANG Dandan. User testing research on scientific data publishing platforms [J]. Information and documentation services, 2017(6): 56-61. [10] MA Jianling, CAO Yuezhen. Research on the development of research data management tools [J]. Library science research, 2014(15): 40-47. [11] TU Zhifang, LIU Ziheng. A review of academic libraries' participation in open access publishing at home and abroad—from the perspective of publishing pathways [J]. Library and information, 2017(3): 61-71. [12] SPARC. Open data [EB/OL]. [2018-11-09]. https://sparcopen.org/open-data/. [13] DRYAD. Terms of service [EB/OL]. [2018-11-04]. https://datadryad.org/pages/policies#publication. [14] DATAVERSE. Dataverse community norms [EB/OL]. [2018-11-04]. https://dataverse.org/best-practices/dataverse-community-norms. [15] DSpace. About DSpace [EB/OL]. [2018-11-04]. https://duraspace.org/dspace/about/. [16] DataONE. What is DataONE? [EB/OL]. [2018-11-04]. https://www.dataone.org/what-dataone. [17] DataConservancy. Building infrastructure for data curation [EB/OL]. [2018-11-04]. https://dataconservancy.org/. [18] ResearchGate. DataStaR: a data sharing and publication infrastructure to support research [EB/OL]. [2018-11-04]. https://www.researchgate.net/publication/267776401_{DataStaR}A_{{{Data}}{{Sh

[19] SAMVERA. How it works [EB/OL]. [2018-11-04]. https://samvera.org/. [20] CKAN. CKAN, the world' s leading open source data portal platform [EB/OL]. [2018-11-04]. https://ckan.org/. [21] CERN. CERN open data terms of use [EB/OL]. [2018-11-05]. http://opendata.cern.ch/docs/terms-of-use. [22] DATAGRAFT. About us, DataGraft development and funding [EB/OL]. [2018-11-04]. https://datagraft.io/about-us. [23] PANGAEA. About [EB/OL]. [2018-11-05]. https://www.pangaea.de/about/. [24] TU Yong, PENG Jie. Research on the application of digital object unique identifier in China's scientific data field [J]. Digital library forum, 2013(8): 31-36. [25] DATAVERSE. Open source research data repository software [EB/OL]. [2018-11-10]. https://dataverse.org/. [26] WIKIPEDIA. Creative Commons license [EB/OL]. [2018-11-10]. https://en.wikipedia.org/wiki/Creative{{{Commons}}_{{license}}}. [27]

WILKINSON M D, DUMONTIER M, AALBERSBERG I J J, et al. The FAIR Guiding Principles for scientific data management and stewardship [EB/OL]. [2018-11-10]. https://www.nature.com/articles/sdata201618.pdf. [28] DATACITE. Citation formatter [EB/OL]. [2018-11-11]. https://www.datacite.org/citation.html. [29] MIT Libraries. Citation management and writing tools: citation management tools [EB/OL]. [2018-11-11]. https://libguides.mit.edu/citewrite/citetools. [30] DataONE. Data life cycle [EB/OL]. [2018-11-11]. https://www.dataone.org/data-life-cycle. [31] SARRETTA A. Research data life cycle [EB/OL]. [2018-11-11]. https://doi.org/10.5281/zenodo.1149049. [32] TU Zhifang, LIU Ziheng. Data publishing quality control practices of foreign data repositories [J]. Library development, 2019(1): 1-8. http://kns.cnki.net/kcms/detail/44.1306.G2.20180818.1836.002.html. [33] GU Liping. Data management services in the transformation of research models: approaches to achieving open access, open data, and open science [J/OL]. Journal of library science in China, 2018(6): 1-15. https://doi.org/10.13530/j.cnki.jlis.180018. [34] XING Wenming, HONG Cheng. Openness as the norm, closed as the exception—interpreting scientific data sharing and utilization in the Measures for the Management of Scientific Data [J/OL]. Library tribune, 2018(3): 5-13. [35] GitHub, Inc. Built for developers [EB/OL]. [2018-11-13]. https://github.com/. [36] Peking University Library. Disciplinary open data navigation [EB/OL]. [2018-11-14]. http://www.lib.pku.edu.cn/portal/cn/fw/sjfw/xuekeshuju/. [37] Lanzhou Documentation and Information Center, Chinese Academy of Sciences. Global research project database: know what you want, offer what you need [EB/OL]. [2018-11-17]. http://project.llas.ac.cn/index.jsp. [38] BAI Rujiang, LENG Fuhai. Research on scientific data integration in the "big data" era [J]. Information theory and practice, 2014, 37(1): 94-99. [39] General Administration of Quality Supervision, Inspection and Quarantine, Standardization Administration of China. Information technology—Scientific data citation [EB/OL]. [2018-11-14]. http://c.gb688.cn/bzgk/gb/showGb?type=online&hcno=A495CA355BAF00D962AA8DD84C3B2C16. [40] QU Baoqiang, WANG Kai. Scientific data peer review from the perspective of data publishing [J]. Library journal, 2017, 36(10): 71-77. [41] DATAVERSE. Data management plans [EB/OL]. [2018-11-15]. http://best-practices.dataverse.org/data-management/.

Author Contributions

Qin Shun: Research topic and framework determination, data investigation, writing and revision. **Wang Quanli:** Research approach revision, guidance on paper writing. **Xing Wenming:** Paper revision suggestions, data proofreading.

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