Research Status and Prospects of Scientific Data Sharing: Postprint

Authors: Bi Datian, Cao Ran, Du Xiaomin

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

Abstract

[Purpose/Significance] This study analyzes the current research status and development trends of scientific data sharing, providing references for advancing domestic theoretical research and practical progress. [Method/Process] By reviewing and synthesizing literature from home and abroad, three major themes related to scientific data sharing research are identified: the development needs of scientific data sharing, the management and sharing models and mechanisms of scientific data, and the sharing degree and service effectiveness of scientific data services. A scientific data sharing management analysis framework is constructed, which reflects the causal relationships among these three themes and emphasizes the rationality, importance, and criticality of scientific data sharing management and services. [Results/Conclusion] Existing research both domestically and internationally still has deficiencies. Future research should strengthen empirical studies, focus on improving management mechanisms, and simultaneously emphasize innovation in sharing service models.

Full Text

Preamble

ChinaXiv Cooperative Journal
Volume 63, Issue 24, December 2019

Research Status and Prospects of Scientific Data Sharing

Bi Datian[1], Cao Ran[1], Du Xiaomin[2]
[1]School of Management, Jilin University, Changchun 130022
[2]Department of Economic Management, Yingkou Institute of Technology, Yingkou 115014

Abstract

[**Purpose/Significance**] This study analyzes the research status and development trends of scientific data sharing to provide references for promoting relevant theoretical research and practical progress in China. [**Method/Process**] Through reviewing and summarizing domestic and foreign literature, we extracted three major themes related to scientific data sharing research: development needs for scientific data sharing, management models and mechanisms for scientific data sharing, and service sharing degree and effectiveness of scientific data services. We constructed an analytical framework for scientific data sharing management that reflects the causal relationships among these three themes, emphasizing the rationality, importance, and criticality of scientific data sharing management and services. [**Result/Conclusion**] Existing research both domestically and internationally remains insufficient. Future research should strengthen empirical studies, focus on improving management mechanisms, and emphasize innovation in service models.

Keywords: scientific data sharing; data management; data services
Classification Number: G250
DOI: 10.13266/j.issn.0252-3116.2019.24.008

Scientific data refers to a category of data generated through scientific research activities or other means that reflect the essence, nature, and changing patterns of relevant objects in scientific research activities and support such activities. With the advent of the big data era and its widespread application across various fields, scientific discovery has emerged as a fourth new paradigm called "data-intensive science," characterized by data-centric and data-driven approaches that discover new knowledge through processing and analyzing massive datasets. The importance of scientific data as a national strategic resource has become increasingly prominent, serving as a strategic resource supporting scientific and technological innovation. Examples include massive DNA data for understanding life, the discovery of the Higgs boson from one trillion events, global carbon monitoring networks, and ARGO ocean buoys [1].

Building upon the important role of scientific data sharing, this study examines, organizes, and summarizes current domestic and international research and practice in scientific data sharing, analyzes its research status and development trends, and proposes reasonable prospects for future scientific data sharing in China.

1 Research Design

1.1 Data Sources

Data were sourced from CNKI and the Web of Science Core Collection, covering the period 2000-2019. The search criteria included topics of "data sharing" or "data reuse" combined with "scientific data" or "research data," yielding 73 documents. Through citation tracing, 156 additional references were identified.

After removing duplicates and irrelevant items, 89 highly relevant documents were ultimately selected.

1.2 Construction of the Analytical Framework

Scientific data sharing constitutes an important component of scientific data management, referring to the public release of scientific research data through various forms for use by others [2]. Scientific data sharing originates from the common demand for scientific data resources driven by scientific and technological innovation activities and in-depth academic exchanges. According to an online survey of 6,344 researchers by A. Mulligan et al. [3], 67% of researchers considered it necessary to obtain supporting data for research, 75% hoped to access others' research data, and 52% expressed willingness to provide data to others. These results indicate that most researchers believe scientific data sharing is significant for scientific research.

Questions about how to effectively promote scientific data sharing, establish and improve management mechanisms, unleash and fully exploit the innovative value of scientific data, and guide and motivate researchers to participate in data sharing have attracted active attention and discussion from scholars. However, existing research is scattered across various industries and platform sharing practices, lacks effective integration, and to some extent neglects reasonable consideration of complete causal relationships, resulting in research themes not being effectively revealed.

Through reviewing and summarizing domestic and foreign literature, we extracted three major themes related to scientific data sharing research: development needs for scientific data sharing, scientific data management models and mechanisms, and scientific data service sharing degree and service effectiveness, aiming to achieve a holistic grasp and deeper understanding of the scientific data sharing process. The analytical framework for scientific data sharing management constructed in this study is shown in Figure 1 [Figure 1: see original paper], which reflects the causal relationships among these three themes and emphasizes the rationality, importance, and criticality of scientific data sharing management and services.

2 Analysis of the Scientific Data Sharing Management Framework

2.1 Scientific Data Sharing Driven by Development Needs

Implementing scientific data open sharing represents a strategic demand for national development and has become an international trend, as evidenced by the successive establishment of institutions oriented toward scientific data management and sharing and the emergence of scientific data centers (or data sharing platforms) [4-5].

Since the 1950s, as the scientific community's demand for exchanging and applying basic scientific data has become increasingly apparent and recognition of the value of scientific data has grown, the fundamental status of scientific data has been gradually established. The International Council for Science (ICSU) organized the establishment of the World Data Center (WDC) in 1957, focusing on earth science, space science, and astronomical data, aimed at collecting, organizing, systematizing, standardizing, and exchanging data in earth science, environmental, and space science fields. To address sharing barriers across different disciplines, ICSU established the Committee on Data of the International Science Council (CODATA) in 1966, covering broader disciplinary areas. In March 2001, CODATA launched the electronic journal *Data Science Journal* focused on scientific data, establishing scientific data as a new discipline [6].

In March 2004, the UK's Joint Information Systems Committee (JISC) initiated the establishment of the Digital Curation Center (DCC), which actively engages in practices related to scientific data preservation, management, and sharing. In March 2013, the European Commission (EC), the U.S. National Science Foundation (NSF), the U.S. National Institute of Standards and Technology (NIST), and the Australian Innovation Department (AID) jointly formed the Research Data Alliance (RDA), creating conditions for researchers to openly share scientific data across technical and disciplinary fields and national borders, and promoting global scientific data exchange.

Under the leadership and promotion of these data sharing management organizations, governments have gradually conducted overall planning at the national level to establish scientific data center systems. For example, the United States has built a massive scientific data center system, including 13 disciplinary World Data Centers and 9 national data centers. The World Data Centers include the Center for Atmospheric Trace Gases, Center for Marine Geology and Geophysics, Seismology Center, Center for Human-Environment Interactions, Meteorological Data Center, Oceanography Center, Remote Sensing Land Data Center, Rocket and Satellite Center, Solar-Terrestrial Physics Data Center, Earth Rotation Center, Solid Earth Geophysics Center, and Paleoclimatology Center. A typical case among national data centers is NASA's "Distributed Active Archive Centers" project, which includes national-level data centers such as NASA, the National Science Foundation (NSF), the Department of Energy (DOE), and the National Institutes of Health (NIH) [7].

China first proposed scientific data sharing at the end of 2001, launching a scientific data sharing engineering project involving 24 fields. In July 2004, the Ministry of Science and Technology initiated the construction of the National Science and Technology Infrastructure Platform, implementing a "people-oriented" philosophy and shifting the focus of national scientific and technological innovation to creating a favorable institutional environment, policy environment, and basic conditions for science and technology. In March 2015, Sugon released the "Data China" strategy, proposing the vision of "enabling the whole society to share data value." In August 2015, the state issued the "Action Outline for

Promoting Big Data Development," which pointed out the need to accelerate government data opening and sharing and promote resource integration [8]. In March 2018, the General Office of the State Council promulgated the "Measures for the Management of Scientific Data," which mandated the mandatory submission of scientific data generated by national science and technology plan projects, reflecting China's increasing emphasis on scientific data sharing [9].

To date, under the influence of user foundations, technical support, and policy advancement, multiple scientific data sharing platforms led by government departments (and their affiliated institutions) have officially entered the operational service stage, such as the National Earth System Science Data Center, National Population and Health Science Data Center, National Agricultural Science Data Center, National Seismological Science Data Center, China Meteorological Data Network (Meteorological Data Sharing Center), and National Forestry and Grassland Science Data Center [10]. Although these platforms have promoted scientific data sharing and reuse to some extent, they still struggle to meet the requirements of rapidly developing research paradigms due to limitations in the timeliness, depth, and breadth of data openness.

In summary, sharing development needs originate from three levels: national strategic development, the scientific community, and research teams. Demands at different levels can stimulate and promote the development of scientific data sharing activities to a certain extent, ultimately reflecting the common goal of pursuing the realization of scientific data value in scientific data open sharing activities. The aforementioned scientific data sharing plans represent overall planning dominated by national development needs, reflecting the state's capacity to integrate resources from various departments and units and create new sharing orders through increased national investment [11]. Since the 1990s, the U.S. government has gradually built a data sharing framework at the national level, incorporating data management activities of governments at all levels, federal departments, commercial companies, and individuals into the national overall planning to form a national-level scientific data management and service system. In 2002, China officially launched its scientific data sharing engineering project, which, under national overall planning and management, incorporated scientific data resources accumulated by various departments and units into a unified national scientific data sharing framework, aiming to form a cross-departmental, cross-disciplinary, and multi-level national scientific data sharing service system, achieve classified and graded sharing of basic and public welfare scientific data accumulated over many years, and enable the full realization and appreciation of the potential value of massive scientific data resources.

2.2 Scientific Data Sharing in the Management Context

Scientific data sharing has a processual nature throughout the scientific data management lifecycle. It is a value-creating activity directly associated with data production, organization, storage, publication, dissemination, and citation, reflecting the synergy among stakeholders including data producers, funders, organizers, publishers, disseminators, managers, and users [12]. For the value-creating activities in the scientific data sharing management lifecycle, two forces promote the efficient completion of scientific data sharing value creation: first, internal dynamic conditions for scientific data sharing, mainly referring to the interconnection among various elements within data management institutions and their adaptation and coordination in response to external environmental changes, which primarily reflects a data management institution's data management capabilities; second, external supporting conditions for scientific data sharing, mainly referring to external environmental factors in which sharing activities occur, including relevant organizations or institutions (stakeholders), the completeness of sharing legal systems, and the information technology environment.

Scientific data sharing management is an important theme in academic and industry research. Existing research has primarily focused on scientific data production and organization, as well as scientific data publication and utilization.

2.2.1 Scientific Data Sharing Management Models Huang Dingcheng argues that the stability of scientific data sharing management system operation depends on the interaction mechanisms among five main controlling factors: development needs, data resources, organizational management, sharing rules, and sharing technology. Data resources, sharing rules, and sharing technology are necessary conditions for sharing management, while data resource managers and users are sufficient conditions for sharing platforms to meet development needs and realize their utility value [11].

Scientific data production and organization have epistemological significance in perceiving scientific data and its movement. In terms of generation methods, scientific data mainly originates from ground observation (monitoring), remote sensing, surveys, statistical analysis, experimental testing, processing, computational simulation, and internet mining. In the United States, scientific data output can be roughly divided into three types from government departments, commercial sectors, and universities/research institutions [13]. Under current circumstances in China, scientific data acquisition and accumulation are mainly completed through government investment, with three primary sources: professional data generated by government department special plans; professional data produced by various national science and technology plans and special scientific and technological activities; and professional databases from research institutes and higher education institutions, international scientific data resources from Chinese branches of international scientific data organizations, and even scientific data held by individual scientists [14].

Due to different generation methods of shareable data resources, research institutions have promoted the integration of distributed and heterogeneous data resources. A. Chilvers et al. [15] believe that metadata is key to scientific data organization and management in a general sense, facilitating the organization and acquisition of scientific data in a unified form and becoming an effective way to achieve multi-dimensional association between scientific data and its supporting literature. Currently, scientific data sharing platforms in operation in China use metadata for data description [10]. The level and quality of research data organization represent a concentrated reflection of understanding research data characteristics and their literature association features. Research data contains a wealth of knowledge and relationships, such as research subjects (personnel, institutions, countries/regions, etc.), research activities (experiments, projects, etc.), research conditions (methods, instruments, funding, journals, etc.), and research outputs (results, patents, standards, etc.). These research entities have interwoven multi-dimensional relationships, such as personnel-related, institution-related, result-related, and theme-related connections. Within the same type of data resources, there also exist relationships of inheritance, subordination, and correlation. At the data organization level, reasonable scientific data classification management and system integration can effectively enhance scientific data value-added, promote data resource circulation and sharing, and highlight the dynamic, systematic, and proactive development style of data resources.

Currently, the main challenges in data organization include system heterogeneity, inconsistent data description syntax, non-uniform scientific data metadata formats, and lack of semantic associations among scientific data [16].

Data publishing refers to the release of research datasets through certain public mechanisms, enabling the public to discover, obtain, evaluate, and apply these datasets according to certain rules [17]. Scientific data publishing is one of the important forms to ensure effective data sharing. Scholars at home and abroad have conducted research on data publishing models [18-21], key issues in data publishing [22-25], and technical application issues in data publishing [26-28]. The ultimate goal of scientific data publishing is to enable data users to conveniently use relevant data. B. Lawrence et al. [29] pointed out that data publishing means data reaches a state of being citable and traceable. Therefore, data publishing requires simplicity, citable data, and convenient data usage [30].

Scientific data citation refers to describing scientific data resources used through methods similar to literature citation [31]. As one of the important ways to achieve data sharing, data citation serves as a bridge connecting data users and data producers, helping to clarify responsibilities and obligations for data sharing and achieving the goal of value-added utilization of research data [32]. Focusing on scientific data citation from a policy perspective is a research hotspot in foreign academia, with the UK, US, Australia, and the EU having formulated policies and implementation plans for research data citation and sharing. Based on a comprehensive investigation of domestic and foreign scientific data citation standards, Shi Yali used case analysis to deeply analyze Australian ANDS data citation practices from four aspects: data citation service implementation, joint promotion of data citation standard implementation guidelines, development and application of data citation systems, and feedback on standard implementation effects [33]. Foreign scholars have explored scientific data citation principles

[34], the joint statement on data citation principles recognized by six publishing platforms including Dataverse (FORCE11) and the FAIR data sharing principles of "Findable, Accessible, Interoperable, and Reusable" [35], relevant standards [36], and data citation technologies [37], such as most publishing platforms using DOI and Universal Numeric Fingerprint (UNF) technology for data citation while following citation principles from the International Council for Science Committee on Data (CODATA) [38].

Domestic scholars have systematically summarized and analyzed research on data citation behavior [39-42], data citation norms [43-46], and data citation strategies [47], pointing out that Chinese researchers' data citation behaviors show significant differences across disciplines regarding cited data types, with non-original data citations still accounting for a large proportion and varying degrees of metadata element information disclosure. Due to lack of cooperation and communication among stakeholders such as publishers, data repositories, and researchers, domestic data citation requirements lag behind foreign counterparts in terms of the number, types, and extensibility of citation elements, lacking uniformity and professionalism with relatively broad citation content. Therefore, strengthening cooperation and communication among publishers, data repositories, researchers, and other stakeholders, and establishing a complete scientific data citation system and mechanism are development directions for promoting standardized scientific data citation.

In summary, when data generation (observation, computation, experimental methods) is scientifically sound, data description information is comprehensive (complete data, standardized metadata), data maintenance is timely (good data update status), data access is more convenient, and data usage meets requirements (reasonable accuracy), users can perceive higher data value and thus demonstrate stronger sharing willingness.

2.2.2 Data Sharing Mechanisms Scientific data sharing mechanisms refer to the interaction processes, methods, and operational patterns among elements of the scientific data open sharing system [48]. Data sharing mechanisms play an important role in creating new orders for data sharing and promoting the healthy development of data sharing platforms. Establishing and improving the main mechanisms for scientific data sharing is closely related to the five basic elements of data sharing. Data resource organization management, establishment of sharing rules, and realization of sharing technology require corresponding guarantees from data resource coordination mechanisms, policy and regulation guarantee mechanisms, technical support mechanisms, performance evaluation and incentive mechanisms, and data security supervision mechanisms.

Research on scientific data sharing was earlier related to technical support for data sharing [49-52]. With the operation of data sharing platforms domestically and internationally, policy and regulatory issues for data sharing rules have become research hotspots [53-56], achieving certain research results. Based on analysis of foreign scientific data open sharing policies, Zhang Xiaoqing et al. [57]

summarized their content into three categories of elements: policy requirements, including requirements for data quality and standards, data acquisition and sharing, data protection and preservation, and data management plans; policy regulations, including regulations on privacy, traditional knowledge, sensitive data, intellectual property rights, or data ownership; other explanations, including descriptions of the purpose, principles, scope, roles and responsibilities, supervision, and implementation of scientific data open sharing. From the perspective of domestic scientific data open sharing practice, although China's scientific data sharing services have achieved initial success, the lack of national or industry-level scientific data open sharing policies has affected the orderly development of scientific data sharing practice in China. Therefore, establishing a complete set of scientific data sharing policies and regulatory systems with interconnected content, complementary systems, and coordinated effectiveness will be the direction of future research efforts.

The scientific data sharing process involves interdependence and mutual promotion among stakeholders including data producers, funders, organizers, publishers, disseminators, managers, and users. Data drives collaboration among relevant stakeholders, builds consensus, and aggregates individual efforts into collective strength. We should gradually establish and improve a scientific data open sharing policy system at the national, industry, and institutional/enterprise levels to provide effective guidance for inter-departmental collaboration (division of labor in data collection scope), central-local collaboration, collaboration among scientists, and collaboration between data producers and managers. Additionally, integrating data sharing mechanisms into the scientific data sharing environment ensures timely, accurate, and comprehensive circulation of data flows while stimulating, guiding, and strengthening researchers' participation in data sharing.

2.3 Service Sharing Degree and Service Effects

Scientific data services are dynamic, cross-boundary, and embedded services that meet national development needs, research team needs, and researchers' needs, as well as the development needs of data service institutions that support research [58]. The fundamental task of data sharing services is to connect data owners and data users, promoting the dissemination and reuse of data resources. In 2011, the Ministry of Science and Technology and the Ministry of Finance formulated the "Performance Evaluation Indicators for National Science and Technology Infrastructure Platform Operation Services," which focuses on examining service quantity and effectiveness of science and technology platforms. The indicators include four first-level indicators—service quantity, service effectiveness, operation management, and resource integration—and 12 second-level indicators [59]. In April 2011, the National Science and Technology Infrastructure Platform Construction Basic Science Data Sharing Network Project Team proposed the "Sharing Service Evaluation Indicator System," which designed six first-level indicators including data resource construction capability, technical capability, organization and management, sharing service, effectiveness, and characteristics, along with 20 second-level indicators [60]. With the launch of China's scientific data sharing engineering project, research and practice on data service sharing performance evaluation have attracted attention from scholars and industry [61-62]. In recent years, influenced by the "user satisfaction-oriented" service performance perspective [63-64], data sharing service evaluation has shifted from focusing on the demand side to "user satisfaction-oriented" service effectiveness [65-66], taking service capabilities that meet user needs as an important standard for evaluating service effectiveness [67].

2.3.1 Service Capability and Sharing Degree Service capability and sharing degree reflect a data management institution's user coverage width, user usage frequency, and service quality in terms of improving data quality, timeliness, and supply promptness. User coverage width and usage frequency are related to data sharing service positioning. The NSF categorizes scientific data into three types: research data, resource data, and reference data. From a service positioning perspective, research data refers to datasets required for research, reflecting the value of research teams and serving only specific research groups (project research teams). Resource data refers to key research datasets and datasets needed by the scientific community, reflecting the value of the scientific community and serving specific research groups (organizations and institutions in a certain field or discipline or cross-system organizations). Reference data refers to datasets important to the nation and society, reflecting social value and serving the entire scientific community. Reference data has large user groups and broad application impact, with higher open sharing requirements than resource and research data [68-69].

From the perspective of data sharing and utilization behavior, the characteristics of research teams' data sharing and utilization behavior aim at utilization in specific research directions for the research team, where data realizes its research value through collaborative development and utilization within the research team, showing strong specificity. Wu Dan et al. [70] investigated scientific data sharing behavior among Chinese medical practitioners, with results showing that most researchers are only willing to share their research data within their teams. The characteristics of the scientific community's data sharing and utilization are oriented toward collaborative innovation goals, integrating advantageous resources of community member organizations and achieving shared utilization, showing strong planning [71].

In terms of timeliness, scientific data resources must come from reliable sources that are authentic and accurate and can be continuously supplemented and updated. To ensure supply promptness, data open platforms can use various dissemination channels and approaches such as retrieval services, search engines, and social networks to ensure that data users can conveniently and quickly locate and obtain openly shared data resources, allowing data resources to circulate rapidly and effectively.

2.3.2 Data Sharing Service Effects Scientific data can only achieve data owners' pursuit of maximum utility through extensive sharing and application. User satisfaction is a core factor in evaluating data sharing service effects, as users have more intuitive feelings about the quality, usability, and value of open data. Huang Xin [67] explored the mutual influence relationships among quality perception, value perception, and user satisfaction in the field of scientific data services from nine aspects: scientific data introduction services, navigation services, acquisition services, metadata creation services, storage services, citation services, publishing services, consulting services, and training services. The research results show that quality perception has a very significant impact on user satisfaction, and value perception also has a very significant impact on user satisfaction. Data resource managers should pay close attention to users' development needs for data sharing, leverage their own advantages to provide data products and services that meet the needs of target groups, and strive to improve service usefulness and ease of use, enabling users to "find what they search for" and "gain what they share."

Driven by user needs and utilizing various technical means, optimizing the interaction design between data management systems and users, improving data sharing service functions, increasing service stickiness and user loyalty, and allowing users to experience the practicality and convenience of scientific data sharing service functions can enhance user perceived satisfaction and form an affinity user group.

The contribution of scientific data sharing services to science, technology, economy, and society is an important standard for evaluating their service effects. The most obvious motivation for sharing subjects to participate in sharing activities is the benefits gained from participation, which can come from three aspects: contributions to supporting various national science and technology plans and government decision-making and producing major achievements; support for enterprise scientific and technological progress and technological innovation, education, and science popularization; and reduction of duplicate investment.

In summary, objective and accurate evaluation of scientific data sharing service effects and user satisfaction is conducive to enhancing the social recognition of data resources as academic achievements, helping to improve the academic status and influence of data workers, and ultimately stimulating the internal motivation of data workers to publish data. User satisfaction with data sharing plays an important role in forming a voluntary data sharing atmosphere where "everyone is both a data user and contributor," promoting social collaboration, improving research investment efficiency, and thereby enhancing scientific and technological levels, economic development, and social benefits.

3 Conclusion and Future Research Prospects

3.1 Conclusion

Based on a review and summary of existing relevant literature, this paper draws the following conclusions:

(1) To deeply understand and effectively reveal research themes on scientific data sharing, we constructed an analytical framework for scientific data sharing management. This framework describes the causal relationships among development needs, scientific data sharing management models and mechanisms, and service sharing degree and effects, emphasizing the rationality, importance, and criticality of scientific data sharing management and services.

(2) Scientific data sharing is not only an important way to obtain data for data-intensive modern scientific research but also an inevitable choice for promoting scientific and technological progress and innovation, and more importantly, a strategic demand for national development. Sharing development needs originate from three levels: national strategic development, the scientific community, and research teams. Demands at different levels can stimulate and promote the development of scientific data sharing activities to a certain extent, ultimately reflecting the common goal of pursuing the realization of scientific data value in scientific data open sharing activities. From the perspective of national development strategic needs, it forms the execution basis for data resource managers to formulate overall plans for scientific data sharing management, an important prerequisite for establishing and improving management mechanisms centered on scientific data resource sharing, and the fundamental guarantee for creating a new order for scientific data sharing.

(3) Scientific data sharing has a processual nature throughout the scientific data management lifecycle, going through five steps: data production, organization, publication, dissemination, and utilization. It is a value-creating activity directly associated with scientific data production, organization, publication, dissemination, and utilization, involving stakeholders including data producers, funders, organizers, publishers, disseminators, managers, and users. For value-creating activities in the scientific data sharing management lifecycle, two forces promote the efficient completion of data sharing value creation: internal dynamic conditions and external supporting conditions for scientific data sharing. From the perspective of sustainable development of data management institutions, it is necessary to attach importance to the construction of management systems and models, strengthen coordination mechanisms, highlight common interests, eliminate individual barriers through joint construction and sharing, and establish a management team with strong service awareness and high-level expertise. This team is a sufficient condition for scientific data resources to achieve maximum utility value. External supporting conditions for sci-

entific data sharing emphasize the operational efficiency and assurance of the sharing system, reflecting the intensity and capacity of scientific data resource sharing allocation. The more complete the external supporting conditions such as relevant organizations, institutions, systems, and information technology infrastructure, the more efficiently sharing activities operate. This also means that promoting the construction and operation of scientific data sharing systems requires comprehensive measures that both stimulate the motivation of sharing subjects to participate in sharing and improve the institutional construction of relevant sharing organizations [74].

(4) The performance evaluation of scientific data management and sharing services reflects the characteristics of data management system output and represents a comprehensive measure of data resource service capabilities and service effects.

3.2 Future Research Prospects

Scientific data sharing research has achieved considerable progress to some extent. However, given the complexity and diversity of scientific data sharing practices, there remains significant room for development in both theory and practice regarding scientific data sharing management and services. Overall, existing research both domestically and internationally is still insufficient. Future research should strengthen empirical studies, focus on improving management mechanisms, and emphasize service model innovation. Specifically:

3.2.1 Strengthening Empirical Research on Scientific Data Sharing Management

First, measuring the value contribution of participants in scientific data value co-creation, establishing key indicators for value contribution evaluation, including measurements of use value and utility value. Second, conducting empirical research on the relationship between participant behavior and value co-creation performance, analyzing the input-output relationship of value co-creation participants. Third, empirically testing user data sharing behavior. From the perspective of users' perceived usefulness of sharing platform services, factors such as data content quality and quantity, and the vividness and interactivity of presentation forms all affect user data sharing behavior. How these factors influence user data sharing behavior are questions with important practical significance for scientific data sharing service practice that await future empirical research.

3.2.2 Improving Scientific Data Sharing Management Mechanisms

Scientific data sharing activities involve various stakeholders including data producers, funders, organizers, publishers, disseminators, managers, and users. How to attract different stakeholders to actively participate in value co-creation of scientific data sharing management systems, how to coordinate interests and contradictions in resource integration during the value co-creation process, and how to ensure the stable and sustainable healthy development of sharing management systems are all core issues facing scientific data management value co-creation. Therefore, how to promote the mutual influence between resource integration and interactive behavior among participants in data sharing value co-creation through mechanism innovation, and how to promote the dynamic mechanism, benefit distribution mechanism, and coordination mechanism for participants in scientific data service systems to co-create value are issues worthy of further attention.

3.2.3 Innovating Scientific Data Sharing Service Models

The value realization of scientific data is closely related to the application of modern information technology. As literature resources develop from digitization to datafication, semanticization, linking, and intelligence through progressive deconstruction and reconstruction, the focus of data resource construction will shift to fine-grained knowledge element revelation and bonding services, and the focus of scientific data sharing services will shift to data-driven fine services and intelligent evaluation. Fine services challenge existing service utilization models while improving user experience. Therefore, scientific data sharing services require new breakthroughs, new developments, and new values, demanding that we rethink the core capabilities of scientific data sharing services and enhance their social contributions. How to innovate scientific data sharing service models to achieve value development and intelligent insights from digital resources through content datafication to data intelligence is worthy of further exploration.

References

[1] Baidu Baike [EB/OL]. [2019-10-02]. http://baike.baidu.com/item/科学大数据/17648505?fr=aladdin#reference-[2]-17764204-w.
[2] Wang Jun. Experience from U.S. Scientific Data Sharing and Its Enlightenment for China's Scientific Funds [J]. China Science Fund, 2016(1): 69-75.
[3] Mulligan A, Mabe M. The effect of the Internet on researcher motivations, behavior and attitudes [J]. Journal of documentation, 2011, 67(2): 290-311.
[4] Si Li, Wang Yuwa. Current Status and Improvement Suggestions for Data Organization of China's Scientific Data Sharing Platforms—Based on Analysis of National Science and Technology Infrastructure Platforms [J]. Library Construction, 2018(10): 52-58.
[5] Qiu Chunyan, Huang Ruhua. Recent Three-Year Progress in International Scientific Data Sharing [J]. Library and Information Service, 2016, 60(2): 6-14.
[6] Fu Xiaofeng, Li Jun, Li Jianhui. Development and Sharing of International Scientific Data [J]. China Basic Science, 2007(2): 30-35.
[7] Investigation and Analysis of U.S. Scientific Data Center Management Models [EB/OL]. [2019-10-06]. https://max.book118.com/html/2018/1001/53332101310013200132.shtm.
[8] Action Outline for Promoting Big Data Development [EB/OL]. [2019-10-06]. https://baike.so.com/doc/10955176-11483156.html.

[9] Liu Yin. China's First Scientific Data Management Measures Released: Scientific Data from National Science and Technology Plan Projects Must Be Submitted [N]. Science and Technology Daily. 2018-04-09(1).
[10] Si Li, Wang Yuwa. Current Status and Improvement Suggestions for Data Organization of China's Scientific Data Sharing Platforms—Based on Analysis of National Science and Technology Infrastructure Platforms [J]. Library Construction, 2018(10): 52-58.
[11] Huang Dingcheng. Theoretical Basis and Sharing Mechanisms of Scientific Data Sharing [J]. China Basic Science, 2003(2): 22-27.
[12] Sheng Xiaoping, Wu Hong. Analysis of Motivations of Different Stakeholders in Scientific Data Open Sharing Activities [J]. Library and Information Service, 2019, 63(17): 40-50.
[13] Council N, Ebrary I. A question of balance: private rights and the public interest in scientific and technical databases [M]. National Academies, 1999.
[14] Chen Jun, Wang Chunliao. Thoughts on Scientific Data Sharing Mechanisms [J]. China Basic Science, 2003(1): 40-43.
[15] Chilvers A, Feather J. The management of digital data: a metadata approach [J]. The electronic library, 1998, 16(6): 365-372.
[16] Bai Rujiang, Leng Fuhai. Research on Scientific Data Integration in the "Big Data" Era [J]. Information Theory and Practice, 2014, 37(1): 94-99, 38.
[17] Kong Lihua, Xi Yan, Zhang Xiaolin. Trends, Mechanisms and Challenges of Data Publishing [J]. China Science Fund, 2019(3): 237-245.
[18] Sayogo D S, Pardo 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.
[19] Huang Guobin, Wang Shu, Qu Yajie. Comparative Study on Scientific Data Publishing Models [J]. Journal of Academic Libraries, 2018, 36(1): 34-40.
[20] Xing Wenming, Liu Ting. Enhanced Publication-Driven Scientific Data Publishing: Motivation, Models and Paths [J]. Chinese Journal of Scientific and Technical Periodicals, 2019, 30(8): 853-860.
[21] Song Lirong, Wang Jian. Thinking on Scientific Data Publishing Models from the Perspective of Data Services [J]. Journal of Intelligence, 2019, 38(5): 154-160.
[22] Nevzorova O, Zhiltsov N, Zaikin D, et al. Bringing math to LOD: a semantic publishing platform prototype for scientific collections in mathematics [M]// Advanced information systems engineering. Berlin: Springer, 2013: 379-394.
[23] Qiu Chunyan. Research Progress on Key Issues in Scientific Data Publishing [J]. Library and Information, 2017(4): 54-60.
[24] Tu Zhifang. Overview of Basic Issues and Identification of Key Issues in Scientific Data Publishing [J]. Library, 2018(6): 86-92, 100.
[25] Tu Zhifang, Liu Ziheng. Research on Data Publishing Quality Control Practices of Foreign Data Repositories [J]. Library Construction, 2018(3): 5-13.
[26] Wu Lizong, Tu Yong, Wang Liangxu, et al. Discussion on Digital Object Unique Identifier in Scientific Data Publishing [J]. China Science and Technology Resources Review, 2010, 42(5): 22-29.

[27] Xu Lifang, Cong Ting. Data-Intensive, Semantic, Visual and Interactive Publishing: Research on Global Science and Technology Publishing Development Trends [J]. Publishing Science, 2012, 20(4): 73-80.
[28] Xu Lei, Pan Ying. Research on Semantic Data of Scientific Publications and Its Application [J]. Chinese Journal of Scientific and Technical Periodicals, 2018, 29(7): 704-710.
[29] Lawrence B, Jones C, Matthews B, et al. Citation and peer review of data: moving towards formal data publication [J]. International journal of digital curation, 2011, 6(2): 4-37.
[30] Smith V S. Data publication: towards a database of everything [J]. BMC research notes, 2009, 2(1): 113.
[31] Wang Xue, Ma Shengli, She Zengli, et al. Research on Scientific Data Citation Behavior and Its Impact [J]. Journal of the China Society for Scientific and Technical Information, 2016, 35(11): 1132-1139.
[32] Wang Dandan. Analysis of Key Issues in Scientific Data Standard Citation [J]. Library and Information Service, 2015, 59(8): 42-47.
[33] Shi Yali. Analysis of Key Issues in Implementing Scientific Data Citation Standards. Modern Information, 2019, 39(4): 34-41.
[34] Starr J, Castro E, Crosas M, et al. Achieving human and machine accessibility of cited data in scholarly publications [J]. PeerJ computer science, 2015, 1(1): e1.
[35] Qin Shun, Wang Quanli, Xing Wenming. Investigation and Enlightenment of European and American Scientific Data Open Access Publishing Platform Services. Library and Information Service, 2019, 63(17): 129-136.
[36] FORCE M M, Robinson N J. Encouraging data citation and discovery with the data citation index [J]. Journal of computer-aided molecular design, 2014, 28(10): 1043-1048.
[37] Park J H, Choi S H, Kim B K. A study on utilization of Korea science citation database (KSCD) based on data mining techniques [J]. Journal of information management, 2012, 43(4): 191-200.
[38] DataCite. Citation formatter [EB/OL]. [2019-10-01]. https://datacite.org/citation.html.
[39] Ding Nan, Yang Liu, Ding Ying, et al. Research on Data Citation Behavior in Chinese Sociology Journals [J]. Library and Information, 2014(6): 88-93.
[40] Ding Nan, Ding Ying, Yang Liu, et al. Analysis of Data Citation Behavior in China's Library and Information Science Field [J]. Journal of Library Science in China, 2014, 40(6): 107-116.
[41] Qiu Junping, He Wenjing. Research on the Interaction Between Scientific Data Sharing and Citation Behavior [J]. Information Studies: Theory & Application, 2015, 38(10): 1-5.
[42] Shi Yali, Si Li. Analysis of Characteristics of Chinese Researchers' Data Citation Behavior. Information Studies: Theory & Application, 2019, 42(6): 36-41.
[43] Wang Weihua, Hu Lianglin, Shen Zhihong. Development of Scientific Data Citation Standards [J]. China Science and Technology Resources Review, 2015, 47(5): 49-54.
[44] Wang Dandan. Analysis of Key Issues in Scientific Data Standard Citation

[J]. Library and Information Service, 2015, 59(8): 42-47, 53.
[45] Zhang Jingbei, Ren Shuhuai. Research on Research Data Publishing Models, Processes and Citation Strategies [J]. Publishing Research, 2014(4): 57-61.
[46] Huang Ruhua, Li Nan. Investigation and Analysis of Foreign Scientific Data Citation Standards [J]. Library and Information Service, 2015, 59(9): 21-27.
[47] Peng Jie, He Defang, Zhang Yingjie. Implementation Path and Strategy Investigation Analysis of Scientific Data Citation in Digital Publishing Environment [J]. Publishing Distribution Research, 2014(4): 57-61.
[48] Sheng Xiaoping, Wu Tong. Review of Research on Scientific Data Open Sharing at Home and Abroad [J]. Library and Information Service, 2019, 63(17): 6-11.
[49] Yati Y, Jain R C. Integration of data mining and operation research [J]. Ultrascientist of physical sciences, 2009, 21(1): 311-313.
[50] Berko Q. Use of ontologies in the processes of semantic data integration: method and framework [J]. Bulletin of the National University Leiviska PolyTechnika, 2009, 6(5): 18-27.
[51] Wang C, Lu J, Zhang G. Web ontology data matching for integration [J]. International journal of Web information systems, 2009, 5(2): 220-238.
[52] Xu Feng, Li Jiming. Conception of Scientific Data Sharing Technology Platform [J]. China Basic Science, 2003(1): 50-55.
[53] Australian code for the responsible conduct of research [EB/OL]. [2019-11-20]. https://research-integrity.uq.edu.au/files/3792/r39_{{{australian}}*{{code}}*}{{{responsible}}*{{conduct}*
[54] Dissemination and sharing of research results [EB/OL]. [2019-11-20]. http://www.nsf.gov/bfa/dias/policy/dmp.jsp.
[55] ESRC research data policy [EB/OL]. [2019-11-20]. http://www.esrc.ac.uk/about-esrc/information/data-policy.aspx.
[56] Lu Peng, Miao Liangtian, Mo Jihong, et al. Policies, Regulations and Legal Norms in the Field of Scientific Data Sharing [J]. Recent Developments in World Seismology, 2008(4): 35-41.
[57] Zhang Xiaoqing, Sheng Xiaoping. Review of Foreign Scientific Data Open Sharing Policies [J]. Library Tribune, 2018(8): 147-154.
[58] Zhang Xiaolin. Disruptive Trends in Digital Libraries [J]. Journal of Library Science in China, 2011(9): 4-12.
[59] Ministry of Science and Technology. Performance Evaluation Indicators for National Science and Technology Infrastructure Platform Operation Services [EB/OL]. [2019-10-13]. http://www.most.gov.cn/mostinfo/xinxifenlei/fgzcfgfxwj/gfxwj2011/201108/W02.
[60] Basic Science Data Network. Sharing Service Evaluation Indicator System [EB/OL]. [2019-10-13]. http://www.doc88.com/p-7784333880914.html.
[61] 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(09): 30-35.
[62] Dong Cheng, Zhao Wei, Tu Yong. Research on Sharing Performance Evaluation of Scientific Data Institutions in China [J]. China Science and Technology Forum, 2007(08): 75-79.

[63] Rust R T, Inman J J, Jia J, et al. What you don't know about customer perceived quality: the role of customer expectation distributions [J]. Marketing science, 1999, 18(1): 77-92.
[64] Oliver R L. Varieties of value in the consumption satisfaction response [J]. Advances in consumer research volume, 1996, 23(1): 143-147.
[65] Wei Caoyuan, Wang Jian, Zhang Guilan, et al. Research on Conceptual Model of Scientific Data Users' Perceived Value Based on Grounded Theory [J]. Journal of Intelligence, 2018, 37(5): 182-188.
[66] Yan Weiwei, Zhang Min. Analysis of Data Sharing and Utilization Behavior Patterns in Scientific Research Collaboration [J]. Information Studies: Theory & Application, 2018, 41(1): 55-60.
[67] Huang Xin. Research on User Satisfaction of Scientific Data Services Based on Service Content [D]. Wuhan: Wuhan University, 2017.
[68] Zhu Yunqiang, Zhu Qi, Feng Zhuo, et al. Research on Open Sharing Mechanisms of Scientific Big Data and Its Enlightenment for Environmental Information Sharing [J]. Chinese Journal of Environmental Management, 2015(6): 38-45.
[69] Wang Jun. Experience from U.S. Scientific Data Sharing and Its Enlightenment for China's Scientific Funds: Taking NSF and NIH as Examples [J]. China Science Fund, 2016(1): 69-75.
[70] Wu Dan, Chen Jing. Investigation and Research on Scientific Data Sharing Behavior of Chinese Medical Practitioners [J]. Library and Information Service, 2015, 59(18): 30-39.
[71] Yan Weiwei, Zhang Min. Analysis of Data Sharing and Utilization Behavior Patterns in Scientific Research Collaboration [J]. Information Studies: Theory & Application, 2018, 41(1): 55-60.
[72] Lee Y W, Strong D M. Knowing-why about date processes and data quality [J]. Journal of management information system, 2003, 20(3): 13-39.
[73] Wang Zhiqiang, Yang Qinghai. Research on Scientific Data Quality and Its Standardization [J]. Standard Science, 2019(3): 25-30.
[74] Guo Bao, Zhuo Xiangzhi. Research on Regional Shared Resource Allocation from the Perspective of Sharing Chain [J]. Journal of Tangshan Normal University, 2019, 41(4): 94-100.

Author Contributions

Bi Datian: Proposed research ideas, wrote and finalized the paper;
Cao Ran: Conducted literature review, wrote and revised the paper;
Du Xiaomin: Collected and organized literature.

A Summary of Researches on Research Data Sharing and Development Trends at Home and Abroad

Bi Datian[1], Cao Ran[1], Du Xiaomin[2]
[1]School of Management, Jilin University, Changchun 130022

[2]Yingkou Institute of Technology, Yingkou 115014

Abstract: [Purpose/significance] Scientific data sharing has become a strategic demand for national development. Analyzing the research status and development trends of scientific data sharing provides reference for promoting relevant theoretical research and practice in China. [Method/process] Through combing and inducing domestic and foreign literature, three major themes related to scientific data sharing research were extracted, namely "scientific data sharing development needs," "scientific data management sharing modes and mechanisms," and "scientific data service sharing degree and service effect." We also built the data sharing analysis framework, which reflects the causal relationship among the three themes, and emphasizes the rationality, importance, and criticality of scientific data sharing management and service. [Result/conclusion] The existing research at home and abroad is still insufficient. In the future, while strengthening empirical research and focusing on the improvement of management mechanisms, we should also pay attention to the innovation of shared service mode.

Keywords: scientific data sharing; data management; data service

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