A Multi-dimensional Analysis of the Research Trajectory of Foreign Scientific Data Management: Postprint

Authors: Li Yu, Liu Hong, Sun Jianjun

Date: 2023-08-27T00:00:00+00:00

Abstract

[Objective/Significance] Under the big data environment, research on scientific data management holds significant theoretical and practical importance. Visualizing and analyzing the evolution of scientific data management research can provide references and insights for domestic research on this topic.

[Method/Process] Based on WoS and Google Scholar, and utilizing analytical tools such as Citespace, Ucinet, and TDA, this study explores the evolutionary trajectory of scientific data management research in the international academic community, comprehensively revealing its history, current status, and development direction.

[Results/Conclusion] The research status of scientific data management is reflected from four dimensions: temporal distribution mapping, spatial distribution mapping, thematic evolution mapping, and knowledge base mapping, with relevant recommendations proposed in conjunction with domestic research and development conditions.

Full Text

Preamble

Li Yu[1,2], Liu Hong[1], Sun Jianjun[1]
[1]School of Information Management, Nanjing University, Nanjing 210023
[2]Northwest University Library, Xi'an 710127

Abstract

[**Purpose/Significance**] In the big data environment, research on scientific data management holds important theoretical and practical significance. Visualizing and analyzing the research trajectory of scientific data management can provide references for domestic research on this topic. [**Method/Process**] Based on Web of Science (WoS) and Google Scholar, this study utilizes analytical tools such as CiteSpace, Ucinet, and TDA to explore the evolutionary process of scientific data management research in the international academic community, comprehensively revealing the history, current status, and development direction of this field. [**Result/Conclusion**] The study reflects the current state of scientific data management research from four dimensions: temporal distribution mapping, spatial distribution mapping, thematic evolution mapping, and knowledge base mapping, and proposes relevant recommendations in combination with domestic research and development conditions.

Keywords: scientific data, data management, knowledge mapping, visualization analysis
Classification Number: G250
DOI: 10.13266/j.issn.0252-3116.2018.13.014

With the continuous and in-depth development of information technology, scientific research has fully entered the "fourth paradigm," and scientific data across various disciplines are becoming increasingly abundant and accumulating rapidly, collectively constituting the big data environment for research activities. As a strategic resource with significant scientific, social, and economic value, the management and utilization of scientific data have been highly prioritized by governments and research institutions at all levels both domestically and internationally. In 2002, China's Ministry of Science and Technology launched the "National Scientific Data Sharing Project" [1], which has, to date, carried out scientific data management and sharing work in multiple disciplinary fields including resource environment, agriculture, population and health, laying a solid foundation for the in-depth development of scientific data management in China. In 2015, the State Council issued the "Action Outline for Promoting Big Data Development" [2], which proposed strategic deployment for big data development in China from a top-level design perspective, serving as a programmatic document guiding the management and utilization of scientific data in the country.

Scientific data are generally defined as fundamental raw data generated by humans through experiments, observations, surveys, and other scientific and technological activities in the process of understanding and transforming the world, as well as datasets and related information processed according to specific needs [3]. In the context of big data, the scope of scientific data has undergone significant changes. Broadly defined, scientific data include not only data from natural science fields but also various types of data generated in humanities and social science research, even encompassing individual researchers' data and various data from the Internet. Scientific data are characterized by broad disciplinary distribution, diverse data structures, and difficulties in data sharing, making research and practice in scientific data management significant for society, research institutions, and individual researchers alike.

In recent years, the domestic and international academic community has conducted research on scientific data management from different perspectives, focusing primarily on two aspects: (1) analyzing and learning from the successful experiences of countries that started scientific data management earlier, such as the United States, United Kingdom, Australia, and Canada. For instance, Liu Guifeng et al. selected three universities—Stanford University, Columbia University, and the University of Washington—as samples to analyze and discuss the practice of library participation in scientific data management in U.S. universities [4]; Si Li and Xin Juanjuan used content analysis to compare and analyze scientific data management and sharing policies issued by 20 universities in the UK and US, and proposed recommendations for formulating relevant policies in China [5]. (2) investigating and analyzing practical aspects such as infrastructure, service needs, funding sources, and business practices of scientific data management. Tang Yanhua used case analysis to explore the main components of scientific data management services in universities and proposed effective recommendations based on these practical elements and the current status of domestic universities [6]; Wei Junchao and Zhang Chunfang selected more than 10 scientific data management platforms at home and abroad, conducted horizontal comparisons across multiple dimensions including construction status, objectives, data sources, and funding, identified gaps between different platforms, and discovered valuable experiences for China's scientific data management practice [7].

A review of existing research reveals that current academic studies on scientific data management are primarily qualitative, with most focusing on surveys and analyses of practices in single or multiple foreign institutions, unable to demonstrate the global research status of scientific data management. Therefore, this study adopts scientometric methods, utilizing multiple data analysis and visualization tools such as CiteSpace, Ucinet, and TDA to more systematically reflect the current research status of scientific data management worldwide from a data perspective, aiming to provide references for research and practice in China.

2. Design of the Visualization Methodology System for Scientific Data Management Research

2.1 Framework of the Visualization Methodology System

To comprehensively analyze the research status in the field of scientific data management, this study employs the methodological framework shown in Figure 1 [Figure 1: see original paper] to intuitively display the main characteristics, hotspot distributions, and changing trends of research in this field from multiple angles and dimensions.

First, based on the research topic, a search query for "scientific data management" was constructed. To comprehensively present the research status of scientific data management, the SCI, SSCI, and A&HCI databases in the Web of Science core collection were selected as data sources. Search terms including "scientific data management," "scientific data curation," and "research data curation" were used as topic keywords for retrieval. The search time was limited to November 20, 2017, and after excluding false positives, a total of 336 research papers on scientific data management were obtained.

Second, data analysis tools such as Thomson Data Analyzer were used to standardize the 336 research papers, including removing invalid keywords, merging similar keywords, constructing keyword co-occurrence matrices, standardizing institution names, and merging different representations of the same institution.

Furthermore, visual display and analysis of knowledge output in the scientific data management field were conducted from four dimensions: temporal distribution mapping, spatial distribution mapping, thematic evolution mapping, and knowledge base mapping.

2.2 Implementation Mechanism of the Visualization Method

Based on the standardized 336 research papers and their citation data, visualization display and analysis were achieved from the following four aspects:

First, temporal distribution mapping of scientific data management research was created to show the research footprint from a time series dimension and analyze the development trend of knowledge output.

Second, spatial distribution mapping of scientific data management was generated to display the spatial distribution of research achievements at both national and institutional levels, identifying the competitive advantages of different countries and research institutions.

Third, thematic evolution mapping of scientific data management was developed. Ucinet was used to conduct co-occurrence analysis of high-frequency keywords to display the knowledge network of scientific data management research; CiteSpace was employed to draw keyword timezone views to grasp the changing trends of research themes over time; content analysis was applied to conduct in-depth analysis of research hotspots and summarize the thematic distribution of scientific data management research.

Finally, knowledge base mapping of scientific data management was constructed. CiteSpace was used to cluster the co-citation patterns of references in this thematic research, revealing the theoretical foundation of scientific data management through knowledge base mapping; Google Scholar was further utilized to conduct in-depth analysis of highly cited references in this thematic research to understand the core knowledge base.

3. Knowledge Mapping Analysis of Scientific Data Management

3.1 Temporal Distribution Mapping

Temporal distribution mapping can reveal the overall landscape of scientific data management research from a time dimension and identify the development history and trends of this thematic research. Therefore, statistical analysis was conducted on the 336 scientific data management research papers, and the temporal distribution mapping is shown in Figure 2 [Figure 2: see original paper] (as the retrieval time was limited to November 20, 2017, the number of papers in 2017 cannot represent the full year).

As shown in Figure 2, academic attention to this theme has continuously increased over time, with knowledge output showing an exponential growth trend. The research footprint of scientific data management can be divided into three stages according to the time sequence:

(1) **Initial Exploration Period** (number of papers < 5): 1970-2002. During this period, knowledge output on scientific data management was relatively small, with fewer than 5 papers published annually. The earliest research on this theme can be traced back to 1970, and for more than 30 years until 2002, exploratory research on scientific data management was conducted in multiple disciplinary fields including aerospace engineering [8], environmental science [9], chemistry [10], biomedicine [11], and computer science [12].

(2) **Steady Development Period** ($5 <$ number of papers < 15): 2003-2011. In October 2003, the Max Planck Society in Germany held the Berlin Conference and released the "Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities," which was signed by multiple countries including China to support open access to research results and raw scientific data. In January 2004, 34 member countries of the Organization for Economic Co-operation and Development (OECD) signed the "Declaration on Access to Research Data from Public Funding" issued by the Committee for Scientific and Technological Policy, clarifying the scope and definition of open scientific data. With the release of a series of scientific data management policies, research in this field entered a steady development period, with knowledge output capacity significantly improved compared to the initial exploration period.

(3) **Rapid Development Period** (number of papers > 20): 2012-2017. In 2010, the U.S. National Science Foundation (NSF) began requiring applicants to submit "Data Management Plans" for NSF-funded projects. The Australian Research Council (ARC) and the National Health and Medical Research Council (NHMRC) proposed management and sharing requirements for research outcomes in 2012 and 2013, respectively. The UK Research Councils (RCUK) required funded projects to share scien-

tific data without harming intellectual property rights in 2011. In 2014, the Chinese Academy of Sciences (CAS) and the National Natural Science Foundation of China (NSFC) also issued open access declarations for research project outcomes. Under the influence of various funding policies, scientific data management research entered a rapid development period, with research output showing explosive growth.

3.2 Spatial Distribution Mapping

Investigating and analyzing the countries and institutions in the field of scientific data management research can identify the competitive advantages of different nations and research institutions from a spatial perspective, understand their cooperation and communication in this field, and provide references for Chinese academia. As shown in Figure 3 [Figure 3: see original paper]:

From the national perspective, the United States leads overwhelmingly in knowledge output on scientific data management, accounting for approximately 52.1% of total papers, which is closely related to the long-term emphasis by the U.S. government, funding agencies, research institutions, and researchers on scientific data policy formulation and resource management. Apart from the United States, the United Kingdom, Germany, and Australia have also published numerous research outcomes in this field. According to investigations, these three countries have significant advantages in the scale of scientific data resource construction, registration or membership in important international scientific data organizations such as the World Data System (WDS) and DataCite, and the construction of global scientific data repository systems (e.g., Re3data.org) [13]. Figure 3 also shows that although China has achieved certain results in scientific data management, there remains a considerable gap compared to early-developing countries such as the United States, United Kingdom, and Germany.

To further understand the development differences between Chinese institutions and those in data-strong countries, core research groups in scientific data management were analyzed from two perspectives: knowledge output capacity and academic influence. Degree centrality measures the extent to which an institution connects with all other institutions in the cooperation network, calculated as follows:

$$C_D(N_i) = \sum_{j=1}^{g} x_{ij} \quad (i \neq j)$$

where $C_D(N_i)$ represents the degree centrality of institution i, g represents the total number of research institutions, and $\sum_{j=1}^{g} x_{ij}$ represents the number of articles in which institution i directly collaborates with other $g-1$ institutions. A higher number of papers indicates stronger knowledge output capacity in scientific data management, while higher degree centrality indicates greater academic influence. Institutions with more than 5 publications were selected for degree centrality calculation, with results shown in Table 1 .

As shown in Table 1, the National Aeronautics and Space Administration (NASA) is the institution with the most publications in scientific data management. NASA established Project Data Management Plans (PDMPs) for each scientific research project as early as March 1993 [14], managing and archiving scientific data at all levels to provide data products and information services for enterprises, academia, and the public. Johns Hopkins University is the university with the most extensive collaboration with other research institutions in this field. Its Johns Hopkins University Data Management Services (JHUDMS) [16] provides personalized services including scientific data management plan consultation, data archiving, and training for researchers and institutions, and has collaborated with multiple institutions including the University of Illinois at Urbana-Champaign, the University of Texas Health Science Center at Houston, Arizona State University, and the U.S. BioFortis company.

Based on Figure 3 and Table 1, the United States and United Kingdom are not only major data producers in scientific data management but also data-strong countries. Driven by various internal and external factors including laws and regulations, funding conditions, and institutional data management policies, they have formed core research groups represented by renowned institutions such as NASA, the University of Sheffield, Oxford University, Johns Hopkins University, and Harvard University. In contrast, only the Chinese Academy of Sciences and Wuhan University in China have published more than 5 papers in international core journals in this field, indicating a significant gap in both knowledge output capacity and academic influence compared to data-strong countries.

3.3 Thematic Evolution Mapping

3.3.1 Keyword Co-occurrence Knowledge Mapping Conducting co-occurrence analysis of high-frequency keywords can clarify the development status of scientific data management research. Using TDA software to clean keywords from the 336 research papers and extract core keywords to construct a co-occurrence matrix, Ucinet was used for visualization analysis, with results shown in Figure 4 [Figure 4: see original paper]. In the keyword co-occurrence knowledge map, node size is represented by degree centrality—larger nodes indicate greater influence of the keyword in the scientific data management research network and represent research hotspots.

As shown in Figure 4, four keywords—data management, research data management, data sharing, and data curation—occupy central positions in the knowledge network and represent the core hotspot issues in scientific data management research. Data management, as the keyword with the highest frequency and centrality, has a very broad research scope covering the entire lifecycle of scientific data activities including data sharing, processing, storage, acquisition, and analysis, and extends to all disciplines including systems biology, information management, remote sensing, social sciences, neuroinformatics, electrophysiology, and interdisciplinary fields. Research data management, as a subcategory of scientific data management, focuses not only on universally applicable data management techniques and methods such as data preservation, curation, and sharing but also approaches research from the perspective of research support including academic libraries, universities, library roles, personal information management, and laboratory information management systems. Scientific data sharing is an important means to promote academic exchange and improve data utilization efficiency. Data sharing is also a keyword with high frequency and centrality, closely associated with open access, data discovery, open data, data literacy, data citation, and data preservation. Data curation is an emerging research hotspot for the long-term preservation, management, and utilization of scientific data in the e-science environment, frequently co-occurring with keywords such as data preservation, digital storage, data integration, archiving, data repositories, and e-science, providing data security for scientific data management research and applications.

3.3.2 Research Frontier Timezone Knowledge Mapping The timezone map of high-frequency keywords can display the evolution and changing trends of research hotspots in scientific data management from a temporal dimension. Using CiteSpace to draw the core keyword timezone map, the results are shown in Figure 5 [Figure 5: see original paper].

As shown in Figure 5, researchers showed different concerns during the three development stages of scientific data management research:

During the **initial exploration period** (1970-2002), scientific data management research had limited knowledge output, focusing on the development and utilization of scientific data management systems based on information technology, such as research on databases, systems, and metadata. Although this stage was in its early development, the research scope expanded fully into multiple natural science fields including environmental science, aerospace engineering, chemistry, and biomedicine.

Entering the **steady development period** (2003-2011), the number of scientific data management research papers increased significantly, and research content further deepened. During this stage, research broke through the limitations of data management systems and began to develop toward the entire lifecycle of scientific data management including data sharing, data curation, and data collection. The research scope also expanded beyond natural sciences, with scientific data management in humanities, social sciences, and interdisciplinary fields beginning to attract researchers' attention.

During the **rapid development period** (2012-2017), scientific data management research not only developed in depth along the scientific research dimension based on the previous two stages but also expanded horizontally toward research support dimensions. For example, while data sharing and data repositories remained key research contents, themes related to scientific data management services such as academic libraries, institutional repositories, and research support also became focal points.

3.3.3 Distribution of Research Hotspot Themes The core keyword co-occurrence knowledge map and timezone knowledge map demonstrate the distribution and evolution of hotspots in the scientific data management field to some extent. To gain deeper and more comprehensive understanding of thematic distribution, through quantitative and qualitative comprehensive analysis of subject terms and disciplinary distribution information, the research in this field was divided into four aspects. Content analysis is used below to discuss the four major research themes in scientific data management:

(1) **Research on Scientific Data Management Based on Management Entities**. Content analysis of titles, keyword frequencies, and abstracts of the 336 research papers shows that analyzing scientific data management activities of various management entities is one of the most concentrated research contents. This includes four levels: international organizations, nations, domestic alliances, and academic institutions. J. Wei et al. investigated how the world's largest social science data center, ICPSR, stores, curates, and shares social science data in the big data environment [17]; J. Kim used questionnaires to survey 190 Korean professors' views on scientific data management and sharing to provide references for policy formulation and practice in Korea [18]; J. Dierkes et al. introduced how the Göttingen E-Research Alliance provides innovative data management services based on the scientific data lifecycle to its member institutions [19]; M. Witt introduced how Purdue University collects and manages scientific data to better support research activities across multiple disciplines [20].

(2) **Research on Scientific Data Management Based on Disciplines**. According to the definition of scientific data, early scientific data were mainly generated in natural science fields, and research on scientific data management also focused on biology, medicine, meteorology, and earth sciences. With the development of big data technology, the scope of scientific data has changed significantly, and research has expanded to humanities, social sciences, and interdisciplinary fields. S. Adcock et al. analyzed problems and challenges in geochemical data management and introduced methods and experiences adopted by the Geological Survey of Canada [21]; P. Fankhauser et al. studied scientific data management issues in humanities research using Germanic linguistics as an example [22]; R. Downs analyzed deficiencies in scientific data management across different disciplines and proposed new models for managing interdisciplinary scientific data [23].

(3) **Research on Scientific Data Management Based on Technology**. In the context of data-intensive scientific research, scientific data exhibit characteristics of explosive growth, heterogeneous and diverse structures, and low value density. Using various information technologies to effectively manage scientific data and achieve preservation, sharing, and utilization has become an important research content. Technology-based scientific data management research involves the entire lifecycle including data collection, transmission, storage, processing, analysis, retrieval, and visualization, covering multiple technical dimensions such as hardware infrastructure, software tool development, and data management platforms.

(4) **Research on Scientific Data Management from the Library Perspective**. Higher education institutions are major research entities that generate vast amounts of scientific data resources across various disciplines. Driven by internal and external factors such as scientific data management policies of funding agencies and researchers' data storage needs, it is imperative for university libraries to develop scientific data management services. In the big data environment, scientific data management services bring new opportunities and challenges to libraries, and the international academic community has conducted extensive discussions on the role of libraries in scientific data management activities, including policy formulation, researchers' willingness for data management and sharing, lifecycle-based management services, and data management education, providing references for Chinese libraries.

3.4 Knowledge Base Analysis

Since its emergence in 1970, scientific data management research has continuously diffused into various disciplines and currently attracts widespread attention in natural sciences, humanities, social sciences, and interdisciplinary fields. The knowledge base of scientific data management consists of the citation collection of all literature on this theme. Through co-citation analysis, the foundational knowledge of scientific data management can be revealed from a citation trajectory perspective. CiteSpace was used to draw the co-citation map of scientific data management research literature, with results shown in Figure 6 [Figure 6: see original paper].

In the knowledge base map, each node represents a citation, and node labels are marked according to citation frequency—larger fonts indicate higher citation frequencies. To deeply understand the key literature influencing scientific data management research, Google Scholar was used for further retrieval and analysis of key nodes. Information on the top 10 co-cited references is shown in Table 2 .

As shown in Figure 6 and Table 2, C. Tenopir's 2011 paper "Data sharing by scientists: practices and perceptions" has the highest citation frequency in the scientific data management field. This study surveyed 1,329 researchers on their scientific data sharing practices, analyzed in detail obstacles such as insufficient time, lack of funding, and inadequate institutional support, discussed impacts of differences in funding agencies, disciplines, researcher age, and geographic location on scientific data management, and proposed the necessity of guidance from national policies and funding agencies. This paper has been cited 29 times within the field and 565 times globally, attracting widespread attention in other thematic areas as well. Additionally, C. Tenopir's 2012 paper "Academic libraries and research data services: Current practices and plans for the future" investigates the current status and future plans of libraries providing scientific data services in data-intensive research environments from the perspective of scientific data management services, and has also been widely cited. Furthermore, T. Hey's 2009 book "The fourth paradigm: data-intensive scientific discovery" explains the challenges brought by data explosion to data acquisition, curation, analysis, and sharing, and discusses the role of cloud computing, collaborative services, and repositories in the fourth paradigm environment. This book has been cited 2,105 times globally and represents an important reference in the scientific data management field.

In summary, the knowledge base of scientific data management research comprises three aspects: (1) In the big data environment, the explosive growth of scientific data has propelled scientific research into the fourth paradigm, with the data-intensive research environment creating external conditions for scientific data management research; (2) Based on objectives such as avoiding data loss, verifying research credibility, improving research impact, and seeking collaboration opportunities, researchers' demands for scientific data management and sharing continue to increase, providing internal drivers; (3) The active participation of university libraries, academic libraries, and other information institutions in scientific data management services continuously promotes theoretical and practical development.

4. Conclusions and Recommendations

In data-intensive research environments, scientific data exhibit big data characteristics including explosive growth, heterogeneous and diverse structures, and low value density. Scientific data management research and practice are significant for data themselves, researchers, research institutions, and society as a whole. Using WoS and Google Scholar as data sources and employing multiple analytical tools including CiteSpace, Ucinet, and TDA, this study comprehensively analyzed the current status of international scientific data management research from four dimensions: temporal distribution mapping, spatial distribution mapping, thematic evolution mapping, and knowledge base mapping. The conclusions are as follows:

(1) **From the temporal dimension**, scientific data management research is currently in a rapid development period. Under the data management policy requirements of funding agencies such as the U.S. National Science Foundation (NSF), Australian Research Council (ARC), UK Research

Councils (RCUK), and the Wellcome Trust, research output has grown rapidly. Currently, few domestic funding agencies require applicants to submit "Data Management Plans" from policy and institutional perspectives, indicating broad development space in scientific data management.

(2) **From the spatial dimension**, countries that started scientific data research earlier, including the United States, United Kingdom, Germany, and Australia, have strong competitive advantages, represented by institutions such as NASA, the University of Sheffield, Oxford University, and Johns Hopkins University. These data-strong countries attach great importance to scientific data management research and practice at both national and institutional levels.

(3) **From the thematic evolution dimension**, scientific data management and sharing are core research contents throughout this field. As research deepens, content has evolved from data management to the entire scientific data lifecycle, and during the rapid development period, has attracted close attention from the library and information science field.

(4) **From the knowledge base dimension**, the development of scientific data management research is built upon three foundations: the data-intensive research environment, researchers' data sharing and management needs, and the active participation of library and information institutions in scientific data management services. Currently, scientific research has entered the fourth paradigm, yet in practice, only a few Chinese institutions such as Fudan University, Wuhan University, Peking University, and the Chinese Academy of Sciences have established data service platforms and developed scientific data management services.

China's data management research started relatively late. Although some strong research teams have emerged at the Chinese Academy of Sciences and Wuhan University, there remains a significant gap compared to data-strong countries. Current domestic research primarily focuses on analyzing and learning from advanced experiences of foreign institutions, with relatively single content that needs expansion from multiple entity levels, disciplinary scopes, and technical perspectives. Specifically, four aspects should be emphasized: (1) Formulate and improve data management policies at national, funding agency, research institution, and publication levels to provide strong policy support; (2) Establish cooperative relationships with experienced foreign research institutions and academic groups to enhance China's knowledge output capacity and international academic influence; (3) Strengthen management and utilization of scientific data in humanities and social sciences, emphasizing its significance in all disciplines while exploring technologies and methods; (4) Recognize the important role of library and information science in scientific data management, with academic libraries, university libraries, and information management schools enriching research content, innovating services, and cultivating talent to promote internationalization.

This study has several limitations: (1) It used SCI, SSCI, and A&HCI in Web of Science as data sources to analyze international core journals but did not comprehensively compare with domestic core journals; (2) While academic papers represent research outcomes to some extent, multi-source data analysis combining with scientific data management practice was not conducted. These will be directions for future research.

References

[1] Zhang Xian'en. National Scientific Data Sharing Project [J]. Scientific Chinese, 2004(9): 11-13.
[2] State Council. Notice of the State Council on Issuing the Action Outline for Promoting Big Data Development [EB/OL]. [2017-12-03]. http://www.gov.cn/zhengce/content/2015-09/05/content_{10137}.htm.
[3] Fu Xiaofeng, Li Jun, Li Jianhui. Development and Sharing of International Scientific Data [J]. China Basic Science, 2007(2): 30-35.
[4] Liu Guifeng, Lu Zhangping, Ruan Lian. Research on Research Data Management Services in American University Libraries [J]. Library Tribune, 2015, 35(8): 137-144.
[5] Si Li, Xin Juanjuan. Investigation and Analysis of Scientific Data Management and Sharing Policies in UK and US Universities [J]. Library Tribune, 2014, 34(9): 80-85, 65.
[6] Tang Yanhua. Research and Recommendations on University Research Data Management Services Practice [J]. Library and Information Service, 2016, 60(24): 130-138.
[7] Wei Junchao, Zhang Chunfang. Comparative Study of Domestic and International Scientific Data Management Platforms [J]. Library and Information Knowledge, 2017(5): 97-107.
[8] Vette J, Karlow N. Data management at National Space Science Data Center [J]. Journal of spacecraft and rockets, 1970, 7(10): 1234-1240.
[9] Campbell W, Smith P, Price R, et al. Advancements in landscape science data management pilot land data system [J]. Science of the total environment, 1986, 56(1): 31-44.
[10] Goffredo M. The role of scientific data management systems in handling laboratory information [J]. American laboratory, 1999, 31(6): 65-69.
[11] Freedland K, Carney R. Data management and accountability in behavioral and biomedical research [J]. American psychologist, 1992, 47(5): 640-645.
[12] Bruegge B, Snider D, Russell K, et al. ISDMs-a scientific data management system [J]. Software practice & experience, 1985, 15(5): 473-487.
[13] Yang Xing, Qu Baoqiang, He Yuntao, et al. Comparative Analysis and Enlightenment of Scientific Data Resource Sharing in Major Countries [J]. China Science and Technology Resources Review, 2016, (6): 18-25.
[14] NASA. NASA mission project data management plans (PDMPs) [EB/OL]. [2017-12-13]. http://nssdc.gsfc.nasa.gov/archive/pdmp/.
[15] NASA. Data & information policy [EB/OL]. [2017-12-13]. http://science.nasa.gov/earth-science/earth-science-data/data-information-policy/.

[16] JOHNS HOPKINS LIBRARIES. Data management services [EB/OL]. [2017-12-13]. http://dms.data.jhu.edu/.
[17] Wei J, He D, Yu C. Social science data repositories in data deluge: a case study of ICPSR's workflow and practices [J]. The electronic library, 2017, 35(4): 626-649.
[18] Kim J. Data sharing from the perspective of faculty in Korea [J]. Libri, 2017, 67(3): 179-192.
[19] Dierkes J, Wuttke U. The Göttingen E-research alliance: a case study of developing and establishing institutional support for research data management [J]. ISPRS international journal of geo-information, 2016, 5(8): 133.
[20] Witt M. Institutional repositories and research data curation in a distributed environment [J]. Library trends, 2008, 52(2): 191-203.
[21] Adcock S, Spirito W, Garrett R. Geochemical data management issues and solutions [J]. Geochemistry exploration environment analysis, 2013, 13(4): 337-348.
[22] Fankhauser P, Fiedler N, Witt A. Research data management in the humanities using the example of germanic linguistics [J]. Zeitschrift fur bibliothekswesen und bibliographie, 2013, 60(6): 296-306.
[23] Downs R, Chen R. Designing submission and workflow services for preserving interdisciplinary scientific data [J]. Earth science informatics, 2010, 3(1/2): 101-110.

Author Contributions:
Li Yu: Paper structure design, content writing, and revision;
Liu Hong: Data processing and foreign literature translation;
Sun Jianjun: Research direction determination and theoretical guidance.

Visualization Analysis of Foreign Scientific Data Management Based on Multi-dimensional Perspectives

Li Yu[1,2], Liu Hong[1], Sun Jianjun[1]
[1]School of Information Management, Nanjing University, Nanjing 210023
[2]Northwest University Library, Xi'an 710127

Abstract: [Purpose/significance] Under the background of big data, the research of scientific data management has important theoretical and practical significance. [Method/process] Based on WoS and Google Scholar databases, the paper utilized CiteSpace, Ucinet, TDA and other analysis tools to explore the evolution of scientific data management research in the international academic community. [Result/conclusion] The paper fully revealed the history, current situation and development direction of scientific data management research from four aspects including time distribution map, spatial distribution map, thematic evolution map and knowledge base map. Furthermore, the paper put forward relevant proposals according to the domestic research status quo.

Keywords: scientific data; data management; knowledge mapping; visualization analysis