

Knowledge mapping of research data in China: a bibliometric study using visual analysis

Knowledge
mapping of
research data

331

Chunlai Yan

Panzhuhua University, Panzhuhua, China

Hongxia Li

Chongqing Technology and Business University, Chongqing, China

Ruihui Pu

Srinakharinwirot University, Bangkok, Thailand, and

Jirawan Deeprasert and Nuttapong Jotikasthira

Rajamangala University of Technology Rattanakosin, Salaya, Thailand

Received 9 November 2020

Revised 14 May 2021

1 August 2021

21 January 2022

Accepted 5 February 2022

Abstract

Purpose – This study aims to provide a systematic and complete knowledge map for use by researchers working in the field of research data. Additionally, the aim is to help them quickly understand the authors' collaboration characteristics, institutional collaboration characteristics, trending research topics, evolutionary trends and research frontiers of scholars from the perspective of library informatics.

Design/methodology/approach – The authors adopt the bibliometric method, and with the help of bibliometric analysis software CiteSpace and VOSviewer, quantitatively analyze the retrieved literature data. The analysis results are presented in the form of tables and visualization maps in this paper.

Findings – The research results from this study show that collaboration between scholars and institutions is weak. It also identified the current hotspots in the field of research data, these being: data literacy education, research data sharing, data integration management and joint library cataloguing and data research support services, among others. The important dimensions to consider for future research are the library's participation in a trans-organizational and trans-stage integration of research data, functional improvement of a research data sharing platform, practice of data literacy education methods and models, and improvement of research data service quality.

Originality/value – Previous literature reviews on research data are qualitative studies, while few are quantitative studies. Therefore, this paper uses quantitative research methods, such as bibliometrics, data mining and knowledge map, to reveal the research progress and trend systematically and intuitively on the research data topic based on published literature, and to provide a reference for the further study of this topic in the future.

Keywords Research data, Scientific data, Library, Data literacy education, Research support services, Visualization

Paper type Research paper

© Chunlai Yan, Hongxia Li, Ruihui Pu, Jirawan Deeprasert and Nuttapong Jotikasthira. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licences/by/4.0/legalcode>

This study was substantially supported by a project (70901080) from the National Natural Science Foundation of China. The research was substantially supported by a project (19SKGH091) from the Chongqing Education Commission of China and project from the Open Fund of Research Centre of Enterprise Management. This study was also substantially supported by a project (2053002) from Chongqing Technology and Business University of China.



Library Hi Tech

Vol. 42 No. 1, 2024

pp. 331-349

Emerald Publishing Limited

0737-8831

DOI 10.1108/LHT-11-2020-0285

Introduction

Research data (also known as scientific data in China) refers to various types of experimental data, personal observation data, Internet data, statistical data, and simulation data, which are obtained by collection, observation, or analysis, and presented in the form of tables, numbers, images, new media, etc. (Wang, 2018). Research data is both the data source and tool for carrying out scientific research innovation and achieving technical foresight, and forms an important knowledge base that supports the country's decision-making. In recent years, as scientific research has entered an era of intensive data-driven research paradigms, international organizations, governmental departments, and research institutions have all increased their focus on and financial support to the scientific research field. For example, the United Nations' Educational Scientific and Cultural Organization (UNESCO) launched the "Global Alliance for Enhancing Access to and Application of Research Data in Developing Countries," and the International Council of Scientific Unions (ICSU) established an international organization to promote global research data sharing: "The Committee on Data for Science and Technology and World Data System" (Si and Xing, 2017). The library is a process monitoring and embedded management organization, and an archiving and educational institution for research data, with an irreplaceable position and role in the management, service, and sharing of research data (Sun, 2016).

However, despite its importance, there are very few literature reviews focusing on research data, and those that exist have a limited scope. Brochu and Burns (2019) reviewed some published studies focusing on the relationship between librarians and research data management (RDM). Grant (2017) conducted a literature-based study on the relationship between research data and record-keeping. Ng'eno and Mutula (2018) also studied the core RDM issues in agricultural research institutes. Similarly, Fuhr (2019) investigated a reviewed study on the RDM skills gap among Canadian information workers in the health sciences field. Chawinga and Zimm (2019) investigated and presented a comprehensive account of the factors hampering data sharing at three levels of the global research hierarchy (individual, institutional and international). Hu and Fang (2021) conducted a systematic review of the relevant literature for the evaluation of the research data. Sheng and Yuan (2021) reviews the influencing factors of the openness and sharing of research data from eight aspects: researchers, policy, data, and technology etc. Liu (2020) sorted out the research status of library research data literacy in China, and focused on analyzing the research theme of research data literacy in the library community. Yan *et al.* (2020) reveals a new paradigm for research data-driven research collaboration and presents future research directions and opportunities. Ruan and Yang (2019) summarized the theories and practices related to information security behavior and research data security management, and commented on the current status and future research direction of research data security behavior. Ma (2019) reviewed the existing research results in the past 10 years from the positioning and function, sharing policy, sharing platform, and sharing strategy, and prospected the future research. The above-cited literature review studies indicate that yet, there is no systematically organized research study that covers all the important aspects of research data.

Research related to research data is an interdisciplinary field, and researchers encounter different issues, according to their respective knowledge backgrounds. Before the advent of bibliometric tools, researchers relied on peer-reviewed articles or documents to quickly obtain a panoramic view of a subject or field of research (Chen and Chen, 2017). Some obvious limitations of this approach are that the research results are ultimately influenced by the vision and subjective judgment of the peers' own knowledge, it cannot completely reveal the critical studies in the field or the emerging research hotspots, and it is controversial. Bibliometric tools offer researchers another possibility however, where the literature review changes from a qualitative research method to a mixed research method, thereby leading to more objective and reliable research results.

This study combined bibliometric research methods with knowledge maps, and then systematically reviewed the published studies on research data. This study provides a panoramic view of research data, and through library informatics, offers a quick understanding of the researchers' collaborative characteristics, institutional collaboration characteristics, hot research topics, evolutionary trends, and research frontiers. The specific research questions are as follows: (1) What are the collaboration characteristics between authors and institutions on the topic of research data in the Chinese library and information field. (2) What are the major research subjects relating to "research data" in the Chinese library and information field? (3) What evolutionary paths have the research hotspots and cutting-edge information followed?

Literature review

Many scholars have used various evaluation methods to try and carry out quantitative analyses and so gain influence for the authors and research institutions. They have also proposed many valuable measurement indicators and evaluation methods, which can be mainly placed in two categories: "Based on Bibliometric Analysis" and "Based on Social Network Analysis" (Chao *et al.*, 2016). The evaluation method using a bibliometric analysis is based on indicators such as the number of publications, CiteSpace and VOSviewer are both important tools for information visualization and bibliometric knowledge map research in recent years, but they are also different in the theoretical algorithm of visual map generation. CiteSpace focuses on expressing the strength of the relationship by graphics and connections, while VOSviewer mainly calculates the relationship by distance. CiteSpace software has certain advantages in revealing the dynamic development law of the discipline and discovering the research frontier. VOSviewer software performs better when the relationship between subject themes is clearly presented, or when the amount of data is very large.

CiteSpace software employs a cosine algorithm to calculate the cooperation intensity of the researchers or institutions. The connection strength between nodes represents the cooperation strength between the researchers or institutions. This is calculated by the cosine distance of the angle between the nodes. Formula (1) is as follows:

$$\text{Cosine}(x, y) = \frac{XY}{[X][Y]} = \text{Cosine}(c_{ij}, s_i, s_j) \frac{C_{ij}}{\sqrt{S_i S_j}} \quad (1)$$

where c_{ij} represents the number of papers published by the co-authors (author i and author j), S_i and S_j represent the number of papers published by author i and author j, respectively, and the value of the cooperation strength is between 0 and 1.

VOSviewer uses the correlation strength algorithm, as shown in formula (2):

$$S_{ij} = \frac{C_{ij}}{W_i W_j} \quad (2)$$

In formula (2), C_{ij} represents the number of papers published by the co-authors (author i and author j), W_i and W_j represent the number of papers published by author i and author j respectively, and S_{ij} represents the similarity between author i and author j. It should be noted that the accuracy of VOSviewer's association strength algorithm can be guaranteed only if author i and author j are independent of each other. Therefore, the association strength algorithm measures the similarity from the perspective of probability.

We use CiteSpace and VOSviewer to generate different maps and compare them. It is found that the maps generated by CiteSpace have richer colors and more beautiful appearance. In addition, we can view the articles involved in a node, the scale and content of clustering, and the average year of clustering from the map. Therefore, we decided to use CiteSpace to analyze the data of this study. By using CiteSpace, this paper is able to draw

visual knowledge maps, and obtain the cooperative relationships present between the author and research institutions, as well as identify the research trends in the field of research data.

China National Knowledge Infrastructure (CNKI) is the largest continuously and dynamically updated full-text database of Chinese academic journals globally, and it is the most authoritative document retrieval tool and network publishing platform for Chinese academic journals. It contains all the academic journals in China and covers the contents of all disciplines. Many databases such as CSSCI, SCI, master's degree papers, doctoral theses are in CNKI. The subject of the paper is Chinese research data, which belongs to the library and information discipline. All the published documents on this subject are included by CNKI. With the data source of CNKI, this paper adopted the advanced professional search, with search formula of Subject = 'data sharing' + 'data management' + 'data curation' + 'scientific data' + 'research data' + AND Subject = 'library', to initially research and obtain 1,260 literature records (the search time was December 13, 2019). The author imported the searched literature into CiteSpace software to automatically check the weight and then manually eliminated the requirements for manuscript collection, forum notification, news reports, popular science essays, and other non-research documents. Finally, 1,238 documents were determined.

Results

Collaboration characteristics of researchers

Scientific research collaboration was defined by scientific metrologists Katz and Martin as follows: research scholars work together for common purposes of jointly producing new scientific knowledge. The attributive information of scientific research collaboration was mainly derived from the research of published authors. Therefore, the author of this paper introduced the data into the software with the authors as the nodes, selected the time span from 1988 to 2018, of which the time slice was four years, and the threshold was top 50 publications at each stage, to carry out the visualization analysis, and finally obtained [Figure 1](#). Moreover, the authors of the top ten publication amounts were listed in [Table 1](#). In [Figure 1](#), the node represents the author, the connecting line represents the partnership, and the thickness of the connecting line represents the strength of the relationship.

It can be concluded from [Figure 1](#) that [Gu \(2018\)](#) corresponds to the largest node and has published 13 articles in the field, mainly focusing on the policies of open access to scientific research data and the rights and interests of data management services. [Shen \(2015\)](#) ranks second in the corresponding node and has published 11 articles, mainly focusing on supervision method of research data, librarians' data literacy connotation, and training system. The statistics in [Table 1](#) show that the top ten authors are in relatively important positions in this field. Their studies can help researchers quickly understand the current status and development of research data. From the perspective of collaborative relationship, there are nine scholars in the largest collaborative area, mainly [Hu \(2015\)](#). The main research content of this team is the process supervision of biomedical data in the era of big data. The collaboration team, ranking second, consists of 7 scholars, including [Meng et al. \(2016\)](#), who have mainly studied the data management system and data literacy of the library. The collaboration team headed by [Hu \(2015\)](#) ranks third, which takes chemistry as an example to discuss relevant policies and services for the publication of scientific research data in subject areas and studies data literacy education in the libraries of foreign universities. Moreover, this team includes three scholars among the top ten publications, indicating that the team is one of the key high-performance teams in the research data field from the library perspective.

For CiteSpace's structural control of the network graphs, the k network can be set to filter out the smaller network structures (Note: k refers to the top k largest network structures); when $k = 1$, the obtained network is the largest subnet of the graphic structure. In order to further clarify the collaborative intensity between the research scholars, the author screened

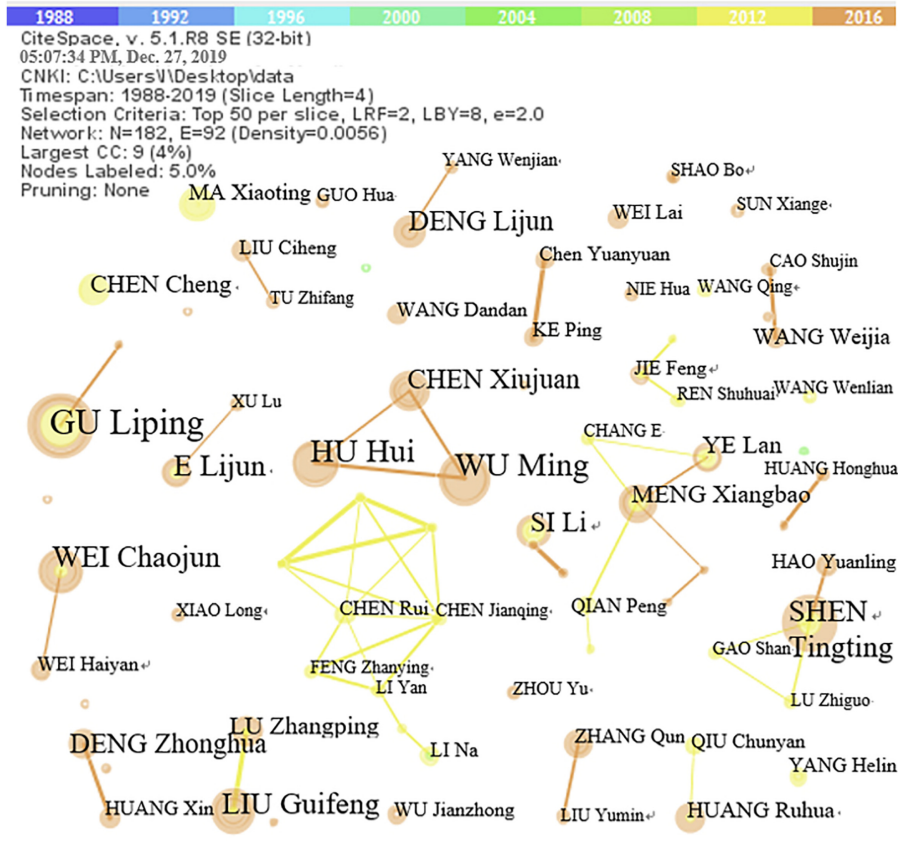


Figure 1. Cooperation graph of scientific research scholars

S/N	Scientific scholars	Amount of publications (ea.)	Time of initial publication (year)
1	Gu Liping	13	2013
2	Shen Tingting	11	2012
3	Wu Ming	10	2016
4	Liu Guifeng	9	2015
5	Hu Hui	9	2016
6	Wei Junchao	9	2014
7	Meng Xiangbao	8	2013
8	Chen Xiujuan	8	2016
9	Ma Xiaoting	7	2013
10	Si Li	7	2013

Table 1. Scholars of top ten publications

the information in Figure 1 and set to display the top five subnet structures and the connecting strength, as shown in Figure 2. Among five largest subnet graph structures, the collaborative strength is 0.84 between Wu and Hu (2016), 1.0 between Jie and Sheng (2016), 0.5 between Meng and Qian (2013), 1.0 among Li *et al.* (2014), and 0.61 between Shen and Hao (2016), these are the collaborative strengths of the largest collaboration teams in the field. In

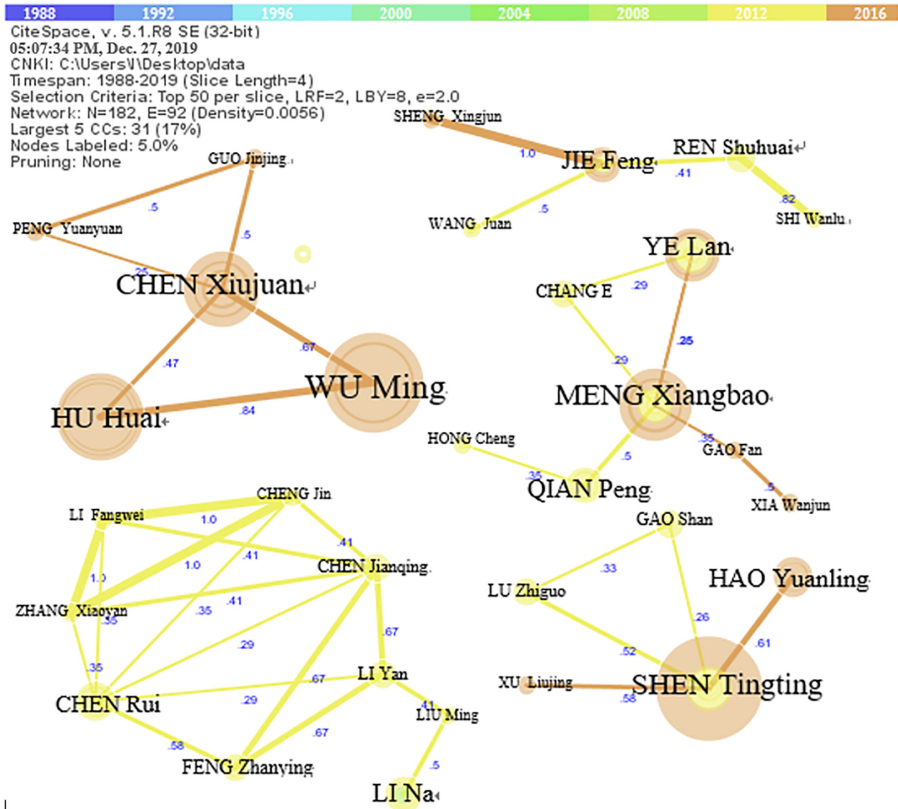


Figure 2.
Cooperative strength
graph of top 5 subnets

summary, the collaborative relationship is relatively strong in values in this field, but the number of publications is low, which should be strengthened further.

Collaborative characteristics of research institutions

In order to analyze the characteristics of the research institutions of the library's research data, the node was set as the operation of institutions to obtain the institutional collaboration graph as shown in Figure 3, and the institutions of top 10 publication amounts were selected to form Table 2. It can be seen from Figure 3 and Table 2 that Wuhan University ranks first and has published the most articles (totally 68) in this field, including School of Information Management, Library, and Center for the Studies of Information Resources of Wuhan University; Shanghai University Library has published 38 articles, ranking second; the Documentation and Information Center of Chinese Academy of Sciences has published 28 articles, ranking third, which is then followed by University of Chinese Academy of Sciences (25 articles), National Science Library of Chinese Academy of Sciences (13 articles), Southeast University Library (12 articles), Medical Library of Chinese PLA (12 articles), National Library (12 articles) and Dept. of Information Management of Nanjing University (11 articles) among top ten publication amounts. It can be seen that the initial publication time was 2004, the earliest, for the School of Information Management of Wuhan University, 2007 for Center for the Studies of Information Resources of Wuhan University, 2009 for National Science

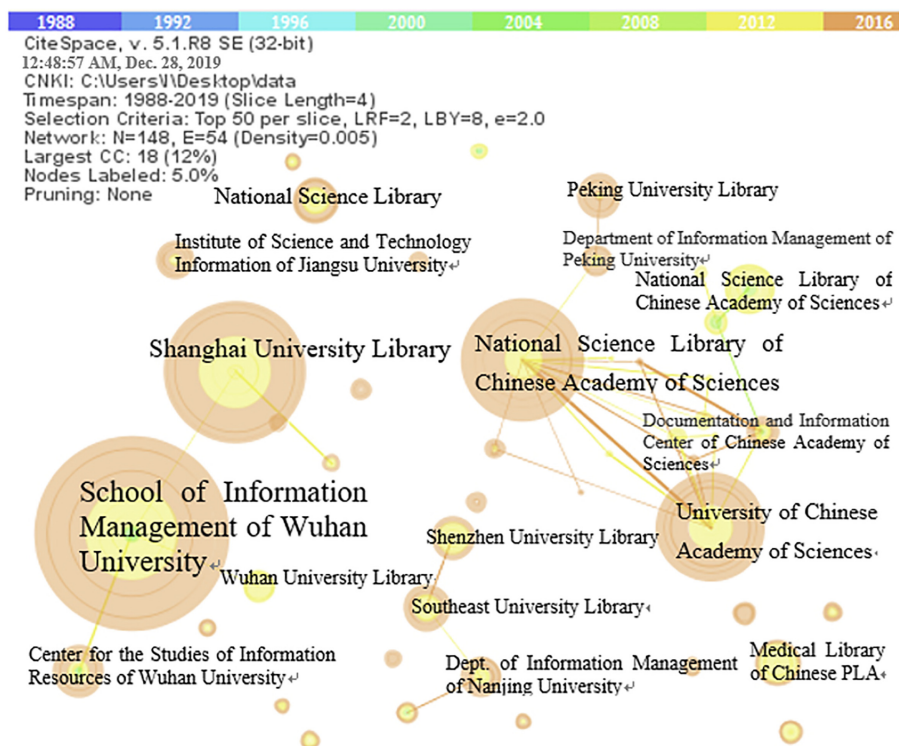


Figure 3. Cooperative Graph for Research Institutions of Library Research data

S/ N	Scientific institutions	Amount of publications (ea.)	Time of initial publication (year)
1	School of Information Management of Wuhan University	45	2004
2	Shanghai University Library	33	2012
3	Documentation and Information Center of Chinese Academy of Sciences	28	2014
4	University of Chinese Academy of Sciences	25	2014
5	National Science Library of Chinese Academy of Sciences	13	2009
6	Center for the Studies of Information Resources of Wuhan University	13	2007
7	Southeast University Library	12	2012
8	Medical Library of Chinese PLA	12	2012
9	National Library	12	2010
10	Dept. of Information Management of Nanjing University	11	2012

Table 2. Institutions with top ten publication amounts

Library of Chinese Academy of Sciences, 2010 for National Library, 2012 for Southeast University, Shanghai University Library, Medical Library of Chinese PLA and Dept. of Information Management of Nanjing University, and 2014 for Documentation and

Information Center of Chinese Academy of Sciences and University of Chinese Academy of Sciences. So, Wuhan University Library, Shanghai University Library and Documentation and Information Center of Chinese Academy of Sciences are high-performance institutions in this field. Meanwhile, the School of Information Management of Wuhan University took the first step early, while the Chinese Academy of Sciences started late but has made very rapid progress in research.

The keywords are concise summaries of the topics and contents of the literature research. It is helpful to know the basic research contents of the literature via correct analysis of the keywords and know the essential hot topics of the subjects, institutions, and research knowledge in a certain period by measuring the number of the keywords (Zhao and Jiang, 2014). In this paper, the author set the node as the keyword, selected the period of 1988–2019 with a slice of 4 years and the top 50 frequent keywords in each stage for visualization, adopted the minimum spanning tree MST to prune the generated graph, and finally clustered the results and extracted them with k (keyword) as the label to obtain Figure 4. In order to visually make out the corresponding frequency and centrality of the keywords, Table 3 was prepared by selecting the top 30 frequent keywords. Clustering was achieved by layering the intimacy and similarity between the research data from high to low. The structure and clearness of CiteSpace clustering were mainly determined by two indicators: modularity (Q -value) and average silhouette (S -value for short). The larger the Q -value was, the better the clustering of the network became; moreover, the Q -value interval was $[0, 1]$; $Q > 0.3$ indicated that the clustering network structure was significant. The S -value could be used to measure the homogeneity of the clustering graph; when it was approaching 1, the homogeneity was

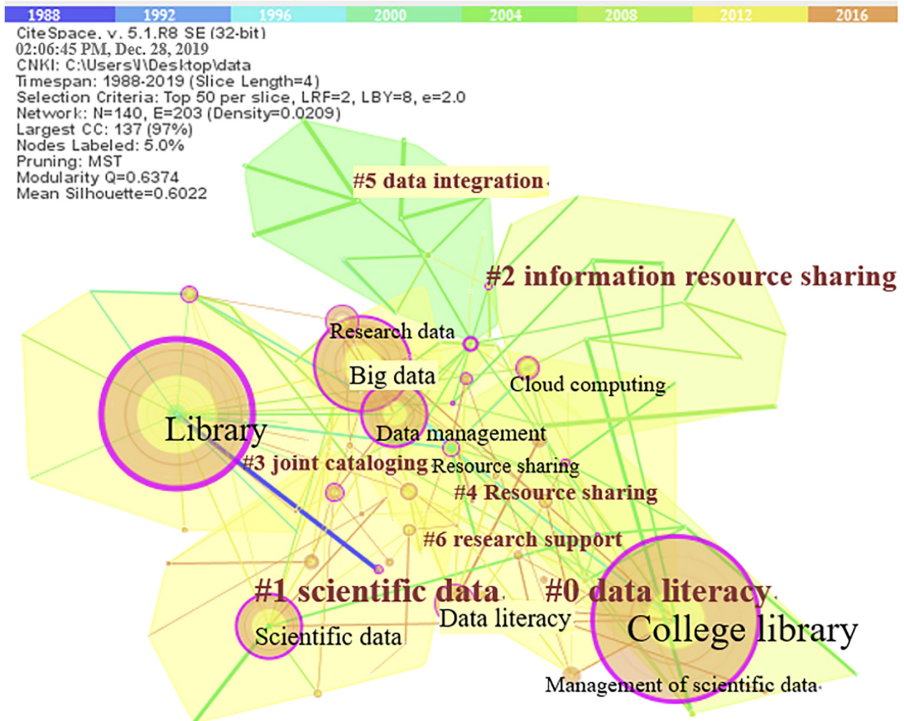


Figure 4.
Keyword
clustering graph

Table 3.
Top 30 frequent keywords

S/ N	Keyword	Freq	Centrality	S/ N	Keyword	Freq	Centrality
1	College library	269	0.47	16	Knowledge service	31	0.04
2	Library	253	0.62	17	Data supervision	29	0.07
3	Big data	154	0.22	18	Data literacy education	28	0
4	Data management	110	0.26	19	Metadata	27	0.26
5	Scientific data	106	0.36	20	Data integration	26	0.05
6	Digital library	79	0.18	21	Subject service	26	0
7	Data literacy	62	0.1	22	Joint construction and sharing	24	0.05
8	Scientific data	58	0.16	23	Information resource	23	0.07
9	Resource sharing	46	0.18	24	Subject librarian	23	0.05
10	Cloud computing	42	0.1	25	Data librarian	22	0.01
11	Scientific data management	38	0.09	26	Institutional knowledge base	22	0.01
12	Data monitoring	36	0.06	27	University library	21	0.14
13	Data service	34	0.1	28	Open access	19	0.09
14	Data sharing	34	0.16	29	Big data era	17	0
15	Scientific data management	33	0.02	30	Research data service	17	0

higher; when it was above 0.5, it was considered that the clustering result was reasonable. As $S = 0.6022$ in Figure 4, it was judged that the clustering structure obtained in this study was clear, and the result was very reliable.

In the graph, each node corresponds to a keyword. The connecting line indicates the co-occurrence relationship between the corresponding keywords, e.g. the co-occurrence relationships between the digital library and cloud computing and between university library and data literacy. The purple edge refers to the point with high intermediate centrality (centrality ≥ 0.1), which is generally considered as a pivot node, such as big data, digital library, data literacy, data management, etc. The flow of knowledge can be judged by the color of each annual ring, e.g. time (color) (the transfer from cool color to warm color means the temporal variation from far to near) (Li and Chen, 2015). It can be visually seen from the graph that the hot research topics in this field are mainly #0 data literacy, #1 scientific data, #2 information resource sharing, #3 joint cataloging, #4 resource sharing, #5 data integration, #6 research support, #7 digital library, etc., of which top five topics will be analyzed in detail in the following paragraphs, the research data, as the subject of this study, will not be discussed further, and #2 information resource sharing and #4 resource sharing are summarized as research data sharing.

Research data sharing

Research data sharing includes resource collaboration, joint construction and sharing, and mutual coordination to meet the needs of users' research activities up to the hilt between university libraries and public libraries or other institutions. #3 scientific data sharing mainly includes 15 keywords, such as digital resources (16, 0.10), library alliance (10, 0.06), document information resources (5, 0.03), information resource sharing system (2, 0.00), data warehouse (4, 0.07), etc. of which the single contour value is 0.753; #4 resource sharing mainly includes 14 keywords, such as cloud computing (42, 0.10), resource sharing (46, 0.18), joint construction and sharing (24, 0.05), literature resources (4, 0.02), etc., of which the single contour value is 0.81. In the era of big data, driven by the data-intensive paradigm, research data sharing has been highly valued in the field of library and information. Among the others,

Wang *et al.* (2008) adopted bibliometrics to analyze the articles of research data sharing from time, journal, and topic in China. Wei and Zhu (2007) and Huang *et al.* (2009) analyzed several measures for the participation of academic libraries in research data sharing. Zhang (2017) discussed the data sharing mode of university libraries from three levels of data research and development, data collection, and data usage. Si and Wang (2018) investigated six research data platforms under the National Basic Science and Technology Condition Platform Project and described the current data organization situation, existing problems, and improvement suggestions for the platforms. In general, the studies on research data sharing in the domestic library and information field mainly focus on dynamic analysis of librarian's research data sharing, dialectical relationship between libraries and research data sharing, new technology of libraries participating in research data sharing platform, and research data sharing mode and practice of libraries.

Data literacy

Data literacy is an extension of information literacy. It mainly includes three aspects of data consciousness, data capabilities (collection and processing, representation and description, discovery and retrieval, selection and evaluation, analysis and utilization, integration and reuse, preservation, and management). Throughout the data lifecycle, data ethics is one of the essential attainments of people in the E-science environment (Huang and Li, 2016). The #0 data literacy mainly includes 21 keywords, including college library (269, 0.47), big data (154, 0.22), data literacy education (28, 0.0), research data service (17, 0), research data (58, 0.16), open access (19, 0.09), research data management (38, 0.09) and so forth, and its individual contour value is 0.674. Through impactful analysis of the hot literature on data literacy research, the author has found that the studies from the library's perspective are mainly focused on data literacy education, which consists of three modules: training data consciousness, cultivating data ability, and establishing data ethics. It is urgently needed by the universities and the society at present and is especially important for researchers and a necessary condition for librarians to participate in research data management and service. Investigations have shown that different groups have significant differences in data literacy. As for university libraries, it is advisable to set data librarian positions, build data service webpages and develop diversified data literacy education to improve the data literacy of high school students and researchers and to effectively improve the efficiency of research data management by improving data literacy of researchers and librarians (Long, 2015). Today, society has entered the era of big data and "Internet +", where information resources are becoming more and more abundant, and higher demand has been posed for data processing capabilities. In the future, library data literacy education can be achieved by traditional literacy education methods, such as training, publicity lectures, and supplemented by library + service practice and data management (Yang, 2015). The research direction at present and even in the future is to improve data literacy by reasonable educational methods and models and practices.

Research data integration management and library's co-cataloging

Data integration is defined to collect, sort, clarify and integrate the research data from different sources to form a new data source. The joint cataloging is mainly to adopt modern technical concepts to integrate and utilize the number and human resources of the libraries at all levels, realize joint construction and sharing of bibliographic resources, and avoid redundant construction of resources. It is one of the important ways of libraries' research data management and an important embodiment of data integration. The #4 data integration mainly includes 12 keywords, such as data integration (26, 0.05), information resource (23, 0.12), text integration (2, 0.01), multi-college integration (5, 0.02), university merger (2, 0.05),

etc. The #3 joint cataloging mainly includes 14 keywords, such as library (253, 0.62), data sharing (34, 0.16), data reference (5, 0.0), catalog sharing (2, 0.03), data conversion (2, 0.03), data integration (2, 0.0), etc. For system transformation of the library or the merger of colleges, the university library needs to implement unified planning and management of the literature resources. [Lu \(2003\)](#) studied the integration of the library bibliography of New Jiangnan University and pointed out that there are problems such as redundancy, waste of human resources, the difficulty of sharing, and low efficiency in the research data before integration. [Zhu and Wang \(2010\)](#) pointed out that the use of cataloging data, data integration of online cataloging organizations, sharing of trans-sectoral cataloging data, and social participation in cataloging mode might effectively promote the explosive growth of libraries' joint cataloging of data resources. The trans-sectoral cataloging and sharing would be an inevitable trend of future development of library cataloging ([Liu and Zhou, 2011](#)).

Research support service

The development of libraries should provide strong support for developing scientific research and continuously optimize the service process to provide researchers with professional research support service. The #6 research support mainly includes ten keywords, such as data monitoring (36, 0.06), research support (10, 0.00), scientific research service (7, 0.07), data librarian (22, 0.01), digital learning (7, 0.00), library service (11, 0.02). The library is the link between the published literature and the research data. It can provide many cross-borders, embedded, and dynamic services for the data support of e-science and e-research, which also lays the foundation for the library to find a foothold in the new era. The research support service has also been one of the requirements for libraries to deepen services in the context of big data in recent years ([Si and Zeng, 2018](#)). The research support service of university libraries is mainly embodied in research data management, open access, academic publishing, research influence measurement, research navigation, research consultation, research tool recommendation. By studying scientific support of American libraries, the scholar [Liu and Chen \(2018\)](#) pointed out that the advanced experience of data librarian training should be learnt from American libraries and continuously improve the career development system of library data librarians in China. [Xia et al. \(2017\)](#) studied scientific research support of the libraries in 40 universities at home and abroad and found through the investigation that foreign research support services had relatively straightforward settings and services and comprehensive contents and were suitable for reference of the libraries at home. The scientific research support of the libraries should be improved in deepening knowledge and service levels ([Song et al., 2017](#)).

Evolution and frontal of research data

CiteSpace V software was adopted to visually analyze the time-zone distribution of keywords to dissect the evolution path of hot scientific data research topics. With the constant time division, the author chose and set the node as keyword, the threshold item as Top 20, and the output as "Time Zone" to obtain the evolution path graph of hot scientific data topics ([Figure 5](#)). In [Figure 5](#), a series of keywords in each time zone represents a hot study topic in this time zone. The study of library scientific data began in the late 1980s and the beginning period was from 1988 to 1995, when related research was very scarce; from 1996 to 2000, the studies mainly focused on the sharing of library research data; from 2001 to 2003, college libraries and digital libraries entered people's field of vision, which was also the development of data sharing research in previous years and made data sharing and use better; from 2004 to 2006, studies mainly focused on scientific data management; from 2007 to 2011, the application of cloud computing became a major research topic; from 2012 to 2019, higher and more extensive demands were placed on the original scientific data due to the arrival of the

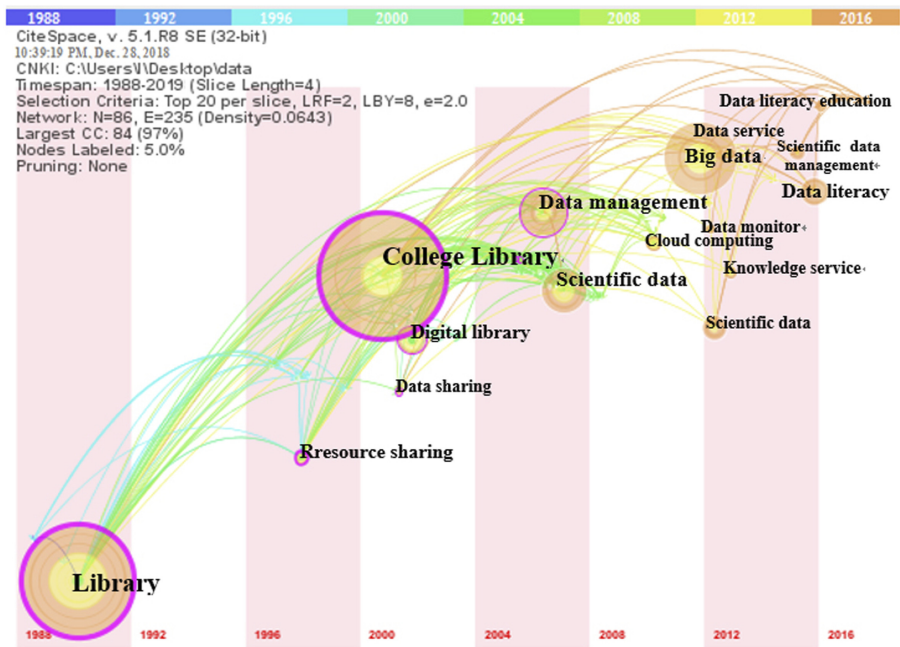


Figure 5.
Keyword time
zone graph

data-intensive era in 2012, and in this period, impactful studies were conducted on data supervision, scientific data management, data service, and knowledge service. The new era has provided researchers with good data resources and research tools and put forward higher requirements for researchers' data literacy. Therefore, data literacy and its education were mainly studied during 2016–2019.

The Burst Detection algorithm was proposed by Kleinberg (2002). It refers to the sudden increase number in a short time, which has an intelligence function and can reveal the frontal aspects of research in this field. Figure 6 is based on the burst detection of keyword click. In this figure, Keywords are the corresponding ones, Year is the time when the search record appears for the first time, strength is the intensity of the burstiness, the beginning indicates the time to become the frontal topic, the ending is the closing time, and the start and end times correspond to one red rectangular block for a year. It can be seen from Figure 6 that the burst keywords up to 2019 are data literacy and its education, of which the corresponding burst strengths are 9.222 and 7.2756%, respectively. The data literacy research has been sustained for four years, while the data literacy education has only lasted for three years. It is a general trend for strengthening data literacy education in the era of big data, and the university library gradually transforms from traditional information literacy education to data literacy education (Zhang, 2018). The development and practice of data literacy education are currently in their infancy. The studies of relevant scholars in China are based on foreign university libraries, such as research course, research team and learning process of the data literacy education practice in Purdue University Library (Xu and Gao, 2018), the New England Data Management Collaborative Program of the University of Massachusetts, the Data Information Literacy Program jointly developed by Purdue University, University of Minnesota, University of Oregon and Cornell University, and the MANTRA Education Program at the University of Edinburgh Library (Hu and Wu, 2016). Some scholars have also

Top 13 Keywords with the Strongest Citation Bursts

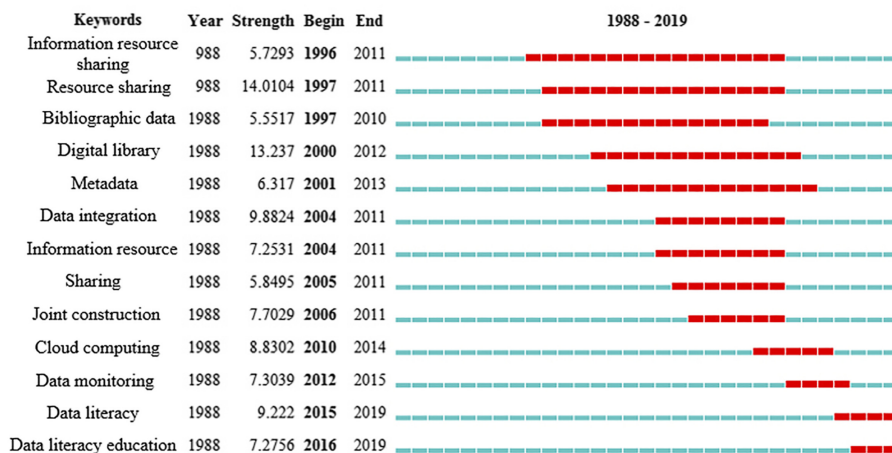


Figure 6. Visualization of keyword burstness

researched current situation in China, but they have not carried out effective practice. Therefore, the impactful study on data literacy and its education practice and application is one of the leading research directions in the future.

Discussion

The study aimed to review the literature on the topic of research data systematically. The researchers selected 1,238 studies fulfilling the inclusion criteria of the study. The reviewed literature revealed that research data is in an immature stage. Comparatively, research on research data is better observed in academic universities than other research institutions. We discuss the data analysis results as follows: (1) The publications of relevant literature in research data are increasing year by year, which indicates that the field has gradually gained attention from academic institutions. With the development of E-science and data-intensive research, theoretical and practical research in the field of research data will grow, and hot topics will change at different stages, (2) the researchers focus through three stages: the construction of research data sharing platform, data management and service, and data literacy education. Most of the literature in the first phase concerns practical cases of research data sharing. It includes four aspects: research data sharing platform construction, heterogeneous data resource integration and access (Liu and Zhu, 2007), the exploration of the influencing factors of sharing platform construction (Huang *et al.*, 2008; Si *et al.*, 2014), and the strategy of promoting research data sharing (Huang *et al.*, 2014). The second stage is that scholars began to pay attention to the management and service of research data. These include research data monitoring service and library role adjustment (Zhu, 2014), big data and research data management, foreign scientific data management and service practice (Zuo and Chen, 2014), and joint cataloging of research data and library literature resources (Zhu and Wang, 2010; Liu and Zhou, 2011). In the third stage, many scholars pay attention to library data literacy education and training (Long, 2015; Yang, 2015). Data monitoring training and transformation of Information literacy education to data literacy education become the main issues considered by librarians and stakeholders, (3) the initial researchers were mainly stakeholders in the study of data. They explored the theory and practice from their respective disciplines and studied the construction of data sharing platforms and data

management in their respective fields. Then librarians gradually participate in the research of research data. They use the research data as an information resource and study it from organization management and information service.

Limitation and future research direction

This study is a systematic literature review, and it is possible that some relevant studies might have been missed. Further, the data were limited to published studies between 1998 and 2018 and further limited to specific databases and sources. This paper uses quantitative research methods to analyze the relevant literature in the field of research data; hence, more studies using the mixed-method approach may be needed to understand the research data in depth.

Implications of the study

Policy implications

Research topics gradually shift from technology to management, service, and policy from the path evolution of research hotspots. However, in addition to the issue, the policy documents such as “promulgation of the Measures for the Management of Research Data” and “Interim Measures for the Sharing of Government Information Resources”, the professional associations, research institutions, universities, and local governments should also formulate supporting implementation policies. Researchers need to continue to study foreign research data management policies, summarize the experience and improve the connection between macro and micro policies combined with the actual situation, and further promote the formulation and implementation of research data management and sharing policies in Chinese libraries.

Practical implications

Academic libraries have studied the theory and practice of research data management and service since 2011, but they are not in the core position in academia. For promoting academic libraries’ research on research data, we put forward three practical implications: (1) Chinese higher education commission or ministry, funding agency, higher education institution and/or research commission should allocate budget to train researchers and librarians to raise their awareness and technical level of research data management services; focus on training librarians data management planning, data processing and analysis, data description, data sharing platform construction, data quality education, (2) academic libraries and stakeholders should pay attention to the transformation of research hotspots in the field of research data and promote the development of library information services, (3) academic libraries should use collaborative governance theory, data life cycle theory, stakeholder theory, data asset theory, research data technology to construct library research data governance model, research data management and structure system. In this way, the relevant theories and technologies of the research data will be combined with innovation.

Because the research data field is still in its infancy in developing countries such as China, the research data literacy and data management awareness of researchers and stakeholders are relatively weak, so the collaborative relationship between scholars is loose and needs to be strengthened urgently. Higher education institutions and/or research boards, funding agencies and higher education commissions or ministries should sit together and make it compulsory while granting funding to researchers to submit their research data in their institutional or subject repositories and publish their work in open access journals. The research literature of academic libraries around research data management and service will

continue to increase in the future. They should cooperate with multidisciplinary researchers to carry out interdisciplinary research data management and service research to expand the boundary of research data.

Conclusion and recommendation

From the perspective of library, this paper establishes a visual knowledge map for the subject literature of Chinese research data by using bibliometrics and social network analysis methods and focuses on the volume and collaboration of authors on the subject in the past 30 years, the volume and collaboration of scientific research institutions, the research status and future research trend of library research data. It can be used as a reference direction for scholars to study all aspects of research data with the library as the starting point. The results show that (1) among the related research on the subject of research data in the library, Gu Liping and Shen Tingting have the most papers, but the collaborative relationship between scholars is loose and needs to be strengthened urgently; (2) the scientific research output institutions of library research data subject research are in the top three: Wuhan University, Shanghai University Library and document Information Center of Chinese Academy of Sciences, and the collaborative relationship between scientific research institutions is not close, and it is necessary to strengthen collaboration; (3) from the cluster analysis of keywords, we get the research hotspots related to the research data, they are the hotspots of library science subject research include the construction of research data sharing platform, librarian data literacy education, research data integration management and library joint cataloging, library research data service mode, library participation in research incentive mechanism and policy support, library to verify the effectiveness of research data; (4) from the path evolution of research hotspots, we know that the theme of research data has experienced three stages of data sharing platform construction, data management and service, and data literacy education, (5) burst detection tells us that in-depth research of data literacy and practical application research of data literacy education are the main research directions in the future.

Through the in-depth interpretation of the related papers about research data, this paper puts forward some suggestions on the future research direction: (1) In the era of big data, cloud computing technology has been gradually applied to the research data sharing platform because of its high security, various data types, fast access, and low energy consumption, which further improves the sharing ability of research data and ensuring its security. In the new era, the data sharing mode of the library will be fully developed, and the construction of the unified data sharing mode, the convenience of data functions on the sharing platform and the improvement of data analysis capabilities will be essential research contents in the future. (2) Based on profoundly analyzing the connotation of data literacy, research data workers should combine the national conditions of China and the future library development plan and deeply study how to adopt effective and operable practical strategies or solutions to embed library data literacy education in every link of other courses or activities. Scholars should study the technical difficulties faced by data literacy work, librarians' ability requirements, and integrating with other library services and education. (3) Joint cataloging is an important and effective way to effectively manage, share, integrate, and develop library data and solve traditional library data problems. Therefore, it is required to continuously improve the joint cataloging of libraries and the trans-sectoral, trans-class, and trans-level cataloging mode to better integrate, manage and share library data resources. (4) The librarians should provide better support services for scientific research. It is necessary to continuously improve the data structure and cultivate the professional quality of the subject librarians and deepen service levels (e.g. fully participating in scientific research, promoting the publication of research data, standardizing data references, and providing research data quality assessment) and carry out impactful studies in other aspects. In addition, scholars should increase systematic empirical

research and strengthen interdisciplinary collaboration to find out the unique value of library research data through joint exploration of research data with other disciplines. In the research direction, scholars should carry out basic research based on library and information discipline characteristics, such as the formulation of scientific data management policy, the professional literacy of relevant personnel, and the intellectual property rights of scientific data.

References

- Brochu, L. and Burns, J. (2019), "Librarians and research data management – a literature review: commentary from a senior professional and a new professional librarian", *New Review of Academic Librarianship*, Vol. 25 No. 1, pp. 49-58.
- Chao, G., Zhen, W. and Li, X. (2016), "PR-index: using the h-index and PageRank for determining true Impact", *PLoS One*, Vol. 11 No. 9, pp. 1716-1725.
- Chawinga, W.D. and Zinn, S. (2019), "Global perspectives of research data sharing: a systematic literature review", *Library and Information Science Research*, Vol. 41 No. 2, pp. 109-122.
- Chen, H. and Chen, L.D. (2017), "Visualization analysis on intellectual structures and research fronts of intercultural communication studies", *Chinese Journal of Journalism and Communication*, Vol. 7, pp. 58-89.
- Fuhr, J. (2019), "How do I do that? A literature review of research data management skill gaps of Canadian health sciences information professionals", *Journal of the Canadian Health Libraries Association/ Journal de L'association Des Bibliothèques de la Santé du Canada*, Vol. 40 No. 2, pp. 51-60.
- Grant, R. (2017), "Recordkeeping and research data management: a review of perspectives", *Records Management Journal*, Vol. 27 No. 2, pp. 159-174.
- Gu, L.P. (2018), "Data management services in the transition of research model: an approach of implementing open access, open data and open science", *Journal of Library Science in China*, Vol. 44 No. 238, pp. 43-58.
- Hu, B. (2015), "Investigation of user information demand in science museums", *Library and Information Service*, Vol. 9 No. 59, pp. 58-63.
- Hu, H.F. and Fang, X., M. (2021), "Review on scientific data evaluation across the World", *Journal of Academic Library and Information Science*, Vol. 39 No. 03, pp. 131-138.
- Hu, H. and Wu, M. (2016), "Research and enlightenment of best practices in data literacy education in foreign libraries", *Journal of Modern Information*, Vol. 36 No. 8, pp. 6678-6774.
- Huang, R.H. and Li, B.Y. (2016), "Data literacy education: expansion of information literacy instruction in the big data era", *Document, Information and Knowledge*, Vol. 1, pp. 21-29.
- Huang, D.C., Li, X.B., Li, J.M., Xu, F. and Wang, J.L. (2008), "Study on sustainable development mechanism of national scientific data center", *China Science and Technology Resources Review*, Vol. 40 No. 5, pp. 3-8.
- Huang, X.J., Zhu, J. and Li, J.N. (2009), "Research on research libraries participating in scientific data sharing service", *Library Tribune*, Vol. 29 No. 6, pp. 177-179.
- Huang, R.H., Wang, B. and Zhou, Z.F. (2014), "A study on countermeasures for promoting scientific data sharing in China", *Library*, Vol. 3, pp. 7-13.
- Jie, F. and Sheng, X., J. (2016), "The center for digital scholarship: service transformation and space change in libraries: a case study of the CDS of academic libraries in North America", *Library and Information Service*, Vol. 60 No. 13, pp. 64-70.
- Kleinberg, J. (2002), "Bursty and hierarchical structure in streams", *Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Edmonton, Alberta, Canada, pp. 91-101, ACM Press.
- Li, J. and Chen, C.M. (2015), *Cite Space: Text Mining and Visualization in Scientific Literature*, Capital University of Economics and Business Press, Beijing.

- Li, F.W., Lv, T., Cheng, J., Zhang, X.Y., Yang, X.R. and Chen, R. (2014), "Biomedical research data service abroad and its enlightenments", *Chinese Journal of Medical Library and Information Science*, Vol. 23 No. 6, pp. 61-65.
- Liu, M. (2020), "Review of library research data literacy in China", *Research on Library Science*, Vol. 12, pp. 17-23.
- Liu, H.J. and Chen, Y. (2018), "Study on data librarian of American academic library [J/OL]", *Library Development*, 1-6 [2018-12-28], available at: <http://kns.cnki.net/kcms/detail/23.1331.G2.20181130.1356.006.html>.
- Liu, X.H. and Zhou, X.L. (2011), "Study on the cross-industry cataloging sharing and its operation", *Library Development*, Vol. 4, pp. 34-37.
- Liu, R.D. and Zhu, Y.Q. (2007), "Explore key issues of scientific data sharing—data sharing network of earth system science as an example", *Progress in Geography*, Vol. 26 No. 5, pp. 118-126.
- Long, Q. (2015), "Quality ability and investigation of present situation of data literacy ability of teachers and students", *Library*, Vol. 12, pp. 5162-5256.
- Lu, H. (2003), "Integration of university library bibliographic data-bibliographic data integration of library computer integration system in former four campuses of new Jiangnan university", *Journal of Modern Information*, Vol. 9, pp. 198-200.
- Ma, H.P. (2019), "A Literature Review of library scientific data sharing from 2010 to 2019 in China", *Research on Library Science*, Vol. 08, pp. 19-26.
- Meng, X.B. and Qian, P. (2013), "International experiences and references of university social sciences data management: taking UKDA and ICPSR for example", *Information and Documentation Services*, Vol. 2, pp. 77-80.
- Meng, X.B., Chang, E. and Ye, L. (2016), "Data literacy research: origins, progress and prospects", *Journal of Library Science in China*, Vol. 42 No. 222, pp. 109-126.
- Ng'eno, E. and Mutula, S. (2018), "Research data management (RDM) in agricultural research institutes: a literature review", *Inkanyiso: Journal of Humanities and Social Sciences*, Vol. 10 No. 1, pp. 28-50.
- Ruan, J.H. and Yang, Y. (2019), "A literature review on scientific data security behavior", *Journal of Modern Information*, Vol. 39 No. 09, pp. 151-159.
- Shen, T.T. (2015), "Data literacy and its Impact on scientific data management", *Library Tribune*, Vol. 35 No. 1, pp. 68-73.
- Shen, T.T. and Hao, Y.L. (2016), "Construction and strategy of data literacy and its training mechanism", *Information Studies: Theory and Application*, Vol. 39 No. 1, pp. 58-63.
- Sheng, X.P. and Yuan, Y. (2021), "A review of influence factors of open sharing of scientific data at home and abroad", *Information Studies: Theory and Application*, Vol. 44 No. 08, pp. 173-179+102.
- Si, L. and Wang, Y.W. (2018), "Current situation and improvement suggestions of data organization of scientific data sharing platform in China - analysis based on national science and technology basic conditions platform", *Library Development*, Vol. 10, pp. 52-58.
- Si, L. and Xing, W.M. (2017), *Theory and Practice of Scientific Data Management and Sharing*, Wuhan University Press, Wuhan.
- Si, L. and Zeng, Y.L. (2018), "Investigation and analysis of library research support services in world-class universities", *Library and Information Service*, Vol. 62 No. 8, pp. 30-41.
- Si, L., Li, Y.T., Xing, W.M., Hua, X.Q., Li, X. and Xin, J.J. (2014), "Empirical research on performance evaluation of China scientific data sharing platform", *Library Theory and Practice*, Vol. 9, pp. 30-35.
- Song, H.Y., Guo, J. and Dong, J. (2017), "Study on process framework and implementation path of deep knowledge service in university library", *Library and Information Service*, Vol. 61 No. 5, pp. 6-13.
- Sun, J.Z. (2016), "Research on the path of scientific data management and sharing in university library under the E-science environment", *Library*, Vol. 5, pp. 66-71.
- Wang, D.D. (2018), *Scientific Data Management Service in Data-Intensive Environment*, Science Press, Beijing.

- Wang, Q.L., Zhong, Y.H. and Jiang, H. (2008), "Bibliometric analysis on scientific data sharing in China", *Journal of Intelligence*, Vol. 7, pp. 128-130.
- Wei, D.Y. and Zhu, Z.Y. (2007), "How to share the scientific data among the professional libraries", *Library Tribune*, Vol. 6, pp. 253-256.
- Wu, M. and Hu, H. (2016), "The actuality of data literacy and data literacy education based on the literature metrology method", *Journal of Modern Information*, Vol. 36 No. 12, pp. 152-159+169.
- Xia, W.J., Liu, Y. and Zhao, Y.M. (2017), "Comparative analysis of university library research services at home and abroad", *Journal of the Library Science Society of Sichuan*, Vol. 3, pp. 25-28.
- Xu, L.L. and Gao, D.W. (2018), "Practice and enlightenment of embedded data literacy education in Purdue university library", *Library World*, Vol. 2, pp. 51-54.
- Yan, J.L., Min, C., Yu, H.Q., Wei, J.P., Jia, T. and Ma, J. (2020), "Presentations, cases, and opportunities for the research collaboration in the context of big scientific data—an overview of the workshop of scientific data driven research collaboration", *Library and Information*, Vol. 03, pp. 127-133.
- Yang, X.F. (2015), "Study of data literacy education in library in the context of "Internet +""", *Library and Information*, Vol. 5, pp. 541-543, 122.
- Zhang, Y.J. (2017), "Study on scientific data sharing modes among university libraries under the E-science", *Journal of the National Library of China*, Vol. 39 No. 2, pp. 56-59.
- Zhang, J.H. (2018), "Research on university library data literacy education under big data environment", *The Library Journal of Henan*, Vol. 38 No. 10, pp. 59-60.
- Zhao, J.F. and Jiang, F. (2014), "Study on bibliometrics and scientific knowledge graph in modern college education", *University Education Science*, Vol. 111 No. 1, pp. 115-123.
- Zhu, C.P. (2014), "The way and contents for university libraries to provide scientific data", *Library and Information*, Vol. 3, pp. 97-99.
- Zhu, Q.Q. and Wang, Y.Q. (2010), "Several ideas on cataloguing sharing mode", *Library Theory and Practice*, Vol. 4, pp. 31-33.
- Zuo, J.A. and Chen, Y., "The analysis on the sharing mode of scientific data in the era of big data", *New Century Library*, Vol. 3, pp. 32-35.

Further reading

- Chen, C. (2006), "Cite space II: detecting and visualizing emerging trends and transient patterns in scientific literature", *Journal of the American Society for Information Science and Technology*, Vol. 57 No. 3, pp. 359-377.
- Chen, Y.Y. and Ke, P. (2017), "A summary of research on scientific research data service in university libraries", *Library Work and Study*, Vol. 10, pp. 16-23.
- Chen, Y. and Liu, Z.Y. (2005a), "The rise of mapping knowledge domain", *Studies in Science of Science*, Vol. 2, pp. 149-154.
- Chen, Y. and Liu, Z.Y. (2005b), "Quietly rising scientific knowledge graph", *Studies in Science of Science*, Vol. 2, pp. 149-154.
- Hu, H., Wu, M. and Chen, X.J. (2016), "Data literacy education in British and American academic libraries", *Library and Information*, Vol. 1, pp. 62-69.
- Huang, R.H. and Qiu, C.Y. (2013), "Review of research of the scientific data sharing in foreign countries", *Information and Documentation Services*, Vol. 4, pp. 24-30.
- Li, D.D. and Wu, Z.X. (2012), "Comprehensive analysis of research data management services", *Research on Library Science*, Vol. 9, pp. 54-59.
- Li, J.M. and Xiong, A.Y. (2004), "Summary of research on meteorological scientific data sharing system", *Journal of Applied Meteorological Science*, No. 1, pp. 1-9.

- Ling, X.L., et al. (2007), "Review on Australian antarctic data management and its implication", *Advances in Earth Science*, Vol. 5, pp. 532-539.
- Qian, J.L. and Liu, G.F. (2017), "A summary of research on foreign scientific research data management", *Information Studies: Theory and Application*, Vol. 10, pp. 130-134.
- Qin, C.J. and Hou, H.Q. (2009), "Mapping knowledge domain - new fields of information and knowledge management", *Journal of Academic Libraries*, Vol. 27 No. 1, pp. 30-37, 96.
- Teng, G.Q., Mou, D.M. and Ren, J. (2014), "Research on the application of social network analysis in the fields of bibliometrics abroad", *Information and Documentation Services*, Vol. 1, pp. 47-51.
- Tu, Z.F. (2018), "A review of fundamental research and identification of key issues on scientific data publishing", *Library*, Vol. 6, pp. 86100-86192.
- Wang, G.H. (2007), "A summary of the research on sharing of land resources scientific data", *Bulletin of Surveying and Mapping*, Vol. 4, pp. 34-37.
- Wang, Q. (2014), "Review on open sharing of scientific data based on bibliometrics (2002-2013)", *Journal of Tianjin R and TV University*, Vol. 18 No. 2, pp. 75-80.
- Zhang, C.T. (2009), "The e-index, complementing the h-index for excess citations", *PLoS One*, Vol. 4 No. 5, pp. 154-169.
- Zhang, X.X. (2016), "A review of scientific data management in universities", *Information and Documentation Services*, Vol. 6, pp. 48-54.
- Zhang, M.X. and Gu, L.P. (2016), "Policy research of data curation", *New Technology of Library and Information Service*, Vol. 1, pp. 3-10.
- Zhou, Z.F. (2016), "The mapping knowledge domains based on hotspots research of scientific data in mainland China", *Journal of Intelligence*, Vol. 35 No. 1, pp. 81-86.
- Zhou, Y. and Liao, S.Q. (2017), "A review of scientific data semantic description", *Library and Information Service*, Vol. 61 No. 12, pp. 136-144.
- Zhou, B. and Qian, P. (2013), "Summary of scientific metadata researches in China", *Research on Library Science*, Vol. 2, pp. 7-10.
- Zhu, L. and Meng, X. (2013), "The comparative study on bibliometric method and content analysis method", *Library Work and Study*, Vol. 6, pp. 64-66.

About the authors

Chunlai Yan is a PhD graduate student at Rajamangala University of Technology Rattanakosin in Thailand and a lecturer of Panzhihua University in China. Her main fields are information management, computer science, information system, information security, etc. Chunlai Yan is the corresponding author and can be contacted at: wingvestige@163.com

Hongxia Li is a doctoral supervisor at Rajamangala University of Technology Rattanakosin in Thailand and a professor of Chongqing Technology and Business University in China. Her main fields are information management, management psychology and behavior, management science, operation management, information system, information security, etc.

Ruihui Pu is a doctoral supervisor at Rajamangala University of Technology Rattanakosin in Thailand. His main fields are human resource management, human resources development, etc.

Jirawan Deeprasert is a doctoral supervisor and an assistant professor at Rajamangala University of Technology Rattanakosin in Thailand. Her main fields are decision-making process, customer relation management, etc.

Nuttapong Jotikasthira is a doctoral supervisor and at Rajamangala University of Technology Rattanakosin in Thailand. His main fields are tourism management, marketing strategy, etc.

For instructions on how to order reprints of this article, please visit our website:

www.emeraldgrouppublishing.com/licensing/reprints.htm

Or contact us for further details: permissions@emeraldinsight.com