ibres volume 26, issue 1, pages 73-88 (2016)

Current Status of Scientific Data Curation Research and Practices in Mainland China

Shiyan Ou
School of Information Management,
Nanjing University, Nanjing, China
oushiyan@nju.edu.cn

Yu Zhou
School of Information Management,
Nanjing University, Nanjing, China
DG1514026@smail.nju.edu.cn

ABSTRACT

With the rapid growth in the body of scientific data, scientific research depends more and more on finding theories and knowledge from the data, and thus data-intensive scientific discovery has become the fourth paradigm of scientific research. Therefore, it is urgent to develop and adopt methods to support the collection, collation, preservation and utilization of scientific data. This paper provides an overview of scientific data curation research and practices in mainland China. Firstly, it reviews Chinese research articles on data curation and outlines the research status and progress in this area. Secondly, it surveys existing scientific data repositories or platforms in mainland China, and analyzes the gaps between China's and other countries' data curation practices.

INTRODUCTION

With the rapid growth in research activities and the resulting body of scientific data, data have become as important a strategic resource as natural resources and human resources. Scientific research depends more and more on finding theories and knowledge from scientific data, and thus data-intensive scientific discovery has become the fourth paradigm of scientific research. Therefore, it is urgent to develop and adopt methods to support the collection, collation, preservation and utilization of scientific data, which are referred to as scientific data curation.

The term *data curation* was first coined by Zorich (1995) in his paper on museum future collection management, to refer to the maintenance and reuse of museum objects and specimen information. At the *Digital Curation: Digital Archives, Libraries and e-Science Seminar*, sponsored by the Digital Preservation Coalition and British National Space Centre and held in London in 2001, the term *digital curation* and *data curation* with their modern meanings were proposed in the context of e-Science which requires the management, preservation, and sharing of a wide range of scientific data (Beagrie & Pothen, 2001). However, the specific definitions of these two terms were not mentioned at the time.

In 2002, Gray, Szalay, Thakar, Stoughton and Vandenberg (2002) mentioned the term *data curation* in their technical report about scientific data publication and preservation. However, they also did not give a clear definition and only listed some related activities, such

Published by WKW School of Communication & Information & NTU Libraries,
Nanyang Technological University

as recording the data gathering process, archiving all versions of programs and event logs, and keeping notes, memoranda and experimental data.

In 2003, the U.K. Joint Information Systems Committee (JISC) gave a clear definition to the term *data curation* for the first time in its e-Science curation report: "the activity of, managing and promoting the use of data from its point of creation, to ensure it is fit for contemporary purpose, and available for discovery and reuse" (Lord & Macdonald, 2003). Here, *data* refer to "digital primary research data generated in academic and scientific research" (Lord & Macdonald, 2003).

After examining the various definitions of *digital curation* from 2003 to 2007, Yakel (2007) concluded that "digital curation has increasingly become an umbrella concept that includes digital preservation, data curation, electronic records management, and digital asset management." However, he did not give clear definitions to these two terms.

The Digital Curation Center (DCC), a U.K.-based center with a focus on building capacity, capability and skills for research data management, gives a formal definition to "digital curation" as "maintaining, preserving and adding value to digital research data throughout its lifecycle" (Digital Curation Centre, 2004). It also made a subtle distinction between "data curation" and "digital curation" as follows: both of them refer to looking after and adding value to data, and moreover the latter also implies creating some new data from the existing to ensure its current and future usefulness (Digital Curation Centre, 2008). In 2011, Walters and Skinner (2011) distinguished between these two terms according to their application scope and deemed that data curation is applied most often in science, engineering and social science fields whereas digital curation is used more frequently to describe digital humanities and the arts environment.

In China, some researchers also had discussions about the intension and extension of these two terms. It has been generally recognized that *data curation* is often used in the U.K. whereas *digital curation* is often used in the USA (Wang, 2014; Xia, 2013; Yang, 2016). *Data curation* is also often used in the natural science disciplines whereas digital curation is often used in libraries, archives and other social science disciplines (Wang & Shen, 2014). No matter how these two terms are defined, they do not have essential differences and can be used interchangeably in most cases. In this paper, we regard *data curation* as the activities that collect, annotate, organize and preserve digitalized scientific data for current or future use.

This paper aims to give an overview of scientific data curation research and practices in mainland China. Firstly, it reviews Chinese research articles on data curation and outlines the research status and progress in this area. Secondly, it surveys existing scientific data repositories or platforms in mainland China and analyzes the gaps between China's data curation practices and those of other countries.

DATA CURATION RESEARCH IN MAINLAND CHINA

To review the research status of data curation in China, we searched for research papers on data curation published in Chinese core journals from 2000 till present. There were 258 research papers retrieved in total. Based on these research papers, we first analyzed the research history of data curation in China, and then identified hot research topics in this field.

Research History of Data Curation in China

Although the concept of data curation was proposed in 2001, there was no research paper specifically about data curation in mainland China until 2011. However, during the ten years

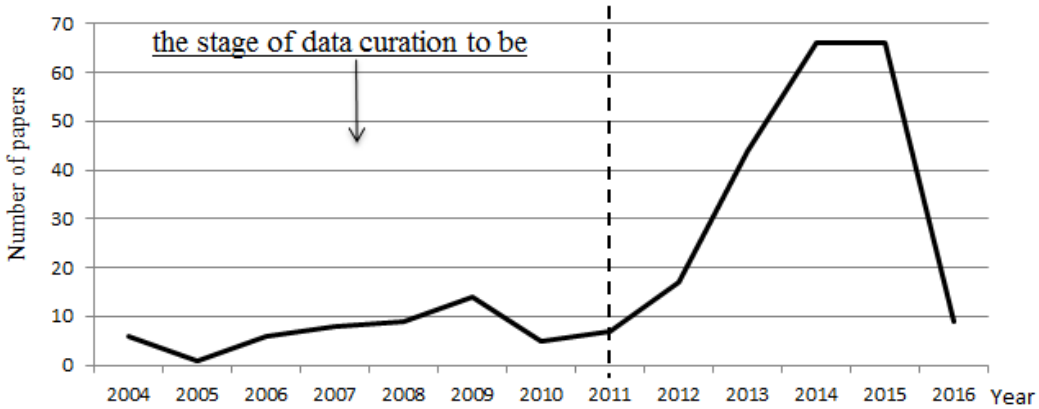

Figure 1. Number of research papers on data curation in mainland China from 2004 to 2016

from 2001 to 2011, almost 50 papers about scientific data have emerged, focusing on exploring the sharing mechanisms and sharing systems of scientific data across multiple disciplines. Chen (2004) published the earliest domestic paper about scientific data in 2004, in which he proposed the importance of scientific data as national strategic resources and provided some effective suggestions to improve public access to scientific data. In 2006, more papers were published to introduce the construction of data sharing platforms or systems in China, including Liu and Yin's paper (2006) about the construction of China's medical science data network, Huang and Li's paper (2006) about the design and implementation of data management and data publishing system for oceanology data of the South China Sea, Li, Shen and Wang's paper (2006) about the construction of a sharing platform for meteorological information, and Chen, Li and Sun's paper (2006) about the development of a welding scientific data sharing platform. In addition, Liu, Cai, Peng, Shan, Tian, Pang, & Znang (2007) introduced the seismological data sharing project sponsored by China's Ministry of Science and Technology in 2007. In the same year, Jin and Zhang (2007) designed and implemented a forging and stamping scientific data sharing system. Although these papers imply nascent ideas about data curation, they are still not regarded as real studies on data curation for three reasons:

1. the depth and breadth of scientific data management is less than that of data curation, since the former places more emphasis on scientific data sharing and sharing platform;
2. the construction of scientific data management systems still follows the way of constructing traditional information systems;
3. the organization of scientific data still uses the same methods as those used in organizing scientific literature, for example, databases or file systems.

Thus we call this research period *the stage of data curation to be*.

In 2011, the first paper which formally introduced data curation in China was published by Yang (2011). From then on, the number of research papers on data curation increased year by year and peaked in 2014 and 2015 (as shown in Figure 1). Research on data curation in China can be divided into three stages: 2011, 2012-2014, and 2015 to the present.

During the first stage, domestic researchers focused on discussing the intension and extension of the *data curation* concept and providing general introductions to data curation. During the second phase, the researchers analyzed and reviewed practical data curation projects in foreign countries to learn experiences and lessons from them. For example, Yang (2012) analyzed DataStaR, a data staging repository for digital research data developed by

Table 1. Top 20 high-frequency keywords in the research articles on data curation

ID	Keywords	Document frequency
1	Scientific data	81
2	Data curation	43
3	Data sharing	37
4	Data management	30
5	University library	27
6	Scientific research data	22
7	Metadata	19
8	Scientific data sharing	18
9	Library	17
10	Scientific data management	16

Cornell University Library; Meng and Qian (2013b) analyzed the U.K. Data Archive (UKDA), the U.K.'s largest collection of digital research data in the social sciences and humanities, and the Inter-university Consortium for Political and Social Research (ICPSR), an international consortium of academic organizations and research institutions in data stewardship; Xu, Liu and Yu (2014) analyzed the Data Observation Network for Earth (DataOne), a community-driven project providing access to data across multiple member repositories and supporting enhanced search and discovery of Earth and environmental data; Huang and Qiu (2014) analyzed the National Evolutionary Synthesis Center (NESCent), a nonprofit science center dedicated to cross-disciplinary research in the field of evolution**.** During the third phase, researchers started to investigate data curation from more perspectives, for example, user studies, talent fostering and high-quality education, policies and regulations, data curation services and performance evaluation, and data organization and integration.

Research Topics on Data Curation in China

In order to identify hot research topics in the field of data curation, we carried out an analysis of the author supplied keywords from the 258 research papers. There are 492 unique keywords in total, and the top 10 high frequency keywords are shown in Table 1. To further locate hot points, we constructed a co-occurrence network of high-frequency keywords based on their co-occurrence matrix to visualize the co-occurrence relationships among these keywords. As shown in Figure 2, keywords such as "scientific data", "data management", "data curation" and "data sharing" are located in the center of the network, which means that these topics are the most popular topics. Keywords such as "university library", "university", and "e-Science" are important nodes of the network, which means that data curation is closely related to universities' scientific data, and university libraries are the main organizations to provide data curation services. Keywords, such as "data publishing," "data policy," "data lifecycle," "linked data," "ontology" and "RDF" are evenly distributed in the network, which means that these topics are also related to data curation and may become potential hot points in the future.

After the co-occurrence network analysis, we performed a manual analysis to discover the research topics in the field of data curation. The 258 research papers were clustered into 9

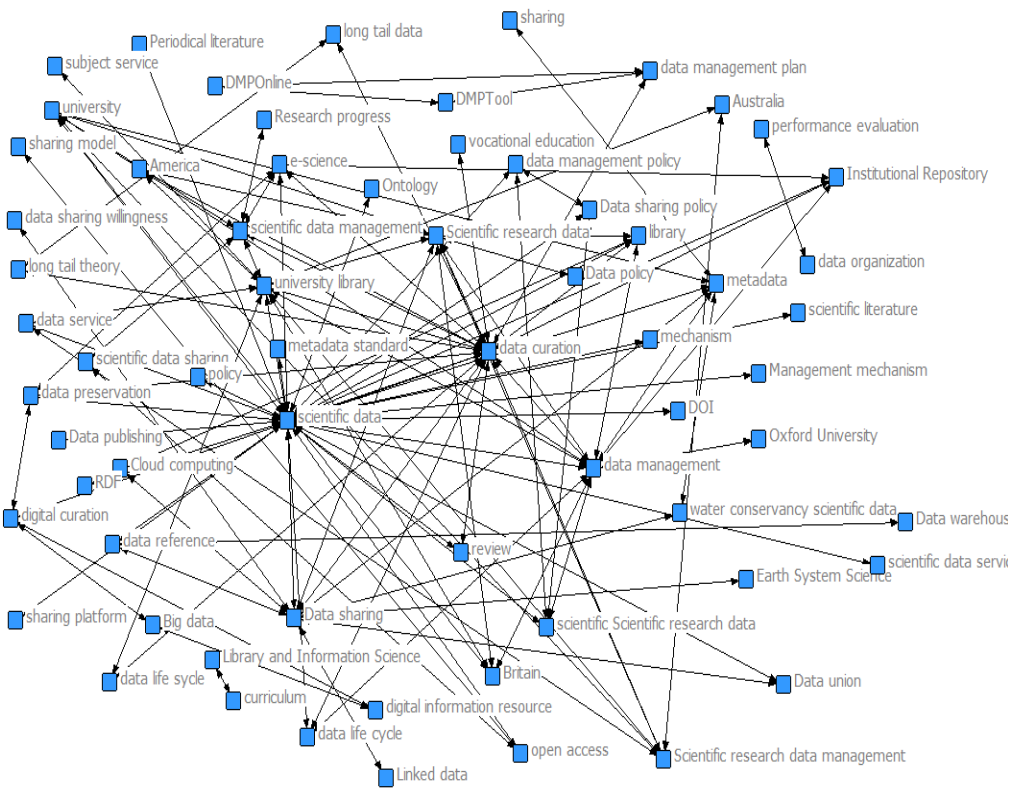

Figure 2. Co-occurrence network of high-frequency keywords in the research papers on data curation

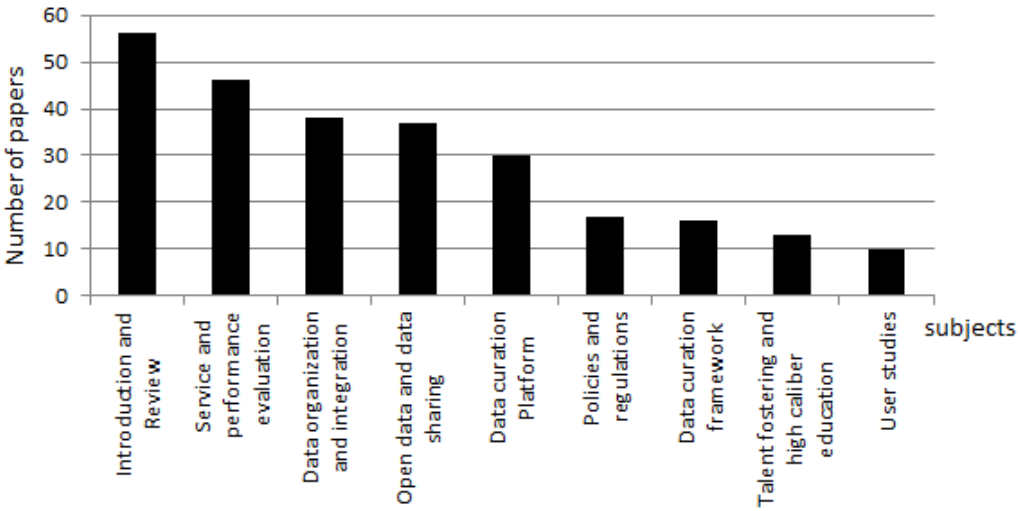

Figure 3. Topic distribution of the research papers on data curation from 2011 till now

groups (as shown in Figure 3). One group includes 56 introduction or review papers which focus on introducing the data curation concept, reviewing the research status in China or other countries, and describing some data curation software tools. The rest of the 8 groups represent the important research topics in this field. We give a detailed introduction to each of them below.

Service and Performance Evaluation

Service innovation is fundamental for libraries to survive in the current data age. In the context of e-science, the depth and breadth of information services in libraries will be further expanded by offering data curation services. Some researchers explored possible service types and service modes of data curation which libraries can provide (Fan, 2014; Li, 2011; Xian, Lai, & Ding, 2013; Xiao & Li, 2012). Furthermore, the performance evaluation of data curation services was discussed, including evaluation metrics, evaluation methods and evaluation process (Dong & Huang, 2008; Si et al., 2014; Ye, 2015).

Data Organization and Integration

The organization and integration of scientific data is very important in data curation services. With the growth of scientific data and the increasing demand for data sharing, some metadata schemas for scientific data have been designed in China, for example, metadata standards for medical and health scientific data, earth system scientific data, and ecological scientific data (Liu, Hu, & Jin, 2008; Standardization Management Committee, 2005; Wu, 2010). With the development and maturity of Semantic Web technologies, there is an obvious tendency to describe and represent scientific data semantically based on ontologies (Ma, Guo, & Wang, 2015; Xu, Wei, & Bi, 2015). However, some problems still need to be solved in annotating different kinds of scientific data, including data entity recognition, data relationship recognition, and automatic annotation (Ding, 2016). Scientific data integration refers to organizing originally discrete, heterogeneous, distributed scientific data and related information resources into an organic whole through logical or physical means to facilitate their management, usage and services. Some studies have been carried out to explore different integration methods, for example, creating semantic links among scientific data, linking scientific data with scientific literature, publishing scientific data as linked data, and integrated publishing of academic papers and scientific data (Fang, 2013; Guo, 2014; Ma, Cao, Wang, Wang & Li, 2014; Si & Li, 2014).

Data Sharing Mechanism

The foundation and preconditions of fully using scientific data and exploiting their value is to open and share these data. Among the impact factors of data sharing, the sharing mechanism is the most important. The traditional sharing mechanism uses all kinds of data sharing platforms. Recently, new sharing mechanisms were also proposed, such as virtual data space sharing model, sharing platform based on cloud computing, and sharing of data journals and data explanatory material in academic journals (e.g., appendices and supplementary information) (Ge, Hu, Liu, Lin & Zuo., 2014; He, Li, & Xu, 2015; Liu & Ma, 2015).

Data Curation Platform

To provide scientific data access, data sharing and data monitoring, it is particularly important to construct a platform for scientific data management services. A data curation platform requires more functions to support the whole process of data curation, rather than simply copying a traditional information system. Some researchers suggested extending institutional repositories to data curation platforms (Liu, 2014; Song & Deng, 2016; Yang, 2012). This way of constructing data curation platforms has three advantages:

1. It can make full use of previous mature software systems and accumulated data assets;
2. Scientific data is the important content of institutional repositories, and it is easy to create

links among scientific data, scientific documents, scientific projects and research scholars;

3. It can provide users one-stop access to scientific data and other institutional knowledge. However, there is still no mature, practical institutional repository system to emerge in China.

Scientific data curation platforms in China are mainly developed by two kinds of organizations: the Chinese Academy of Sciences and its branches, and universities. The scientific data systems developed by the Chinese Academy of Sciences were earlier than the data curation movement, and have accumulated a certain amount of data. However, these systems need to be extended in functions to cater for data curation requirements. The systems developed by universities are founded on newer conceptions of data curation, and can be constructed based on mature ideas and experiences of the data curation movement with the advantage of hindsight. However, these two kinds of systems face problems of data interlinking and integrating with each other.

Policies and Regulations

Policies and regulations related to data curation provide the foundation and guidelines to perform data acquisition, data archiving, data integration and data sharing, to construct data curation facilities and infrastructures, and to provide related services. However, scientific data sharing policies at the national level have not been developed. Existing related policies and regulations in China focus more on traditional library, information and archiving services, rather than on modern data curation activities (Liu & Peng, 2010).

Data Curation Framework

The data curation framework and processes are continuously evolving in tandem with developments in technology and management. A conceptual framework is required to logically represent the relationships among scientific data practices, data characteristics, data curation activities, etc. There is no unified framework in China. Most of the researchers preferred the lifecycle-based models and divided the modules into two categories: core modules and auxiliary modules (Li & Cui, 2015). The former includes requirements evaluation, policies and regulations, data organization, quality control, data storage, and data services, while the later includes personnel organization, user training, and performance evaluation (Chen, 2013; Wu & Chen, 2015). All the modules need to be logically integrated to *be most effective.*

Talent Fostering and High-quality Education

The promotion and development of data curation cannot be separated from the development of high-quality expertise and users. The former refers to the specialists who are engaged in the work related to data curation, including data scientists, data managers, data creators and data librarians. These specialists often grow out of subject librarians. The latter refers to the researchers who are major users of scientific data. By reviewing and analyzing the lessons and experiences of other countries, China's researchers have proposed improvements to library and information science curricula and on-the-job training programs to support the development of data curation (Huang & Ji, 2015; Meng & Qian, 2013a & 2013c; Ye, 2013 & 2013).

User Studies

Scientific researchers are the producers and consumers of scientific data, and thus become the main users of data curation activities. It is necessary and useful to investigate the researchers' data demand, and their intention to acquire and share data. A set of user studies have been done in China to construct a data sharing intention model—identifying the factors that influence the researchers' intention to acquire scientific data, and to understand the researchers' demand for scientific data management services (Hong & Qian, 2014; Hu & Liu, 2013; Zhang, 2013; Zhang, 2015).

Overall, Chinese research on data curation exhibits two obvious characteristics:

1. It integrates previous data sharing theories into a research framework of data curation, and applies of open data, big data and cloud computing technologies to data curation;
2. It regards data curation as a role of libraries, and perceives data curation as an extension of library reference services.

However, data curation research in China is still at an initial stage, mainly digesting and absorbing learning experiences and achievements from other countries (especially Europe and the USA).

DATA CURATION PRACTICES IN MAINLAND CHINA

With the growth of scientific data, government agencies and research institutions increasingly realize the importance of scientific data in scientific discovery and policy making. The Chinese government launched the National Scientific Data Sharing Project in 2002. Afterwards, a number of data management and sharing platforms (systems) were constructed to provide data archiving and data sharing services, for example the National Scientific Data Sharing Platform for Population and Health (National Science & Technology Infrastructure, 2009), the National Data Sharing Infrastructure of Earth System Science (National Science & Technology Infrastructure, 2003), the Geological Scientific Data Sharing Net (Chinese Academy of Geological Sciences, 2004), and the Chinese National Survey Data Archive (National Survey Research Centre, 2003). Some platforms encapsulated embryonic ideas of data curation. We examined four representative platforms and analyzed their data sources, key technologies and functions.

Fudan University Dataverse Network

The Fudan University Dataverse Network (illustrated in Figure 4) is a data exchange and sharing platform developed by the Fudan University Institute for Social Research. The objective of this network is to provide data submission, preservation, management and sharing services for teaching and research in the whole university. It was the first data platform constructed by a Chinese university.

From 2012, the developers started to investigate popular social science data management and sharing systems and categorized them into three types:

- Self-built software: Michigan University's Inter-university Consortium for Political and Social Research
- Open-source software: MIT and HP Labs' DSpace, DuraSpace's Fedora, and Harvard's Dataverse
- Commercial software: Nesstar from Norway.

Figure 4. The homepage of Fudan University Dataverse Network

The Inter-university Consortium for Political and Social Research, a unit within the Institute for Social Research at the University of Michigan, was established in 1962 and developed its own data management system to store, curate, preserve, and provide access to scientific data. However, the system does not have any portability. DSpace and Fedora are both general-purpose digital asset management systems rather than being focused on social science data. Thus, they only provide very simple descriptions of social science research products and research data, and there is no online analysis and online commenting functions which are necessary for social science data. Nesstar is a commercial software owned by the Norwegian Centre for Research Data. In this platform, only internal administrators can publish data whereas external users have no such rights. Furthermore, as Nesstar's source code is not open, it is impossible to do further development on it. Dataverse is an open-source data repository framework to share, preserve, cite, explore and analyze research data. It allows further development and has good portability. Based on the above analysis, the developers finally selected Dataverse as the architecture of the network and signed a cooperation agreement with Harvard University in March 2013. After three months, the Chinese version of Dataverse was released by Fudan University.

Based on Dataverse, the network provides the following functions which can be categorized into four groups (Zhang, Yin, Zhang, Guo & Zhang, 2015):

- Data management: submission, authentication and publication of research data, data-based research products and derived publications, data format conversion etc.
- Data services: data retrieval, data browsing, data download, navigation, statistical analysis and visualization etc.
- Data exchange: data harvest and exchange based on OAI-PMH with the data platforms of Harvard University and University of Michigan;
- Data curation: assign different authorities to different users, assign different access restrictions to different datasets or files, versioning, data preservation etc.

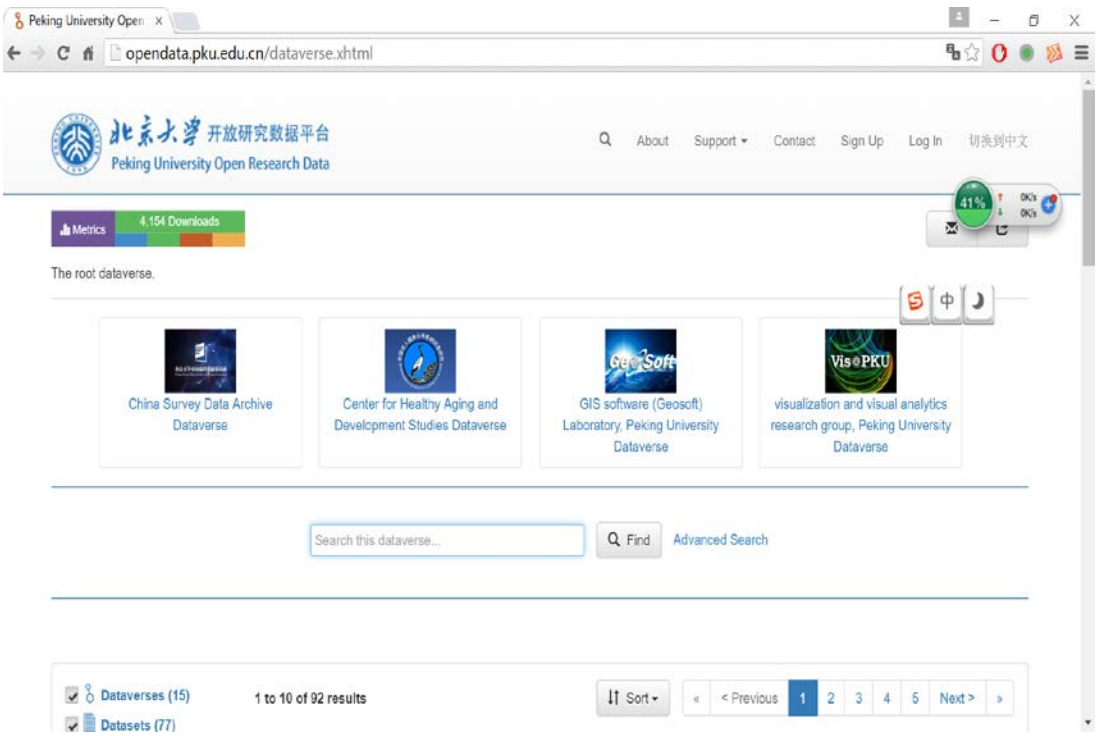

Figure 5. The homepage of the Peking University Open Research Data Platform

The network has published 40 Dataverses, containing 65 datasets, 687 research projects and 1902 data files. Furthermore, based on this network, it is possible to obtain data integration among different datasets and different research projects. For example, Chinese demographic data, consumption data and carbon emission data were incorporated into the Chinese Population Geographic Information System to show the status of energy distribution, flow and carbon emission in different regions of China (Zhang et al., 2015).

Peking University Open Research Data Platform

Peking University (PKU) Open Research Data Platform (illustrated in Figure 5) is a data management platform to provide preservation, management and publication services of research data for the researchers in the university. This platform was mainly built by Peking University Library from 2014, and its demo version was launched online at the end of 2015. Its objectives are to encourage data producers to share data, as well as to satisfy the requirements of data consumers to browse, search and download data, and thus to promote the dissemination, reuse and standardized citation of research data.

At the initial stage of development, four open-source data repository software packages, Dataverse, Data Conservancy (developed mainly by Johns Hopkins University), Comprehensive Knowledge Archive Network (CKAN) (maintained by the Open Knowledge Foundation), and DSpace were reviewed. The evaluation criteria included:

1. standardized metadata and good interoperability;
2. administration rights assignment and flexible access control;
3. dataset publication based on DOI and version
4. online analysis and visualization.

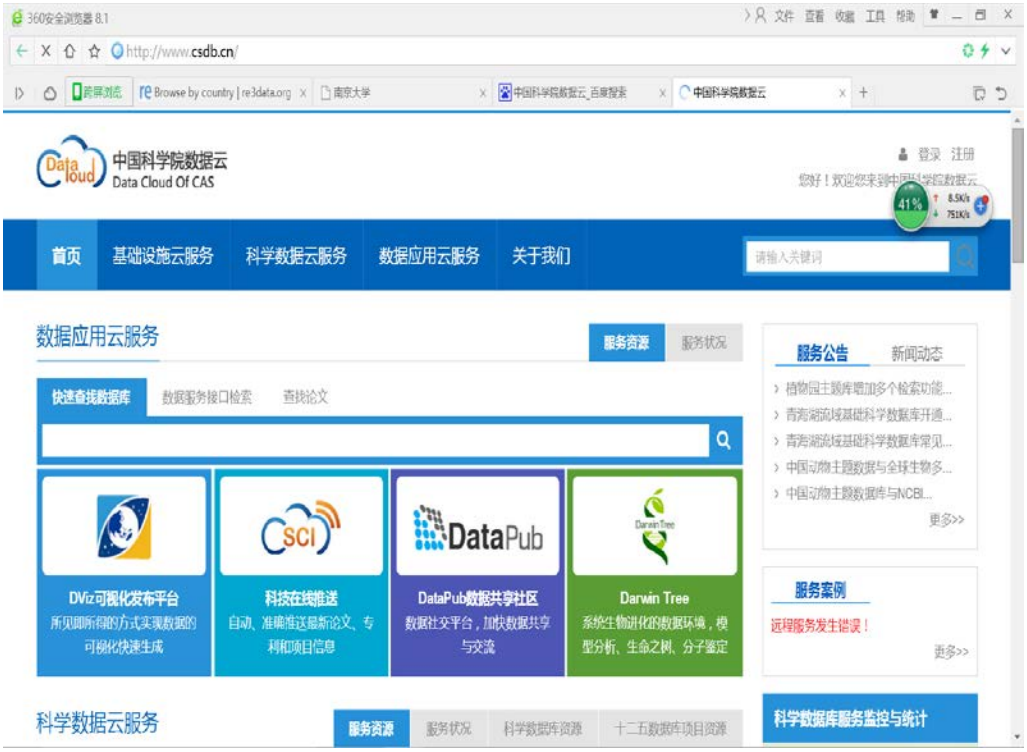

Figure 6. The homepage of Data Cloud of the Chinese Academy of Sciences

Finally, Dataverse was selected as the base of the platform (Luo, Zhu, Cui & Nie, 2016). The main functions of Dataverse include data management, user management, rights management, data index and retrieval, multiple APIs for searching, depositing and accessing data, online analysis, and DOI registration (Luo et al., 2016). In addition, the developers also expanded some new functions, such as user classification, user authentication, customized user groups, statistical analysis of data usage, and Chinese-English bilingual user interface (Luo et al., 2016). This platform has collected 73 open-source datasets and 252 data files from China Survey Data Archive (CSDA), PKU Visualization and Visual Analytics Group, and PKU Bioinformatics Data Center (Luo et al., 2016). However, it is still on its testing stage. The developers are improving the system and collecting more databases.

Data Cloud of the Chinese Academy of Sciences

From 1983, the Chinese Academy of Sciences (CAS) started the construction of scientific databases. Until 2005, 503 scientific databases were constructed, and the total quantity of data has reached 16.6 terabytes**.** In 2008, the project *The Construction and Services of Data Application Environment* was launched, and Data Cloud (illustrated in Figure 6) was the product of this project.

Data Cloud was developed by the Computer Network Information Center of the Chinese Academy of Sciences. It is based on cloud computing architecture and has the following three layers of cloud services (Li et al., 2015):

- *Infrastructure as a Service (IaaS)*: the lowest layer which refers to hardware facilities, such as computer servers, network devices and storage devices, to provide four services:
    - Cloud storage: a distributed storage system which consists of one national-level central node and eleven regional-level sub-central nodes;

  - Cloud computing;
  - Cloud archiving: remote archiving and backup services for scientific data that are not frequently used;
  - Cloud disaster backup: data recovery services within 30 minute.
- *Data as a Service (DaaS) and Platform as a Service (PaaS)*: the middle layer which provides data services and data processing including four services:
  - VDBCloud: used to quickly construct and publish a database;
  - GSCloud: provides various free or commercial geographic datasets, online services of some geographical models (e.g., MODIS and LANDSAT), data integration, and data visualization;
  - Scientific data publication, includes two systems: CSData, a system to support the submission, evaluation, publication and retrieval of data articles associated with scientific data; SciDB, an online data repository to support the storage, preservation, publication, open access and reuse of scientific data;
  - Linking scientific data with scientific literature: using Linked Data principles to create links among scientific literature and scientific data to support knowledge discovery.
- *Software as a Service (SaaS)*: the top layer which provides data applications with software packages:
  - DViz: a platform to provide data visualization service;
  - Datapub: a data sharing and communication platform;
  - DarwinTree: a data analysis platform for biological data.

China Academic Scientific Data Service

The China Academic Scientific Data Service (illustrated in Figure 7) is a sub-project of the China Academic Library & Information System (CALIS). Its objective is to collect distributed scientific data from Chinese universities and provide long-term preservation, and thus support data sharing and reuse.

The service is mainly developed by Wuhan University Library. It is still at its initial stage, and only the Scorpion Species and Toxins Database is available. DSpace was selected as the basic architecture of the service since it has a friendly user interface and it is easy to do further development. The service has the functions of metadata management, data submission, data retrieval, user management and rights control, as well as sequential data analysis (Hong & Xian, 2013).

CONCLUSION

This paper has surveyed data curation research and practice in mainland China from 2001 to 2016. We divided the history of data curation research in China into four stages: open data and data sharing ("data curation to be"), introduction of the data curation concept, digestion and absorption of foreign experiences and achievements of data curation, and the blooming of domestic research from multiple perspectives. Hot research topics include data curation services and performance evaluation, data organization and integration, data sharing mechanism, data curation framework, data curation policies and regulations, talent fostering and high-quality education, user studies, and so on.

In China, the main organizations performing data curation are government agencies and universities. A number of data repositories or data management platforms (systems) have been constructed, often based on open-source software packages (e.g., Dataverse and

Figure 7. The homepage of China Academic Scientific Data Service

DSpace). Some systems have adopted the idea of data curation to support data preservation, data integration, data analysis, data visualization, etc. With the development and popularity of e-science, more and more scholars and organizations have realized the importance of scientific data and have an increasing demand for data curation. However, there is still a big gap between mainland China and well-developed countries. In the Registry of Research Data Repository, only 25 indexed repositories are from China, whereas 797 are from the USA (re3data.org, 2016). Thus, in scientific data curation, China still has a long way to go.

ACKNOWLEDGMENTS

This project is sponsored by the 2016 Qian Lan Project of Jiangsu Province, China and the National Social Science Foundation of China (Grant No. 15ZDB126).

REFERENCES

Beagrie, N., & Pothen, P. (2001). *Digital curation: Digital archives, libraries and e-science seminar*. Retrieved from http://www.ariadne.ac.uk/Issue30/Digital-Curation/

Chen, C. (2004). The mechanism for public access to scientific data: Features, barriers and suggestion for optimization. *China Soft Science*, *2*, 8-13.

Chen, D. (2013). Study on data management service framework in foreign academic universities. *Journal of Academic Libraries*, *31*(6), 10-17.

Chen, Y., Li, X., & Sun, J. (2006). Design of welding scientific data sharing system and metadata research. *Welding & Joining*, *6*, 56-60.

Chinese Academy of Geological Sciences. (2004). *Geological Scientific Data Sharing Net*. Retrieved from http://www.geoscience.cn/

Digital Curation Centre. (2004). *What is digital curation?* Retrieved from http://www.dcc.ac.uk/digital-curation/what-digital-curation

Digital Curation Centre. (2008). *DCC approach to digital curation*. Retrieved from http://twiki.dcc.rl.ac.uk/bin/view/old/dccapproachtocuration

Ding, P. (2016). Semantic-oriented association between scientific data and scientific literature in fine-grained level. *Library Tribune*, *36*(7), 24-33.

Dong, C., & Huang, D. (2008). Fundamental appraisal threads and theoretical support of performance assessment of national scientific data service organizations. *Science of Science and Management of Science and Technology*, *29*(10), 43-46.

Fan, J. (2014). Study on the role of libraries in scientific data management. *Library and Information Service*, *58*(6), 37-41.

Fang, X. (2013). Research on the scientific data organization of university libraries based on linked data. *Library Development*, *10*, 31-34.

Ge, J., Hu, C., Liu, X., Lin, W., & Zuo, H. (2014). Virtual data space sharing model towards domain-specific science data. *Journal of Chinese Computer Systems*, *35*(3), 514-519.

Gray, J., Szalay, A., Thakar, A., Stoughton, C., & Vandenberg, J. (2002). *Online scientific data curation, publication, and archiving*. Retrieved from http://research.microsoft.com/pubs/64568/tr-2002-74.pdf

Guo, X. (2014). Link study of scientific data and scientific literature based on citation analysis. *Information Science*, *32*(4), 59-62,125.

He, H., Li, S., & Xu, C. (2015). Study on data sharing for scientific investigation in the South China Sea based on cloud computing. *Journal of Tropical Oceanography*, *34*(6), 49-56.

Hong, C., & Qian, P. (2014). Investigation and analysis on the scientific data needs and utilization behaviors of graduate students: Taking southeast university for example. *Journal of the National Library of China*, *1*, 16-21.

Hong, Z., & Xian, Y. (2013). Construction of university scientific data management platform based on Dspace: A case study of Scorpion Species and Toxins Database. *Library and Information Service*, 57(6), 39-42, 84.

Hu, Y., & Liu, Y. (2013). Demand analysis on scientific data management in universities based on user survey. *Library and Information Service*, *57*(6), 28-32, 78.

Huang, L., & Li, S. (2006). Design and realization of data management and publication system for oceanologic database of South China Sea. *Journal of Tropical Oceanography*, 25(1), 71-75.

Huang, R., & Ji, C. (2015). Current situation and enlightenment of the data curation education in University of Illinois at Urbana-Champaign. *Library & Information*, *1*, 61-65.

Huang, R., & Qiu, C. (2014). Metadata management of dryad repository. *Library Journal*, *1*, 68-73.

Jin, H., & Zhang, Q. (2007). Study on designing forging & stamping data-sharing system. *Forging & Stamping Technology*, *32*(5), 118-121.

Li, J., Yu, L., Zhang, B., & Wu, Z. (2015). Analysis on the science data cloud of Chinese Academy of Sciences. *China Education Network*, *10*, 33-34.

Li, J., Shen, W., & Wang, G. (2006). Meteorological information sharing service platform and its key technologies. *Journal of Applied Meteorological Science*, *17*(5), 621-628.

Li, W., & Cui, Y. (2015). The research data management lifecycle model and application in service evaluation. *Information Studies: Theory & Application*, *38*(9), 38-41.

Li, X. (2011). Research data management and service pattern in libraries. *Journal of Library Science in China*, *37*(5), 46-52.

Liu, D., & Yin, L. (2006). Medical science data sharing practice in China. *Computer Applications and Software*, *23*(10), 1-3.

Liu, J. (2014). The study on institutional repositories of universities based on data curation. *Library Work and Study*, *8*, 109-112.

Liu, J., & Ma, J. (2015). Three approaches of opening and sharing scientific data. *Journal of Intelligence*, *34*(10), 146-150.

Liu, L., Hu, K., & Jin, S. (2008). The study about metadata specification of health information dataset. *Chinese Journal of Health Statistics*, *25*(4), 363-365.

Liu, R., Cai, J., Peng, K., Shan, X., Tian, L., Pang, L., & Zang, A. (2007). The project of seismological data sharing. *Earthquake*, *27*(2), 9-16. Retrieved from http://en.cnki.com.cn/Article_en/CJFDTOTAL-DIZN200702001.htm

Liu, R., & Peng, J. (2010). Construction and prospect of scientific data sharing policy and regulation in china. *Science and Technology Management Research*, *30*(13), 40-43.

Lord, P., & Macdonald, A. (2003) *e-Science curation report: Data curation for e-Science in the UK: An audit to establish requirements for future curation and provision.* Retrieved from http://webarchive.nationalarchives.gov.uk/20140702233839/http:/www.jisc.ac.uk/media/documents/programmes/preservation/e-sciencereportfinal.pdf

Luo, P., Zhu, L., Cui. H., & Nie, H. (2016). The construction of Peking University Open Research Data Platform based on Dataverse. *Library and Information Service*, *60*(3), 52-58. Retrieved from http://en.cnki.com.cn/Article_en/CJFDTotal-TSQB201603012.htm

Ma, Y., Guo J., & Wang F. (2015). Study on scientific data semantic organization model and framework under e-science circumstance. *New Technology of Library and Information Service*, *Z1*, 48-56.

Ma, J., Cao, Y., Wang, S., Wang, N. & Li, H. (2014). Studies in the integrated publishing of academic papers and scientific data. *Information and Documentation Services*, *2*, 82-86. Retrieved from http://en.cnki.com.cn/Article_en/CJFDTotal-QBZL201402019.htm

Meng, X., & Qian, P. (2013a). A review of overseas practice and research of data curation education. *Journal of Library Science in China*, *39*(6), 63-74.

Meng, X., & Qian. P. (2013b). International experiences and references of university social science data management: Taking UKDA and ICPSR for example. *Information and Documentation Services*, *34*(2), 77-80.

Meng, X., & Qian, P. (2013c). Position setting and management mechanism of data librarian in academic library at abroad. *Library & Information*, *4*, 12-17.

National Survey Research Center at Renmin University of China. (2003). *Chinese National Survey Data Archive*. Retrieved from http://www.cnsda.org/

National Science & Technology Infrastructure. (2003). *National data sharing infrastructure of earth system science.* Retrieved from http://www.geodata.cn/

National Science & Technology Infrastructure. (2009). *National scientific data sharing platform for population and health*. Retrieved from http://www.ncmi.cn/1

re3data.org. (2016). re3data.org: Registry of Research Data Repository. Retrieved from http://www.re3data.org/

Si, L., & Li, X. (2014). Research on the scientific data integration and sharing based on linked data: An example of BIO2RDF project. *Research on Library Science*, *21*, 51-55.

Si, L., Li, Y., Zhuang, X., Xing, W., Hua, X., Li, X., & Xin, J. (2014). An empirical study on the performance evaluation of scientific data sharing platforms in China. *Library Theory and Practice*, (9), 30-35.

Song, X., & Deng, Z. (2016). The study on institutional repository based on data curation. *Research on Library Science*, *2*, 44-48.

Standardization Management Committee. (2005). *Metadata Standard for Geography Data* (Gb/T 19710-2005).

Walters, T., & Skinner, K. (2011). *New roles for new times: Digital curation for preservation*. Washington, DC: Association of Research Libraries.

Wang, F., & Shen, J. (2014). Advances in data curation abroad: Research and practice. *Journal of Library Science in China*, *40*(4), 116-128.

Wang, Q. (2014). Review of Data Curation Research in China. *Information and Documentation Services*, *5*, 33-38.

Wu, B. (2010). Study on ecological metadata and its standard. *Journal of Central South University of Forestry & Technology*, *30*(12), 75-79.

Wu, J., & Chen, Y. (2015). Research on scientific data curation system oriented to the fourth paradigm of science. *Library and Information Service*, *59*(16), 11-17.

Xia, Y. (2013). The best practices of foreign data curation and related revelations. *Journal of Information Resources Management*, *1*, 53-58.

Xian, Y., Lai, J., & Ding, N. (2013). Research on practice of university library scientific data management service: An example of Wuhan University Social Scientific Data Management. *Information Studies: Theory & Application*, *36*(12), 89-93.

Xiao, X., & Li, J. (2012). Research on embedded scientific data service in library. *Research on Library Science*, *21*, 85-89.

Xu, K., Wei, X., & Bi, Q. (2015). Study on scientific data ontology construction. *Library and Information Service*, *59* (17), 120-126.

Xu, X., Liu, T., & Yu, S. (2014). Revelation of data one project for data curation in China. *Library & Information*, *6*, 109-116.

Yakel, E. (2007). Archives and manuscript digital curation. *OCLC Systems & Services: International Digital Library Perspective*, *23*(4), 335-340.

Yang, G. (2016). Research hotspots and progress of data curation in China. *Library and Information Service*, *60*(2), 129-136.

Yang, H. (2011). Data curation: A new development of university libraries in the U.S. *Journal of Academic Libraries*, *29*(2), 18-21, 41.

Yang, H. (2012). Data curation and new thoughts in developing institutional repositories in U.S. Universities: The Datastar Example. *Journal of Academic Libraries*, 30(2), 23-28.

Ye, L. (2013). Data curation education and career development at abroad. *Journal of Academic Libraries*, *3*, 22-28.

Ye, L. (2015). Review and analysis of capability maturity model for research data management. *Document, Information & Knowledge*, *2*, 115-123.

Ye, L. (2013). The setting up and requirement of data curation positions in libraries at abroad. *Journal of Academic Libraries*, *5*, 5-12.

Zhang, J. (2013). Intention to acquire scientific data of college scientific researchers. *Journal of Intelligence*, *32*(6), 70-75.

Zhang, J. (2015). The study on university library scientific data service based on user's requirement. *Research on Library Science*, *19*, 74-79.

Zhang, J., Yin, S., Zhang, Y., Guo, Y., & Zhang.Y. (2015). Social scientific data sharing and serving - an example of Fudan University social scientific data platform. *Journal of Academic Libraries*, *1*, 74-79.

Zorich, D.M. (1995). Data management: managing electronic information: data curation in museums. *Museum Management and Curatorship*, *14*(4), 430-432.