The Electronic Library

Research data management in universities of central China: Practices at Wuhan University Library

Xia Liu Ning Ding

Received 24 April 2015
Revised 4 September 2015
Accepted 5 October 2015

Research data management in universities of central China

Practices at Wuhan University Library

Xia Liu and Ning Ding
Wuhan University, Wuhan, China

Abstract

Purpose – Revealing research data's production and use, the status of research data management (RDM) and researchers' service requirements in universities of Central China; this study aims to investigate the feasibility of university libraries in providing RDM services without any supporting policies from governments or funding agencies.

Design/methodology/approach – Using a stratified sampling method, faculties and graduate students from 11 universities were investigated. Four pilot subjects at Wuhan University (WHU) were chosen for whom a pilot RDM platform was to be constructed.

Findings – Research data at Chinese universities are small, sporadic and discontinuous. Such data are intensively or dispersedly under researcher's management, with some unresolved problems regarding data security, data sharing and utilisation efficiency. Researchers' needs for data services are strong. University libraries in China can develop RDM systems and provide related services. To realise this, more work should be done on service mechanism, service promotion, software development and staff training.

Research limitations/implications – The user survey covered 11 universities in central China, which may not reveal the real RDM status of researcher in different areas of China.

Practical implications – The practice at WHU could provide reference to other university libraries in China or other developing countries.

Social implications – The practice at WHU could provide reference to other university libraries in China or other developing countries.

Originality/value – The user survey is designed to be as comprehensive as possible and cover 902 researchers from 11 different types of Chinese universities. The practice at WHU is one of the first RDM initiatives led by university library in China.

Keywords China, Data management, Research data, Pilot study, University library, User investigation

Paper type Research paper

Introduction

Recently, research data management (RDM) has attracted the attention of higher education institutes (HEI) internationally. University libraries in the USA, Britain, Australia and other countries are increasingly involved in data management services. Universities play an important role in scientific research in China. From 2005 to 2010, HEI in China held over 50 per cent of the national major scientific research projects and more than 80 per cent of research awards from the National Science Foundation of China

The research is supported by the China Academic Library and Information System, the biggest library consortia in Chinese higher education.

The Electronic Library
Vol. 34 No. 5, 2016
pp. 808-822

0264-0473
DOI 10.1108/EL-04-2015-0063

(MOE, 2015). Research data, which can be a substantial amount, are produced or obtained in the process of research through laboratory experiments, social investigations, field observations, internet mining and in other ways. However, governments or funding agencies have implemented few policies for RDM until recently. Hence, it is necessary to investigate how to utilise the existing achievements in RDM to promote data submission, data management and data sharing in Chinese universities.

Wuhan University (WHU) is a comprehensive and key national university with full support for construction and development from the central and local Government of China. WHU offers 119 undergraduate programmes, 347 master's degree programmes, and 293 PhD programmes, which cover almost all disciplines from natural sciences to humanities and social sciences. There are 3,639 full-time faculty members and over 50,000 full-time students, with half of new enrolments being post-graduates. The university receives significant funding from the government and companies, with the total funding for scientific research accounting for approximately 30 per cent of WHU's revenue since 2004 (WHU, 2015).

In 2011, with the support of the China Academic Library and Information System (CALIS), Wuhan University Library (WUL) started a study on RDM platforms and service mechanisms in Chinese HEI, aiming to explore the following three aspects: to reveal the production of research data, researchers' attitudes and behaviour regarding data management and their requirements for data platforms and services in Chinese universities; to construct a pilot data management platform for multiple disciplines at WHU; and to provide and promote data services for researchers at the university without any supporting policies from the government, funding agencies or the university itself. This study describes the user investigation findings from 11 universities and articulates the RDM practices at WHU. Several problems and suggestions regarding service mechanism, software development, service promotion and staff training are analysed and summarised to provide reference for other university libraries.

Literature review

University user surveys

A greater understanding of researchers' needs will better inform the approach and nature of data services offered to researchers by the library. Many related studies have been conducted at universities internationally. Some studies were conducted via online questionnaires to obtain a wide-scale response and draw broad conclusions, whereas others were conducted by focus interviews to acquire insight into researchers' data curation behaviour.

Scaramozzino *et al.* (2012) conducted a survey at California Polytechnic State University, with 82 participants who were selected from 131 eligible faculty members, to provide insight into current data curation practices for Small Science, including researchers' data curation awareness; behaviour; attitudes towards and needs for services; and education. To explore the nature of data sets and researcher practices and needs for data curation, a research team at Colorado State University conducted five 90-min focus groups with 31 self-selected faculty members, research scientists and research associates (McLure *et al.*, 2014). Beagrie *et al.* (2009) conducted a survey using an online questionnaire combined with in-depth qualitative interviews to explore

researchers' views on and practices for preservation and dissemination of research data at four UK universities.

Regarding Chinese researcher data curation behaviour, a few studies have been published in Chinese papers. Zhang (2013a, 2013b) conducted an online questionnaire survey to investigate HEI researchers' intentions to acquire and share research data. The findings of 151 participants are as follows: intentions of researchers to acquire scientific data are directly affected by their perceived usefulness, knowledge growth, trust and their willingness for data sharing is affected by their perception of self-worth, anticipated reciprocity, interpersonal trust, and image (Zhang, 2013a, 2013b). Zhang (2013c) conducted a questionnaire survey of 50 researchers from three universities in Jilin City. He found that most of the participants acknowledge the importance of data curation and are willing to share their data whether the data are preserved personally in electronic or paper formats. Hong and Qian (2014) conducted a questionnaire survey on all the steps of RDM to recognise the data needs and utilisation behaviour of graduate students. There were 265 participants from Southeast University, and their needs and behaviour varied with their disciplines. Most of them hoped to receive data services from the library, although they knew little about the library's data services.

Previous studies on RDM activities in China ranged from attitudes on data acquiring and sharing to managing behaviour on practical data. However, these studies often have low response rates, thus making it difficult to generalise the conclusions. Hence, a more systematic survey with more participants from various disciplines and universities is necessary.

Practices at university libraries

It is a common perspective in library and information science (LIS) that libraries and librarians are well suited for work in data curation because of "their long stewardship of collections and support of research and scholarly communication; their more recent implementation of institutional repositories; and their expertise", Steinhart notes, in "archival practices, cataloguing and indexing, development of platforms for discovery and distribution, and education and user support" (as cited in McLure *et al.*, 2014, p. 140). From a global perspective, many university libraries have provided services for data management to users on and/or outside campus. Among the 87 libraries of the top 100 universities investigated, 50 libraries offered research data services, including research data introduction (provided by 47.1 per cent of the libraries), data curation and storage service (43.7 per cent), data management guidelines (42.5 per cent), data management reference (41.4 per cent), resource recommendation (41.4 per cent) and data management training (24.1 per cent) (Si *et al.*, 2015). RDM platform services are high-cost activities; some are provided by libraries themselves, such as DSpace@Cambridge (UCL, 2015); some are cooperative services through library and university IT departments, such as ScholarSphere (PSU, 2015); and others are provided by multiple institutes from in or outside the university, which are still led by the library, such as the DataConservancy (DC, 2015).

Concerning Chinese universities, some research data are externally available via the internet; however, they are mainly provided by different types of centres for social sciences, such as the Centre for Social Survey of Sun Yat-sen University, the Institute of Social Science Survey of Peking University, the Wuhan University Institute of Quality Development Strategy and the National Survey Research Centre at Renmin University

of China. Most of these universities list the survey data on their home page and provide download options for the public (CSS, 2015; ISSS, 2015; WHUIQDS, 2015); however, Renmin University of China developed the Chinese Social Survey Data Archive (NSRC, 2015).

In 2011, Fudan University established a scholarly institution called the Fudan Institute of Social Research (FISR, 2015) at the university level to realise the unified planning, managing, accessing and sharing of social science data. Moreover, the Fudan University Library may be the only university library in China that is involved in data management activities as an important member. Nonetheless, no interdisciplinary platforms have been developed to preserve and discover different types of research data at the university, whether provided by the library or other institutes. It seems that libraries are more sensitive to the importance of RDM than that of funding agencies or university administrators in China. Under these circumstances, it would be beneficial for the library to develop a suitable platform for a variety of research data at the university and provide services for data management.

User investigation

Methodology

To reveal the exact production and use of research data, the status of data management and the requirements for researchers in Chinese universities, a questionnaire was designed to include 19 questions, divided into three parts. The first part (four questions) includes the basic status of research data, such as data sources, sizes of data files and the frequency of data produced or obtained. The second part (13 questions) includes the cognition of data management and the attitude regarding data sharing; technical means, tools and standards used to record, preserve and describe research data; data storage life; the methods and persons responsible for data management in the project team; the probability of data loss; whether to retain background information related to data; whether to share data with others, etc. The last part (two questions) concerns researcher expectations regarding data services and the needs for data management platforms. Using a stratified sampling method, faculty members and graduate students were selected with a 1:3 ratio from 11 different types of universities in Wuhan City, and questionnaires were distributed through face-to-face meetings and via e-mail. Finally, from November 2011 to March 2012, 1,200 questionnaires were sent out, and 902 valid forms were returned.

Sample analysis

From the perspective of age, 81 per cent of the participants are graduate students or faculty members aged below 35 and 17 per cent are aged between 36 and 50, with only 2 per cent of the participants above the age of 50. Regarding education level, more than 86 per cent of the participants are researchers with master's or doctoral degrees. The results of both the groups are identical in terms of expectation. In terms of professional titles, faculty members with senior professional titles account for 60.7 per cent of the total 242 faculties. Additionally, the subject distribution indicates that participants in social sciences, natural sciences and engineering are more than 20 per cent each and those in information sciences, medical sciences, agricultural sciences and humanities as fewer (12, 10, 4 and 3 per cent, respectively). This is because of the lack of balance in the distribution of the subject distribution at the sample universities. Considering the

differences in data quantity and the frequency of different disciplines, the subject distribution was controlled when distributing or sending the questionnaire.

Findings

Among all the 19 questions, 9 are single-selection questions and 10 are multi-selection questions. All selections for each question from the 902 answer sheets were recorded and compiled in Microsoft Excel. The following three findings depict the results.

Research data at universities: small, sporadic and discontinuous

The results show that research data at universities are acquired from different sources (Figure 1), with 29 per cent from controlled laboratory experiments, 26 per cent from internet mining, 15 per cent from social investigations and 13 per cent from field work. Regarding the frequency of data production (Figure 2), 14 per cent of researchers are involved daily, 37 per cent are involved weekly, 26 per cent are involved monthly and 15 per cent are involved quarterly, similar to the data acquirement frequency from outside the research group. The sizes of the data file also vary (Figure 2), with 34 per cent between 1 and 100 MB and 26 per cent between 100 MB and 1 GB. Compared with Big Science fields, such as topography and astronomy, most subjects at universities are small, with almost 70 per cent less than 1 GB (Figure 3).

According to the survey, research data are complex in source, and their generation or attainment is discontinuous. Differences in source will cause differences in data progress and formats, leading to differences in data collecting, data organising and data preserving. Moreover, many research teams will not collate or produce certain data

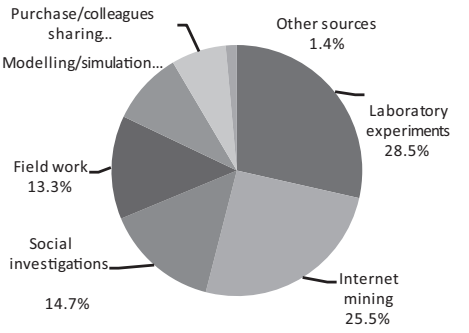

Figure 1.
Data sources

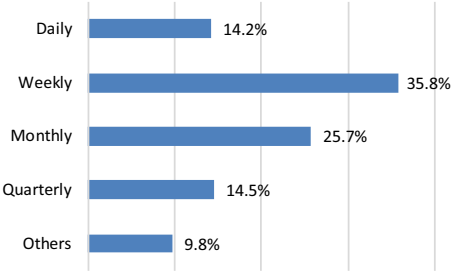

Figure 2.
Frequency of data production

persistently, unless they have a long-term funded project. Therefore, the continuity of data and research is challenged.

Research data are under management to a certain extent: researchers' data literacy needs improvement

The results show that some researchers have sound knowledge of data management and that most of the research data have been under management by some means or other. However, many researchers do not have sufficient data literacy, which leads to some unresolved issues regarding data security, data sharing, efficiency of data utilisation, etc.

As previously mentioned, research data at universities tend to be small in size, which makes it easy for researchers to manage and store data by themselves. In the survey, 70.2 per cent of the data are recorded in an electronic format, whereas 29.8 per cent are recorded on paper or other formats (Figure 4). Figure 5 shows that 82 per cent of researchers store their data on PCs (45.5 per cent) or mobile memory devices (36.5 per cent), and more than 65 per cent of researchers have lost their data once, indicating that most researchers lack knowledge about storing data appropriately and safely. Research data at universities are dispersive (Figure 6): 39.1 per cent of data are saved separately by different researchers and post-graduates, whereas post-graduates often have high mobility, so they are not effective in ensuring the integrity and continuity of data. Meanwhile, more than 50 per cent of the researchers preserve their data only for two years and never consider permanent preservation (Figure 7), resulting in data deletion or loss after a certain period of time, which can prove ineffective for subsequent research. The background information of 73 per cent of research data cannot be accessed, and 14 per cent of researchers never record the background information that is a requisite part of research data to help the researcher understand former research.

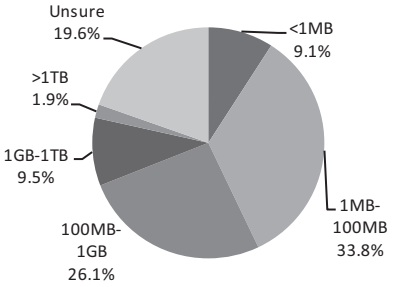

Figure 3.
Data size

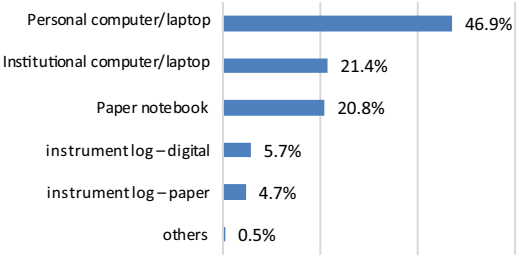

Figure 4.
Data recording mode

Regarding data sharing, most researchers declare that they are willing to share data with others for free (49.6 per cent) or for a price (32.9 per cent), and 60.1 per cent have actually shared data. Common or disciplinary standards and software/platforms to describe, organise and manage data to improve data sharing and academic communication should be used. However, 50 per cent of researchers use their own standards to describe their data (Figure 8), which makes it difficult to integrate such standards into one platform or share them with other research teams. Regarding management systems, 32.3 per cent of researchers select commonly used software, such as SPSS, Access, SQL Server and Oracle, and 21.8 per cent use disciplinary platforms, whereas the remaining 45.9 per cent use their own software/platforms, which also presents a technical barrier to data sharing.

Although many issues and shortcomings emerge from current data management, 60 per cent of researchers are still satisfied with their current skills and methods for data

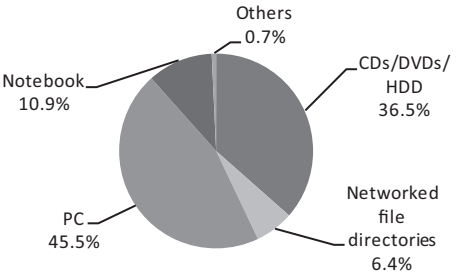

Figure 5.
Storage medium

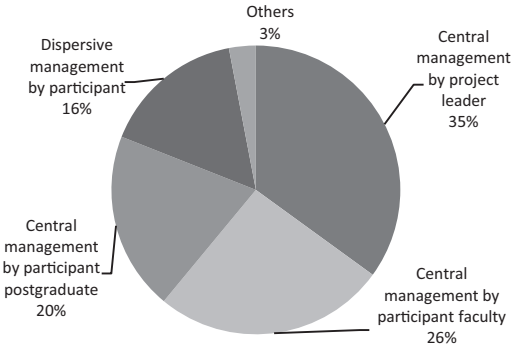

Figure 6.
Supervisor mode

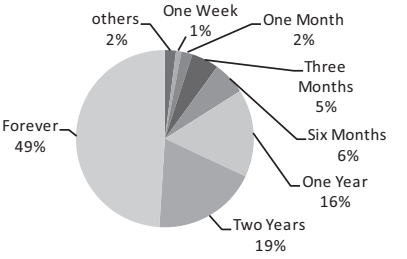

Figure 7.
Storage life

management; furthermore, 47 per cent of researchers disagree or are unaware that data management can promote the research itself. Measures need to be taken to improve the data literacy of researchers.

Strong need for data services and data management platforms

The questionnaire provides several options for data services; additionally, it offers some choices for the functions of data management platforms that allow multiple selection. The results show that every option provided was chosen at a high percentage (Figures 9 and 10). A total of 61.5 per cent of researchers need a library to develop a data management platform; 57.2 per cent need data management consultation services; and 69.8 per cent of researchers think that browsing, searching and downloading data are the most important functions that they need. These are basic functions for a data management platform. Some advanced functions, such as data integrity, analysis and reuse, are not currently in strong demand.

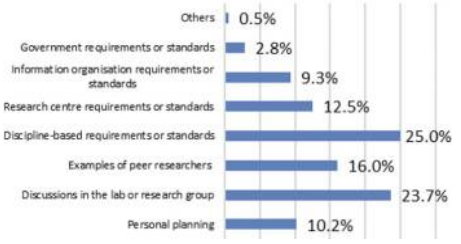

Figure 8.
Data description

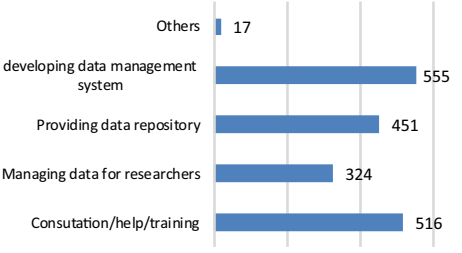

Figure 9.
Data service needed

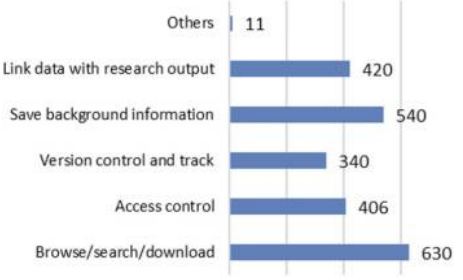

Figure 10.
Function needed for RDM

Practices at Wuhan University Library

Based on this large-scale investigation, several pilot subjects were examined at WHU for which a pilot RDM platform was to be constructed, and data service was provided and promoted to researchers at the university. The practices at WUL can be summarised in four steps.

Formation of the project team

To receive more support from the university's research management institutes, as well as more professional advice from the scholars of LIS, the project leader/library director invited various experts from different institutions of WHU to comprise the project team. These experts included two associate deans from two research management institutes for Natural Sciences and Social Sciences and Humanities, respectively, one professor from the School of Information Management (SIM) and librarians from different departments of WUL. Additionally, a Chinese professor who is an expert in data curation from Syracuse University in the USA was invited as a consultant. There are three working groups within WUL. The metadata group is responsible for developing the metadata for research data. The technical support group is concerned with software selection, local system settings and development. The subject librarians group is responsible for communicating with researchers in different schools, investigating user needs, seeking pilot subjects and working with researchers on data management.

The project team proved to be successful. The two associate deans from the research management institutes supported this project by leading subject librarians in their interviews of faculty members, helping pilot subject selection and holding related seminars on data service promoting. The professor from SIM provided suggestions for platform construction, and the expert from the USA provided suggestions on questionnaire design, software selection and metadata design. Library staffs from different departments formed the core of the project. It was easy to expand the project team to cover more related departments or staff when needed; furthermore, it was easy for the project team to combine project activities with their daily work.

Pilot selection

Based on the above survey, the principles for pilot subject selection could be determined, namely, a preference for teams/schools with a larger quantity or higher frequency of data production, teams/schools with a stronger demand for RDM and younger researchers who need more help from the library. Subjects that already had national or international disciplinary data sharing platforms or required higher data security were not considered.

According to these principles, the project team screened potential faculty members and research teams. The subject librarians then visited each group or individual personally. The upcoming platform was defined as a pilot service and was clearly explained to the researchers. After several focused interviews, four pilots were chosen: the "Scorpion toxin project" from the Life Science College, the "Social investigation programs" from the Department of Sociology, the "Transmission dynamic of micro blog project" from the Information Management School and the "User investigation programs" from WUL itself.

Process of platform construction

User requirements analysis. User requirements analysis includes two aspects: one is the target analysis and content analysis of the data management system and the other is the requirement analysis for system functions. According to the feedback from several interviews with the pilot teams, the target and basic functions substantially agree, namely, to describe and index data/data sets according to the characteristics of their subjects; to effectively store, manage and release data; to grade management and access control to metadata and data/data sets; and to provide the services of browsing, searching and downloading via the internet available throughout the day.

Because of the differences in the disciplines of the teams, data set types, quantity, size and formats and system requirements are not identical, special functions will need to be designed according to different requirements. Take the "Scorpion toxin project", as an example; besides browsing, retrieval, uploads and downloads of species data of scorpions, genetic data and genetic protein data, BLAST needs to be embedded into the system to realise the comparison between protein sequences and nucleotide sequences.

Metadata design

There are some existing metadata standards that could be used to document research data and several discipline-specific standards that have been used. The metadata system includes common standards to describe the common characteristics of data/data sets and the specialised standards for disciplinary data. The three steps for metadata design are as follows:

(1) The first step is differentiating the possible data types and formats for each pilot. In the scorpion resources programme, for example, there are image data (images of scorpion species), sequence data (scorpion protein and nucleic acid sequence measurement data) and text data (research project information and relevant literature).

(2) The second step is to determine connections between the data of each pilot subject and to design the metadata structure.

(3) The third step is to design metadata schemes for each type of data.

The objects of RDM platforms are data sets with *polytype* digital content: these are experimental, observational and statistical data; descriptive information of programmes; and references for publications in different data sets. The following three types of metadata were defined: collection description; general and disciplinary metadata, which is used to describe data collections, projects and literature; and digital objects. General metadata are based on Dublin Core, whereas social investigation, species and sequence data are disciplinary data, referring to existing, normative and admissive schemes. For example, considering that the scorpion resources group has uploaded a part of their data to GenBank, the GenBank metadata system is used to describe the gene and protein sequence data to ensure availability, openness and matching of the former data. However, a specific standard was designed for the scorpion species data, based on the investigation of existing species description standards, which can also be used to describe other species data.

Function design and system selection. The RDM platform of WHU is designed as a two-level distributed system: a one-stop data sharing portal and a disciplinary RDM system. The disciplinary RDM system includes three modules: data display, data

management and user management interface. The data display interface should satisfy the demands of browsing, searching, downloading data and other advanced functions required for different disciplines. The data management interface is for team members to submit and manage their data. The user management interface is responsible for ensuring that different users and data are within corresponding rights. The data sharing portal is to realise a one-stop retrieval function, which can integrate metadata from disciplinary RDM systems. Data could be integrated into this portal by metadata harvesting or by establishing a sub-disciplinary system directly under it. Full-text and digital objects are directly obtained from the sub-disciplinary systems.

After defining system structure and function, system selection is defined next. As different disciplines may have different demands, a diverse and flexible software is needed, which could be easily developed secondarily. This software must be able to process multidisciplinary data of several types and formats, and the input cost must also be considered. After comparing DSpace, Fedora, EPrints, Plone, Dataverse and Digital Commons, DSpace was found to be most suitable for preserving and sharing almost every type of data. Additionally, some secondary developments could be made depending on demands. Chinese language support, Chinese retrieval support, data submission interface and metadata management are realised by local parameter settings. Personalised interface design, special data submission, data retrieval and display and unified authentication must be developed secondarily.

User evaluation and feedback. The pilot projects benefited from communication with researchers who were satisfied with the interface and functions of the platform. They agree that the platform has provided a suitable place to store data and that it solved some security problems caused by decentralised management. Furthermore, the researchers agreed that the platform is beneficial for data reuse, publication of research findings and the promotion of science communication and corporation. Some issues, such as durative data submission and data analysis, still persist. The scorpion database, for example, requires professional knowledge to organise its data and needs researchers to sort through the data manually, which is very time-consuming. Without data sharing policies and direct benefits, no strong motivation exists for researchers to invest sufficient time to organise data files and submit them to the system. Moreover, the platform can only provide basic functions. The functions for data reuse and analysis would be the next step.

Research data management services. RDM services at WHU started along with the project. Lectures on RDM were organised, which publicised the importance of data management to researchers by visiting and interviewing them. This is currently ongoing. After completing the platform construction, the library and research management institutions jointly organised two seminars for researchers to introduce the platform and obtain feedback from them. Data services that the library could provide were also determined: consultation and training of data management skills for researchers, developing disciplinary RDM systems, helping researchers manage their data and investigating medium- to long-term preservation services for research data. Simultaneously, data services have been changed from project development to library daily services, with the system department being responsible for technical support and platform supervision, the cataloguing department being responsible for metadata design and quality control and the reference department being responsible for consulting, training and promotion.

Discussion

The success of the pilot RDM platform constructed at WHU proves that research libraries have the ability to develop RDM systems and provide data services for their patrons. However, some problems have emerged during the system development and data service promotion. To provide more efficient, widely received and tailored data services, more measures should be implemented regarding service mechanism, services promotion, software development and staff training.

The establishment of service mechanism on research data management

A successful RDM first needs to be planned at the university level to define the RDM process and the responsibilities of all related units to ensure normal operation and sustainable development of data management (Jones *et al.*, 2013). WUL's practice only involved cooperation on the project level. Data management is not included in the university's research management plan, and no related policies were established by the university. Therefore, there is a lack of ability to stimulate researchers to submit and share data and to maintain long-term cooperation with relative institutions. Likewise, there is a lack of financial and institutional guarantees for sustainable system developments.

To achieve these goals, it is necessary to demonstrate the importance of RDM and open access to the university administrators and senior managers. Scholarly communications are evolving into a natively digital system, and free online availability has proven to substantially increase the impact of research (Lawrence, 2001), which can benefit both the individual researcher and his/her institution. It would be a new role for universities in management, preservation and stewardship of digital materials and would establish institutional repositories as new university-based publishing platforms (Brown, 2007). In this respect, Fudan University provides a good example. However, for universities without such a scholarly entity, a feasible solution could be referencing the practice of legal publication deposits in many Chinese universities, that is, headed by libraries or research management institutions combined with other relative departments, establishing a university policy to regulate and clarify the responsibility of the relative stakeholders and the progress of data management.

Further stimulation for user requirements

According to the investigation, despite many emerging and potential problems, 60 per cent of participants were satisfied with their existing RDM methods and more than 40 per cent of researchers had no clear understanding that data management can play an active role in research. Furthermore, regarding data sharing, although over 80 per cent of participants expressed a positive attitude, most researchers, in fact, are positive in principle but almost universally reluctant in practice (Ward and Freiman, 2011).

Unlike developed countries, almost no specific requirements exist for RDM from the Chinese Government or funding agencies, except for some major national projects of science and technology, which would be the primary and external driving forces for researchers to manage research data. Concerning interior requirements, publishing in international scholarly journals and improving scholarly impact are the most important things for Chinese researchers under the existing research evaluation mechanism. Providing relative support services will be appreciated by researchers, such as a

platform to release their related data as supporting material for publication and improving their scholarly impact through open access of research data.

According to the feedback from several seminars and interviews, most faculty members are still in an "eliminating conceptual illiteracy" and "uncertain demand" period. Therefore, libraries need to consider how to effectively communicate with researchers, involve themselves in research activities and establish a complete solution for understanding, accessing and providing data services. Some important findings by Ward and Freiman (2011) provided some guiding principles. They noted from conversations with researchers that most of them do not understand the process of digital curation and that they are not familiar with terms such as *preservation* or *digital repository*. Additionally, it is better for data curation to start early in the research life cycle and in researchers' careers to help researchers access the data when they need it and to tailor support to help researchers make informed choices.

An example of cultivating and stimulating users' demands is the initiative of Monash University Library in Australia, which offers a set of completed data management training systems (MUL, 2014). The library not only set up a special lecture for graduate students but also prepared related PPTs and training programmes on data management, intellectual property and academic exchanges for teachers as part of their teacher certification system. This type of systematic training can help all teachers and students rapidly understand data management.

Continuous improvements of system functions

Developing back-ups, long-term preservation and sustainability are issues that every data management system has to consider (Peer and Green, 2012). The interviews at WHU revealed that after realising the basic function of data storage and retrieval, the focus of the researcher was on big data uploads, online computation, data mining and knowledge extraction or similar concerns. The changes in user requirements provide the external motivation for system improvements. Libraries should analyse these requirements and focus on upgrading their common functions. Simultaneously, libraries must cooperate with relevant schools and IT departments to resolve issues, such as the integration of data storage, data analysis and high-performance calculations so that users can manage and utilise data conveniently.

Additionally, providing efficient personal data services for researchers provides internal motivation for system improvement. The DataStaR project offers a successful model for this. Files uploaded to a data set are stored in the Fedora repository, and DataStaR customises Vitro to define and specify the relationships between data sets, individuals and organisations (Khan *et al.*, 2011). It combines Digital Record Object Identification with Simple Web Service Offering Repository Deposit to reduce manual operations so that semi-automatic metadata indexing and automatic or semi-automatic transmissions among different platforms can be realised.

Gradual advancement of librarians' data literacy

Based on the practices at WUL, it is necessary for librarians from several related departments to work together to provide appropriate data services, which means that all of them should continue to learn and track new developments in data management. Additionally, as direct communicators with researchers, subject librarians' consulting and serving abilities usually directly determine whether researchers trust the library's

data management service. Therefore, subject librarians must understand the characteristics of disciplinary data and also have a clear understanding of the entire data management process, which makes an even higher demand of subject librarians' abilities.

To meet these new demands of data service, university libraries would typically recruit personnel trained in this field or establish relevant departments when starting their data service (Li, 2011). Simultaneously, professional education has formed a curriculum system containing undergraduate, master and doctoral courses and on-the-job training (Wu, 2012). Professional education in China lags behind, and, therefore, on-the-job training is even more urgent. Learning from CALIS subject librarian training courses, hiring experts and establishing a series of data management service training courses should be rapid and provide effective ways to upgrade the university libraries' data service ability.

References

Beagrie, N., Beagrie, R. and Rowlands, I. (2009), "Research data preservation and access: the views of researchers", *Ariadne*, No. 60, available at: www.ariadne.ac.uk/issue60/beagrie-et-al/ (accessed 15 January 2015).

Brown, D. (2007), *Scientific Communication and the Dematerialization of Scholarship*, Proquest-CSA Discovery Guides, Ann Arbor, MI.

Center for Social Survey (CSS) (2015), "Center for social survey, Sun Yat-sen University", available at: css.sysu.edu.cn/Data (accessed 15 January 2015).

DataConservancy (DC) (2015), "Data conservancy", available at: dataconservancy.org (accessed 15 January 2015).

Fudan Institute of Social Research (FISR) (2015), "Fudan University dataverse network", available at: dvn.fudan.edu.cn/dvn/ (accessed 15 January 2015).

Hong, C. and 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*, Vol. 23 No. 1, pp. 16-21.

Institute of Social Survey (ISSS) (2015), "Institute of social survey of Peking University", available at: www.isss.edu.cn (accessed 15 January 2015).

Jones, S., Pryor, G. and Whyte, A. (2013), *How to Develop Research Data Management Services – A Guide for HEIs*, DCC How-to Guides, Digital Curation Centre, Edinburgh, available at: www.dcc.ac.uk/resources/how-guides (accessed 15 July 2014).

Khan, H., Caruso, B., Corson-Rikert, J., Dietrich, D., Lowe, B. and Steinhart, G. (2011), "DataStaR: using the Semantic Web approach for data curation", *The International Journal of Digital Curation*, Vol. 6 No. 2, pp. 209-221.

Lawrence, S. (2001), "Free online availability substantially increases a paper's impact", *Nature*, Vol. 411 No. 6837, p. 521.

Li, X. (2011), "Research data management and service pattern in libraries", *Journal of Library Science in China*, Vol. 37 No. 5, pp. 46-52.

McLure, M., Level, A.V., Cranston, C.L., Oehlerts, B. and Culbertson, M. (2014), "Data curation: a study of researcher practices and needs", *Libraries and the Academy*, Vol. 14 No. 2, pp. 139-164.

Ministry of Education (MOE) (2015), "The twelfth five-year plan for science and technology of HEI, Ministry of Education of China", available at: www.gov.cn/zwgk/2012-03/27/content_2100852.htm (accessed 15 January 2015).

Monash University Library (MUL) (2014), "Research data skills development, Monash University Library", available at: www.monash.edu.au/library/researchdata/skills/ (accessed 11 July 2014).

National Survey Research Center (NSRC) (2015), "Chinese social survey open database", available at: www.cssod.org (accessed 16 January 2015).

Peer, L. and Green, A. (2012), "Building an open data repository for a specialized research community: process, challenges and lessons", *The International Journal of Digital Curation*, Vol. 7 No. 1, pp. 151-162.

Penn State University (PSU) (2015), "Scholar sphere", available at: scholarsphere.psu.edu (accessed 18 January 2015).

Scaramozzino, J.M., Ramírez, M.L. and McGaughey, K.J. (2012), "A study of faculty data curation behaviors and attitudes at a teaching-centered university", *College & Research Libraries*, Vol. 73 No. 4, pp. analysis 349-365.

Si, L., Xing, W., Zhuang, X., Hua, X. and Zhou, L. (2015), "Investigation and of research data services in university libraries", *The Electronic Library*, Vol. 33 No. 3, pp. 417-449.

University College London (UCL) (2015), *DSpace@Cambridge*, available at: www.repository.cam.ac.uk (accessed 17 January 2015).

Ward, C. and Freiman, L. (2011), "Making sense: talking data management with researchers", *The International Journal of Digital Curation*, Vol. 6 No. 2, pp. 265-273.

Wu, M. (2012), "Digital curation: an emerging research field of library and information science", *Library Journal*, Vol. 31 No. 3, pp. 8-12.

Wuhan University Institute of Quality Development Strategy (WHUIQDS) (2015), *China Quality Survey*, available at: 59.175.148.71:2000/index.aspx (accessed 17 January 2015).

Wuhan University (WHU) (2015), *Wuhan University Home Page*, English Version, available at: en.whu.edu.cn/index.htm (accessed 12 July 2014).

Zhang, J. (2013a), "Intention to acquire scientific data of college scientific researchers", *Journal of Intelligence*, Vol. 32 No. 6, pp. 70-75.

Zhang, J. (2013b), "Intention to share scientific data of college scientific researchers", *Information Studies: Theory & Application*, Vol. 36 No. 10, pp. 25-30.

Zhang, Q. (2013c), "Research into scientific data curation of colleges and universities", *Information Science*, Vol. 31 No. 5, pp. 42-45.

Corresponding author
Xia Liu can be contacted at: liuxia@lib.whu.edu.cn