Research Data Management Policy and Practice in China

Yingshen Huang  Andrew M. Cox
Peking University  University of Sheffield
Laura Sbaffi
University of Sheffield

Abstract

On April 2, 2018, the State Council of China formally released a national research data management (RDM) policy “Measures for Managing Scientific Data”. Literature review shows that university libraries have played an important role in supporting Research Data Management at an institutional level in countries in North America, Europe and Australasia. The aim of this paper is to capture the current status of RDM in Chinese universities, in particular how university libraries have involved in taking the agenda forward.

This paper uses mixed methods: a website analysis of university policies and services; a questionnaire for university librarians; and semi-structured interviews. Findings from website analysis and questionnaires indicate that RDS at a local level in Chinese Universities are in their infancy. On the whole there is more evidence of activity in developing data repositories than support services. Despite the existence of a national policy there remain significant barriers to further service development, such as the lag in the creation of local policy, insufficient funding for technical infrastructure, shortages of staff skills in data curation, and language barriers to international data sharing and open science. RDS in Chinese university libraries are still lagging behind the English-speaking countries and Europe.
Introduction

On April 2, 2018, the State Council of China formally released a national research data management (RDM) policy “Measures for Managing Scientific Data” (The State Council of China, 2018). Measures was the first attempt to define the responsibilities of administrative institutions such as the Ministry of Science and Technology and provincial technology departments, as well as of individual research institutions and research data centres. The policy makes it clear that local institutions should establish their own policy and create research data services (RDS) to improve RDM.

There has, of course, been work around RDM before in China. Since at least 1984, there has been spontaneous, informal academic exchange activity, sharing and ideas about research data in the country (CN-CODATA, n.d.). In 2001, the “Meteorological Data Sharing Management Regulation” was issued, which was the first data resources management policy in China, focusing on data sharing (China Meteorological Administration, 2008). Nevertheless, considering China’s importance to global scientific production, Measures comes relatively late compared to developments in RDM policy in North America, Europe and Australasia.

In the context of this major government initiative, this paper seeks to examine the changing status of RDM in China, by exploring two specific questions:

1. What is the current status of policy, practice and services in Chinese universities?
2. How are Measures impacting local policy, practice and services?

Given the centrality of academic libraries in developing research data services internationally, this paper examines these questions particularly from the library perspective. The analysis is based on analysis of web sites, a survey and interviews. Data from previous work on international development of RDS by libraries conducted in 2014 and repeated in 2018 is used to provide comparative context (Cox, Kennan, Lyon, and Pinfield, 2017; Cox, Kennan, Lyon, Pinfield, and Shaffi, 2019).

The Emerging Policy Context in China

The last 15 years has seen the gradual emergence of recognition of the importance of RDM in government and funder policy, triggered by the OECD Principles and guidelines for access to research data from public funding (2007). In response to this emerging policy framework, institutional services have advanced at the local level (Tenopir, Pollock, Allard, and Hughes, 2016; Tenopir et al., 2017; Cox, Kennan, Lyon, et al. 2017, 2019).

Measures reflects this trend, but there has been previous work around RDM in China. Prior to 2018, there were policies specific to certain natural sciences that focus on measurement, with an emphasis on data sharing, submission and long-term preservation (CMA, 2008; MOST, 2004). The national research institution, the Chinese Academy of Sciences has built a scientific data cloud offering distributed mass storage environment (Li, Yu, Zhang, Liu, and Wu, 2015). There has also been activity at university level, with some institutions creating data platforms for sharing and reuse (Liu and Rao, 2013; Zhang, Yin, Zhang, Guo, and Zhang, 2015; Luo, Zhu, Cui, and Nie, 2016). In 2014 the
China Academic Library Research Data Management Implementation Group, committed to promoting the development of RDM was jointly established by some high-ranking universities’ libraries (Yin and Wang, 2014). However, as in other countries, there have been significant barriers to developing RDS such as lack of policy norms, inadequate technical support and skills gaps (Zhou, Duan, and Song, 2017). Research data services nationally appear to be in their infancy.

Analysing Measures there are some differences between Chinese and national policy elsewhere. Most EU policy, for example, is advisory, but the Chinese policy, as an executive and governmental order issued by the highest research management department of China, is compulsory and mandatory (SPARC Europe, 2017; UKRI, 2016). Nevertheless, it will not be straightforward to translate it into practice. Measures as a national guideline sets out the responsibilities of institutions at various levels, but does not tell how to address these responsibilities. Furthermore, in setting out differing responsibilities, Measures only names high level stakeholders such as national and provincial bodies, research institutes that generate and manage data, and data centers that focus on data curation. Measures does not mention or define the role of other stakeholders, such as the researchers, the funding organizations, the publishers and the data professionals (Erway, 2013). The definition of research data used in Measures is a bit ambiguous, but later parts of the text which mention the range of applications of the policy, do imply that it is relevant to all disciplines.

**Methodology**

This paper is based on three forms of data: a website analysis of university policies and services; questionnaire results; and semi-structured interviews. The research approach has been reviewed and approved by the University Research Ethics Committee (UREC) of The University of Sheffield.

The website analysis sought to identify the main aspects of current RDS with data being collected from January 1, 2019 to April 1, 2019. Specifically, the analysis examined 1) Policy: rules, regulations, or plans. 2) Which departments are involved in RDS. 3) Services: advisory services, technical services. The scope of the analysis was the 137 Double-class universities which are approved by China Ministry of Education as key universities (MOE, 2017), 11 universities in Hong Kong who are eligible for doctoral degrees, and 4 universities in Macau.

In order to deepen the understanding of the RDM status, a follow up questionnaire was sent to the directors of Chinese universities. To improve comparability of results a Chinese version of the questionnaire used by Cox, Kennan, Lyon et al. (2019) was developed and piloted, and then distributed to the target universities’ libraries via an invitation email sent directly to library directors from June to November 2019. Because Chinese academic library staff contact details are not always published, some invitations were sent to the library’s public mailbox. Thus, the final number of libraries that the invitation email reached was 122 libraries (107 in China mainland, 12 in Hong Kong, three in Macao) and received 63 valid responses. The data from the questionnaire was analysed through descriptive statistics and some factor and comparative analysis.

The literature review and early questionnaire results suggested that most of the Chinese universities have not yet developed RDS. During October and December 2019, ten semi-structure interviews were conducted with librarians who were interested in RDM or Open Science via remote video or voice call, in order to understand the drivers and challenges for RDS and capture the changing scene. Some of the target
Interviewees were selected from the questionnaire respondents representing those with a greater interest in RDM and the others were selected from the institutions who did not respond to survey invitations. The purpose of this data collection was to enable the creation of case studies of pathfinder institutions who are leading the way in developing RDS in China. This paper focuses on reporting the survey results.

Findings

Website Analysis

Policy
At the time the website analysis was conducted, there was only one university – Hong Kong University – that had a policy for Research Data and Records Management which has been released on 2015. It was an adapted version of Oxford University’s Policy of 2012 (The University of Hong Kong, 2015). Some universities had announced the national Measures or had a provincial notice about the release of Measures on their website. However, none of other universities in the sample appeared to have a policy in place. This may be because there is the possibility of the creation of a national data service, though this is very much in doubt (Yuan, 2018).

RDM practice and service
Although there was no formal policy in place in any of the 137 Double-class universities in China, nine universities did have their own data platforms containing data or reports from research projects. According to the types and nature of data collected, the data platforms can be divided into social science data platforms and comprehensive ones. Seven of them are social science ones, which store and make open data including statistical data, social survey data and social projects outcome data, (Fudan, Renmin, Huazhong university of Science and Technology, East China Normal, Sun Yat-Sen, Hunan and Tsing Hua University). Platforms run by Peking and Wuhan universities contained data from a more comprehensive range of subjects. The Peking University Open Research Data Platform set up in 2016 and Hunan University Economic Data Research Center set up in 2013 have a user guide and usage rules similar to a data policy, but are not strictly data management policies. A common feature of these platforms is that the data they store is almost always the social science data, and the rate of deposit of material remains low and users rarely submit data (Liu and Zeng, 2017).

The website analysis also revealed that the libraries of Peking, Fudan and Wuhan University provide research data services, including advisory services as well as having a data platform. Training, courses, presentations and workshops about RDM are being organized by five academic libraries, enabling other stakeholders, such as researchers, the research management office, IT departments etc., to learn more about RDM, clearly distinguishing between managing and fully open data. Though the academic library plays an important role in supporting research, RDM was more normally being led by Research Management Offices.
Questionnaire Results

Response rate and the respondents
As of the end of November 2019, the questionnaire had received 63 valid responses, 42 completed the whole questionnaire and 21 partly answered, the response rate is 52% (n=122) which is not as high as expected, but may in itself reflect the low development of RDS in China.

More than half valid respondents were from universities located in Beijing and Hong Kong. As intended by the method of circulation of the survey, 75% of participants were the senior management team of the library, 50% responses were from the library’s directors who are likely to be responsible for the overall future planning of their library at a strategic level and might be thought to understand the priorities for university development.

RDM policy
We asked questions about the RDM policy in the university and which departments were involved in developing the policy. 8% respondents state that their institution has an RDM policy (Figure 1) but there was no formal policies or rules or guidelines can be found on the university’s website except the Hong Kong University. Only a further 23% of institutions planned to have a policy. A rather larger number had no plans. The result is consistent with the website analysis but with a higher rate of respondents saying that they have or plan to have a policy, perhaps due to the lack of transparency of university business on the internet.

![Figure 1. Current or planned RDM policy (n=63).](image)

At the universities having or planning an RDM policy, libraries are highly involved in policy development, followed by the Research Management Office, the IT services.
However we can see from the Figure 2 that Research Management Offices are usually taking the leading responsibility.

Figure 2. Departments/units involved in development of the RDM policy (n=19).

Auditing institutional data and researchers’ attitudes
There were also a few universities that had undertaken an audit of institutional research data (26%, n=47). In the universities that undertook the audit of institutional research data, most of the libraries did participate but not take the leading role. This is suggestive that libraries are taking a less dynamic role than seems to have happened in other countries.
Figure 3. The auditing and libraries’ participation (n=47).

That only 13% (n=46) participants have undertaken a survey of faculty/academic staff' attitudes to RDM, suggests there is a lack of awareness of RDM in the university level and most are waiting for specific mandatory requirements or detailed policy in national or provincial level. Open text comments relating to this question suggest that the results of researchers' attitudes survey might have less influence on the policy making.

Research data services (RDS)

About 42% of the respondents said they provide some research data related services and a further 26% have planned to provide them, see Figure 4. The services refer to any kind of service that relate to research data, such as advisory support, technical support, institutional repository and data platform, etc.
Library is a service providing institution within the university, so although even where there was no policy from the university or funding organization, some libraries intended to widen their service range, create new services or new roles according to the emerging RDM agenda and try to perform a good practice in supporting research. So libraries participate heavily in developing RDS with 90% planning to do so and with more than half planning to participate in a leading role (Figure 5).

Figure 4. Research data related services (n=50).

Figure 5. Participants in research data service delivery (n=26).
Development of advisory and technical services

The questionnaire investigated RDS development through a matrix of choices on a wide range of services offered by libraries, with no service = 0, basic service = 1 and well developed or extensive service = 2. Figure 5 shows the current development for each service type. Compared with the previous survey conducted by Cox, Kennan, Lyon et al. (2017, 2019), the technical services are more developed than advisory services in China, especially the Run a data repository, where almost all respondents considered that they had reached the Basic service level.

The strategic priority given to RDS was evaluated via the same matrix catalogue. Figures were calculated on the basis of scoring low priority = 0, a mid-level priority = 1 and top priority = 2.
External cooperation

While providing RDS, about 65% (n=29, 17 respondents replied Yes and two replied No but planned) libraries do already or plan to cooperate with external organizations and use commercial products to deliver RDS. As there was an open text box for respondents to answer with more detail, we know from the comments that libraries intend to link the RDS to existing or planned Institutional Repository which are institutional research outputs and mainly the publications.

Librarians’ responsibilities and skills

Measures is the first formal policy issued by the nation and every research institution should have some activities in response sooner or later. But researchers manage their data all the time whether there is a formal RDM or not, so although there is no policy, there still have some research support from libraries. Figure 8 shows how the library has organized RDM support: there are two thirds libraries would distribute RDS tasks to a specific research data team or existing research support team with 37% to the existing team.

![Figure 8. The primary leadership responsibility for plans and programs of RDS.](image)

Chinese university libraries have always provided some research support including literature retrieval, innovation check of research program and discipline competitiveness evaluating reports etc., meaning that subject librarians have a close relationship with department and researchers. Meanwhile, IT librarians have some relevant skills such as maintaining the Institutional Repository that research data services also need. When faced with newly emerging needs, it is more economic and easier to develop a new kind of service based on the existing staff or team, so reducing the training, learning and time cost.

Knowledge or skills need development to offer RDS

There were 35 respondents who participated in this question which sought to find out what skills the library thinks is most in needed for delivering RDS whether the library has already had the services or plans to (Figure 9). All the respondents thought knowledge of a variety of research methods (e.g. data analysis, data visualisation) are necessary for delivering RDS. The second highest necessary skill was Data curation. A
high concentration on these two options is similar to the previous 2018 survey, but the mean percentage of needed skill in the Chinese responses are very high – up to 81% – and indicates the lack of knowledge and skills. Needs for subject or discipline knowledge differed a lot between the two surveys, the reason might be that subject librarians in China are thought to have higher professional skills than just a general librarian with a subject background and to have deeper understanding of the research of their subjects/majors.

Drivers and challenges
The survey asked respondents to reflect on the drivers and challenges for libraries working in RDM in two open questions. There were 20 respondents who wrote about 650 words of comments responding to the question asking about the major drivers, 34 themes were identified in this text (Table 1).

Table 1. Major drivers of RDM service (from 20 respondents).

<table>
<thead>
<tr>
<th>Codes</th>
<th>Number</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Awareness of the value of research data and the benefits of RDM</td>
<td>8</td>
<td>24%</td>
</tr>
<tr>
<td>Library role – having the skills/need to stay relevant</td>
<td>6</td>
<td>18%</td>
</tr>
<tr>
<td>Funders' requirements</td>
<td>3</td>
<td>9%</td>
</tr>
<tr>
<td>Needs of researchers</td>
<td>3</td>
<td>9%</td>
</tr>
<tr>
<td>Codes</td>
<td>Number</td>
<td>Percentage</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>--------</td>
<td>------------</td>
</tr>
<tr>
<td>Awareness of the value of research data and the benefits of RDM</td>
<td>8</td>
<td>24%</td>
</tr>
<tr>
<td>Needs of university administrative department (research manage office, principal office)</td>
<td>3</td>
<td>9%</td>
</tr>
<tr>
<td>Openness of research data</td>
<td>2</td>
<td>6%</td>
</tr>
<tr>
<td>Reputation of university</td>
<td>2</td>
<td>6%</td>
</tr>
<tr>
<td>Researchers’ attitudes</td>
<td>1</td>
<td>3%</td>
</tr>
<tr>
<td>Integrity</td>
<td>1</td>
<td>3%</td>
</tr>
<tr>
<td>Publishers</td>
<td>1</td>
<td>3%</td>
</tr>
<tr>
<td>Institutional repository</td>
<td>1</td>
<td>3%</td>
</tr>
<tr>
<td>Sharing of research data</td>
<td>1</td>
<td>3%</td>
</tr>
<tr>
<td>Government’s mandatory requirements</td>
<td>1</td>
<td>3%</td>
</tr>
<tr>
<td>Institutional policy</td>
<td>1</td>
<td>3%</td>
</tr>
</tbody>
</table>

Though the amount of comments and reviews that the open questions collected may not be statistically significant, some underlying factors still are apparent from this summary.

There are no some common drivers mentioned in all the comments, but nearly half of the respondents to this question emphasized RDM awareness of university’s leadership, library’s director and researchers. Top down requirements or policy are thought to be more efficient than bottom up ones.

‘The increased awareness by the university of the importance and value of research data, the researchers’ attitudes to the openness of research data.’

‘Whether the library undertakes RDS or not and how to undertake them chiefly depends on the library director’s awareness, depends on the director’s thought that these are the things library should do for school or not.’

‘The University's attitude and policies towards RDM.’

The university’s library is an institution to provide support and meet the needs of academic staff and students, so it is difficult for the library to tell researchers what and how to deal with their data without the policy or requirements from funders or university. Researchers tend to keep their own way of managing their data and have no extra time to share or prepare to share data if there is no direct impact on their funding.

The Library’s new role or responsibility were mentioned many times in the drivers for working in RDM. As the number of users physically coming into the library goes down gradually every year and users can access resources online conveniently, many people even librarians themselves might doubt about their value.

‘Crisis in the awareness of the library.’
‘The library should commit the responsibility of information organization and knowledge management.’

‘Improvement of library’s research service.’

‘Driven by the expansion of the range of library’s services.’

‘To broaden and deepen of library service capabilities.’

Some key university’s libraries have begun to focus on supporting research and study, so Research Data Service is a new chance for them to strengthen their important presence and the value of their service. Although there is no institutional policy in place, but libraries actively try to get involved in RDM to prove their relevance.

Funders’ requirements was also mentioned but not as much as expected nor as much as in the previous survey about other countries.

When asking about challenges, even more comments were collected. 22 respondents answered the open text question which contains about 730 words and 50 items were coded, because text often mentioned multiple points (Table 2.).

<table>
<thead>
<tr>
<th>Codes</th>
<th>Number</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skills or knowledge</td>
<td>10</td>
<td>21%</td>
</tr>
<tr>
<td>Acceptance of data sharing</td>
<td>6</td>
<td>13%</td>
</tr>
<tr>
<td>Acceptance in the institution</td>
<td>4</td>
<td>9%</td>
</tr>
<tr>
<td>Lack institutional policy</td>
<td>4</td>
<td>9%</td>
</tr>
<tr>
<td>Lack mandate/rewards</td>
<td>4</td>
<td>9%</td>
</tr>
<tr>
<td>Engagement of academic staff</td>
<td>3</td>
<td>6%</td>
</tr>
<tr>
<td>Resourcing – staffing</td>
<td>3</td>
<td>6%</td>
</tr>
<tr>
<td>Infrastructure</td>
<td>3</td>
<td>6%</td>
</tr>
<tr>
<td>Awareness on RDM of researchers</td>
<td>2</td>
<td>4%</td>
</tr>
<tr>
<td>Resourcing – financial</td>
<td>2</td>
<td>4%</td>
</tr>
<tr>
<td>Legal issues</td>
<td>2</td>
<td>4%</td>
</tr>
<tr>
<td>Acceptance of the need for RDM</td>
<td>1</td>
<td>2%</td>
</tr>
<tr>
<td>Data – scale, variety</td>
<td>1</td>
<td>2%</td>
</tr>
<tr>
<td>Acceptance of RDM role in the library</td>
<td>1</td>
<td>2%</td>
</tr>
<tr>
<td>Preservation</td>
<td>1</td>
<td>2%</td>
</tr>
</tbody>
</table>

Skills or knowledge (21%) are mentioned as a barrier in about half of the answers (10 times mentions out of 22 answers). Research data management implies a complicated work flow, involving several stakeholders and the need for professional skills and knowledge to support, especially considering the scale of data and differences.
among disciplines. Kinds of service which focus on different needs are based on various skills and it is not an easy work to gain the specialists in time and keep relevant skills and knowledge up to date. It is not hard to predict the difficulties in supporting researchers from a wide range of disciplines which are the characteristics of comprehensive university.

‘Skills, skills, skills. We just can't find people with the right ones.’

‘No policy guides, no funding, severe shortage of personnel and skills.’

‘NO relevant knowledge.’

‘… (2) Knowhow and skills of library staff…’

Challenges such as Acceptance of sharing data (13%), Acceptance in the institution (9%), Lack institutional policy (9%) and Lack mandate/rewards (9%) are closely related to the awareness to RDM across the university and researchers which were thought to be the major drivers in the previous open question and they are inseparable and affect each other. High awareness in the university of its own research data heritage can promote the development of policy which can place requirements on researchers.

‘The difficulty is that the university does not pay attention to nor value the management of research data, and researchers are not willing to open their own research data. ’

‘Now: Awareness of RDM; Future: Culture and incentive for researchers to spend effort in RDM.’

‘No one will follow if no mandates.’

‘… (3) University’s support in terms of budget, human resources and a policy that makes RDM a requirement; (4) Faculty’s willingness to share data.’

‘The sharing of research data, sensitive data.’

Discussion

According to the completed part of study presented here, we find that Chinese RDS are in their initial stages of development. Compared with the maturity model of RDS
proposed by Cox, Kennan, Lyon et al. (2017; 2019), Chinese RDM seems to be following a different path in which the technical services are established earlier than the basic advisory services and in advance of development of policy and more advanced advisory services. From 2009 on, some universities began to set up data platforms or data repositories to preserve the research data collected from certain social science projects, even though the Measures had not yet been issued. Those data platforms or repositories have the basic functions of searching, accessing and reusing of data. The reason for this path of development may be that there are similarities between an institutional research outputs repository and a data repository, so it might be an economic way that build up the two kinds of repository together instead of building up separately. However, this unique model also has its own challenges, lack of policy support and awareness of the RDM concept, data life cycle and FAIR principles etc. among various stakeholders would reduce the reliability and sustainability of the data platform or repository.

The findings about whether there is a policy in place from website analysis and questionnaire were same suggesting that there is almost no institutional level policy, more than one and a half years after the national policy Measures has been issued. Only Hong Kong University has an RDM policy, which came out at 2015 and there has been no revised or updating after the national policy has published. There were four other respondents who said that we have a policy now, but we were not able to confirm this through viewing the website of university after collecting the questionnaire. The reasons for this deviation between the website analysis and questionnaire might be 1) the respondents have their own understanding of policy, rules or instructions of data platform, introduction of RDM or RDS might be thought to be institutional policy; 2) the lag on the open information of university and the relatively weak promotion of policy and services.

According to the open questions asking about major drivers and challenges in library working in RDM, the awareness of university, librarian and researchers to RDM has been frequently mentioned as a challenge, which indicates that research data management might not be the greatest priority in university or lack of driver in university level. Research activities in university are different from research institutions, they are part of the major work in university which still have important responsibility of education, so university has to balance the resources distribution in terms of financial and staffing supports in kinds of needs. Besides, most research projects within universities have been funded directly by the national funding organization, the university is only helping the national funder to organize those research activities.

In contrast to western countries, it is rare for commercial, civil society or personal funding to support research activities in universities and most projects in university are funded by the national or provincial government. It needs lots of time and huge amount of work to localize and develop the provincial or institutional policy, rules or requirements from such a general guiding national policy which does not give further details of how to manage research data. Libraries have been doing preparation work to the extent of their capabilities, such as providing pilot services, building up data platforms and training etc., while waiting for the policies or requirements that not depend on them. As for the university level, they might focus on how to help the academic staff to get more funds, raise the university’s global reputation, and believe it is the researchers or research teams’ responsibility to look after their data, and not be going to invest more on library or elsewhere to promote the RDM work or not take RDM as a high priority task. So more universities are inactive towards RDM and waiting for the requirements or policy from the higher administrative level.
Conclusion

RDS at a local level in Chinese Universities are in their infancy. The current study is timed to capture trends in a rapidly developing context. Despite the existence of a national policy there remain significant barriers to RDS development, such as the lag in the creation of local policy, insufficient funding for technical infrastructure, shortages of staff skills in data curation, and the language barriers to international data sharing and open science. RDS in Chinese university libraries are still lagging behind the English-speaking countries and Europe.

Reference


