ADVANCING **DISCOVERY**

Research Data

CHALLENGES AND OPPORTUNITIES FOR DATA SHARING IN CHINA

White paper

Contents

Authors
Mithu Lucraft, Katie Allin,
Grace Baynes, Roza Sakellaropoulou

January 2019

Introduction

This report summarises the findings of a survey with more than 2,000 researchers working in China, looking at the challenges and opportunities for data sharing in China. The full dataset is available openly on Figshare[1].

Springer Nature undertook this analysis as part of a series of global initiatives to better understand the challenges in data sharing. As a major research publisher, we are committed to developing collaborative solutions, for example on data management and education. This survey of researchers in China follows a global survey in 2017 with more 7,000 researchers worldwide[2], asking specifically about data sharing at the point of submitting an article for publication. The level of respondents from some regions and countries – notably China and Japan – meant that we could not do detailed analysis, so in 2018, we extended our research to capture feedback from researchers working in Japan and China.

The goals of the survey were to:
- Understand if researchers create **data management plans** and what, if any, barriers prevent them from doing so.

- Understand what researchers do with data management plans where these are created.

- Understand if researchers are **sharing data** and what, if any, barriers prevent them from doing so.

- Understand what data types are produced and who owns these.

Springer Nature worked in close collaboration on this survey with the National Science Library, Chinese Academy of Sciences. The survey was run online in English and Chinese, with invitations to participate sent to registrants of Springer Nature email marketing lists in China, to CAS institutions by CAS library, and social media. As an incentive to complete the online survey, respondents were offered the chance to enter themselves into a prize draw for a gift card worth 1,000 Yuan when they reached the end of the survey. In total, 2,202 responses were received from active researchers (academic, scientific or clinical) in China, representing all major research disciplines and career stages.

As with any research that is survey-based, the respondents have self-selected to an extent, as such researchers who have experience with data management and data sharing are more likely to have taken part in this survey and could have led to a certain bias in the responses.

1 Research, Nature; Allin, Katie; Baynes, Grace; Lucraft, Mithu; Penny, Dan; Chong, Steven; et al. (2018): Research data: challenges and opportunities for Chinese researchers-Springer Nature survey data. figshare. Dataset. https://doi.org/10.6084/m9.figshare.7321604

2 Stuart, David; Baynes, Grace; Hrynaszkiewicz, Iain; Allin, Katie; Penny, Dan; Lucraft, Mithu; et al. (2018): Whitepaper: Practical challenges for researchers in data sharing. figshare. Paper. https://doi.org/10.6084/m9.figshare.5975011.v1

Data management plans

The good news is that 93% of researchers in China say they have created a data management plan (DMP) before, with 58% saying that they create DMPs for half or more of their research projects. This compares to a global average of 70% who have created DMPs before, according to similar research conducted in 2018[3], and is notably higher than the 56% of researchers in Japan who say they have made a DMP before[4].

However, for 36% of researchers working in China, DMPs are created only rarely (n=1,827). For those researchers who have never created DMPs (7%), 50% say they have not heard of a DMP before, and 40% say they do not know how to create one (n=126). 69% of researchers in China are 'extremely likely' or 'likely' to create a DMP in the next two years (n=1,747), with those who have created one in the past saying they are more likely to create them again in the future. This suggests that further education on how to create DMPs would increase the proportion of researchers who create these regularly as part of their research.

Support from the research office or other researchers can best help researchers achieve the aims of their DMPs. Nearly half of respondents (48%) said the research office at their institution could best support them in fulfilling the objectives outlined in future data management plans. Others in the same research group was the second most popular answer (37%), and supervisors third (36%).

The most common details included in DMPs relate to data storage (71%) and data collection (62%). Data sharing is included by only 21% of researchers (n=1,490).

The main reasons given for creating DMPs centre on good data practice: 67% of respondents said "to ensure efficient and effective management of data", and 58% said DMPs were "good practice when undertaking research" (n=1,645). Respondents cited requirements from funders or institutions significantly less as drivers for creating DMPs (18% and 8% respectively). Compared with results from similar research in Japan and comparable global data, this is a notable difference, where institutional and funder requirements were cited more frequently as a driver for creating DMPs in these other results.

93% of researchers have created a Data Management Plan (DMP) before. Of these:

58% of researchers create DMPs for half or more than half of their research

36% create DMPs only rarely

Of those who haven't created a DMP before:

50% had not heard of a DMP before

40% do not know how to create one

(n=126)

3 Digital Science; Hahnel, Mark; Fane, Briony; Treadway, Jon; Baynes, Grace; Wilkinson, Ross; et al. (2018): The State of Open Data Report 2018. figshare. https://doi.org/10.6084/m9.figshare.7195058.v2

4 Research, Nature; Allin, Katie; Baynes, Grace; Lucraft, Mithu; Penny, Dan; Chong, Steven; et al. (2018): Research data: challenges and opportunities for Japanese researchers- Springer Nature survey data. figshare. Dataset. https://doi.org/10.6084/m9.figshare.6328952.v1

The most common details in a DMP were:
• **Data storage (71%)**
• **Data collection (62%)**

Data sharing is included by only **21%** (n=1490)

69% **are extremely likely or likely to create a DMP in the next two years**
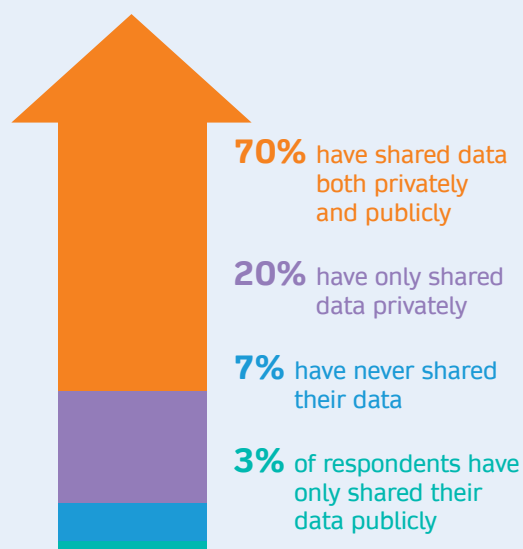
(n=1747)

Data sharing

Data sharing is important to the majority of researchers in China who responded to the survey. 79% of respondents rated the discoverability of their data as being at least somewhat important to them (score of 6 or above on an eleven point scale) and the average score to the question was 7.99, which is comparable to the global average of 7.8[2]. Only 7% of researchers in China have never shared their data either privately or publicly. 70% have shared both privately and publicly, which is similar to global findings, where public and private sharing is carried out by 71% of researchers[3]. This is much higher than reporting of public data sharing by researchers in Japan where 35% share only privately, and 59% both publicly and privately (n=975)[4]. 20% of researchers in China share only privately, and 3% only publicly (n=1645).

The top two reasons why researchers were motivated to share their data were "to progress research in their field" (46%) and "increased visibility for their research" (44%) (n=1,579). Lack of journal requirements was cited by 35% as the main reason why researchers had not shared their data (n=108), whilst "concerns about misuse of my data"(48%), and being "unsure about copyright and licensing" (32%) were the main concerns researchers faced with sharing data (n=1,560). In similar research, concerns about misuse of data was the highest concern selected by researchers in Japan (49%)[4], and globally (38%)[3]. Globally, "not receiving appropriate credit or acknowledgment" was the second largest concern (n=1356), only selected by 17% of researchers in China. Increased journal requirements, and more communication around copyright and licensing might encourage increased data sharing amongst researchers in China.

79% of respondents rated the discoverability of their data as being at least somewhat important to them

Private sharing of data is more common than public sharing of data

70% have shared data both privately and publicly

20% have only shared data privately

7% have never shared their data

3% of respondents have only shared their data publicly

The three most common methods of private sharing were:

USB or flash drives (49%)

email (43%)

PC hard drive (30%)

Most sharing is with immediate colleagues and collaborators:

share privately with colleagues from their institution (61%)

share privately with known peers (55%)

think it is important to share with other researchers in China (64%)

considered it important to make data openly available to anyone (45%)

Greater importance is placed on sharing data with immediate colleagues and collaborators rather than wider sharing, with private sharing either directly with colleagues from their institution (61%) or known peers (55%) the most common responses. Significantly more Chinese researchers consider it important to share their data with other researchers in China (64%) than more widely: only 45% considered it important to make data openly available for anyone to use. One-to-one sharing of data via physical drives is still very prevalent among researchers, with 49% of private sharing via USB/flash drives, 43% via email and 30% via PC hard drive (n=1,441). In similar research with researchers in Japan, the majority of private sharing was by email (65%), USB/flash drives (41%), and file sharing services such as Dropbox or Google (39%)[4].

Submitting data as supplementary information with journal articles is the most common way of sharing data publicly (42%, n=1119), followed by sharing in a data journal (33%). This percentage is high given that there are relatively few dedicated data journals available, and Springer Nature estimates that less than 1% of the world's 1.8 million research articles published annually are currently published in data journals. The 2018 State of Open Data Report reports that 18% of respondents had published in a data journal[3], a figure that also seems high compared to global research outputs. More work is needed to check researcher's understanding of a "data journal", in China and globally, to ensure that we interpret this finding correctly. Differences were observed by subject discipline, with more biological sciences researchers depositing their data in subject specific repositories (34%) and more engineering researchers using code sharing services (19%). Awareness and use of community repositories appears to be higher in China than Japan (42% of Chinese researchers were aware of and had deposited in dedicated community repositories compared with 30% of Japanese researchers)[4]; however this is lower than the global average of 60% of researchers aware of and using these repositories[2].

The type of data shared and timing of sharing differs for public and private sharing. Privately, negative and positive data are shared at the same levels (around 84%), whereas for negative data, a smaller proportion were willing to share publicly (46%, compared to 65% for positive data). A high number of respondents commented that data relating to unpublished findings (75%) and data relating to national security (70%) or personal data (70%) would not be shared publicly (n=1,196). The timing of sharing also varies: for private sharing, researchers share their data throughout the research process, whereas publicly the majority of research is shared after (46%), or at the point of (33%), publication (n=1479).

Perceived ownership of data is different before and after publication (n=1,902). There is a big decrease in researchers thinking they own their data after publication (56%) vs. before publication (79%). There is also a large increase in researchers thinking that data is owned by publishers after publication (43%) compared to ownership before publication (4%). This is comparable to perceptions globally on data ownership.

A high number of respondents commented that data relating to unpublished findings (75%) and data relating to national security (70%) or personal data (70%) would not be shared publicly (n=1,196)

Summary and conclusions

Overall a large number of researchers are creating DMPs in China (93%), although the frequency of this planning varies widely, with only 58% doing this for half or more than half of their research. Effective education on the value of DMPs would likely see this figure increase, with those researchers who have created one in the past saying they are more likely to create them again in the future. DMPs in China should include more detail and information on plans for data deposition and sharing, which could be addressed with more information about repositories and support to deposit and share data.

Data sharing is important to researchers, with the discoverability of data important to 79% of researchers and 93% having shared their data either privately or publicly. Private sharing is still more common than public, as is sharing with immediate colleagues and collaborators. Most data sharing is sub-optimal, either on USB or email privately, or as supplementary information to a journal article publicly. With lack of journal requirements, concerns about misuse of data, and lack of knowledge about copyright and licensing being the main concerns researchers faced with sharing data, increased journal requirements, and more communication around copyright and licensing might encourage increased data sharing best practice amongst researchers in China.

Journal publishing clearly has a role to play in encouraging and incentivising best practice, including sharing data in repositories, and routine inclusion of data citation and data availability statements. In this survey, the number one reason for not sharing data is it has not been a journal requirement, and journal/publisher requirement ranks in the top 3 most motivating factors for sharing data. The results of this survey suggest that to encourage data sharing and best practice in China, a connection to publications is helpful. In addition to journal policy requirements, it may be worth considering a focus on encouraging publication of data descriptors and data note articles. Data articles provide an established credit mechanism – a citable publication – while making datasets easier to find, access and reuse.

While sharing data as supplementary information is better than not sharing data at all, it is a sub-optimal solution. Data deposited in a repository is more findable and accessible. More effort is needed globally to convince researchers of the benefits of sharing data in repositories, and to make it easier for them to do so. In China, this will be particularly important for researchers, who currently report lower levels of data deposition in repositories than their global counterparts. "The Measures for Managing Scientific Data" released by the General Office of the State Council in March 2018 requires data deposition in repositories ahead of publication in international journals.

Lastly, we recommend emphasizing the known benefits in terms of publication output and citation advantage of research data best practice, including sharing data in repositories, and including data availability statements/data citations in articles.

Recommendations

- Provide information and support on how to create and implement DMPs, with a focus on data sharing

- Education for researchers on the value of DMPs

- More research is needed to understand researcher concerns about "misuse of data"

- Clear information about copyright and licensing options

- Stronger journal policy requirements, including encouraging data deposition, data citation and data availability statements

- Faster, easier routes to deposit data in repositories, and clear explanations of the benefits of data deposition in community repositories

- Encourage publication of data articles to provide credit for good data practice

- Promote the benefits of data management including publication output and citation advantage

Appendix

Number of respondents by subject area

Total	100%	2128
Biology	19%	410
Medicine	14%	293
Engineering	13%	272
Earth and Environmental Science	11%	237
Physics	11%	231
Materials Science	10%	204
Chemistry	7%	157
Other (please specify)	6%	128
Social Sciences	6%	118
Arts & Humanities	2%	33
Business/Investment	1%	29
Astronomy and planetary science	1%	16

Number of respondents by job title

Total	100%	1499
First stage researcher Inc. PHD Student, Post Doc, Technician/Research Assistant, Student	55%	824
Recognised researcher Inc. Research Scientist, Staff Scientist, Associate Professor/Lecturer	25%	367
Established researcher Inc. Senior Scientist, PI	5%	78
Leading researcher Inc. Consulting/Fellow/Attending Physician, Lab Director/Head, Medical Professional, Professor and Research Director/VP of Research	13%	190
Other	3%	40