10-2022

Academic libraries and research data management: a case study of Dataverse global adoption

Hsin-liang (Oliver) Chen
Philadelphia College of Osteopathic Medicine, hsinliachen@pcom.edu

Tzu-Heng Chiu

Ellen Cline

Academic libraries and research data management:

A case study of Dataverse global adoption

ABSTRACT

Purpose: The purpose of this case study is to examine the development of Dataverse, a global research data management consortium. The authors examine specifically the institutional characteristics, the utilization of the associated datasets, and the relevant research data management services at its participating university libraries. This practical, evidence-based approach is essential for understanding the current state of research data management practices in the global context.

Design/methodology/approach: The data were collected from 67 participants' data portals between December 1, 2020 and January 31, 2021.

Findings: Over 80% of its current participants joined the group in the last five years, 2016-2020. Thirty-three Dataverse portals have had less than 10,000 total downloads since their inception. Twenty-nine participating universities are included in three major global university ranking systems and 18 of those university libraries offer research data services.

Originality: This project is an explorative study on Dataverse, an international research data management consortium. The findings contribute to the understanding of the current development of the Dataverse project as well as the practices at the participating institutions. Moreover, they offer insights to other global higher education institutions and research organizations regarding research data management. While this study is practical, its findings and observations could be of use to future researchers interested in developing a framework for data work in academic libraries.

Keywords Research data management, Dataverse, Open science, Open data, Scholarly communication, Academic libraries

Introduction

Research data management has become increasingly important to researchers as major funding agencies have started requiring data sharing and data management plans for funded projects (Zhang and Chen, 2015). Accompanying this, academic libraries are seen as curatorial liaisons of data due to their long-standing history, credentials and commitments (Fox, 2013; Heidorn, 2011; Lyon, 2012; Schubert *et al*., 2013), academic institutions as well as government agencies are increasingly making their data repositories available to the public.

Many academic libraries have developed research data management services to meet these new needs (Buys and Shaw, 2015; Kellam and Thompson, 2017). Due in part to such demands, new data management systems have been developed to support data management on campus (e.g., Purdue University Research Repository, PURR). Several open-source data portals are also available (e.g., Dataverse, Mendeley Data, Open Data Repository, Open Science Framework, Zenodo).

Darch *et al*. (2020) discovered that different curatorial practices and related services in data management systems have an impact on the possibilities for data reuse. Their discovery led the authors of this case study to survey thirteen top U.S. research universities to see which data portals are used by their university libraries. In August 2020, the authors selected top ten U.S. research universities from two categories: *2017 total R&D expenditures* and *2016 total federal obligations*, based on the latest data from the National Science Foundation (NSF, n.d.). A total of thirteen universities were selected: Columbia University, Duke University, Harvard University, John Hopkins University, Stanford University, the University of California-Los Angeles, the University of California-San Diego, the University of California-San Francisco, the University of Michigan, the University of Pennsylvania, the University of Pittsburgh, the University of Washington and the University of Wisconsin-Madison. The authors were interested in knowing what these universities offered in terms of data management services and

programming at their university libraries. Based on the information obtained from the library websites, the authors discovered several common features:

Data storage and public access

- Independent data portal: 3
- As part of institutional repository: 7
- Use of Dataverse, an open data consortium: 3

Professional support

- Dedicated data service unit: 11
- Part of digital service unit: 2

Data service programming

- All 13 university libraries offered various data management services including emerging tools, discovery and evaluation to process and analysis, share and archive, etc.

Based on those findings, the authors became interested in Dataverse and wanted to understand the level of adoption of Dataverse by its members. Chen and Zhang (2014) pointed out that such understandings helped organizations implement an open-source system based on common practices with a similar purpose. Additionally, Dataverse recently became a part of the Generalist Repository Ecosystem Initiative (GREI), the goal of which is to support data repositories that house biomedical- and NIH-related datasets, and to encourage their finding and use (NIH Office of Data Science Strategy, 2022).

The purpose of this research is to examine common data management practices among institutions participating in Dataverse, an emerging research data portal worldwide. To better understand the current development of data repositories at Dataverse members, as well as the practices of creating and maintaining a data portal in general, the following four research questions are addressed:

- RQ 1: What characteristics are common to Dataverse member institutions?

- RQ2: What is the current state of dataset development and usage at these Dataverse member institutions?
- RQ3: What characteristics are common to Dataverse member universities that are highly ranked academically?
- RQ4: Are research data management services offered at the libraries of these Dataverse member universities?

Literature review

Scholarly Communication and Research Data

As emerging technologies have transformed the creation, dissemination, evaluation and preservation of scholarly communication, stakeholders have taken note of research data as an essential component of scholarly communication. Borgman (2015) emphasized the importance of research data in relation to scholarly communication. Mooney (2017) highlighted the impact of digital technologies on data sharing in multiple modalities and in new forms of scholarship, as well as how academic librarians can contribute to this emerging area of library service. In the meantime, many scholars have applied the Open Access (OA) concept to research data as well (Pampel and Dallmeier-Tiessen, 2014). A fundamental part of this concept is open data (De Silva and Vance, 2017).

Even though data sharing and reuse seem beneficial to the scholarly community, scholars from different disciplines have demonstrated a range of attitudes regarding sharing and reusing data (Jiao and Darch, 2020; Johnson *et al.*, 2016). In order to facilitate the alignment between scholarly communication and research data management, academic libraries and librarians are contributing greatly to their field through myriad services and programming (Schmidt and Shearer, 2016).

Academic Libraries and Research Data Management

Academic librarians are aware of the need for data management support and associated services on campus (Buys and Shaw, 2015; Kellam and Thompson, 2017). According to Buys and Shaw's 2015 survey at Northwestern University, the major challenges were: finding the right data storage size, finding storage at a local level, a lack of long term preservation, historically limited data sharing, and general awareness of data management requirements and policies. The survey respondents expressed their desire to see more data management services and programming at the library.

The results of the Jisc survey in the UK echoed these findings (Johnson *et al*., 2016). Noted challenges included low use of data management plans, limited data sharing practices for various reasons, various data storage volumes and sizes, and long-term preservation and data security, among others. Most respondents were not aware of their institutional data management services.

Additionally, Houtkoop *et al*. (2018) surveyed psychology researchers to identify differences in perceptions regarding data sharing. They found that these difficulties included things like data sharing being seen as an uncommon practice, data sharing only upon request, the extra work involved, and lack of training. Darch *et al*. (2020) studied research data curation and associated services at two university libraries, and discovered that the different curatorial practices and related services have an impact on the possibilities for data reuse. Recently, Huang, Cox and Sbaffi (2020) reported that research data services were very limited at over 150 Chinese universities based on their analysis of the university library websites. According to their surveys and interviews, most libraries focused on the development of their data portals rather than on data management policy development. They found that a lack of national infrastructure for research data management, professional training for librarians, and advocating for open research data sharing were key issues in China. These previous studies led the authors to focus on top research universities and the data management services at their libraries.

Development of Dataverse

Similar to DSpace and CKAN (Comprehensive Knowledge Archive Network), Dataverse is an open source repository that enables data storage and data sharing. It started as the Dataverse Network at Harvard in 2006, as part of the Institute for Quantitative Social Science (Altman *et al*., 2015). It was built on the foundations laid by the Virtual Data Center, a collaboration between Harvard and other entities (Altman *et al*., 2015; Crosas, 2011). Today, 67 institutions form its global community of data archives and research (Dataverse, n.d.).

A dataverse is defined as a digital archive, which can contain datasets, files, and collections of data. Figures 1 and 2 are examples from the Dataverse project at the University of North Carolina-Chapel Hill (UNC). Figure 1 shows the top-tier structure at the university level. Under the university, multiple research centers, institutions and researchers (the second-tier) can host different dataverses. Figure 2 shows that one second-tier dataverse can have multiple dataverses as well.

Figure 1.

Image capture December 11, 2020, from https://dataverse.unc.edu/

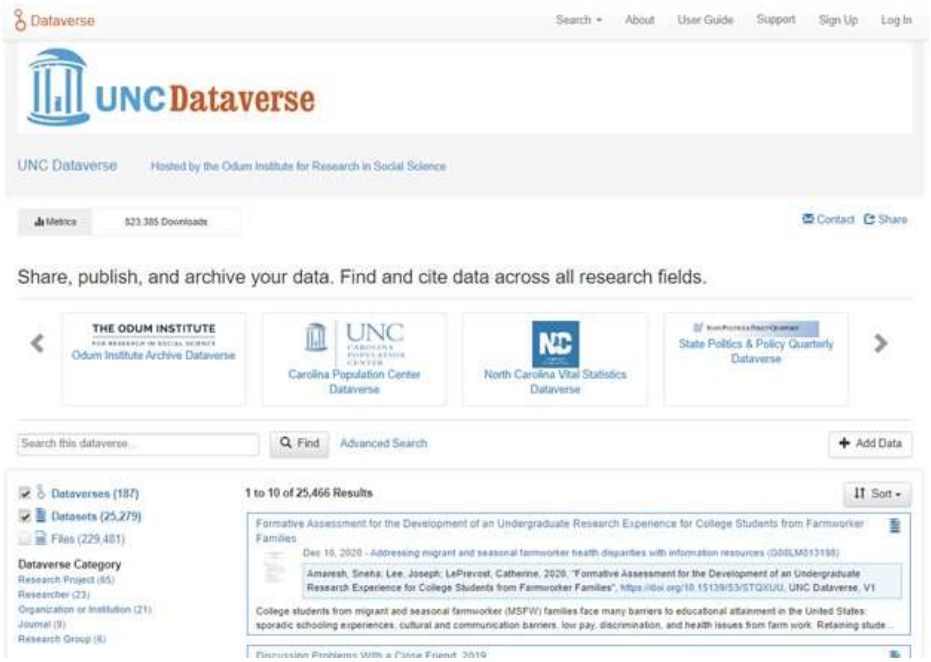
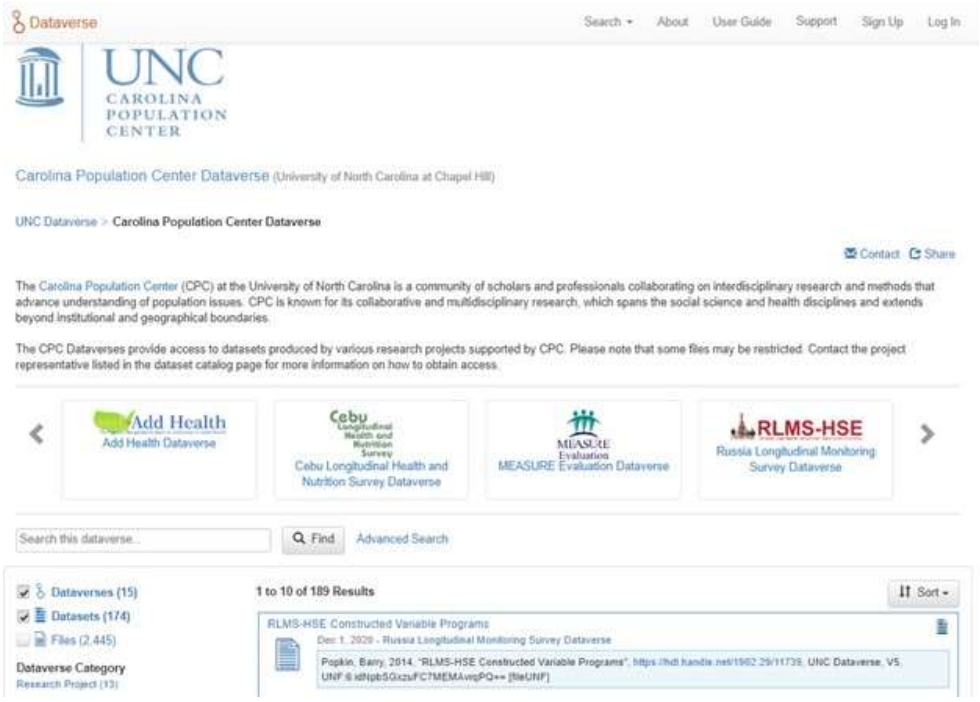


Figure 2.

Image capture December 11, 2020, from https://dataverse.unc.edu/dataverse/cpc



A major feature of Dataverse is the enabling of reliable citation methods for data. Crosas (2011) explains how persistent identifiers and numerical fingerprints that are attached to Dataverse citations permit the sharing of data, while ensuring they remain updated and usable regardless of file format. She underlines persistent citation, as well as increased visibility and ease of access, as ways of helping scholars share and use data in our data-saturated environment. Other features include search, browsing, and capacity for data storage.

The authors used a practical, evidence-based approach to examine institutional characteristics, utilization of associated datasets, and relevant research data management services at participating Dataverse member libraries. The findings from this analysis could be used in future work to better understand research data management, metadata implementation, organizational adoption, and/or the development of data service guidelines and policies.

Methodology

A list of the participating institutions (N=67) was obtained from the Dataverse project website (https://dataverse.org/) during December 1, 2020 - January 31, 2021. The authors collected the data from each institution's Dataverse portal, then used spreadsheets to record and analyze the institutional data. For the purpose of the data collection and reporting, the authors called these participating institutions "Dataverse members" (Appendix A). The following elements were collected for each Dataverse portal to answer the first two research questions:

- continent and country
- institution affiliation
- the languages used, defined as the default language of the landing webpage
- institution type
- the number of dataverses / datasets / files
- the number of downloads per dataset
- the percentage of collection growth (from the initial implementation to January 31, 2021).
- the years of membership

For RQ3, Elsevier's SciVal, a web-based analytics tool, was used to collect data on the university members between 2015 and 2020. For rankings, SciVal collects information from three global systems:

- *Quacquarelli Symonds (QS) World University Rankings*
- *Times Higher Education (THE) World University Rankings*
- *Academic Ranking of World Universities (ARWU)*

Since these three systems rank universities based on different factors, each university may receive various rankings. In this study, the authors only used the highest rank from the three systems for each university for analysis.

For academic output, the authors focused on

- the number of the publications

- the number of citations
- H5-index, an author-level metric that measures both the productivity and citation impact of the publications of a scholar. For example, a university with a h5-index of 90 means that 90 publications published by university affiliates in 2015-2020 have received at least 90 citations.
- Field-weighted citation impact (FWCI), a measurement developed by SciVal. FWCI looks at the publication level to compare the relative impact of a publication to similar ones. For example, a FWCI value of 1 means the publication has an equal impact to its peers.
- International collaboration, which is defined by co-authorship. SciVal calculates the percentage of international collaboration at each university.

For RQ4, the authors collected available information on any research data management services at the libraries of the university members.

Not all 67 members offer comprehensive information available at their data portals. The authors reported the number of members where the information was available according to each study element.

Research limitations

This paper is meant to be the first part of a project focusing on the organizational characteristics of Dataverse member libraries as well as the current state of data portal implementations at their institutions.

This first part of the project is an exploratory study and the authors relied on the Dataverse participating institutions' self-reporting figures to answer the proposed research questions. A more in-depth study would verify those figures directly. For example, a survey on those participating institutions could focus on their research data storage options, institutional data requirements, or policies concerning research data services. Additionally, survey data would be beneficial for the development of guidelines on research data management.

Findings

Common Institutional Characteristics Shared by the Member of Dataverse

Geographic distribution (N=67)

Table 1 depicts the geographic distributions of Dataverse members. The Americas lead with 32 members, of which the United States, Brazil, and Canada are the top three countries in terms of Dataverse membership. Europe is next in the ranking. France, Germany, and the Netherlands are the top three countries with the most members, respectively. Asia is the number three continent in terms of number of members, of which both China and Singapore have three members. In Africa, Botswana and Kenya have one member each, while Australia is the only member from Oceania.

Table 1.

Geographic Distribution of Dataverse Members (N=67)

Continent	**n**	**%**
Asia	11	16.42
Africa	2	2.99
Americas	32	47.76
Europe	21	31.34
Oceania	1	1.49

Languages (N=67)

English is the major language used by the members (n=46, 68.7%), followed by Portuguese (n=7; 10.5%), French (n=5, 7.5%), Spanish (n=3, 4.5%), and Chinese (n=2, 3%). Polish, Russian, and Indonesian were each used once (n=1, 1.5%).

Types of member institutions (N=67)

In terms of organizational types, universities make up the majority of the Dataverse membership (n=29, 43.3%); 26 are independent research institutions (38.8%); seven are university affiliated research centers (10.5%). The remaining five members (7.5%) are placed in the “other” category, representing networks, consortia or alliances.

Our findings reveal some common characteristics of Dataverse members: though mostly diverse in location and language, they are more likely to be part of research universities or institutions. The majority of Dataverse members are inthe Americas and Europe. The higher rates of participation in the West may be related to mandates for open research data by government agencies and funding organizations. English is one of the world’s major languages, so it stands to reason that it is the main language used by around 70% of Dataverse members.

Current state of Dataverse Development and Usage at these Dataverse Member Institutions

Years of Dataverse membership (N=41)

Only 41 members have stated their year of establishment in their portals. Among the 41 members, thirty-three (80.5%) Dataverse portals were established in the last five years, 2016-2020. This shows that the Dataverse project started adding more members after 2015.

As the open access movement continues to capture attention in academia, so too has membership in open data projects such as Dataverse grown; many of the 67 Dataverse members had created their Dataverse instances recently. Over 80% of its current members joined the group in the last five years, 2016-2020.

Dataset size (N=64)

Even as more members have joined the Dataverse project, the growth of the datasets at its member institutions has been flat. If a member institution had not added any dataset to its

Dataverse portal since 2020, the authors treated that institution as “inactive.” Five members are in the “inactive” category.

Only three members (4.69%) have more than 10,000 datasets, while seven members (10.94%) have datasets that number between 1,000 and 9,999. The top three members that have more than 10,000 datasets are: Data INRAe (https://data.inrae.fr/) of the National Research Institute for Agriculture, Food and the Environment (France), UNC Dataverse (https://dataverse.unc.edu/) of the Odum Institute for Research in Social Science at the University of North Carolina Chapel Hill, and Harvard Dataverse (https://dataverse.harvard.edu/dataverse/harvard) of Harvard University. The majority of members (n=54, 84.37%) have fewer than 1,000 datasets. Among those 54 members, 31 have fewer than 100 datasets.

Collection growth (N=63)

The authors determined the collection growth based on the number of datasets in 2020 divided by the total number of the datasets in the first year of participation. To illustrate collection growth, one institution joined the Dataverse project and added 400 datasets in 2019. In 2020, it reported the total number of datasets were 900. Hence the collection growth is 225% (900/400=2.25) between 2019 and 2020. Eight member institutions (12.70%) have seen zero collection growth since the initial year. Most of the members (n=28, 44.44%) had grown their collections less than 10 times (Table 2). However, as stated earlier, 31 members have fewer than 100 datasets in total.

Table 2.

Collection growth (N=63)

Collection growth	**n**	**%**

0%	8	12.70
Below 1000%	28	44.44
1000~9999%	16	25.40
Over 10000%	11	17.46

Use of datasets (N=61)

Only 61 members offer data download information. One of the 61 members has seen no download since its inception. More than half of the members (n=33, 54.01%) have a total download count between 1 and 10,000, while the other 21 members (34.43%) have a total download count somewhere between 10,001 and 100,000. Only seven members (11.48%) have downloads that total more than 100,000 (Table 3). The top five portals are Harvard's Dataverse (https://dataverse.harvard.edu/dataverse/harvard) at Harvard, UNC's Dataverse (https://dataverse.unc.edu/), managed by the Odum Institute for Research in Social Science , Fudan University's Social Science Data Repository (https://dvn.fudan.edu.cn/home/), Scholars Portal Dataverse (https://dataverse.scholarsportal.info/), a service of the Ontario Council of University Libraries, and Texas Data Repository (https://dataverse.tdl.org/) of Texas Digital Library, respectively.

Table 3.

Use of the dataset (N=61)

Number of downloads	**n**	**%**
0	1	1.64
1~10,000	33	54.01

10,001~100,000	21	34.42
>100,000	7	11.48

Download per dataset (N=60)

Furthermore, the authors investigated the usage of each dataset at 60 members where the information was available. Only 30% of members (n=18) experienced more than 100 downloads per dataset (Table 4). In this category, the minimum downloads per dataset is 0.78; the maximum is 1,431.78; the mean is 128.87 among 60 reporting members.

Table 4.

Download per dataset (N=60)

Number of downloads per dataset	**n**	**%**
0	1	1.67%
1-9	11	18.33%
10-49	23	38.33%
50-99	7	11.67%
Over 100	18	30.00%

The purpose of Dataverse is to facilitate the dissemination of data for use in research and other projects. Due perhaps to the relatively young membership, 31 members (about 50%) have fewer than 100 datasets, while eight members have not added new datasets since the first

year of implementation. *Reuse* is another major purpose of the Dataverse portals. However, 33 portals have had less than 10,000 total downloads since their inception. Only 18 portals have reached over 100 downloads per dataset. These findings may indicate that universities and research institutions have established their research data portals to meet open data mandates and support their researchers, but research data management at those portals is still in its infancy in terms of collection size and reuse.

Though membership has continued to grow, dataset and collection growth were found to be mostly minimal or flat. If a data portal at a Dataverse member institution is fairly new, it stands to reason that researchers and affiliates of that institution might not yet be aware of its existence. There may also be a connection between a lack of growth of Dataverse collections and the noted reluctance of some researchers when it comes to sharing their data (Pampel and Dallmeier-Tiessen, 2014). Those members that did see significant use of their data portals and grew their collections were few, though significant. Two of the three largest dataverses were established several years ago, thus giving the institutions more time to promote and grow their collections.

Future studies may wish to do a citation analysis of datasets stored in different dataverses. A combined analysis of downloads (as counted by an institution's dataverse) and citations across the academic literature may be useful indicators of the impact of different institutional dataverses.

Characteristics of the Participating Dataverse Universities that are Highly Ranked Academically

The purpose of the Dataverse project is to make research data available as well as to facilitate potential research activities. With this in mind, the authors were interested in learning more about the characteristics of Dataverse participating universities with high academic rankings and scholarly output. The authors used Elsevier’s SciVal to examine the academic rankings, the number of publications, and the other scholarly activities of the 29 members that

are universities in the Dataverse project from 2015 to 2020. (Appendix B). Four of the 29 universities are not included in Elsevier's SciVal (Table 5). Those remaining 25 universities (86.2%) are included in several global university ranking systems, and also have established their Dataverse portals for research data management to promote potential data sharing.

Sixteen ranked universities from the Americas are the leading group among the 29 members, and four of them are top global universities. Four Asian universities are in the top category as well. Most of the members (n=16, 55.17%) are ranked but not in the top category.

Table 5.

Academic rankings of the university members (N=29)

Rankings	**n**	**Continent**	**Country**
1-49	9	Americas: 4	Canada: 1 U.S.A.: 3
		Asia: 4	China: 3 Singapore: 1
		Europe: 1	Germany: 1
50-99	0		
Over 100	16	Americas: 12	Brazil: 2 Canada: 3 Chile: 1 Columbia: 1 U.S.A.: 5
		Europe: 4	Belgium: 1 Germany: 1 Italy: 1 Portugal: 1
Not ranked	4	Americas: 1	West Indies: 1
		Asia: 2	Singapore: 2
		Europe: 1	France: 1

Source: SciVal (data range: 2015-2020) (retrieved date: 2021/1/22)

Twenty-five university members are included in three major global university ranking systems, which demonstrates that those members are reputable academically and have substantial scholarly achievements. Many of them were in Canada, China, and the United States. An initial analysis revealed that supporting research data services were offered at the libraries of those top global universities. Future studies are needed to investigate specific research data services at those universities.

Table 6 illustrates the academic output of the 25 universities. In terms of number of publications, 60% of the universities (n=15) have publications ranging from 10,000-49,999. One Canadian, two Chinese, and three U.S. universities are leading universities in this category.

Regarding the number of citations, 84% of the universities (n=21) have more than 100,000 citations between 2015 and 2020. Eight universities (three from the United States, two from China, one each from Canada, Singapore, and Germany) are the leading universities in the citation category.

The results of the h5-index analysis indicated that one Canadian and three U.S. universities are the top performers. On the other hand, the FWCI measure revealed that four American, one Chinese and one German university outperformed the other 19 members.

In terms of international collaborations, Canada dominated in this category with three universities when Chile, China, Singapore, Germany, Portugal had one university each.

Table 6.

Academic output of the university members (N=25)

Academic output	**n**	**Continent**	**Country**
Number of publications			
<10,000	4	Americas: 4	Brazil: 1 Canada: 1 Columbia: 1 U.S.A.: 1

10,001-49,999	15	America: 8	Brazil: 1
			Canada: 2
			Chile: 1
			U.S.A.: 4
		Asia: 2	China: 1
			Singapore: 1
		Europe: 5	Belgium: 1
			Germany: 2
			Italy: 1
			Portugal: 1
>50,000	6	America: 4	Canada: 1
			U.S.A.: 3
		Asia: 2	China: 2
Number of citations			
<100,000	4	Americas: 4	Brazil: 1
			Canada: 1
			Columbia: 1
			U.S.A.: 1
100,000-500,000	13	America: 8	Brazil: 1
			Canada: 2
			Chile: 1
			U.S.A.: 4
		Asia: 1	China: 1
		Europe: 4	Belgium: 1
			Germany: 1
			Italy: 1
			Portugal: 1
>500,000	8	America: 4	Canada: 1
			U.S.A.: 3
		Asia: 3	China: 2
			Singapore: 1
		Europe: 1	Germany: 1
H5 index			
Below 100	9	Americas: 7	Brazil: 2

			Canada: 1 Chile: 1 Columbia: 1 U.S.A.: 2
		Europe: 2	Germany: 1 Portugal: 1
101-199	12	Americas: 5	Canada: 2 U.S.A.: 3
		Asia: 4	China: 3 Singapore: 1
		Europe: 3	Belgium: 1 Germany: 1 Italy: 1
Over 200	4	Americas: 4	Canada: 1 U.S.A.: 3
FWCI			
<1	1	Americas: 1	Columbia: 1
1-1.99	18	Americas: 11	Brazil: 2 Canada: 4 Chile: 1 U.S.A.: 4
		Asia: 3	China: 2 Singapore:1
		Europe: 4	Belgium: 1 Germany: 1 Italy: 1 Portugal: 1
>2	6	Americas: 4 Asia: 1 Europe: 1	U.S.A.: 4 China: 1 Germany: 1
% of international collaboration			
Below 50%	16	Americas: 12	Brazil: 2 Canada: 1

			Columbia: 1 U.S.A.: 8
		Asia: 2	China: 2
		Europe: 2	Germany: 1 Belgium: 1
Over 50%	9	Americas: 4	Canada: 3 Chile: 1
		Asia: 2	China: 1 Singapore: 1
		Europe: 3	Germany: 1 Portugal: 1

Source: SciVal (data range: 2015-2020) (retrieved date: 2021/1/22)

Number of Dataverse University Members' Libraries Offering Data Management Services

Among the 29 university members, 62.07 % (n=18) of their libraries offer some kind of research data management services. These 18 libraries are in the following countries: the United States (n=8, 44.45%), Canada (n=3, 16.67%), and China (n=2, 11.11%). Brazil, Colombia, Singapore and Germany have one library each (5.56%).

These university libraries use a wide variety of terms to describe their associated data management services. Seven of eight U.S. university library websites have dedicated sections for data management services: four with the name of *Research Data Management*, the other three with the name of *Research Data Services*, *Data Services*, and *Data Management and Planning*. Two of three Canadian university library websites use *Research Data Management*, and one uses *Data Management*. In China, the Peking University Library uses *Research Data Services* with additional information on *research data management*. The Hong Kong University of Science and Technology Library uses *Research Support* with information on *data management plans*. In the meantime, the titles of the librarians providing these services are

reflective of their duties. For example, *Research Data Program Manager* and *Research Data Services Librarian* are used at the Harvard Library.

Eighteen university libraries offer research data services at various levels, according to the information available on their websites. This observation echoes key results from the studies by Buys & Shaw (2015) and Kellam & Thompson (2017). This finding also underscores the importance of research data policies, as suggested by Cox *et al*. (2019) and Huang *et al*. (2021), as a driver of research data services demand.

Implications

As research data management and related services are emerging needs at research and academic libraries, understanding how institutions and researchers work with data is a key part of helping future library users in those areas. Dataverse is one of the major institutional data repositories; taking note of common practices among its members is also important for the future of data and data management. These observations may prove useful to scholars interested in further investigating the behavior of researchers as it relates to data, or to institutions interested in learning best practices from the field.

Conclusion

The Dataverse project is continuously growing, with many new members joining in the last five years (2016-2020). Its membership is mainly research-oriented universities and institutions. Even though the Dataverse membership is growing, the growth and reuse of the data collections at most members' data portals is relatively low, particularly at those younger members. Most of its Dataverse university members are highly placed by three major global university ranking systems, which indicates those universities are interested in disseminating scholarly results and outcomes produced by their affiliates. Additionally, 18 of the 29 university

libraries are offering various research data services on campus. Based on those findings, the authors propose the following recommendations for future studies:

- Research data discovery and metadata implementation: due to the low rates of downloads per dataset, the authors will explore potential barriers in terms of discovery functions and metadata elements adopted by the Dataverse members. The available search functions have direct impact on users' search behaviors when the metadata elements and descriptions may influence the discoverability of the datasets.
- Library research data services and research data management policy: previous studies emphasized that research data management policy and research data services mutually support each other. The authors will continue to study the research data services available at the university libraries and the potential impact of the services on the development of their data portals. Related professional development for library professionals, faculty and practitioners will also be addressed in the future.

As the open access movement is getting more attention from government agencies and funding organizations, sustaining, managing, and disseminating research data have become one major focus at academic libraries. The lessons learned from the Dataverse project will assist other research data initiatives as well as academic library services.

References

Altman, M., Castro, E., Crosas, M., Durbin, P., Garnett, A. and Whitney, J. (2015), "Open journal systems and Dataverse integration – helping journals to upgrade data publication for reusable research", *Code4Lib Journal*, Vol. 30, available at: https://journal.code4lib.org/articles/10989.

Borgman, C.L. (2015), *Big Data, Little Data, No Data*, MIT Press, Cambridge, MA.

Buys, C.M. and Shaw, P.L. (2015), "Data Management Practices Across an Institution: survey

and Report", *Journal of Librarianship and Scholarly Communication*, Vol. 3 No. 2, p. eP1225.

Chen, H. and Zhang, Y. (2014), "Functionality Analysis of an Open Source Repository System: current practices and implications", *The Journal of Academic Librarianship*, Vol. 40 No. 6, pp. 558–564.

Cox, A.M., Kennan, M.A., Lyon, L., Pinfield, S. and Sbaffi, L. (2019), "Maturing research data services and the transformation of academic libraries", *Journal of Documentation*, Vol. 75 No. 6, pp. 1432–1462.

Crosas, M. (2011), "The Dataverse Network®: an open-source application for sharing, discovering and preserving data", *D-Lib Magazine*, Vol. 17 No. 1/2, available at: https://doi.org/10.1045/january2011-crosas.

Darch, P.T., Sands, A.E., Borgman, C.L. and Golshan, M.S. (2020), "Library cultures of data curation: adventures in astronomy", *Journal of the Association for Information Science and Technology*, Vol. 71 No. 12, pp. 1470–1483.

Dataverse. (n.d.). "The Dataverse Project", available at: https://dataverse.org/.

Fox, R. (2013), "The art and science of data curation", *OCLC Systems & Services: International Digital Library Perspectives*, Vol. 29 No. 4, pp. 195–199.

Heidorn, P.B. (2011), "The Emerging Role of Libraries in Data Curation and E-science", *Journal of Library Administration*, Vol. 51 No. 7–8, pp. 662–672.

Houtkoop, B.L., Chambers, C., Macleod, M., Bishop, D.V.M., Nichols, T.E. and Wagenmakers, E.-J. (2018), "Data Sharing in Psychology: a survey on barriers and preconditions", *Advances in Methods and Practices in Psychological Science*, Vol. 1 No. 1, pp. 70–85.

Huang, Y., Cox, A.M. and Sbaffi, L. (2021), "Research data management policy and practice in

Chinese university libraries", *Journal of the Association for Information Science and Technology*, Vol. 72 No. 4, pp. 493–506.

Jiao, C. and Darch, P.T. (2020), "The role of the data paper in scholarly communication", *Proceedings of the Association for Information Science and Technology*, Vol. 57 No. 1, available at:https://doi.org/10.1002/pra2.316.

Johnson, R., Parsons, T., Chiarelli, A. and Kaye, J. (2016), "Jisc Research Data Assessment Support - findings of the 2016 data assessment framework (DAF) surveys", available at:https://doi.org/10.5281/ZENODO.177856.

Kellam, L.M. and Thompson, K. (Ed.s). (2017), *Databrarianship: The Academic Data Librarian in Theory and in Practice*, ACRL.

Lyon, L. (2012), "The Informatics Transform: re-engineering libraries for the data decade", *International Journal of Digital Curation*, Vol. 7 No. 1, pp. 126–138.

Mooney, H. (2017), "Scholarly communication and data", in Kellam, L.M. and Thompson, K. (Ed.s), *Databrarianship: The Academic Data Librarian in Theory and in Practice,* ACRL, pp. 195–218.

NIH Office of Data Science Strategy. (2022). "NIH Office of Data Science Strategy Announces New Initiative to Improve Access to NIH-funded Data." available at: https://datascience.nih.gov/news/nih-office-of-data-science-strategy-announces-new-initiative-to-improve-data-access.

NSF. (n.d.). "Ranking by total R&D expenditures".

Pampel, H. and Dallmeier-Tiessen, S. (2014), "Open Research Data: from vision to practice", *Opening Science*, Springer International Publishing, Cham, pp. 213–224.

Schmidt, B. and Shearer, K. (2016), "Joint Task Force on Librarians' Competencies in Support

of E-Research and Scholarly Communication Librarians' Competencies Profile for Research Data Management", *Positioning and Power in Academic Publishing: Players, Agents and Agendas. Proceedings of the 20th International Conference on Electronic Publishing*, pp. 1–7.

Schubert, C., Shorish, Y., Frankel, P. and Giles, K. (2013), "The evolution of research data: strategies for curation and data management", *Library Hi Tech News*, Vol. 30 No. 6, pp. 1–6.

De Silva, P.U.K. and Vance, C.K. (2017), *Scientific Scholarly Communication*, Springer International Publishing, Cham, available at:https://doi.org/10.1007/978-3-319-50627-2.

Zhang, Y. and Chen, H. (2015), "Data management and curation practices: the case of using DSpace and implications", *Proceedings of the Association for Information Science and Technology*, Vol. 52 No. 1, pp. 1–4.

Appendix A: Dataverse Members (N= 67, as of February 28, 2021)

No.	Name	URL	Country
1	Abacus	https://abacus.library.ubc.ca/	Canada
2	ACSS Dataverse	https://dataverse.theacss.org/	Lebanon
3	ADA Dataverse	https://dataverse.ada.edu.au/	Australia
4	ASU Library Research Data Repository	https://dataverse.asu.edu/	USA
5	AUSSDA Dataverse	https://data.aussda.at/	Austria
6	Botswana Harvard Data	http://dataverse.bhp.org.bw/	Botswana
7	CIDACS	http://dataverse.intracidacs.org/	Brazil
8	CIFOR Dataverse	https://data.cifor.org/dataverse/s	Indonesia
9	CIMMYT Research Data	https://data.cimmyt.org/	Mexico
10	CIRAD	https://dataverse.cirad.fr/	France

11	Dartmouth Dataverse	https://dataverse.dartmouth.edu/	USA
12	DaRUS	https://darus.uni-stuttgart.de/dataverse/darus	Germany
13	Data INRAe	https://data.inrae.fr/	France
14	Data Suds	https://dataverse.ird.fr/	France
15	data.sciencespo	https://data.sciencespo.fr/dataverse/sciencespo	France
16	DataRepositoriUM	https://datarepositorium.uminho.pt/	Portugal
17	DataSpace@HKUST	https://dataspace.ust.hk/	China
18	Dataverse e-cienciaDatos	https://edatos.consorciomadrono.es/	Spain
19	DataverseNL	https://dataverse.nl/dataverse/root	Netherlands
20	DataverseNO	https://dataverse.no/	Norway
21	Datos	https://datos.cedia.edu.ec/dataverse/root/?q=	Ecuador
22	DR-NTU (Data)	https://researchdata.ntu.edu.sg/	Singapore
23	Florida International University Research Data Portal	https://dataverse.fiu.edu/	USA
24	Fudan University	https://dvn.fudan.edu.cn/dataverse.xhtml	China
25	Göttingen Research Online	https://data.goettingen-research-online.de/	Germany

26	Harvard Dataverse	https://dataverse.harvard.edu/dataverse/harvard	USA
27	HeiDATA	https://heidata.uni-heidelberg.de/	Germany
28	IBICT	http://repositoriopesquisas.ibict.br/	Brazil
29	ICRISAT	http://dataverse.icrisat.org/	India
30	ICWSM	https://dataverse.mpi-sws.org/	USA
31	Ifsttar Dataverse	https://research-data.ifsttar.fr/	France
32	IISH Dataverse	https://datasets.iisg.amsterdam/	Netherlands
33	Institute of Russian Literature Dataverse	https://dataverse.pushdom.ru/	Russia
34	International Potato Center	https://data.cipotato.org/	Peru
35	Johns Hopkins University	https://archive.data.jhu.edu/	USA
36	Jülich DATA	https://data.fz-juelich.de/	Germany
37	Libra Data	https://dataverse.lib.virginia.edu/	USA
38	LIPI Dataverse	https://data.lipi.go.id/	Indonesia

39	Maine Dataverse Network	http://dataverse.acg.maine.edu/dvn/	USA
40	MELDATA	https://data.mel.cgiar.org/	Lebanon
41	NIE Data Repository	https://researchdata.nie.edu.sg/dataverse/root;jsessionid=e3becb2bccece55ce1f024e072c5?q=&types=datasets&sort=dateSort&order=desc&page=1	Singapore
42	NIOZ Dataverse	https://dataverse.nioz.nl/dataverse/root	Netherlands
43	Open Data @ UCLouvain	https://dataverse.uclouvain.be/	Belgium
44	Open Forest Data	https://dataverse.openforestdata.pl/dataverse/root	Poland
45	Peking University	https://opendata.pku.edu.cn/dataverse/CSDA;jsessionid=e12c9d8bc10834e38c42363c11b2	China
46	Pontificia Universidad Católica del Perú	http://datos.pucp.edu.pe/	Peru
47	QDR Main Collection	http://data.qdr.syr.edu/	USA
48	Repositório de Dados de Pesquisa UNIFESP Dataverse	https://repositoriodedados.unifesp.br/	Brazil
49	Repositório de Dados de Pesquisa da UFABC	http://dataverse.ufabc.edu.br/	Brazil
50	Repositório de Dados de Pesquisa do ILEEL	http://dataverse.ileel.ufu.br/	Brazil

51	Repositorio de datos de investigación de la Universidad de Chile	https://datos.uchile.cl/	Chile
52	Repositorio de Datos de Investigación Universidad del Rosario	http://research-data.urosario.edu.co/	Colombia
53	Repositório Institucional de Dados para Pesquisa da Fiocruz	https://dadosdepesquisa.fiocruz.br/	Brazil
54	Repositórios Piloto da Rede Nacional de Ensino e Pesquisa	https://dadosabertos.rnp.br/dataverse/root/?q=	Brazil
55	Scholars Portal	https://dataverse.scholarsportal.info/dataverse.xhtml	Canada
56	SODHA	https://www.sodha.be/	Belgium
57	Texas Data Repository Dataverse	https://dataverse.tdl.org/	USA
58	UAL Dataverse	https://dataverse.library.ualberta.ca/	Canada
59	UCLA Dataverse	https://dataverse.ucla.edu/	USA
60	UNB Libraries Dataverse	https://dataverse.lib.unb.ca/	Canada
61	UNC Dataverse	https://dataverse.unc.edu/	USA

62	Università degli Studi di Milano	https://dataverse.unimi.it/	Italy
63	University of Manitoba Dataverse	https://dataverse.lib.umanitoba.ca/	Canada
64	UWI	https://dataverse.sta.uwi.edu/	Jamaica
65	VTTI	https://dataverse.vtti.vt.edu/	USA
66	World Agroforestry - Research Data Repository	https://data.worldagroforestry.org/	Kenya
67	Yale-NUS Dataverse	https://dataverse.yale-nus.edu.sg/	Singapore

Appendix B: University Ranking and Academic Output (N= 29, as of February 28, 2021)

No	Name of Dataverse	University	QS	THE	ARWU	No. of pub	No. of citation	h5 index	FWCI	% of Int'l coll
1	Peking University	Peking University	23	23	49	89280	1110488	199	1.59	33.1
2	Fudan University	Fudan University	34	70	100	60996	702604	159	1.58	30.1
3	DataSpace@HKUST	Hong Kong University of Science and Technology	27	56	301-400	17609	280075	132	2.00	72.4
4	NIE Data Repository	National Institute of Education, Singapore	NA							

5	DR-NTU (Data)	Nanyang Technological University (NTU)	13	47	91	47348	712758	191	1.91	65.2
6	DataRepositoriUM	University of Minho	591-600	801-1000	401-500	16487	142122	84	1.28	50.3
7	Università degli Studi di Milano	University of Milan	301	351-400	151-200	40759	489988	151	1.79	45.2
8	HeiDATA	Heidelberg University	64	42	57	43243	719455	195	2.04	55.1
9	DaRUS	University of Stuttgart	333	351-400	301-400	15368	120712	75	1.33	38.1
10	Ifsttar Dataverse	Université Gustave Eiffel	NA							
11	Open Data @ UCLouvain	Université catholique de Louvain	189	164	151-200	19538	256297	128	1.76	62.0
12	UWI	University of the West Indies	NA							
13	Florida International University Research Data Portal	Florida International University	751-800	401-500	401-500	15822	168629	93	1.55	39.5
14	ASU Library Research Data Repository	Arizona State Univ	220	184	101-150	33663	361386	127	1.71	35.5
15	UCLA Dataverse	University of California at Los Angeles (UCLA)	36	15	13	74805	1154926	232	2.03	40.5
16	Libra Data	University of Virginia	217	117	151-200	30803	413757	143	1.93	32.6
17	Johns Hopkins University	Johns Hopkins University	25	12	15	87062	1408096	257	2.13	40.7

18	Scholars Portal	University of Toronto	25	18	23	112225	1646940	275	1.98	52.5
19	Harvard Dataverse	Harvard University	3	3	1	192685	3432389	355	2.26	45.4
20	Dartmouth Dataverse	Dartmouth College	203	101	201-300	15705	227052	121	2.01	31.5
21	Maine Dataverse Network	University of Maine	NA	NA	401-500	4654	45618	59	1.28	35.6
22	UNB Libraries Dataverse	University of New Brunswick	NA	801-1000	NA	5477	40968	49	1.31	52.0
23	UAL Dataverse	University of Alberta	119	131	101-150	47351	556316	150	1.65	52.0
24	University of Manitoba Dataverse	University of Manitoba	601-650	351-400	301-400	19685	218727	106	1.60	46.6
25	Repositorio de Datos de Investigación Universidad del Rosario	Universidad del Rosario	751-800	1001+	NA	2362	13677	35	0.91	47.7
26	Repositorio de datos de investigación de la Universidad de Chile	University of Chile	180	801-1000	401-500	19207	161425	93	1.16	54.6
27	Repositório de Dados de Pesquisa UNIFESP Dataverse	Federal University of São Paulo (Universidade Federal de São Paulo, Unifesp)	420	601-800	601-700	17362	149775	81	1.18	34.4

28	Repositório de Dados de Pesquisa da UFABC	ABC Federal University (Universidade Federal do ABC, FUABC)	NA	1001+	NA	5755	54706	64	1.35	45.5
29	Yale-NUS Dataverse	Yale-NUS College	NA							

Source: SciVal database (data range: 2015-2020, retrieved date: 2021/1/22)