THE UNIVERSITY OF BRITISH COLUMBIA

Three UBC Research Data Management (RDM) Surveys: Science and Engineering, Humanities and Social Sciences, and Health Sciences

Summary Report

UBC Library & UBC Advanced Research Computing
July 2017

Contact: Eugene Barsky, Research Data Librarian, eugene.barsky@ubc.ca
Analysis assistance by Dmitriy Ryabika, Research Analyst, Office of the Vice President Research

Key Findings:

Storage and Retention of Data: Most researchers (63 percent) estimate using less than 50 GB in their average research project, while only 12 percent are big data users, consuming more than 1 TB in their average research project.

Data Management Plans: When asked, 90 percent of the respondents stated a need for assistance to draft a data management plan as part of a grant application.

Data Sharing Issues: When asked about privacy restrictions and embargoes, 33 percent of respondents noted that privacy, confidentiality or ethics considerations may prevent sharing their research data. One respondent commented that consent forms may not have included information for research participants about data sharing. 81 percent of the respondents were interested in assistance with issues associated with data preservation and/or sharing.

Vulnerable Data Storage: To store research data from their current projects, 46 percent of the respondents use Cloud-based solutions (e.g., Dropbox, Google Drive, Amazon Cloud, Microsoft Cloud) and 39 percent use flash drives. These data storage options are considered vulnerable and data security risks. Only 9 percent use external data repositories (e.g., Dryad, Abacus Dataverse, Protein Data Bank, GenBank). When asked to rate their interest in research data management service, 78 percent of the respondents were interested or very interested in data storage and backup during active research projects. In their comments, some respondents urged the university to provide access to secure cloud storage and data collection tools for the researchers.

Data Management Education: 88 percent of the respondents were interested in workshops on best practices in data management for faculty, and 84 percent were interested in personalized consultations on data management practices for specific research groups or projects. Comments stated that data management workshops should target specific groups, research areas, or methodologies, and that a series of modules be offered to cover more details. Moreover, 86 percent were interested in communication about funding and journal requirements regarding research data.

Introduction:

Canadian academic institutions have been expecting changes in the Canadian public funding landscape for research, with new requirements around sharing and preserving research data and the submission of data management plans. In June 2016, Canadian federal funding agencies introduced the *Tri-Agency Statement of Principles on Digital Data Management*[1] , which advocates for developing data management plans (DMPs) and making data available for future research.

In preparation, librarians and other service providers at University of British Columbia surveyed researchers about their RDM practices and needs in three phases, each targeting different disciplines: (1) Sciences and Engineering (fall 2015)[2], (2) Social Sciences and Humanities (fall 2016), and (3) Health Sciences (spring 2017)[3]. Together, these surveys illuminate disciplinary differences in RDM, and inform the University in developing infrastructure and services to support researchers in RDM.

Demographics:

In total, there were 290 respondents to the three UBC RDM surveys.

- Health Sciences - total of 89 respondents completed the survey, most of whom were faculty members (Assistant/Associate/Full professors, clinical faculty).

- Social Sciences and Humanities - This survey had 101 respondents, most of whom were ongoing Faculty members. 20.7 percent were at the rank of Full Professor with 37 percent at the Assistant or Associate Professor rank. The largest single category of respondents was Graduate Students at 27.2 percent.

- Sciences and Engineering - Altogether 100 people responded to this survey. Most of the survey responses came from ongoing faculty members, with 33 percent being Full Professors and 23.7 percent being Assistant or Associate Professors; 21.6 percent were postdoctoral fellows.

[1] Tri-Agency Statement of Principles on Digital Data Management - http://www.science.gc.ca/eic/site/063.nsf/eng/h_83F7624E.html?OpenDocument

[2] Barsky, E. (2015). Research Data Management Survey : Science and Engineering, Report. doi:http://dx.doi.org/10.14288/1.0348069

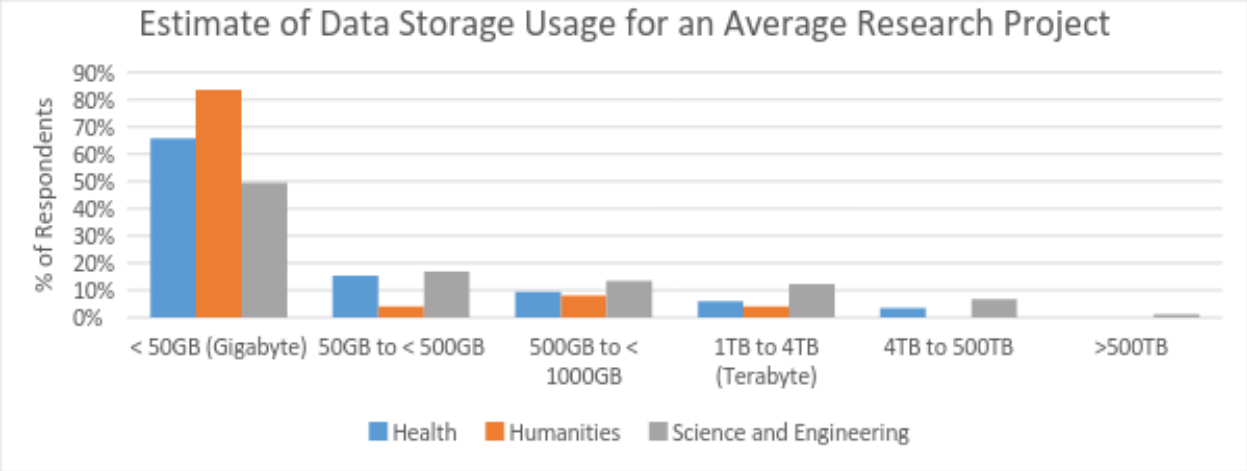
[3] Barsky, E., Brown, H., Ellis, U., Ishida, M., Janke, R., Menzies, E., … Vis-Dunbar, M. (2017, May). UBC Research Data Management Survey : Health Sciences : Report. doi:http://dx.doi.org/10.14288/1.0348070

Key Findings

How much data UBC researchers work with:

Planning for possible infrastructure needs was an important consideration when designing this survey. We suspected that most researchers did not work with large datasets, and indeed only around 4 percent in Social Sciences and Humanities (Figure 1) disciplines created more than 1TB in their work. The number of researchers working with large datasets was notably higher in the Health Sciences (10 percent) and especially Science and Engineering (19 percent) surveys, but most researchers across the institution still work within the long tail of research data, with 63 percent using less than 50GB of data for an average project:

Figure 1: Data Storage Use for an Average Project:



Research Data Storage:

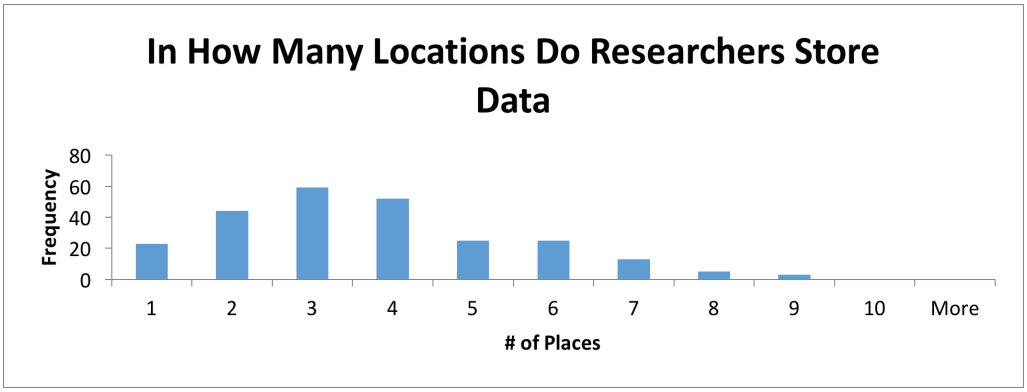
The results indicate that UBC researchers use a variety of storage options for their data (Figure 2). Unfortunately, we were not surprised to see 46 percent of survey respondents using Cloud-based solutions (e.g., Dropbox, Google Drive, Amazon Cloud, Microsoft Cloud) and 39 percent using flash drives. These data storage options are considered vulnerable and privacy and data security risks.

It was also obvious that computer and laptop hard drives (59 percent), shared university or departmental drive (53 percent), and external hard drives (51 percent) are the most popular storing media for research data. Again, it is alarming to see 39 percent of all respondents storing their data on flash drives and only a small minority using a data repository (9 percent) and a grid/high performance computing centre (3 percent).

Figure 2: Research Data Storage Options:

Moreover, Figure 3 below shows in how many different places researchers are typically storing their data. The findings may reflect both an understanding of the need for reliable backups, but also something of the fragmented nature of resources available to researchers.

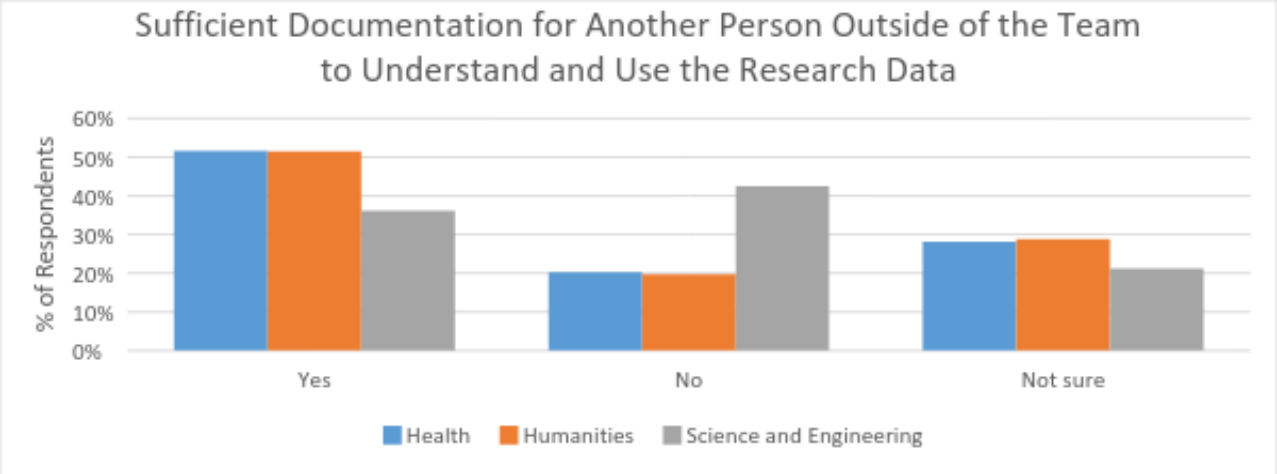
Figure 3: How Many Locations Researchers Use to Store Their Data



Data documentation and metadata:

It seems that our researchers in the Humanities and Health are divided equally about the quality of their documentation (Figure 4), with 52 percent being happy with their metadata practices. However, there is greater negativity about the quality of documentation and metadata within Science and Engineering with only 43 percent reporting less optimal documentation and only 36 percent reporting good documentation. Overall, only 46 percent of our respondents were happy with their data description for other researchers' understanding.

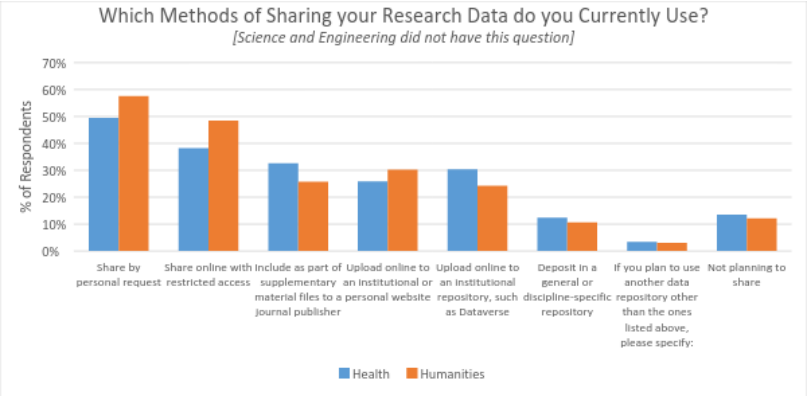
Figure 4: Data Description and Metadata:



Sharing Research Data

The survey suggests that attitudes towards open sharing of data are mostly positive (Figure 5). Many researchers are sharing data when requested (53 percent). And interestingly, only a minority are not planning to share their data - 13 percent.

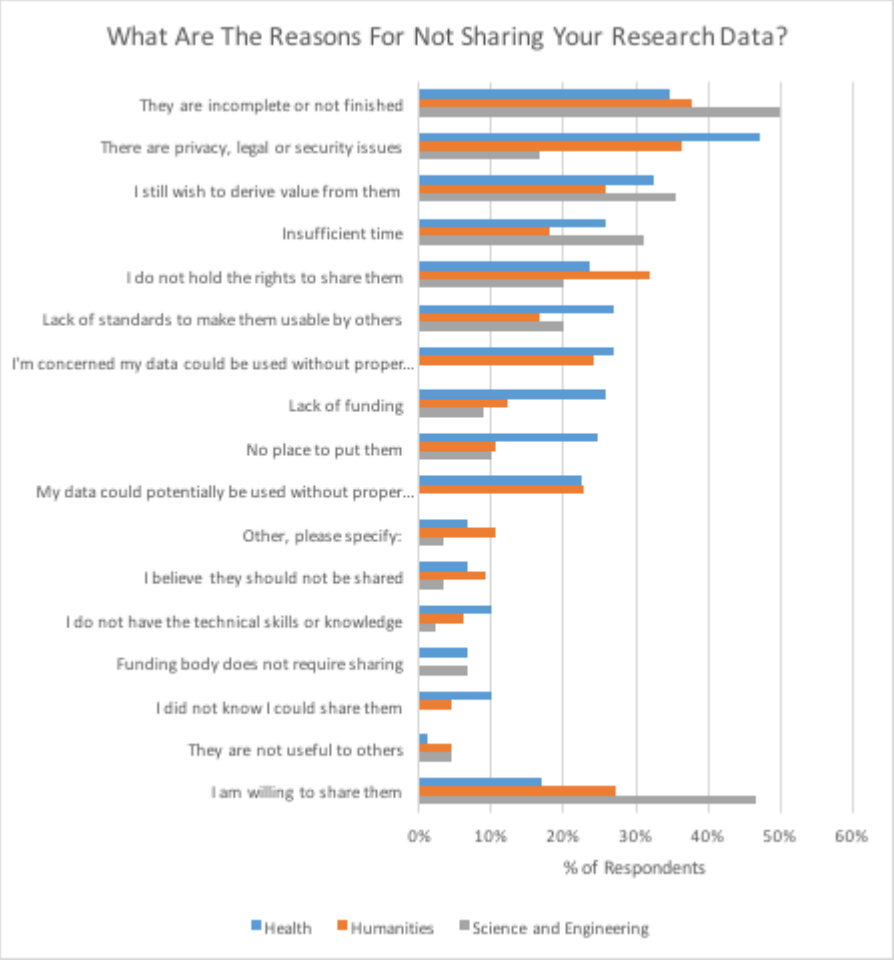
Figure 5: Data Sharing Methods



Obstacles for Data Sharing:

Most of the researchers are willing to share their data if not affected by restrictions or embargoes (Figure 6). Researchers need to publish their research and get credit for it before sharing their data. The sensitivity of data was cited as the main reason why researchers do not share their research data, especially in the Health Sciences (47 percent).

Figure 6: Reasons for not sharing Research Data



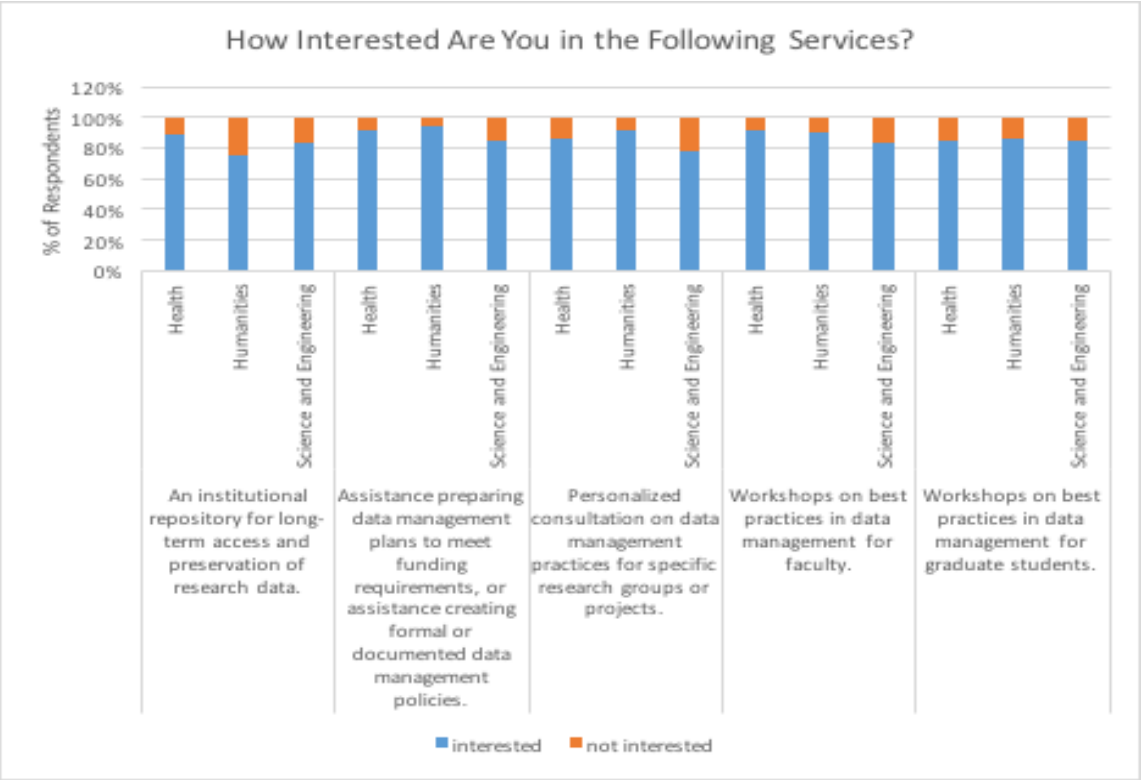
Focus on Services:

When asked when data management plans would be made part of grant applications from funding bodies such as SSHRC, CIHR, and NSERC, how interested would researchers be in the following services, we found that most services that we have in mind are of significant interest to UBC researchers (more than 80 percent for each service listed below in Figure 7).

Specifically:

- Help with preparing Data Management Plans - 90 percent
- Workshops for Faculty and Graduate students on data management - 87 percent
- Personalized consultations for data management best practices - 84 percent
- Access to an institutional data repository - 83 percent

Figure 7: Interest in RDM Services:

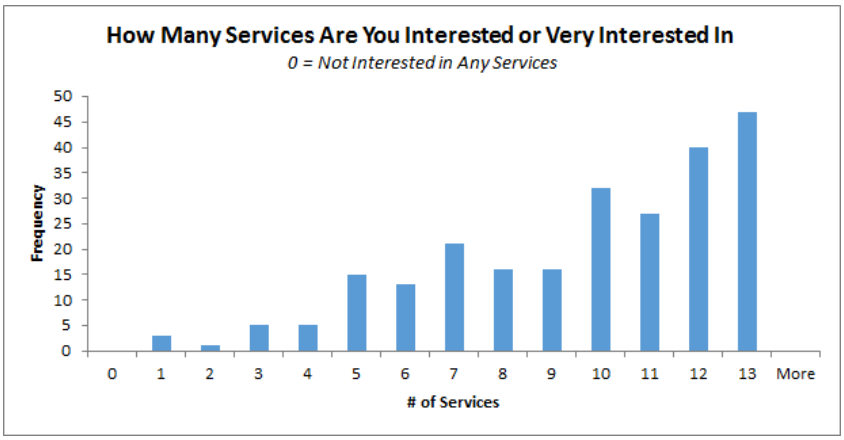


Moreover, the following services were also of significant interest:

- Data Preservation services - 81 percent
- Data Storage and Backup services - 78 percent
- Data Documentation and Metadata services - 77 percent
- Digital Object Identifiers (DOIs) services - 69 percent

Additionally, UBC researchers are interested in most or all services we had in mind, showing a dire need for help in UBC in the research data management realm (Figure 8).

Figure 8: How many RDM Services researchers are interested in



Summary:

The results presented here are from the series of surveys conducted by librarians and ARC at UBC to determine the current practices and needs of researchers with respect to research data management. As the Tri-Council granting bodies prepare to include Research Data Management Plans and access to research data repositories as requirements for grant proposals, the information gathered provides a solid base from which to develop a suite of services supporting the complete management of research data, from the planning and grant application phase through to preserving valuable data for long term access. As one respondent said "*I like the idea of making data available, but we need to make this process as simple as possible to facilitate this practice (e.g., funded as part of the grant applications).*"