

University of Lethbridge



Report on the University of Lethbridge Research Data Management Survey, October 2018

Prepared by Sheilah Ayers, Emma Scott and Rhys Stevens

October 2019

All data in the report, including all survey questions, are available here:

Stevens, Rhys; Scott, Emma; Malacrida, Claudia, 2019, "Research Data Management Survey, University of Lethbridge", <https://doi.org/10.7939/DVN/2KABSX>, UAL Dataverse, V1, UNF:6:U1gCf/UcjsFjL7zSoebQlw== [fileUNF]

Table of Contents

Summary of Survey and Reports.....	2
Findings.....	2
Recommendations.....	3
Section 1: Demographic and General Questions.....	4
1.1 Survey Demographics.....	4
1.2 Funding Sources.....	5
1.3 Digital Humanities.....	6
Section 2: Working with Research Data.....	7
2.1 Volumes of research data.....	7
2.2 Types of research data.....	8
2.3 Data storage, data documentation and retention.....	8
Section 3: Sharing Research Data.....	11
3.1 Methods of data sharing.....	11
3.2 Factors limiting the ability to share data.....	12
Section 4: Funding Mandates and Research Data Management Services.....	16
4.1 Interest in DMP Services.....	16
4.2 Interest in other proposed data management services.....	16
Comments Received.....	19
Conclusions.....	21

SUMMARY of SURVEY AND REPORT

In October 2018, the University of Lethbridge Library conducted a Research Data Management (RDM) survey of faculty, post-docs and graduate students in collaboration with the University's Office of Research and Innovation Services (ORIS). The survey is a local customization of three generic instruments provided by the Portage Canadian RDM Survey Consortium covering engineering and science, health and medical sciences, and humanities and social sciences.

The survey was conducted to explore:

- How researchers manage and share their data
- Differences in RDM practices and needs across disciplines and sub-disciplines
- How the University of Lethbridge Library and ORIS can help researchers in this process

The research team created three separate but substantially similar surveys that were distributed in a manner that matched each potential respondent to their particular area of study, allowing them to give more nuanced responses. The survey was divided into three categories: **sciences, social sciences, and humanities**. These categories were representative of the major disciplines at the University of Lethbridge and were chosen in an attempt to fully understand the data management practices of the entire University of Lethbridge research community.

This survey reached academic staff, adjuncts, post-docs and graduate students from across all disciplines at the University of Lethbridge. Out of a potential 1,215 responses, 102 were received, equaling 8.4% of those invited to participate. While this response is smaller than some other universities received, there is value in the information gathered. In particular the graduate student response rate was quite high, composing 55% of the responses. Participants were eligible to enter a draw to win one of five prizes sponsored by the University Library.

FINDINGS

How researchers manage their data

The majority of participants (63.4%, n=52) were involved in one to two projects. The storage needs of all projects ranged across the disciplines with 45.1% (n=37) using under 50 GB, 9.8% using between 50GB – 500 GB (n=8), and 12.2% (n=10) using between 500 GB – 1000GB. In addition, 14.6% (n=12) used over 1 TB. The University of Lethbridge Information Technology (IT) Services provides 200 GB of research file storage. Projects requiring over this amount of storage are required to undergo assessment and approval (<http://www.uleth.ca/information-technology/services/research-file-storage>). With 26.8% of respondents to the survey indicating they use over 500 GB of data storage, this may need to be addressed.

Social Sciences and Humanities researchers were split almost evenly between currently sharing research data (n=25) and not currently sharing (n=26). A similar split was observed in Health Sciences with five participants not currently sharing data and six currently sharing. A larger difference was evident in the Sciences, where 24 participants were currently sharing data and 16

were not. The most common methods of sharing data across all disciplines included sharing by personal request only (38.2%), sharing as part of publisher supplementary material (14.7%), sharing on a personal website (12.7%) and sharing online with restricted access (11.7%).

Reasons that prevented researchers from sharing their data included data that was incomplete (29.4%), improper citation or acknowledgment (23.5%) and researchers not holding the rights for their research data (17.6%). Only four respondents (3.9%) indicated that they believed their data should not be shared.

A majority of participants indicated they would prefer or need more information and guidance in understanding the requirements of the Tri-Council regarding research data management as well as instruction in research data management best practices (87.5%, n=83).

How the University of Lethbridge can help

In previous years the Library has offered workshops on creating data management plans. Feedback has indicated that the sessions were appreciated though the reach of these workshops has been limited. Focusing on offering more targeted workshops and online options is expected to increase the effectiveness of these workshops.

Additionally, an institutional strategy regarding research data management will raise the profile of RDM on campus while identifying how the University can better support its researchers.

Recommendations

- Ensure that relevant stakeholders on campus are aware of developments in Canadian RDM and have access to national training resources such as Portage
- Work with stakeholders such as ORIS, Grad Studies, and Information Technology Services to meet the needs identified in the survey, such as:
 - Arranging targeted introductory workshops for faculty and graduate students
 - Meeting the various data storage needs of researchers
 - Informing researchers of upcoming Tri-Council requirements
 - Providing access to online tutorials or guides for all researchers both on campus and off campus

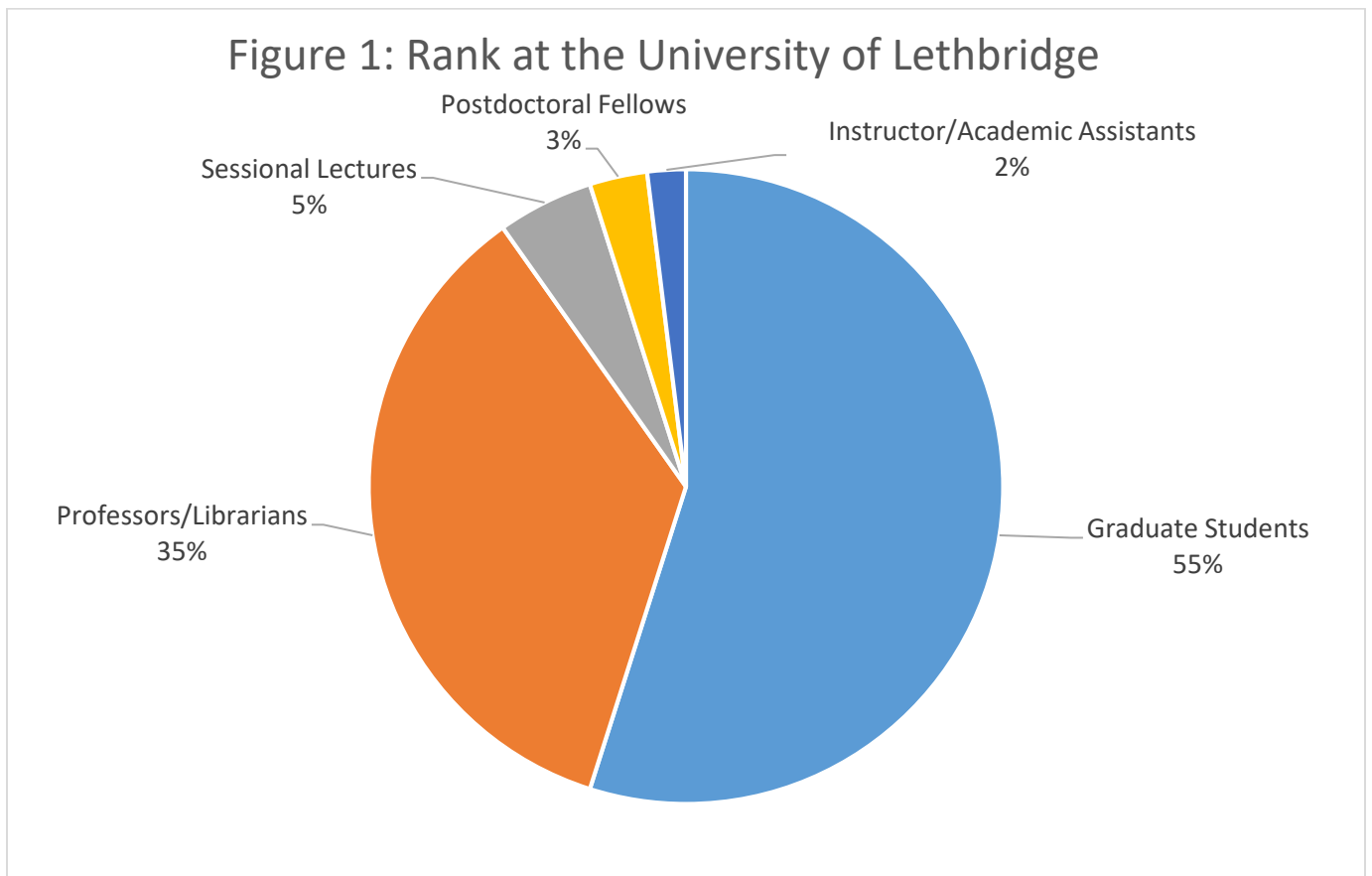
The survey has provided very useful information for understanding how researchers at the University of Lethbridge manage their data and what needs are present. While efforts have been made to support researchers, more targeted approaches in outreach will address these gaps.

SECTION ONE – DEMOGRAPHIC AND GENERAL QUESTIONS

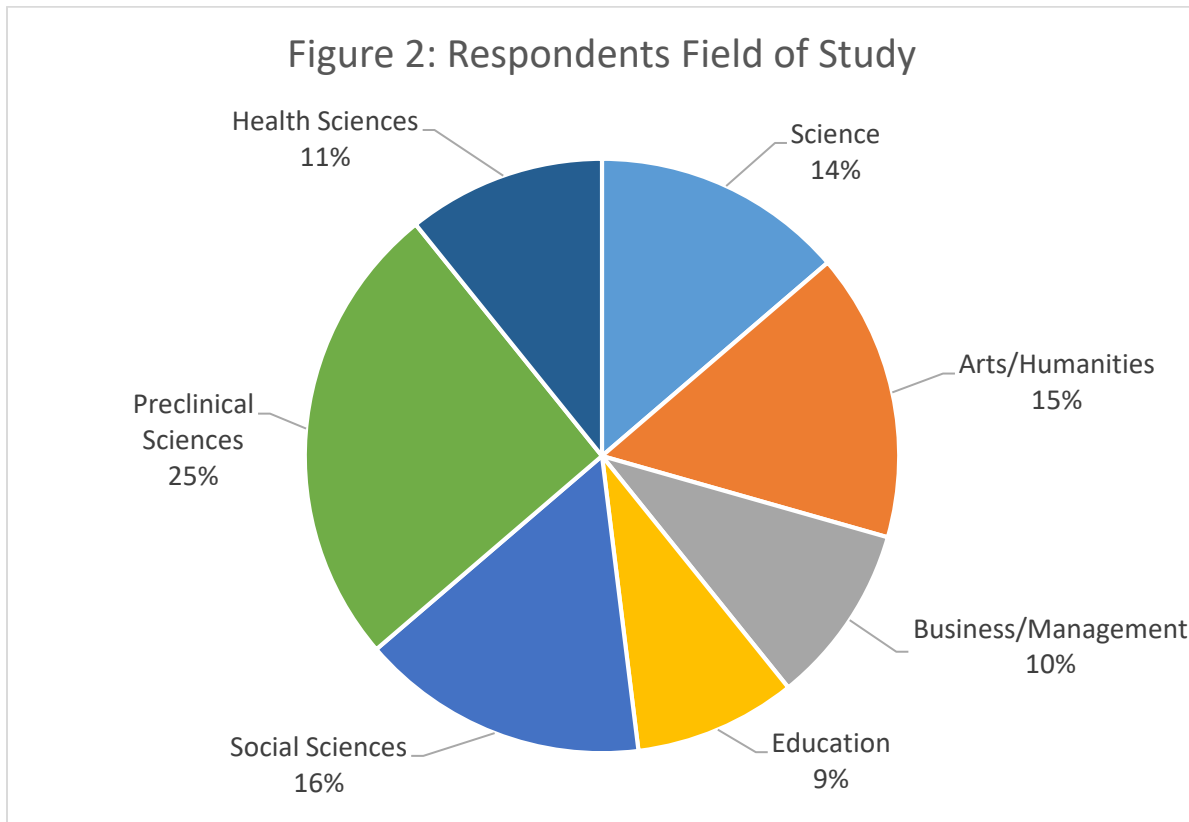
1.1 Survey Demographics

The RDM online survey was sent via email to academic staff, adjuncts, post-docs and graduate students in the disciplinary areas of Social Sciences, Humanities, Sciences, Management, and Public Health at the University of Lethbridge. The goal of the survey was to understand the research data management habits of the University community, making it essential to generate a list that gave equal opportunities for all of the University's researchers to participate.

Question 2.2 asked respondents to indicate their rank at the University of Lethbridge (see Figure 1). It is interesting to note that the highest percentage of responses came from graduate students.



Question 2.3 asked respondents to select their home department at the University of Lethbridge. From this, it was determined that the field of study with the highest percentage of respondents (n=102) was preclinical sciences (25%). Preclinical sciences is defined here as including Biochemistry, Microbiology, and Neuroscience. This category was included in order to standardize the University of Lethbridge's data set to match the National dataset. The groupings are based on NLM Classification, but modified for the surveys. The Health Sciences category includes Kinesiology, Nursing, and Rehabilitation Sciences.



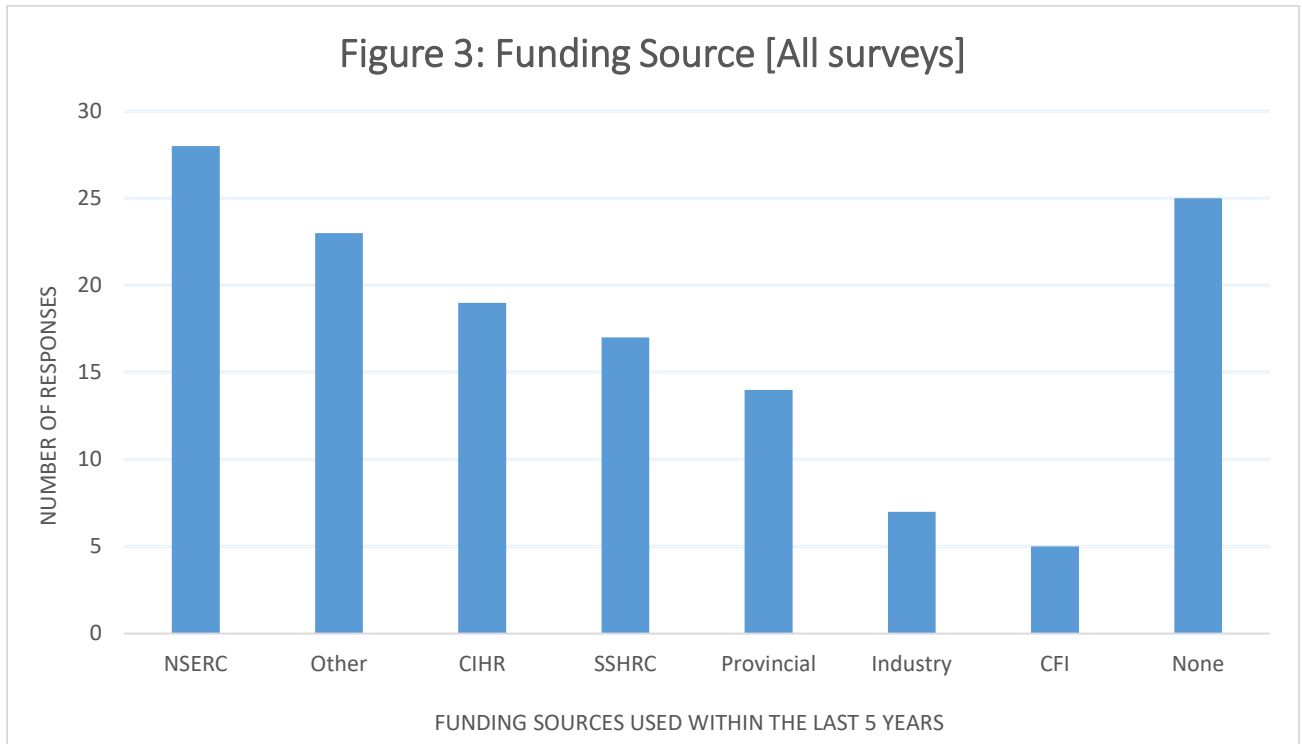
1.2 Funding Sources

Question 2.6 asked respondents (n=102) to select all funding sources they had used within the past five years, or were planning to use in the next five years. Options included the Natural Sciences and Engineering Research Council (NSERC), Social Sciences and Humanities Research Council (SSHRC), Canadian Institutes of Health Research (CIHR), Canadian Foundation for Innovation (CFI), provincial grants and industry grants.

Twenty-three percent of the survey's respondents received research grant funding from other sources. For the respondents from the Sciences, this included funding from the Pacific Institute for Mathematical Sciences, CREATE, the Canadian Space Agency, among others.

For the respondents from the Health Sciences, this included funding from the University of Lethbridge Research Fund and Western North-Western Region Canadian Association of Schools of Nursing, among others.

For the respondents from the Social Sciences and the Humanities, funding was granted from the Board of Governor's Research Chair, the Teaching Development Fund, and from professional association grants, among others.



1.3 Digital Humanities

Question 9.2 of the Social Sciences and Humanities survey asked respondents whether they felt their research fell under the definition of Digital Humanities or Digital Scholarship. Of the 50 responses received, 36% answered 'yes', while 38% of all respondents answered 'no', and 26% of all respondents were 'not sure'.

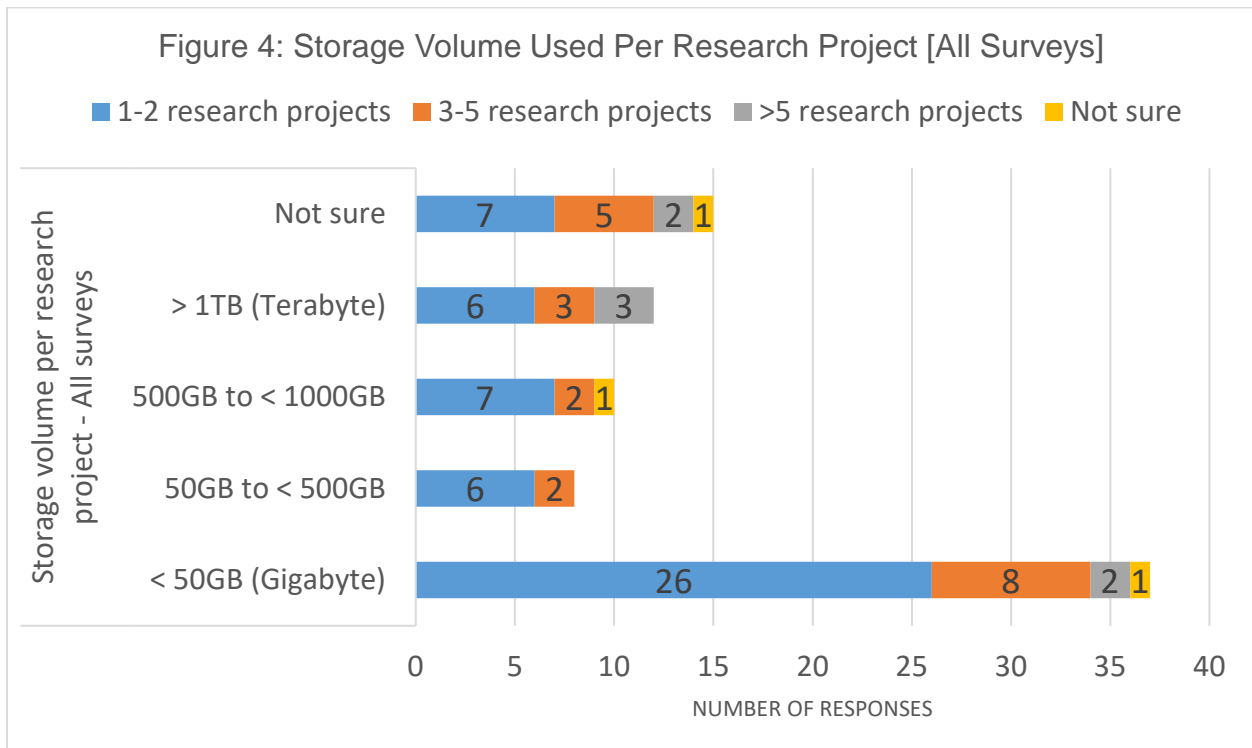
SECTION TWO – WORKING WITH RESEARCH DATA

2.1. Volumes of Research Data

Figure 4 below illustrates the relationship between storage volume required for a research project and the number of projects led by the researcher in the past year. The responses from two questions were included (“How many research projects did you lead in the last year?” and “How much data storage do you estimate you use in a research project?”). Eighty-two responses were recorded.

Figure 4 shows that the majority of respondents (63.4%, n=52) were involved in one to two projects, with 24.4% (n=20) leading 3- projects, and 8.5% (n=7) leading more than 5. Three respondents (3.7%) indicated they were unsure of the number of projects they were involved in but did indicate an estimated volume of storage used for their projects.

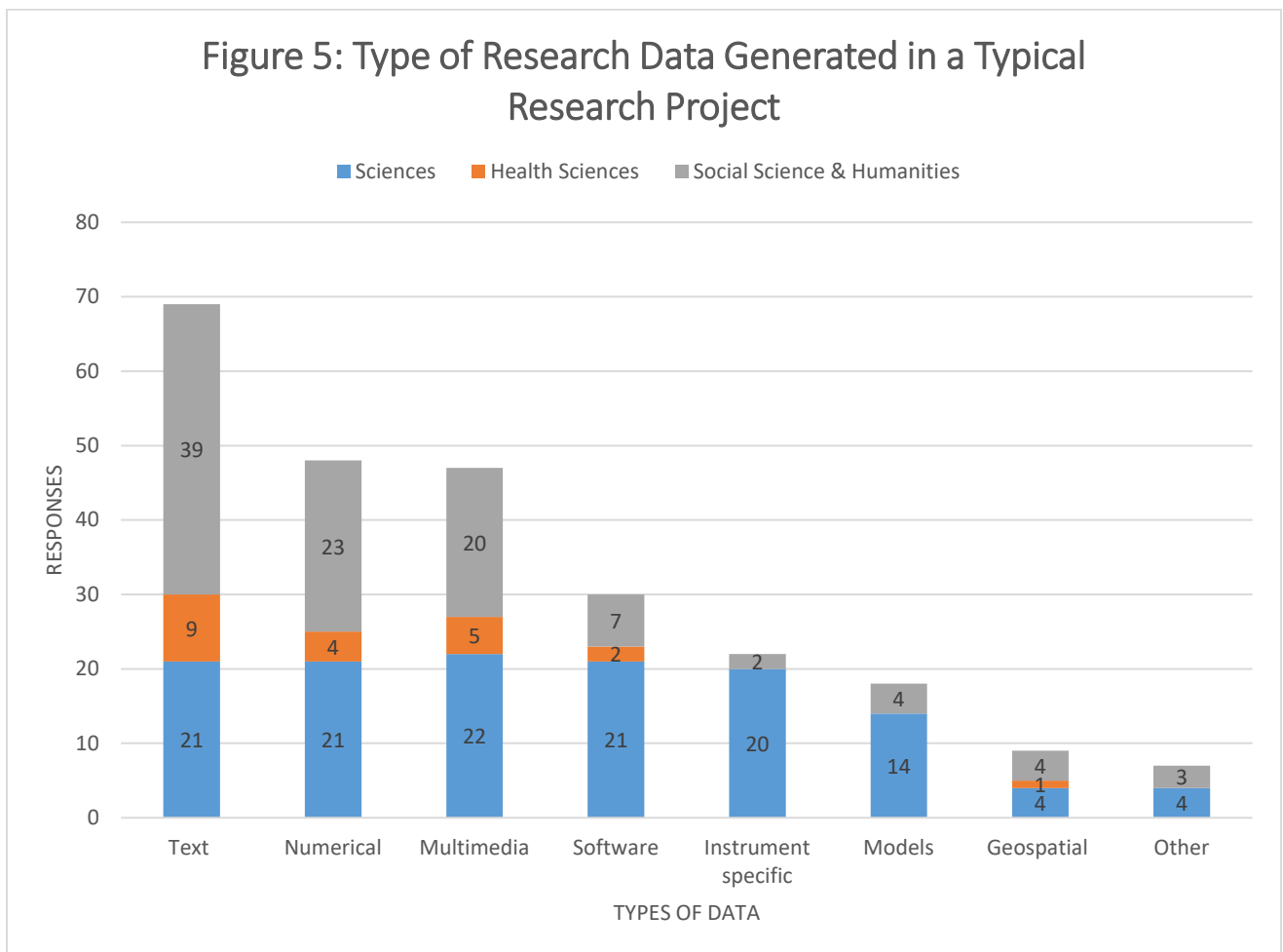
While 45.1% of respondents indicated they used under 50 GB of storage per project, the remaining respondents used between 50 GB and over 1 TB of data storage (see Figure 4). This is an important figure when compared with what the University of Lethbridge provides for its researchers. The University of Lethbridge IT Services provides 200 GB of research file storage. Projects requiring over this amount of storage are required to undergo assessment and approval (<http://www.uleth.ca/information-technology/services/research-file-storage>). With 26.8% of researchers in the survey indicating they use over 500 GB of data storage, this may need to be addressed.



2.2. Types of research data

Q3.4/Q6.4/Q9.5

Each of the surveys asked the question “Which of the following best describes the type of research data you generate or use in a typical research project?” Participants were invited to select all that apply. 250 answers were selected, with a number of respondents selecting more than one answer. The most common types of data used were text (n=690), numeric (n=48) and multimedia (n=47). Under the category ‘other’ (2.8%, n=7), responses included instrument or discipline-specific types of data such as AUC, analysis of genomic data from DNA/RNA sequencer, FITS, various forms of raw binary data, local database content, QSR NVivo, hand-drawn movement notations, and maps.



2.3. Data storage, data documentation and retention

Q3.6/Q6.6/Q9.7

Survey participants were also asked to indicate where they store research data from current projects. Respondents (n=329) were asked to select all options that applied to them. The top seven storage options chosen were:

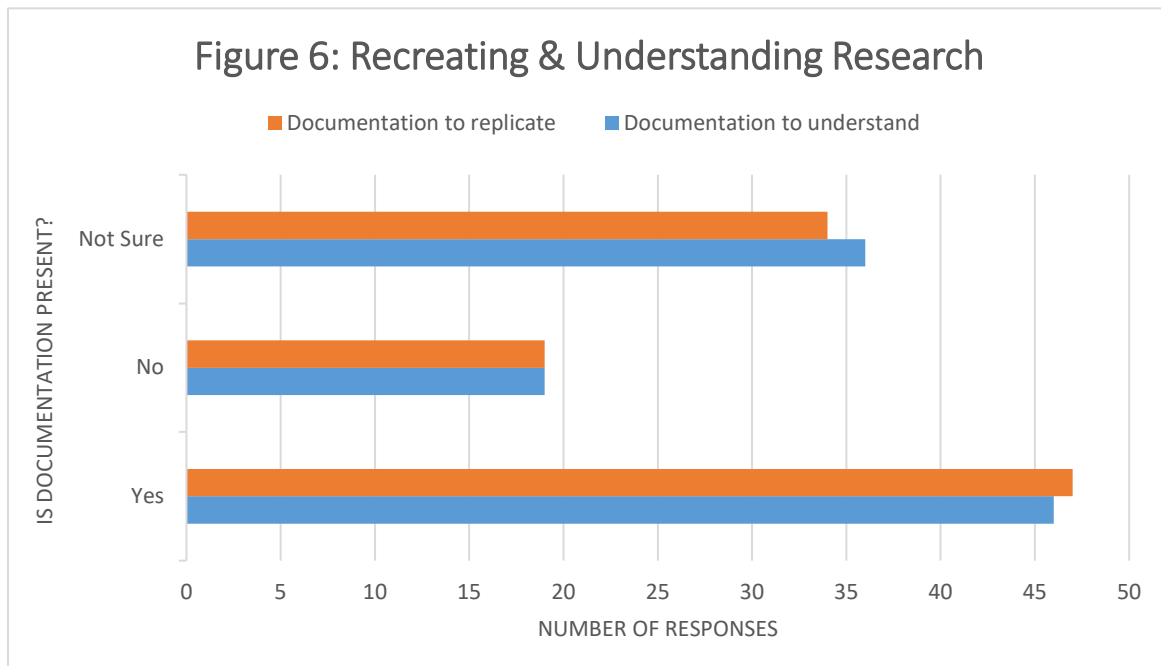
- Laptop hard drive, 17.6% (n=58)
- Computer hard drive, 17.0% (n=56)
- External hard drive, 13.1% (n=43)
- Flash drive/USB, 13.1% (n=43)
- Cloud/web based solution, 12.5% (n=41)
- Shared drive/ university server, 10.3% (n=34)
- Physical copy retained, 9.7% (n=32)

Other options included hard drive of the instrument/sensor (2.7%, n=9), external data repository (1.5%, n=5), grid/high performance computing (HPC) centre (0.9%, n=3), and CD/DVD (0.3%, n=1). Four responses indicated 'other'.

Respondents were also asked two questions in relation to their documentation of data (n=102). The questions are as follows:

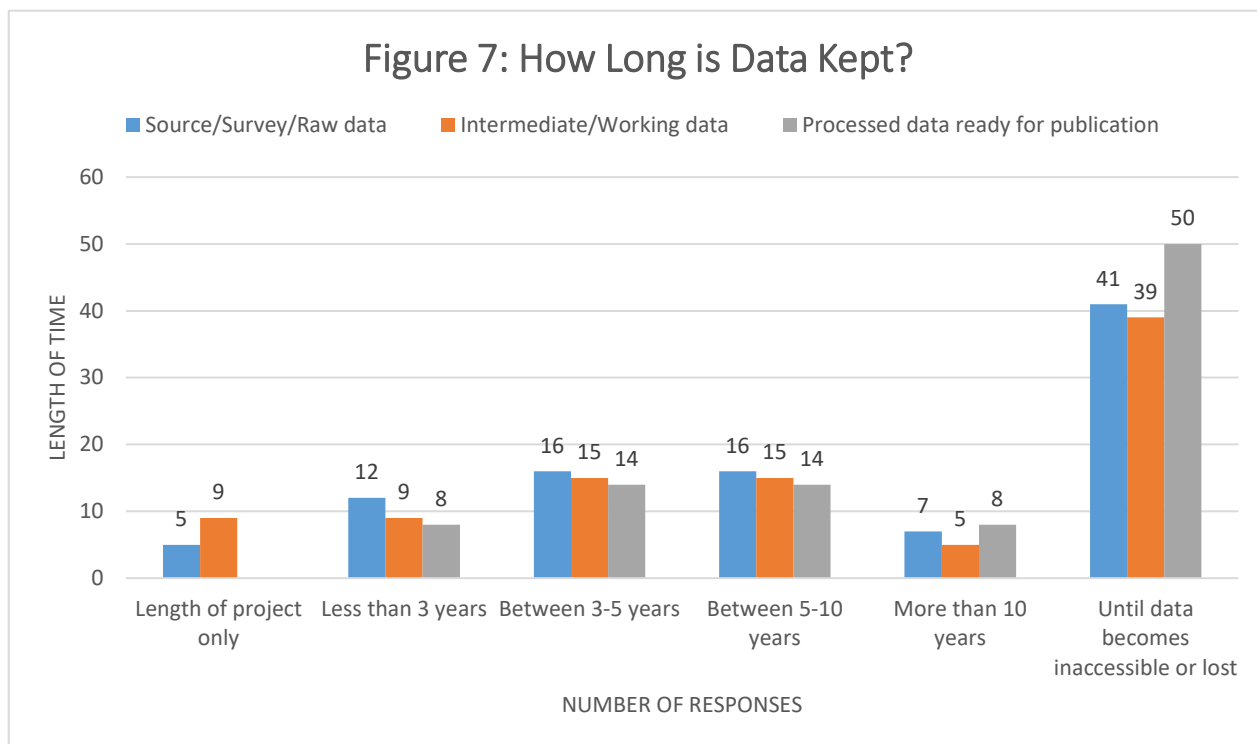
“Is there sufficient documentation and description (for example, file naming, cells & values, defined parameters, scripts to run) for another person outside your lab to understand and use the research data?” Q3.7/6.7/9.8

“Is there sufficient documentation and description (for example, file naming cells & values, defined parameters, scripts to run) retained in the same file, folder or document as the research data for another person outside your lab to replicate the methodologies that produced the data?” Q3.8/Q6.8/Q9.9



In regards to documentation to understand the research 45.1% of responses said yes (n=46), while 18.6% said no (19) and 35.3% were unsure (n=36). For replicating and recreating research 46.1% said yes (n=47), while 18.6% said no (n=19) and 33.3% were unsure (34).

Respondents were also asked to indicate the length of time after project completion that they typically intentionally keep each type of research data. [Q3.9/6.9/9.10]



Roughly half of respondents indicated that they had no set retention date. They opted to keep all forms of data, raw or processed, until it became inaccessible or was lost. For example, 42.3% said they keep raw data such as source material and/or survey results, 42.4% keep intermediate working data, and 53.2% keep all processed data ready for publication, which may include supporting information such as metadata and documentation, spectra, or synthesis methods.

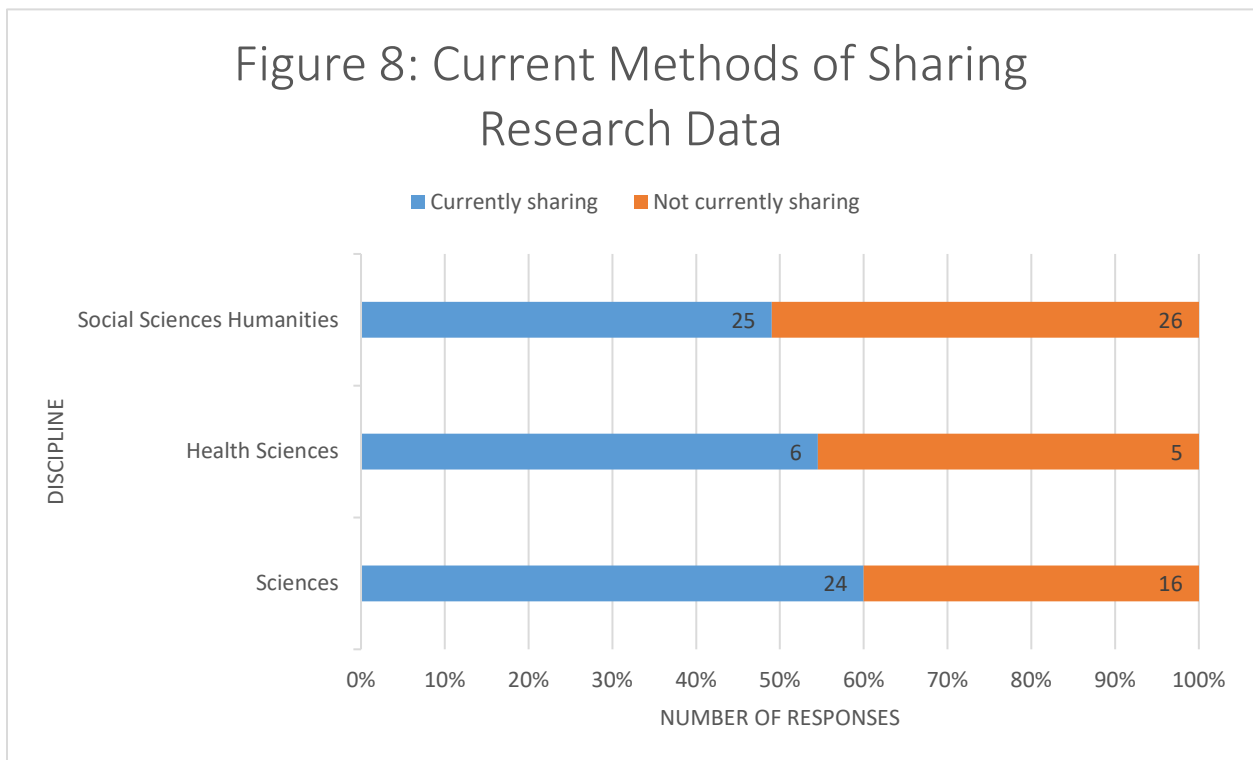
SECTION THREE – SHARING RESEARCH DATA

3.1 Methods of data sharing

All of the surveys (n=102) included questions about the sharing of research data. Specifically, questions 4.2, 7.2, and 10.2 asked respondents about their current methods of sharing research data. Fifty-five (53.9%) of the respondents were currently sharing their data. Forty-seven (46%) of the respondents were not currently sharing their data.

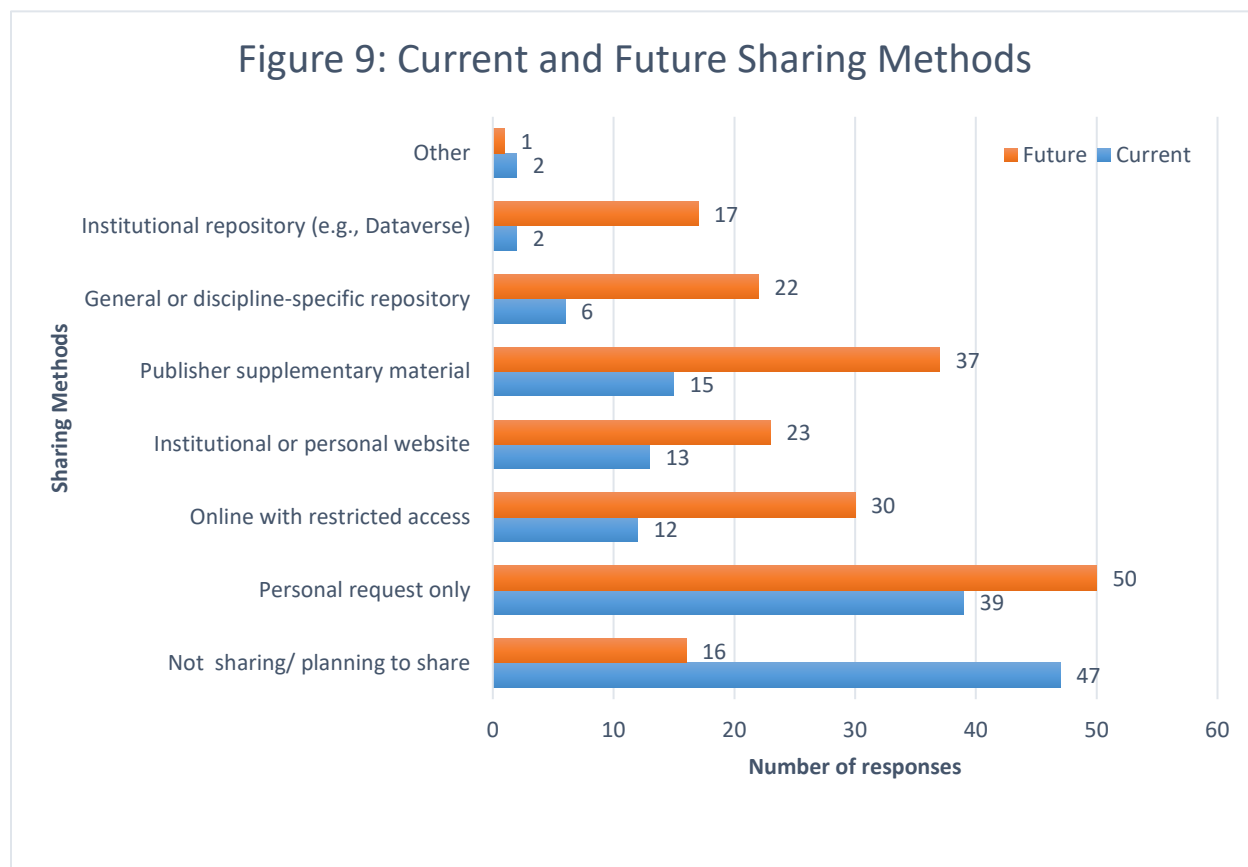
Of the 55 (53.9%) respondents that were currently sharing their data, 25 (24.5%) of them were from the Social Sciences and Humanities. Six (5.8%) of the respondents were from the Health Sciences. Twenty-four (23.5%) of the respondents were from the Sciences.

Of the 47 (46%) respondents that were not sharing their data, 26 (25.4%) of them were from the Social Sciences and Humanities. 5 (4.9%) of the respondents were from the Health Sciences. Sixteen (15.6%) of the respondents were from the Sciences.



Questions 4.2, 7.2, and 10.2 also asked respondents to indicate all methods of sharing their data they currently use. Questions 4.4, 7.3, and 10.3 also asked respondents which hypothetical methods of sharing their research data would they consider using in the future, and were able to check all methods that applied to them. Forty-seven (46%) respondents were not currently sharing or planning to share their data while only 16 respondents had no plans to share their

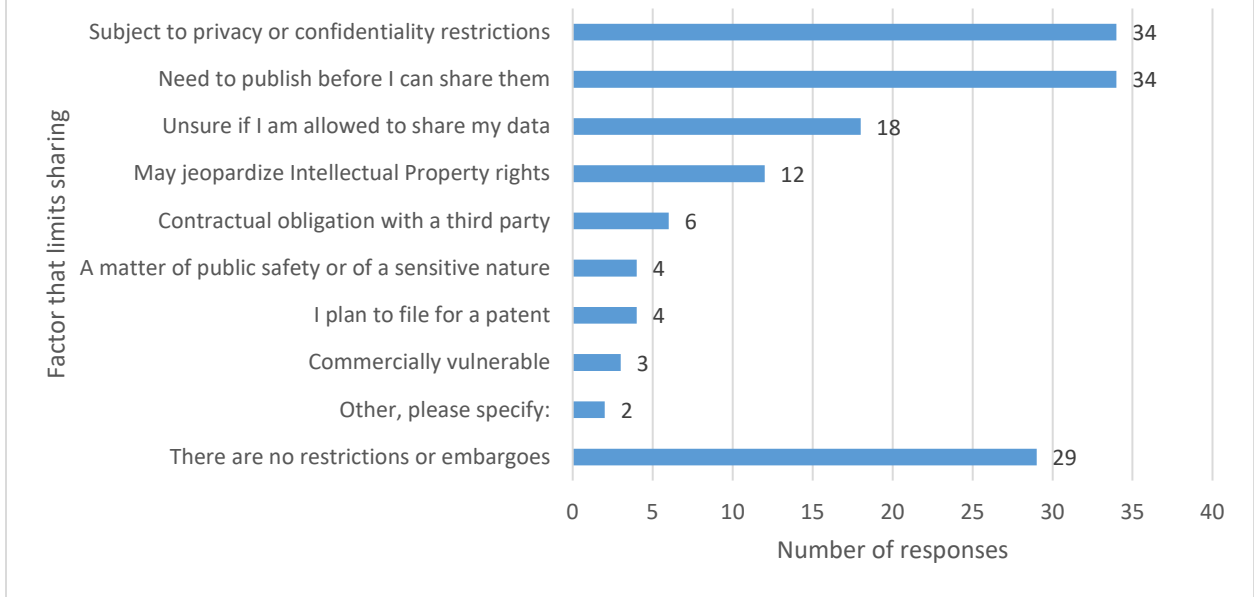
work in the future. Thirty-nine (38.2%) respondents were currently sharing their data by personal request only and 50 (49%) respondents were planning on sharing their work by personal request only. An additional 37 responses indicated researchers plan on including their data within publisher supplementary material, which would not necessarily meet the Tri-Council’s requirements for open data.



3.2 Factors limiting the ability to share data

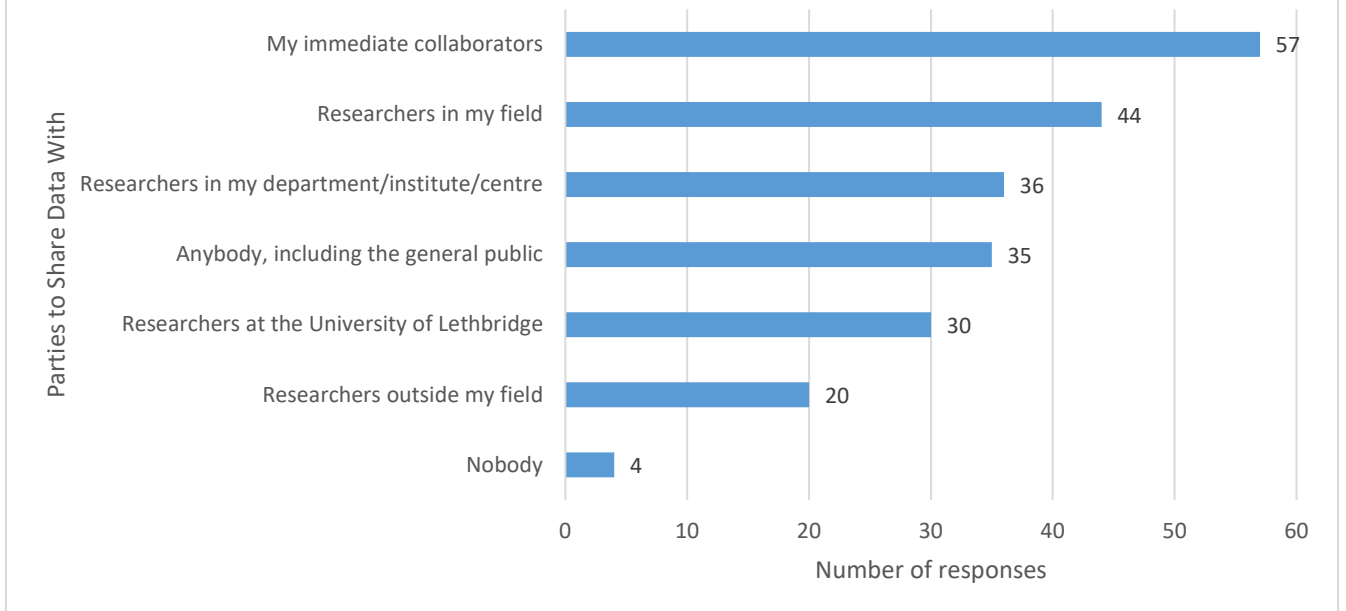
Questions 4.6, 7.4, and 10.4 asked all respondents about restrictions or embargoes that may have limited their ability to share their data with others. Respondents were able to select all the responses that applied to them. If there were no restrictions or embargoes, they were able to choose that as their response. Two main restrictions related to data being subject to privacy or confidentiality (33.3%) and needing to publish their data before sharing it (33%). While a variety of other reasons prevented researchers from sharing their data (see Figure 10), 29 (28.4%) respondents pointed out that there were currently no restrictions or embargoes on sharing their data with other parties.

Figure 10: Embargoes or Restrictions Limiting Ability to Share Research Data

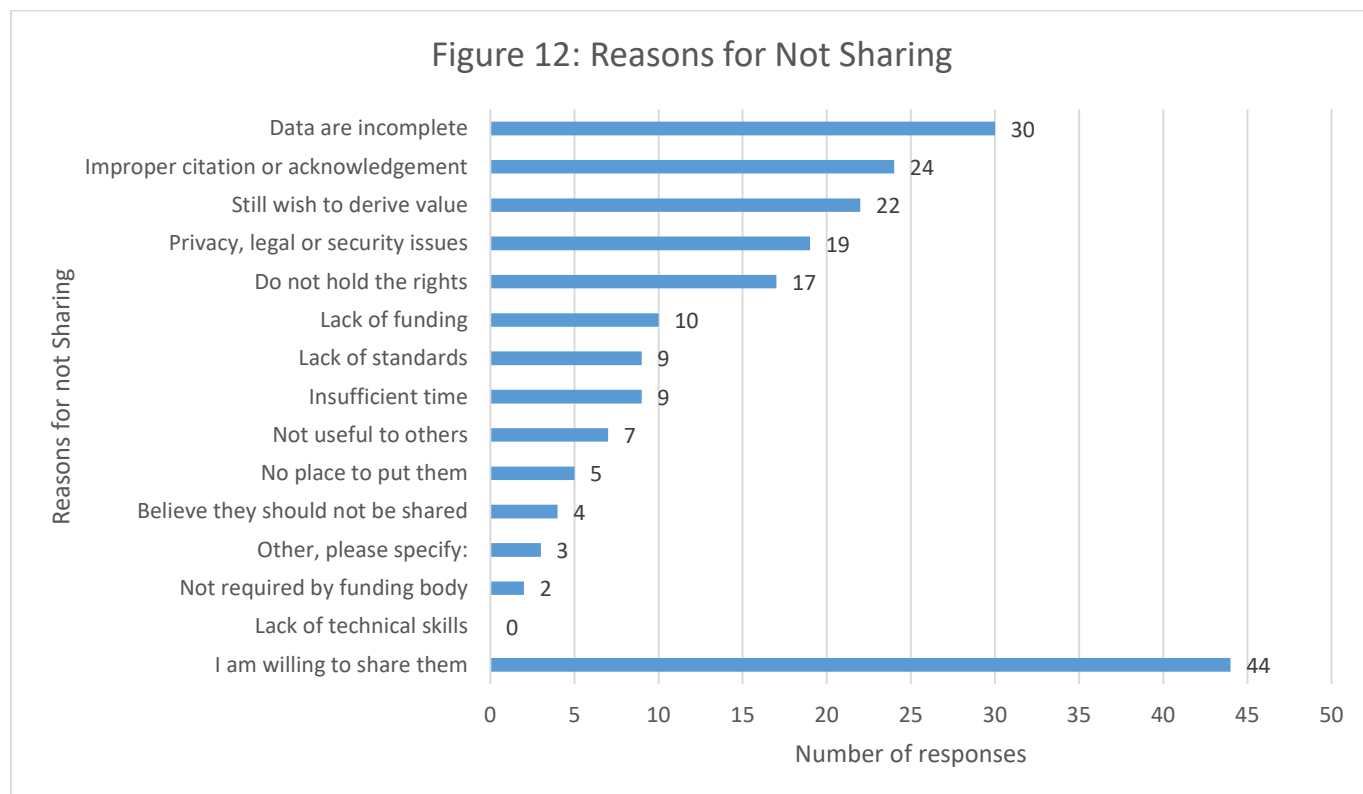


Questions 4.7, 7.5, and 10.5 asked all survey respondents who they would be willing to share their research data with if no embargoes existed (see Figure 11). They were able to select all the answers that applied to them. While the majority of responses indicated that sharing amongst collaborators (55.8%) and fellow researchers was a high priority (43.1%), 35 (34.3%) respondents said they would share their research data with anybody, including the general public.

Figure 11: Willing to Share Research Data With

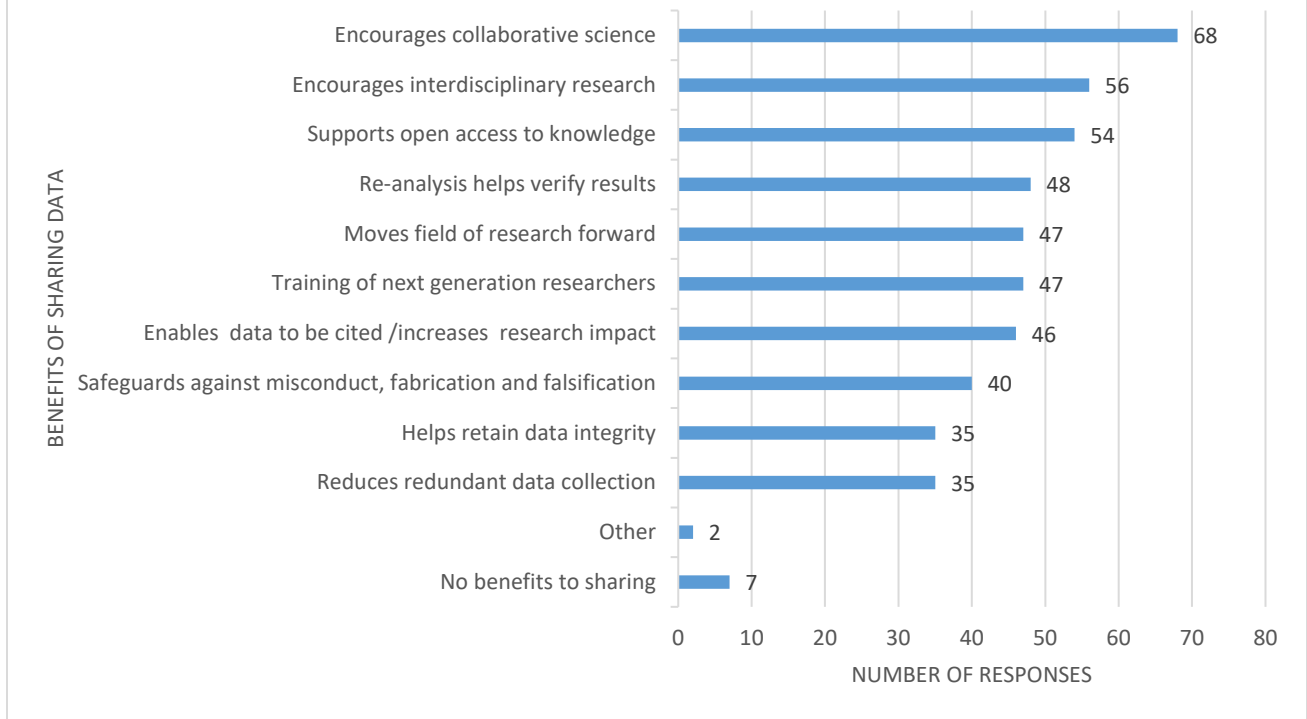


Questions 4.8, 7.6, and 10.6 asked respondents from all surveys to share their reasons for not sharing their research data (see Figure 12). Forty-four respondents indicated that they were willing to share their data. Having incomplete data (n=30) along with concerns over improper citation or acknowledgment (n=24) and a wish to continue to derive value from the data (n=22) were top reasons for not sharing data. For those that chose ‘other’, their reasons for not sharing research data included previous “egregious plagiarism” where their data and text was published by others as their own, and when their data was used without acknowledgement. Other reasons included research data sharing not being a common practice in their field, so it had never really been considered, as well as the lack of tools or algorithms developed to do so properly.



Questions 4.9, 7.7, and 10.7 asked respondents from all surveys to comment on the perceived benefits of sharing of research data. The top response was that researchers believed sharing research data encouraged collaborative science (n=68). Seven respondents did not think there were any benefits to sharing data, and two respondents cited “other benefits” including reducing the cost on licensing or alternatively, they did not know what the benefits would be for their particular data.

Figure 13: Perceived Benefits of Sharing Research Data



Respondents were also asked if they were aware of any discipline-specific research data repositories related to their field, and were asked to list them (Q 4.10, 10.8). Answers included DINAA and various corpora of language. Respondents also indicated that they were unaware of institutional repositories other than the University of Lethbridge’s research repository, OPUS.

For those responding to the science-specific survey, additional questions included asking about the respondent’s awareness of discipline-specific research data repositories related to their field. Responses included: BioModels, GIT, Synapse, GenBank, PDB, Brainstorm, and Human Metabolome Database, among others.

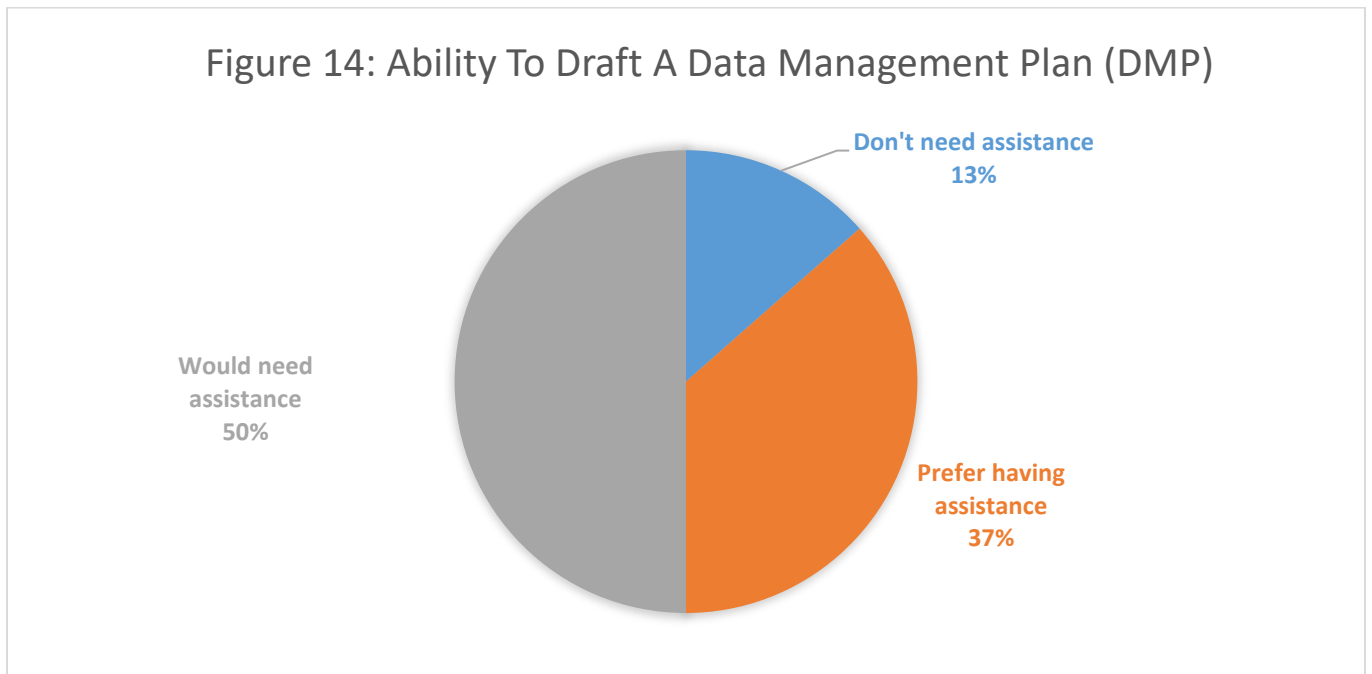
SECTION 4: FUNDING MANDATES AND RESEARCH DATA MANAGEMENT SERVICES

4.1 Interest in DMP Services

In anticipation of Tri-Council requirements to include data management plans (DMPs) with funding applications, there is value in gauging interest in the provision of data management services for the academic community. The University of Lethbridge Library has been providing workshops for the last two years aimed at helping faculty begin using data management best practices.

Q.22 asked respondents *“If you were asked to draft a data management plan as part of a grant application, which of the following statements would best describe your situation?”* (n=96).

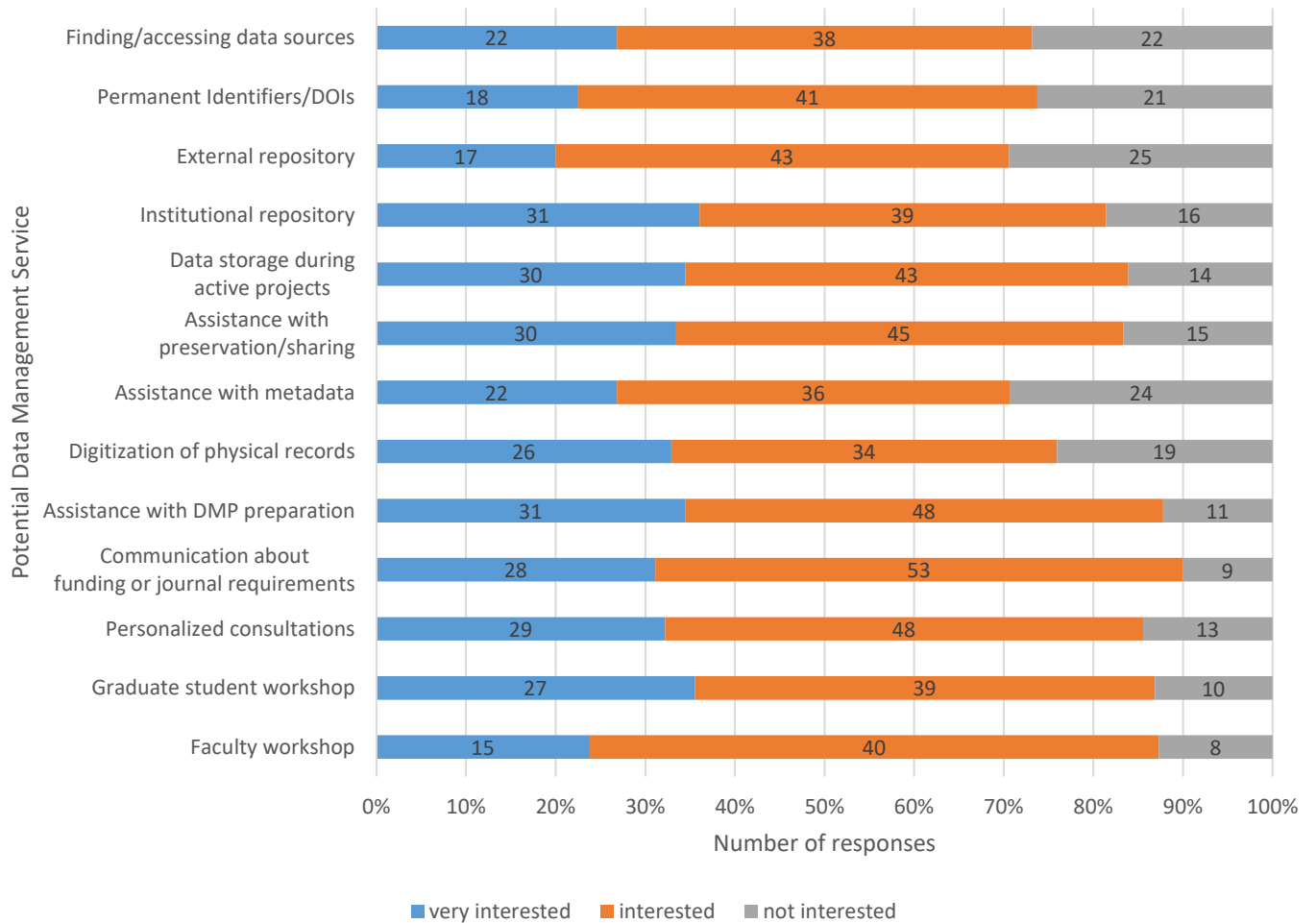
Only 13.5% (n=13) of all respondents indicated they would be able to create a DMP without assistance. The majority of respondents (87.5%, n=83) stated that they would prefer or need assistance and guided documentation to ensure the success of their application to a funding agency.



4.2 Interest in other proposed data management services

In order to gain a clearer understanding of needs related to research data management, participants were asked to rate a range of possible data management services (Q. 5.3, 8.5, 11.5). Ratings were on a three point scale that included *very interested*, *interested*, and *not interested*.

Figure 15: Data Management Plan (DMP) Assistance & Services
[All Surveys]



To simplify, the number of respondents who were ‘interested’ and ‘very interested’ were combined (‘n’) to generate a combined “interested” statistic (n divided by total responses for each service). Proposed services are ranked by interest in the following table:

Rank	Possible Data Management Supports	n	% ¹
1.	Communication and information about funding requirements and journal requirements regarding research data	90	90.0%
2.	Assistance preparing DMPs or assistance creating formal or documented data management plan	90	87.8%
3.	Workshops on best practices in data management for faculty	63	87.3%

¹ Percentage of responses for each service.

4.	Workshops on best practices in data management for graduate students	76	86.8%
5.	Personalized consultation on data management practices for specific research groups or projects	90	85.6%
6.	Data storage and backup during active research projects	87	83.9%
7.	Assistance with data preservation/sharing	90	83.3%
8.	Digitization of physical records	79	75.9%
9.	Assignment of permanent digital object identifiers (DOIs) for datasets	80	73.8%
10.	An institutional repository for long-term access and preservation of research data	95	73.7%
11.	Assistance in finding and accessing data sources	82	73.2%
12.	Assistance with documenting and describing data (i.e. metadata creation)	82	70.7%
13.	Assistance with depositing research data in external data repositories	85	70.6%

Interest among all the options of possible data management services was relatively high with a range 'interested' falling between 70.6% to 90%.

Other comments relating to potential data management services: Q5.4., Q8.6 Q 11.6

"I run linux servers for my lab data storage and would like an easy way to backup my data periodically on campus-managed storage (but would continue to use my local storage for day-to-day use)."

"Funding for ongoing management of open data bases. For example, who is going to answer questions about the database, how will we pay to host the database online, who will be responsible for general data maintenance and updates and how will this be paid for, who is going to deal with privacy issues/concerns with data? How will all of this extra work/institutional structures be funded? From my perspective making data open/public is a great idea in theory but will only be made possible if significant funds are allocated to covering the costs (including personnel costs) associated with preparing, managing, and maintaining public databases. This is not a small task that can be offloaded on researchers. The preparation and ongoing management of a public database is a full-time job in and of itself."

"I am aware of the tri-council RDM program and have started using it, I think it's great.

"Technologies to ensure data security."

"probably a discipline specific discussion on the appropriateness of data sharing and related considerations"

"Unlimited storage space for all data pertaining to academic research activities. Storage space should not be a limiting factor for research."

COMMENTS RECEIVED

The end of survey invited comments from the respondents, including feedback on the survey. In addition, respondents were able to leave an email address for follow up.

SCIENCE

- As a pure mathematician, I really don't deal with large quantities of data, ever. At most I write small computer programs to perform tedious computations; the results of these computations are not really the point of the research. I'm not sure if I should really be taking this survey...
- As first year graduate student, I am not sure my contribution is helpful in this survey.
- Good Survey, although it went a little over my head in terms of what it was asking
- I guess it's hard to answer some of these questions generally because we have so many types of data collection, from labbook, to image to multiple formats of data files. I had discussions in the past about digital labbooks and while they would have some benefits, I cringe at the thought of doing such a switch and worry about the additional limitations this puts on accurate documentation during experimentation. Overall an important concept to flesh out and support. I don't think it replaces the individual researchers responsibility to maintain their own data for possible future use/re-visits by self or other researchers.
- I think it's great that the University of Lethbridge is looking into the issue of data management and storage. We have a machine that can create 48Tb of data in 48 hours. This is a huge issue and is directly applicable to my research.
- I would just like to say that I believe that although sometimes data collection seems redundant, as with my own experience, you may find the person/company who collected the data the first time did not do so correctly. This also discourages me from using other peoples data and why most times I would opt for just collecting my own.
- If you are going to send future general surveys out to graduate students, please better define areas which are most applicable to graduate students.
- Limited storage and cost of required storage is an issue in Neurosciences.
- Other kinds of data include physical archives: museum specimens, etc
- Questions were redundant. Above my level of understanding.
- Thanks for the survey. I think this survey acts as a bridge between the researchers and the university. The University will have a clear idea of what the researchers actually need.
- The question made me stop and think about my data management practices. It's good to be mindful of how the data is stored and whether other researchers would be able to navigate folders with ease.

- While I am ideologically in favour of data sharing, I am concerned that the survey does not consider how competitive the research environment is, and how closely guarded most data is prior to publication.
- You likely get this from everyone, but my situation is unique and I may not need some of the more basic levels of support that others do. If there are external resources that I have ready access to, and can save me time and provide more data security, I would be very happy to use these.

HEALTH SCIENCE

- Completing the survey would have been easier if questions were on separate pages.
- I suggest you reconfigure your definition of research in your preamble since not all research is based on deductive, hypothesis testing methodologies..
- the questions seem to tap into data management of quantitative data - not many questions fit with qualitative work

SOCIAL SCIENCE & HUMANITIES

- Funding for ongoing management of open data bases. For example, who is going to answer questions about the database, how will we pay to host the database online, who will be responsible for general data maintenance and updates and how will this be paid for, who is going to deal with privacy issues/concerns with data? How will all of this extra work/institutional structures be funded? From my perspective making data open/public is a great idea in theory but will only be made possible if significant funds are allocated to covering the costs (including personnel costs) associated with preparing, managing, and maintaining public databases. This is not a small task that can be offloaded on researchers. The preparation and ongoing management of a public database is a full-time job in and of itself.
- I am aware of the tri-council RDM program and have started using it, I think it's great.
- probably a discipline specific discussion on the appropriateness of data sharing and related considerations
- Technologies to ensure data security.
- Unlimited storage space for all data pertaining to academic research activities. Storage space should not be a limiting factor for research.

CONCLUSIONS

Though the way researchers' data management needs vary across disciplines, especially in terms of storage needs and the desire to share their data openly, similar needs for training and guidance exist in each research area. Eighty-three respondents (87.5%) indicated that they would prefer or need more information and guidance in understanding research data management (RDM) best practices to navigate the requirements of the draft [Tri-Council Research Data Management Policy](#). The Library and other partners in research, such as ORIS, have an opportunity to increase RDM engagement for faculty and graduate students across the University of Lethbridge by creating opportunities for communication and training as well as understanding the storage and privacy needs of different faculties.

The responses gathered from the University of Lethbridge RDM Survey have been submitted to the National dataset. We look forward to consulting with University of Lethbridge researchers regarding available options and best practices for managing, preserving and sharing their research data.