

Report on the Faculty & Postdoctoral Survey on Research Data Management at the University of Ottawa

Report prepared by:
Melissa Cheung, Principal Investigator

Co-investigators:
Jessica McEwan
Patrick Labelle
Lindsey Sikora

March 2020



uOttawa

Bibliothèque
Library

Acknowledgements

This study would not have been successful without the help and support of colleagues. The survey team would like to thank Catie Sahadath, former Data Librarian, for her work on this project. Many thanks to Felicity Tayler, E-Research Librarian, and Chantal Ripp, Research Librarian (Data), for their feedback and guidance in preparing this report, and to library colleagues who provided suggestions to improve the survey questionnaires. We are also grateful for the support from past University Librarian Leslie Weir, from current University Librarian Talia Chung as well as from members of the Library's senior leadership. We would also like to acknowledge the contributions of Dave Weatherall, former Communications and Marketing Officer.

Executive Summary

Introduction

This report provides an overview of the Faculty and Postdoctoral Survey on Research Data Management conducted by librarians at the University of Ottawa (uOttawa). The survey was designed to:

- determine how uOttawa researchers manage and share research data beyond the duration of their project
- determine how the uOttawa Library might help to facilitate data management activities
- understand some of the differences in research data management practices and needs across disciplines and sub-disciplines

Survey participants were asked general questions related to research data management (RDM) as well as specific questions about potential RDM services that could be offered by the Library. The bilingual (English/French) survey was deployed in two phases targeting different disciplines. The surveys for Science and Engineering, and for Humanities and Social Sciences were available from November 14 to 25, 2016. The survey for Health and Medical Sciences was available from March 20 to March 31, 2017.

Key Findings

- Researchers need RDM services and infrastructure that facilitate collaboration, including international collaborations that may be subject to multiple RDM policies.
- Researchers indicated their intent to share their data in the future and would benefit from assistance in identifying appropriate methods of data sharing that comply with applicable policies related to research ethics. In particular, there is an interest in support for using institutional repositories to share research data.
- Researchers could benefit from assistance with data documentation for appropriate reuse of their data and to foster reproducibility in research to meet Tri-Agency expectations for excellence in data stewardship.
- Researchers need or would prefer assistance and/or guided documentation to prepare data management plans (DMPs).
- Researchers indicated they were not teaching RDM topics, which suggests a lack of education and training in the essentials of research data management for the next generation of researchers.

Background

As public funding agencies and publishers recognize the value in preserving research data, many are developing policies around data management, sharing, and preservation. In the US and the UK, funding mandates require that research groups submit a research data management plan (DMP) in order to secure funding. Similar mandates are expected in Canada, following the release of the Tri-Agency Statement of Principles on Digital Data Management¹ in 2016 and the draft Tri-Agency Research Data Management Policy² in 2018.

Librarians from universities across Canada, including uOttawa, are collaborating to conduct a series of surveys to better understand data management practices and needs in the natural sciences, engineering, social sciences, humanities, and health sciences.

Methodology

Survey Design

The survey instrument was initially developed by the University of Toronto Libraries for researchers in science and engineering. Members of the Canadian RDM Survey Consortium³ worked together to modify the survey, creating one version targeting researchers in the humanities and social sciences and another version for researchers in the health and medical sciences. The uOttawa research team further customized these three surveys to reflect the departmental structure and unique characteristics of the uOttawa academic community, including translating the surveys into French to address the bilingual nature of the University. As a result, the uOttawa survey included six questionnaires: one for each discipline, and each with separate English and French versions.

The science and engineering (SciEng) survey consisted of 22 questions. The humanities and social sciences (HUSS) survey included additional questions related to the digital humanities for a total of 24 questions. The health and medical sciences (HSM) survey consisted of 23 questions due to differences in how the questions were entered and numbered in the survey tool.

The questions in all three surveys were organized into four sections:

1. Demographics and general questions
2. Working with research data
3. Data sharing
4. Funding mandates and RDM services

Survey Distribution

The surveys were administered using FluidSurveys. The SciEng and HUSS surveys were deployed at the same time and were available from November 14 to 25, 2016. The HSM survey, which was developed at a later date, was available from March 20 to March 31, 2017.

A distribution list containing names, email addresses and language preferences of faculty members and postdoctoral fellows in each department was requested from uOttawa Human Resources. Personalized invitations to participate in the survey were sent by email from the University Librarian's Office to

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

² Draft Tri-Agency Research Data Management Policy - http://www.science.gc.ca/eic/site/063.nsf/eng/h_97610.html

³ Canadian RDM Survey Consortium - <https://portagenetwork.ca/network-of-experts/network-of-expertise/rdm-survey-consortium/>

3,213 faculty members and postdoctoral fellows. Researchers were invited to complete the appropriate disciplinary survey that corresponded to their faculty⁴ and in their preferred language. Completing the survey was entirely voluntary and anonymous.

Limitations of Research Methodology

The results of this survey provide insights into the RDM practices and attitudes of uOttawa faculty members and postdoctoral fellows who completed the survey. However, there are limitations in the survey design. Participants who completed the survey were self-selected, which may have led to bias in the results. Additionally, survey questionnaires were developed at different times and were further modified to reflect characteristics of the institution administering the survey. This resulted in inconsistent terminology appearing in the survey.

Furthermore, the timing of survey deployment may have contributed to a low response rate. The SciEng and HUSS surveys were distributed six months after a Library collections survey that was launched to inform the decision-making process of a cost-cutting exercise. This may have led to survey fatigue and a lack of trust in the Library's motivations for conducting another survey. At least one faculty member declined to participate in the RDM survey citing concerns that the Library would use survey results to cancel more services and resources as part of cost reduction. Meanwhile, the HSM survey was distributed at the end of the academic year before final exams, which may have led to a low response rate from the Faculty of Health Sciences and the Faculty of Medicine.

Given these limitations, caution must be taken when interpreting the results. The data discussed in this report are only representative of individuals who completed the survey and cannot be applied to the larger uOttawa academic community without further research and analysis.

Results

Survey Data

Research data, survey questionnaires, and relevant documentation are available at:

Cheung, Melissa; Sahadath, Catie; Labelle, Patrick; McEwan, Jessica; Sikora, Lindsey, 2017, "Faculty & Postdoctoral Survey on Research Data Management at the University of Ottawa", <https://doi.org/10.5683/SP/L1H3SS>, Scholars Portal Dataverse, V6.

Survey Status

Table 1 shows the status of survey responses obtained from FluidSurveys after the survey was closed. The total number of responses to the survey was 260 with 177 complete survey responses and 83 incomplete surveys. Of the 177 complete responses, 48 responses were in French and 129 were in English; 43 were received from the SciEng survey, 101 from the HUSS survey and 33 from the HSM survey.

A survey was counted as completed if the respondent answered all required questions and clicked the submit button. Participants were permitted to skip questions, except for the two required questions pertaining to rank and affiliated faculty, institute or department. An incomplete survey refers to a survey

⁴ Researchers in the Faculty of Education, Faculty of Law and Telfer School of Management were invited to complete the HUSS survey.

where some questions were answered but the respondent never finished the survey by clicking the submit button.

Based on the number of completed surveys, the response rate was only 5.5% (177/3213).

Table 1. Survey Status reported by FluidSurveys.

Survey	Invites	Responses	Incomplete	Completed
HSM EN	994	33	7	26
HSM FR	184	10	3	7
HUSS EN	898	96	29	67
HUSS FR	566	50	16	34
SciEng EN	381	57	21	36
SciEng FR	100	14	7	7

Survey Findings

The findings reported here are based on the 177 completed surveys. At least one individual from each Faculty at uOttawa completed the survey. Given the low response rate, especially from Law and Management, the survey data are not statistically significant and the findings cannot be generalized to the entire uOttawa academic community. However, the results may still reveal trends in researcher needs and/or gaps in knowledge that could be useful for future discussions regarding RDM support or services.

Section 1: Demographics and General Questions

As presented in Table 2, 133 of the survey respondents were professors (assistant, associate and full) and clinical colleagues, 7 were adjunct professors, 7 were part-time professors, 13 were lecturers, 1 was a professor emeritus, 1 was a research associate, and 13 were postdoctoral fellows.

Table 2. Respondent rank. Survey participants were asked to indicate their rank at the University of Ottawa (n=177).

Rank	Number of responses	% of respondents
Adjunct Professor	7	4%
Assistant/Associate/Full Professor/Clinical Colleague	133	75%
Lecturer	13	7%
Part-time Professor	7	4%
Research Associate	1	1%
Professor Emeritus	1	1%
Postdoctoral Fellow	15	8%

Table 3. Number of distributed surveys, sample responses and percentage of responses by faculty (n=177).

	Arts	Education	Engineering	Health Sciences	Law	Management	Medicine	Science	Social Sciences	Not Declared
# distributed surveys	519	152	194	236	198	137	942	286	549	
Sample Responses	39	7	20	10	3	6	23	22	46	1
% of Responses	8%	5%	10%	4%	2%	4%	2%	8%	8%	

Survey participants were asked to select their home department at uOttawa from a list of options. The responses were aggregated by faculty and shown in Table 3. The Faculty of Social Sciences had the highest number of respondents, representing 26% (46/177) of the total number of respondents, followed by the Faculty of Arts (39/177=22%). Meanwhile, the Faculty of Law (3/177=2%) and the Telfer School of Management (6/177=3%) had the lowest number of respondents. One respondent who completed the SciEng survey did not specify their department or faculty.

It is interesting to note that, overall, the Faculty of Engineering did not represent the largest number of survey respondents. However, they represented the largest sample of the target population for their faculty. On the other hand, respondents from the Faculty of Medicine, which included medical doctors employed at affiliated hospitals, represented the smallest sample of the target population.

Section 2: Working with Research Data

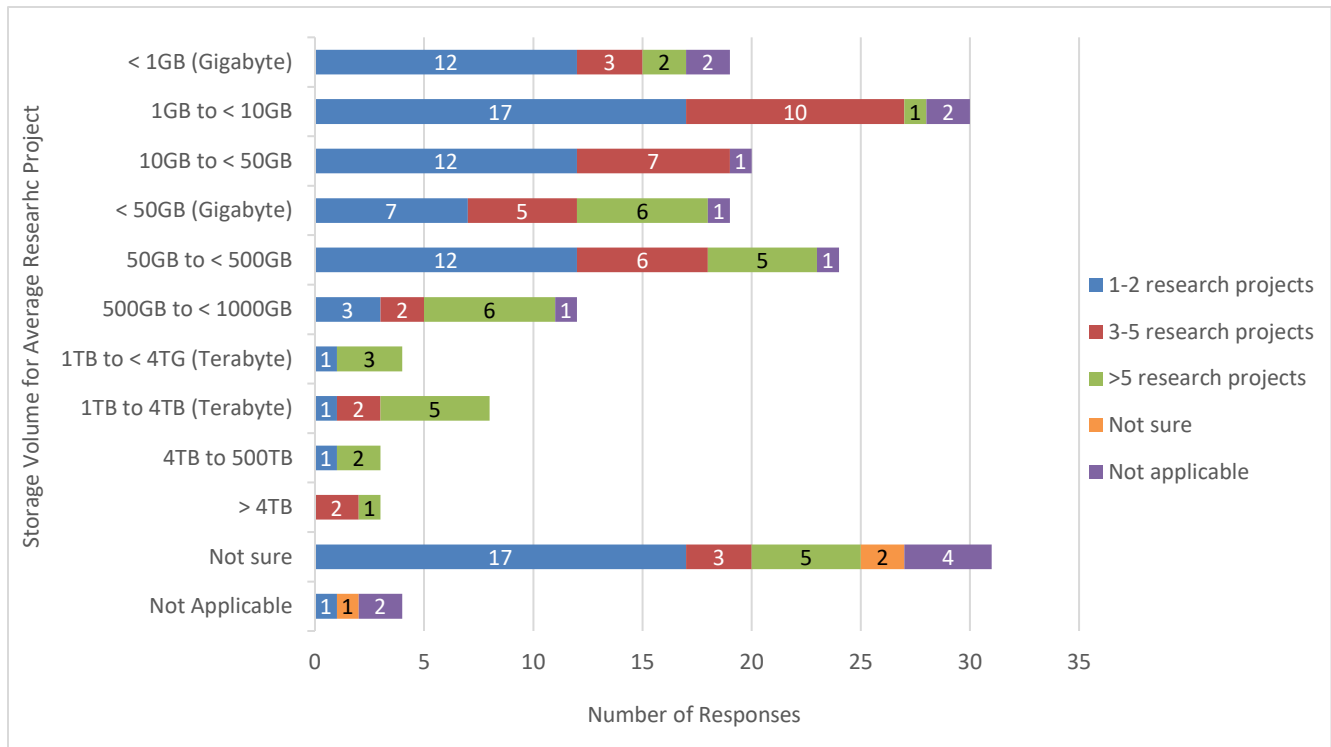


Figure 1. Amount of storage in relation to number of research projects. Survey participants were asked to estimate the amount of storage they use in an average research project as well as the number of research projects they lead in the past year (n=177).

The relationship between storage volume required for the average research project and the number of projects led by the researcher in the past year is shown in Figure 1. At the time the survey was conducted, most research projects required less than 1,000 GB of storage (124/177=70%). However, approximately 18% (31/177) of respondents were not sure how much storage they used in an average project.

Two respondents answered “Not Applicable” to both of these questions because they did not work with the kind of research data defined in the survey introduction or they were not advising research projects at the time of the survey. One respondent noted they were “Not sure” how many research projects they lead in a year and they did not require storage (“Not Applicable”) because they did not use the type of data defined in the survey introduction. One respondent, who led 1-2 research projects in the past year, noted that they did not require storage (“Not Applicable”) because they stored their data at home.

Researchers in Arts and Social Sciences represented most of the respondents who were unsure of the amount of storage they used in an average research project (survey responses by faculty are included in the [Appendix](#)). Meanwhile, respondents from the Faculty of Science and the Faculty of Engineering all provided an estimate for the amount of storage they used; the most frequent response was less than 50 GB. These results suggest that researchers in Arts and Social Sciences may require more support in identifying the amount of data storage they need to develop an appropriate data management plan.

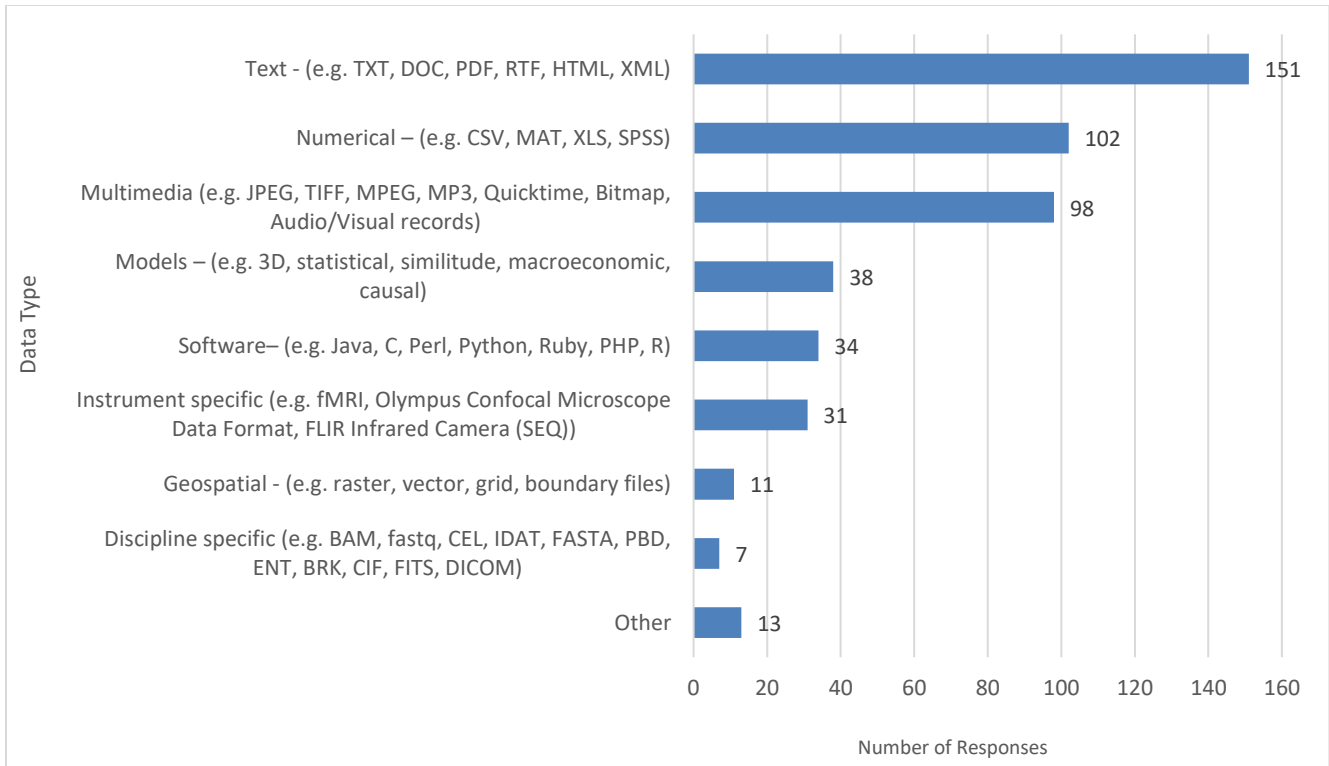


Figure 2. Type of research data generated or used. Survey participants were asked to indicate all of the applicable types of research data they generate or use in a typical research project (n=177).

As Figure 2 illustrates, the most frequent data types generated or used by survey respondents were text, numerical and multimedia. When given examples, respondents in the Faculty of Medicine were the only ones to select discipline specific (e.g. BAM, fastq, CEL, IDAT, FASTA, PBD, ENT, BRK) as a data type. Respondents in other disciplines characterized their research data using the more common types ([Appendix](#)).

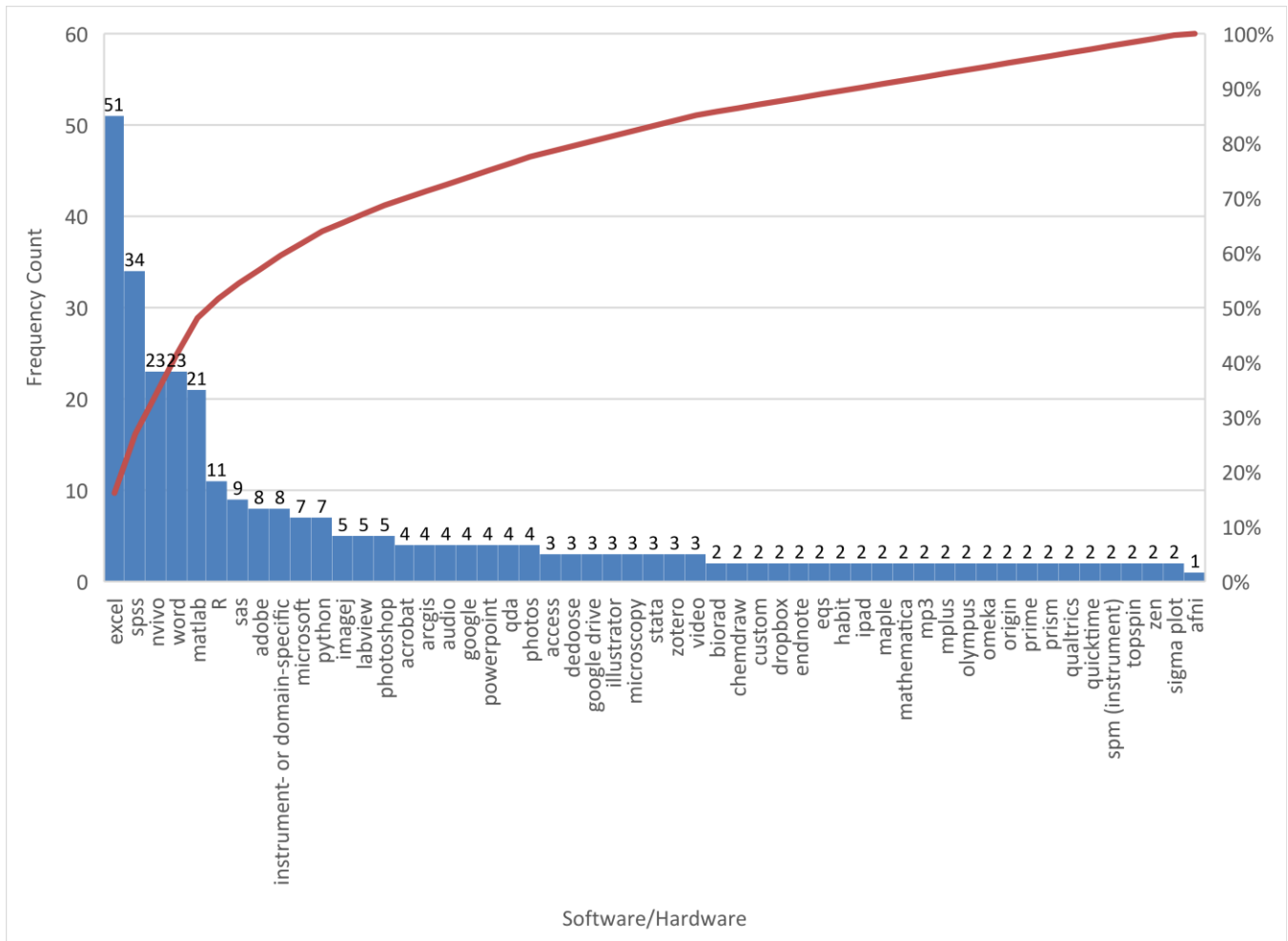


Figure 3. Software or hardware mentioned most frequently in response to the question “Please list any software and/or hardware used for the collection, analysis, or manipulation of your research data, if applicable”.

Survey participants were asked to list any software and/or hardware they used for the collection, analysis, or manipulation of research data. The text responses received, consisting of 202 terms, were analyzed using Voyant Tools⁵ to count the number of times a particular software/hardware was named⁶. Figure 3 shows the software/hardware that were mentioned more than once in the responses received.

At the time of the survey, Excel (51), SPSS (34), NVIVO (23), Word (23) and MatLab (21) were the most frequently named software/hardware used for the collection, analysis or manipulation of research data. These results are unsurprising, given that text and numerical data were identified as the most commonly generated or used data type (Figure 2). It should also be noted that nearly half of the survey respondents were from the Faculty of Arts and the Faculty of Social Sciences (Table 2), which may have skewed the results in Figure 3 towards software/hardware used most frequently in those disciplines. Therefore, the frequency of software/hardware used in other disciplines may be underrepresented.

⁵ Sinclair, S.& G. Rockwell. (2019). Trends. Voyant Tools. Retrieved December 22, 2019, from <http://voyant-tools.org>

⁶ The number of times statistical computing software R was named was counted manually since Voyant Tools did not recognize it as a word or term.

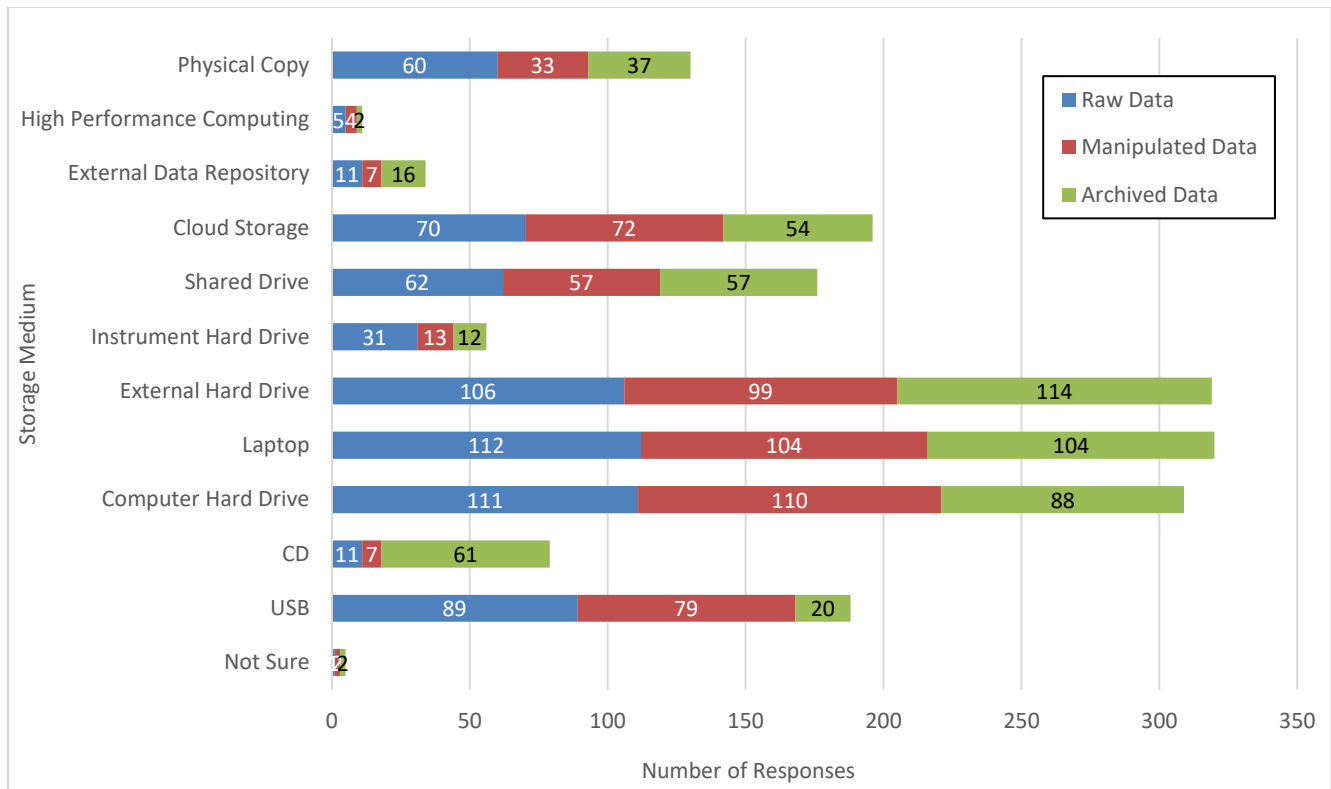


Figure 4. Data storage. Survey participants were asked to select all the applicable options for where they store raw data, manipulated data (e.g. converted, curated, processed), and archived data (e.g. long-term storage or preservation) from current project(s) (n=176).

The survey results illustrated in Figure 4 show that researchers used a variety of options to store their research data at different stages of the research data lifecycle⁷. Survey respondents indicated that they stored their data primarily on external hard drives, laptops and computer hard drives. On the other hand, external data repositories (e.g. Protein Data Bank, Cambridge Structural Database, GitHub, Dryad, Figshare) were not used often as a data storage option⁸. Flash drives/USBs were used primarily to store raw and manipulated data, while CDs/DVDs were generally used to store archived data.

At the time of the survey, researchers in the Faculty of Arts and the Faculty of Social Sciences were the only ones who were “Not Sure” where they stored their research data ([Appendix](#)). This suggests that they would benefit from additional support in data management planning and in identifying appropriate data storage options for their research. Meanwhile, Engineering, Medicine, Science and Social Sciences were the only disciplines that made use of grid/high performance computing (HPC) centres for storage. Only one respondent in the Faculty of Social Sciences selected grid/HPC centres to store archived data.

⁷ <https://biblio.uottawa.ca/en/services/faculty/research-data-management/what-research-data-management>

⁸ The question “Are you aware of any discipline-specific data repositories related to your field? Please list. If you are not aware of any discipline-specific data repositories related to your field, please say ‘none’.” was asked in the Data Sharing Section of Survey. Text responses to this question were removed from the data set and are not included in this analysis.

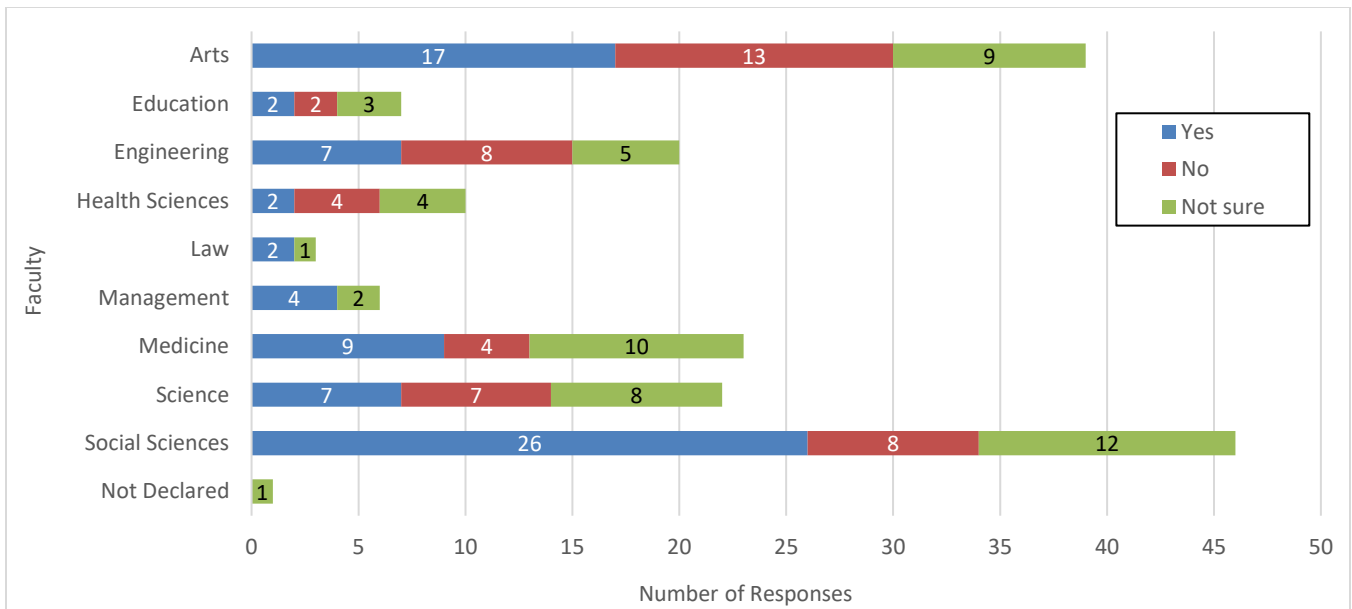


Figure 5. Documentation for data reuse. Survey participants were asked whether there is sufficient documentation and description for another person outside the research team to understand and use their research data (n=177).

When asked whether they were creating sufficient data documentation for another person to understand and make use of their research data (Figure 5), 43% of respondents said “yes” (76/177), 26% of respondents said “no” (46/177), and 31% of respondents were “not sure” (55/177). Social Sciences was the only discipline with a greater number of positive responses over negative responses compared to other disciplines. These results suggest that researchers require additional support in creating documentation for data reuse, which is an expectation of the Tri-Agency Statement of Principles on Digital Data Management.

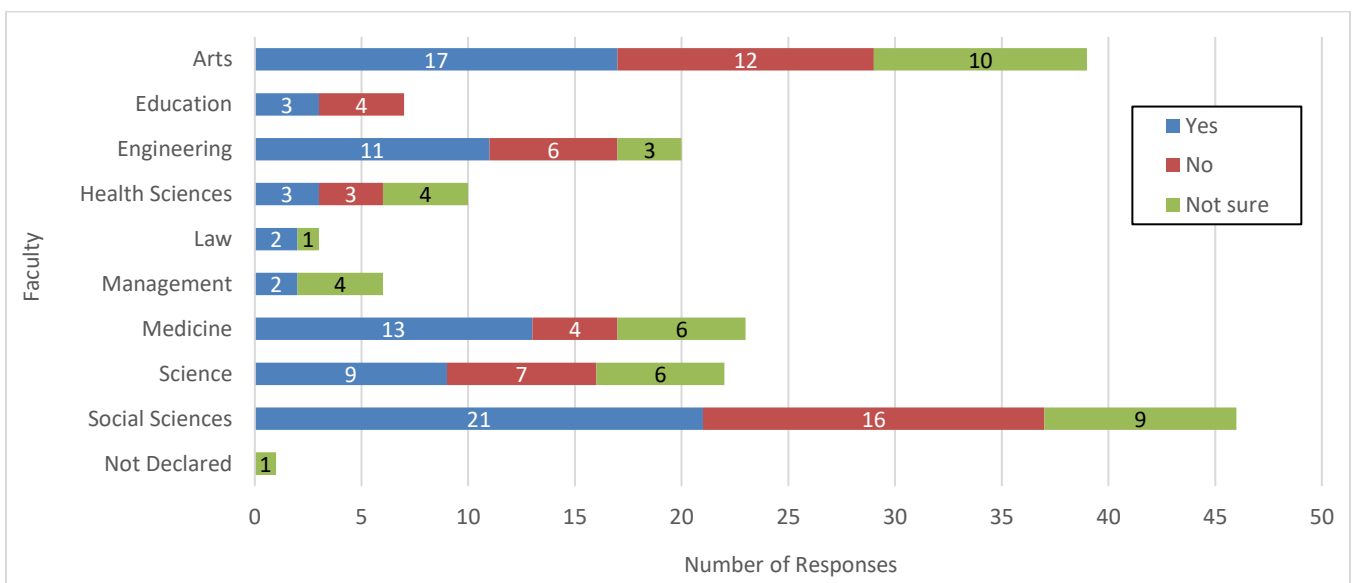


Figure 6. Documentation for data reproducibility. Survey participants were asked whether there is sufficient documentation and description for another person outside the research team to replicate the methodologies that produced their research data (n=177).

When asked whether they were creating sufficient data documentation for another person to replicate the methodologies that produced their research data (Figure 6), 46% of respondents said “yes” (n=81), 29% of respondents said “no” (52/177), and 25% of respondents were “not sure” (44/177). Researchers in Social Sciences were less confident that they had produced sufficient documentation for reproducibility compared to data documentation produced to understand and make use of data (shown in Figure 5). Meanwhile, researchers in Science and Engineering were slightly more confident that they had produced sufficient data documentation for reproducibility compared to documentation for understanding and making use of their data.

These results suggest that researchers could benefit from additional support in creating documentation for data reproducibility to meet the Tri-Agency’s expectation that experiments and studies be replicable to meet standards in research excellence as described in the draft Tri-Agency RDM policy.

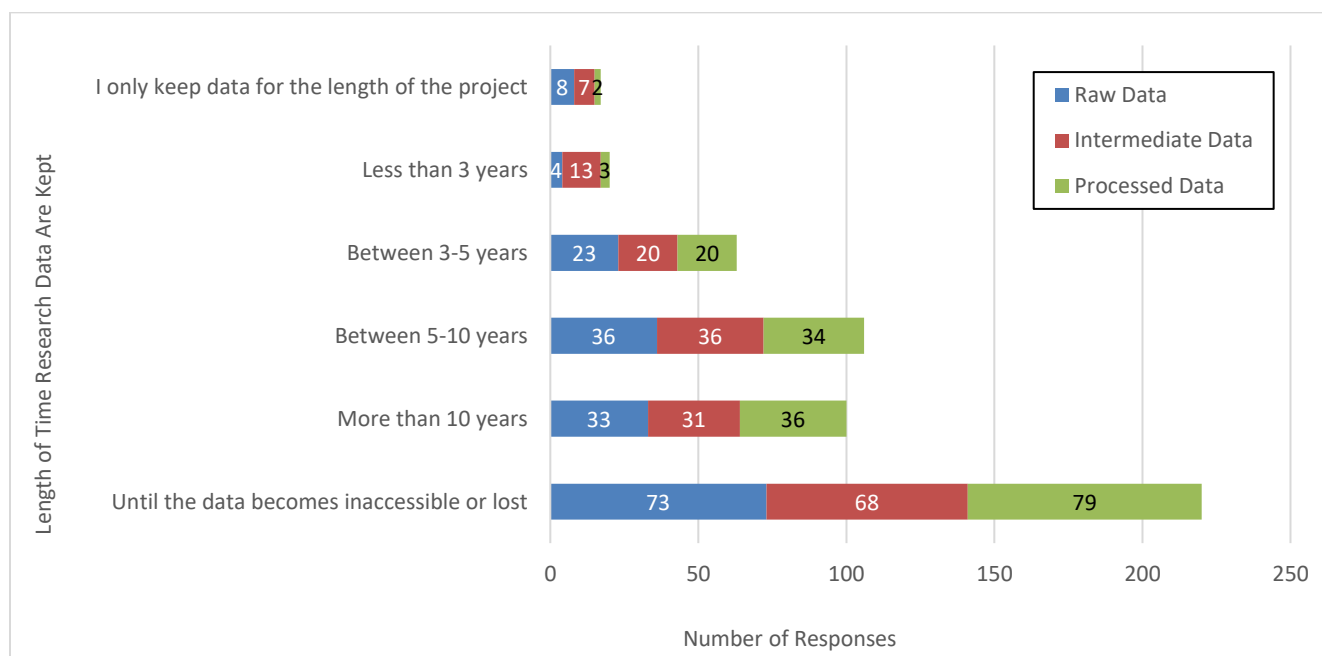


Figure 7. Length of Time Research Data Are Kept. Survey participants were asked to indicate how long they typically intentionally keep “source material/survey results/raw data” (n=177), “intermediate/working data” (n=175), and “processed data ready for publication” (n=174) after project completion.

The majority of respondents indicated they typically intentionally kept their research data until the data becomes inaccessible or lost (Figure 7). However, survey participants were not asked about the data curation practices they used to actively preserve their data for as long as possible. Therefore, the intention to keep research data until it becomes inaccessible or lost may be the default practice among researchers who do not take action to dispose or destroy their data (e.g. compliance with research ethics requirements).

Researchers in Arts, Management and Social Sciences intentionally kept their raw data for the length of their project; researchers in all the other disciplines intentionally kept their raw data beyond completion of their project ([Appendix](#)). Researchers in Engineering and Science indicated that they kept raw data less than 3 years, while Education, Law, Health Sciences and Medicine kept raw data for longer than 3 years. Researchers in Arts, Engineering, Health Sciences and Social Sciences specified that they kept

intermediate data only for the length of the project; all other disciplines kept data beyond completion of project. Notably, Law kept intermediate data for at least 3 years, while Education kept intermediate data for more than 5 years.

When it comes to processed data ready for publication, only researchers in Social Sciences intended to keep processed data for the length of the project; all other disciplines intended to keep processed data beyond the completion of the project. In particular, Engineering and Health Sciences indicated that they kept processed data for less than 3 years, while all other disciplines intended to keep processed data longer than 3 years.

Section 3: Data Sharing

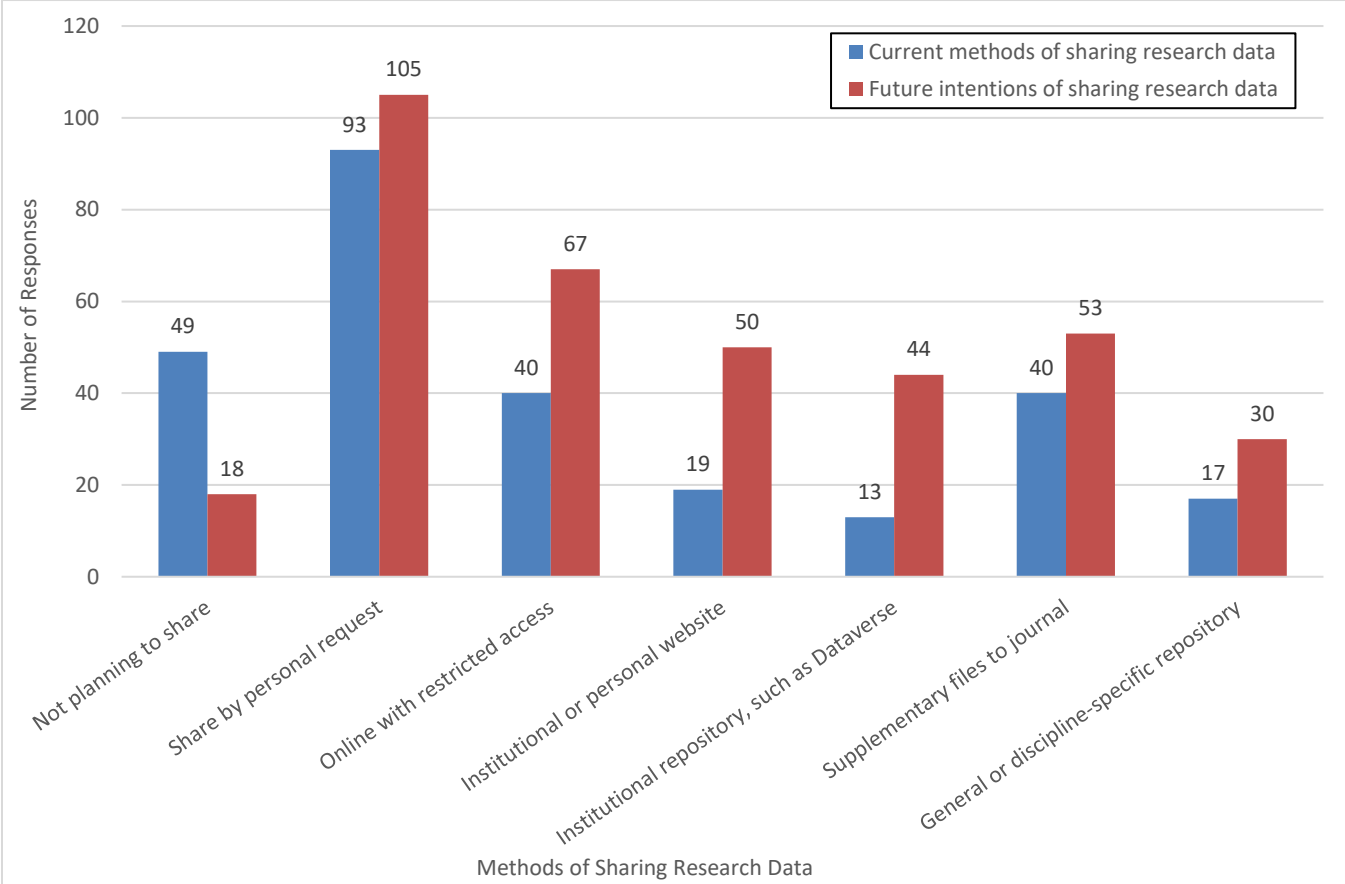


Figure 8. Comparison of current methods and future intentions of sharing research data. Survey participants were asked to select all the applicable methods of sharing research data they currently use (n=177) and would consider using in the future (n=176).

As Figure 8 illustrates, at the time of the survey, 28% (49/177) of respondents were not currently sharing their data. However, only 10% (18/176) were not planning to share their data in the future, which indicates that more researchers would consider sharing their data in the future. Share by personal request was the most frequent method of data sharing with 53% (93/177) of respondents currently sharing their data using this method and 50% (105/176) of respondents reporting they would consider sharing their data in the future using this method.

In general, data repositories were not frequently used to share research data. Researchers in Management and Education were not using data repositories (institutional, general, or discipline-specific), while Science and Engineering made use of general or discipline-specific repositories, but not institutional repositories ([Appendix](#)). Researchers in Engineering were also more likely to share by personal request only over any other method of sharing. In Science and Medicine, respondents indicated they were more likely to share as supplementary files to journal publication over any other method, which is not surprising since this is a fairly well-established practice in these disciplines.

Comparing current practices at the time of the survey and future intentions of data sharing, the results showed an increased interest in using a variety of mechanisms for data sharing. Notably, 25% of respondents (44/176) indicated willingness to consider using an institutional data repository, such as Dataverse, in the future while only 7% (13/177) were using an institutional repository to share their data at the time of the survey. Notably, in the context of this survey, “institutional repository” referred to an institutional data repository (i.e. uOttawa’s instance of Dataverse) and not a repository for self-archiving of research publications and/or conference materials (ie uOttawa’s uOResearch). The survey results may therefore reflect confusion over the meaning of “institutional repository”. That said, the results suggest there is interest from researchers in using an institutional repository for data sharing to achieve compliance with data deposit requirements.

Researchers in Education, Law, and Science all indicated intent to share their data in the future, whereas some researchers in other disciplines responded that they were not planning to share their data in the future ([Appendix](#)).

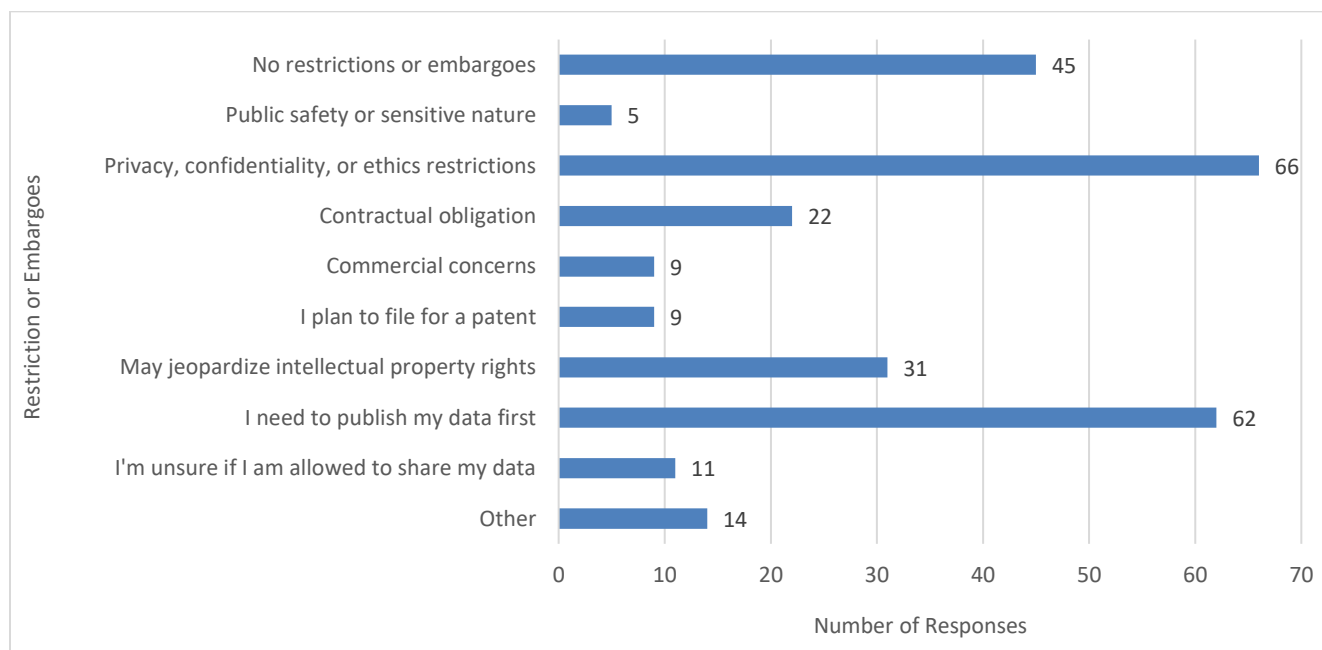


Figure 9. Restrictions or embargoes on data sharing. Survey participants were asked to select all applicable restrictions or embargoes that may limit their ability to share their data with others (n=176).

When asked about restrictions or embargoes that may limit the ability to share their data with others, “privacy, confidentiality, or ethics restrictions” was the most frequently chosen response, followed by the need to publish before sharing the data (Figure 9). Meanwhile, 25% (45/176) of respondents

indicated that there were no restrictions on embargoes that may limit the ability to share their data with others, and 6% (11/176) of respondents were unsure if they were allowed to share their data.

An analysis of restrictions or embargoes identified by each faculty ([Appendix](#)) revealed that researchers in all disciplines were affected by privacy, confidentiality, or ethics restrictions. Arts and Social Sciences were the only disciplines that had restrictions because their data were a matter of public safety or of a sensitive nature. They were also the only disciplines to indicate that they were unsure if they were allowed to share their data at the time of the survey; researchers in all other disciplines were aware of whether or not there were any restrictions on sharing their data. These results suggest that researchers in Arts and Social Sciences may require more support in navigating policies related to data sharing and research ethics requirements compared to other disciplines.

Interestingly, researchers in Management, Law, and Education were the only disciplines that indicated they were not concerned about the need to publish their research before sharing their data (i.e. being scooped).

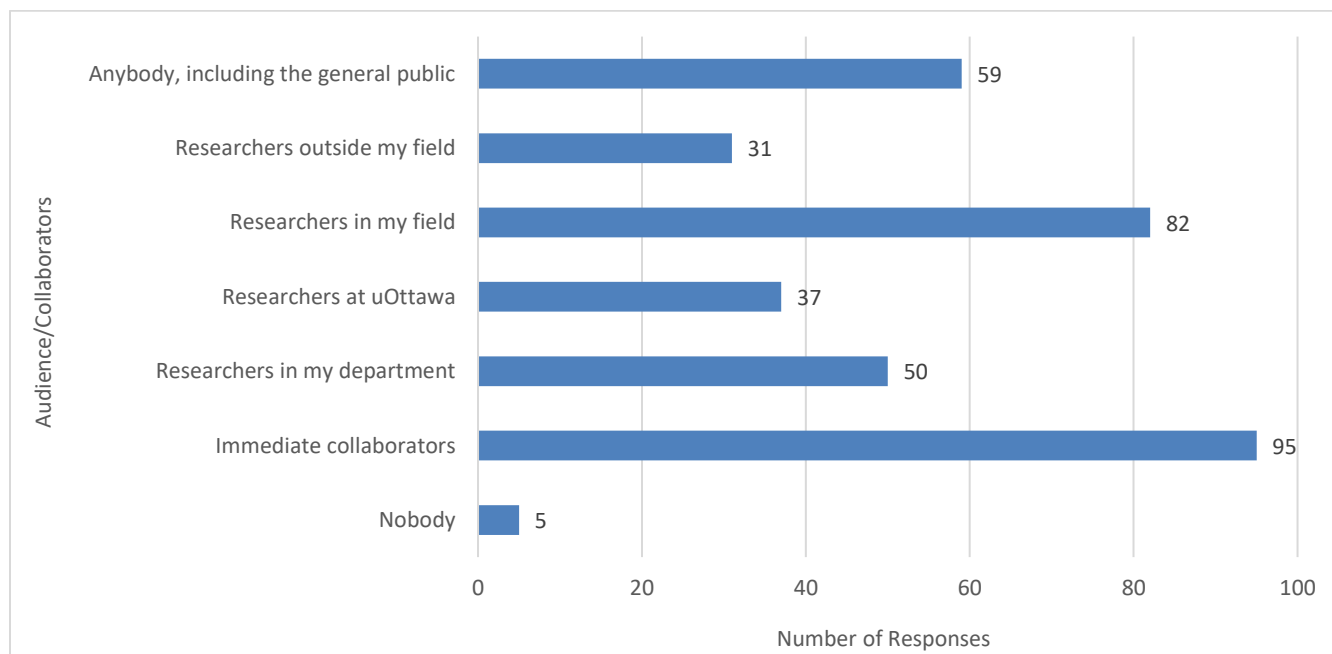


Figure 10. Audience/Collaborators for data sharing. Survey participants were asked to select all the applicable audiences and/or collaborators with whom they would be willing to share their data with if they were not affected by restrictions or embargoes (n=173).

If their data were not affected by restrictions or embargoes, survey participants indicated they would be most willing to share their data with immediate collaborators, followed by researchers in their field (Figure 10). Only 5 participants indicated they were not willing to share their data (“nobody”). As the survey did not ask respondents to specify where their immediate collaborators were located it is not possible to deduce whether survey respondents had collaborators outside Canada. However, given the rise of inter-disciplinary and inter-institutional research collaborations, it is highly possible that researchers at uOttawa may need to comply with multiple RDM policies, mandates, and/or practices, including those of international collaborators, and this will have implications on the development of RDM services to support researchers in the uOttawa context.

Researchers in Arts were most willing to share with anybody, including the general public, compared to researchers in other disciplines ([Appendix](#)), followed by researchers in Engineering and Science. Survey respondents from the Telfer School of Management (n=6) indicated that they would only be willing to share their data with specific audiences or collaborators; none of them indicated they would be willing to share with anybody.

Researchers in Medicine were more likely to share with other researchers in the same department (12/23=52%) compared to other disciplines, which may suggest a higher degree of departmental collaboration than other disciplines.

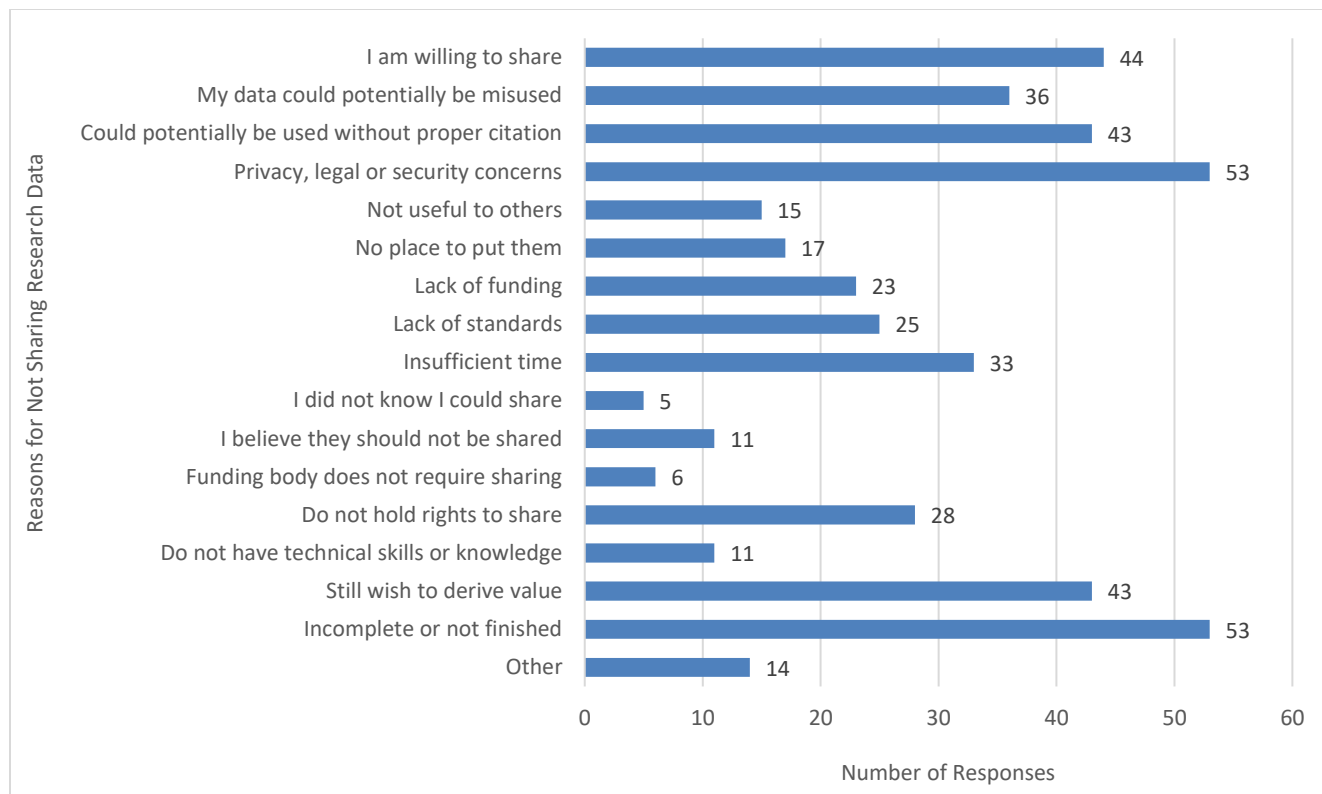


Figure 11. Reasons for not sharing research data. Survey participants were asked to indicate all applicable reasons why they would not be willing to share their research data and associated methods/tools (n=173).

When asked about reasons why they would not be willing to share their research data and associated methods/tools, survey respondents cited incomplete or not finished data and privacy, legal or security issues as the primary reasons (Figure 11). Respondents also indicated that they still wished to derive value from their data and were concerned that their data could potentially be used without proper citation or their data could be potentially misused. On the other hand, 25% (44/173) of respondents were willing to share their data.

At the time of the survey, 55% of the 20 respondents from the Faculty of Engineering were most willing to share their data ([Appendix](#)). However, they indicated that the most frequent reasons for not sharing were due to incomplete data or they still wished to derive value from their data. These results suggest that researchers in Engineering could benefit from training and support to identify ways to share their

data to comply with the anticipated Tri-Agency RDM policy requirements regarding data deposit, while mitigating the risks of being scooped.

Researchers in Social Sciences were least willing to share their data with only 3 of 46 (7%) respondents indicating that they were willing to share their data. Researchers in Education and Social Sciences were the most concerned about privacy, legal or security restrictions on data sharing. Researchers in Arts and Medicine also frequently cited privacy, legal and security concerns as a reason for not sharing data.

Researchers in Arts, Health Sciences, Medicine, Science and Social Sciences were more frequently concerned about improper citation compared to other disciplines. Additionally, researchers in Health Sciences, Science and Social Sciences were more frequently concerned about misuse of data compared to other disciplines. These results suggest that researchers in these disciplines could benefit from training and increased awareness of topics such as proper data citation and improving data documentation practices to lower the risks of data misuse.

Although the sample responses from the Faculty of Law (n=6) were too low to determine any trends, respondents indicated that the reasons for not sharing their data included incomplete data, a lack of standards to make them usable by others, a lack of funding and nowhere to deposit their data. These results suggest there may not be a culture of data sharing in Law at the time of the survey, but there is an interest in developing one because respondents indicated they were interested in sharing their data in the future ([Appendix](#)).

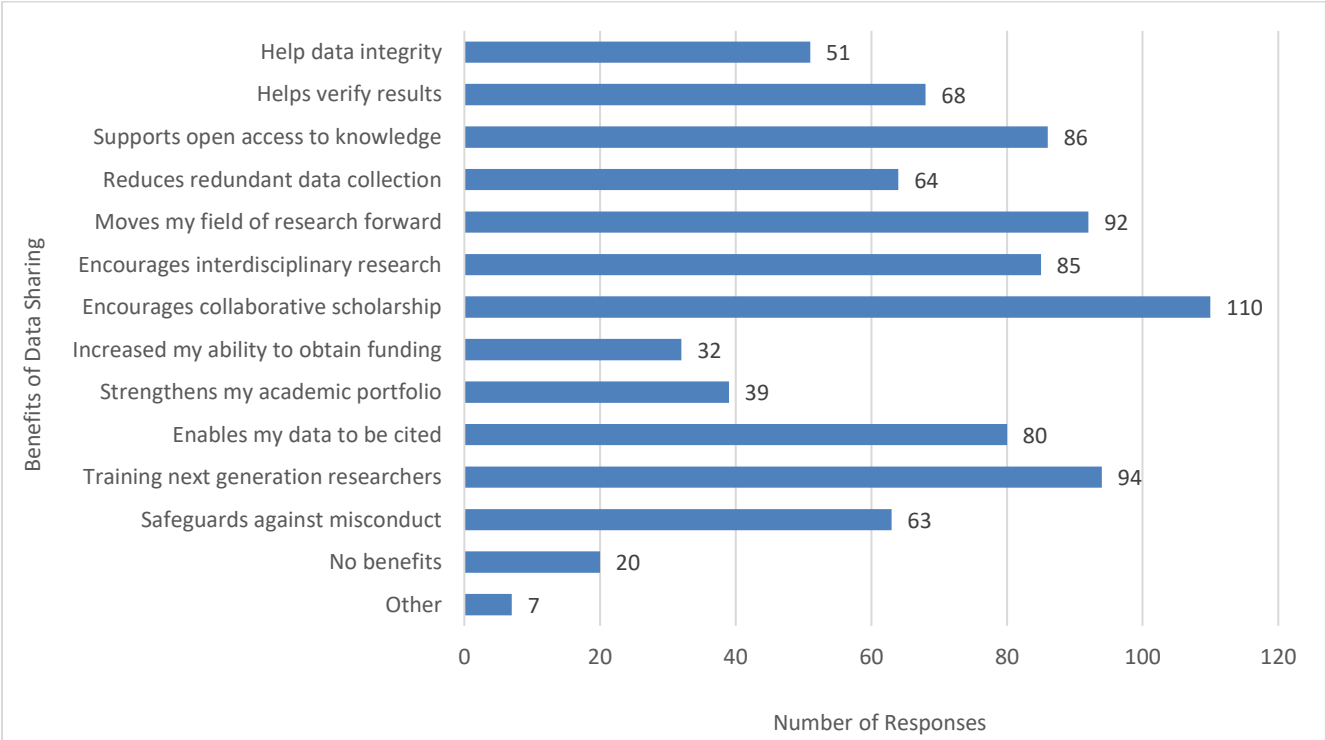


Figure 12. Benefits of sharing research data. Survey participants were asked to select all the applicable benefits they see to sharing their research data (n=174).

When asked what benefits they see to sharing their research data, “encourages collaborative scholarship”, “training the next generation of researchers” and “moves my field of research forward” were the most frequently selected responses as illustrated in Figure 12. These results correspond to the results illustrated in Figure 10, which suggest that the primary reason researchers share their data is to collaborate on their research projects. Therefore, RDM services and infrastructure should facilitate research collaborations, including international collaborations that may be subject to multiple RDM policies.

Although most respondents saw some benefits to sharing their data, twenty respondents (20/174=11%) indicated that they saw no benefits to sharing their data. These respondents were from disciplines other than Science and Law ([Appendix](#)).

Section 4. Funding Mandates and Research Data Management (RDM) Services

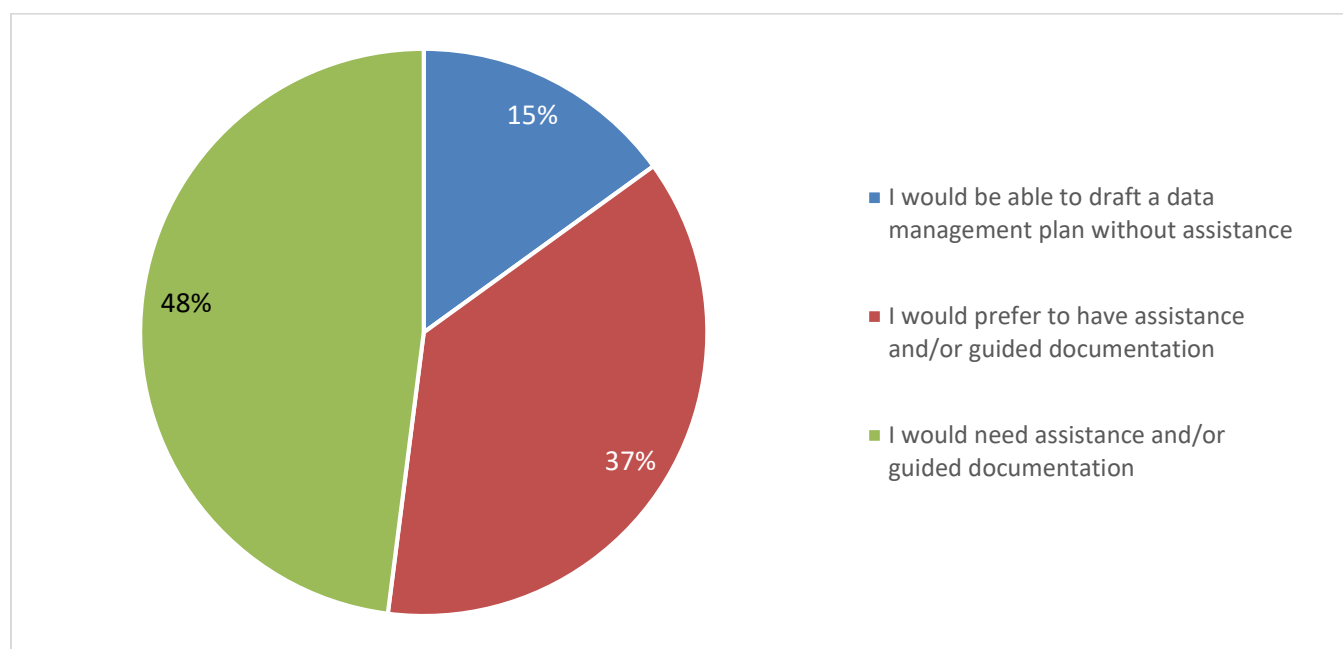


Figure 13. Ability to draft a data management plan. Survey participants were asked whether they would be able to draft a data management plan (DMP) as part of a grant application (n=173).

Figure 13 shows that 15% of respondents (26/173) indicated they would be able to draft a data management plan (DMP) without assistance; 37% (65/173) would be able to draft a DMP, but would prefer to have assistance and/or guided documentation; and 48% (83/173) would need assistance and/or guided documentation to appropriately address the sections of a DMP. These results clearly indicate an interest in assistance and guidance for preparing DMPs to meet anticipated Tri-Agency RDM requirements and should be a priority in the development of RDM services.

Researchers in Science and Engineering were more likely to need assistance to draft a DMP compared to other disciplines ([Appendix](#)). Responses from researchers in other disciplines were more evenly divided between “need assistance” and “prefer assistance”. These results suggest that researchers in Science and Engineering may not be accustomed to planning and/or describing their data management strategies. Whereas researchers in other disciplines working with data that are subject to research ethics policies

may be more familiar with certain aspects of planning adequate data management strategies, such as privacy, legal or security concerns ([Appendix](#)).

Taken together, our survey results suggest that researchers in all disciplines would benefit from guidance to address different sections of a DMP. For example, researchers in Arts and Social Sciences may need more assistance to identify storage requirements ([Appendix](#)) and to determine data sharing permissions ([Appendix](#)), while researchers in Science and Engineering may need more assistance in articulating the specifics of their data management strategies in a DMP.

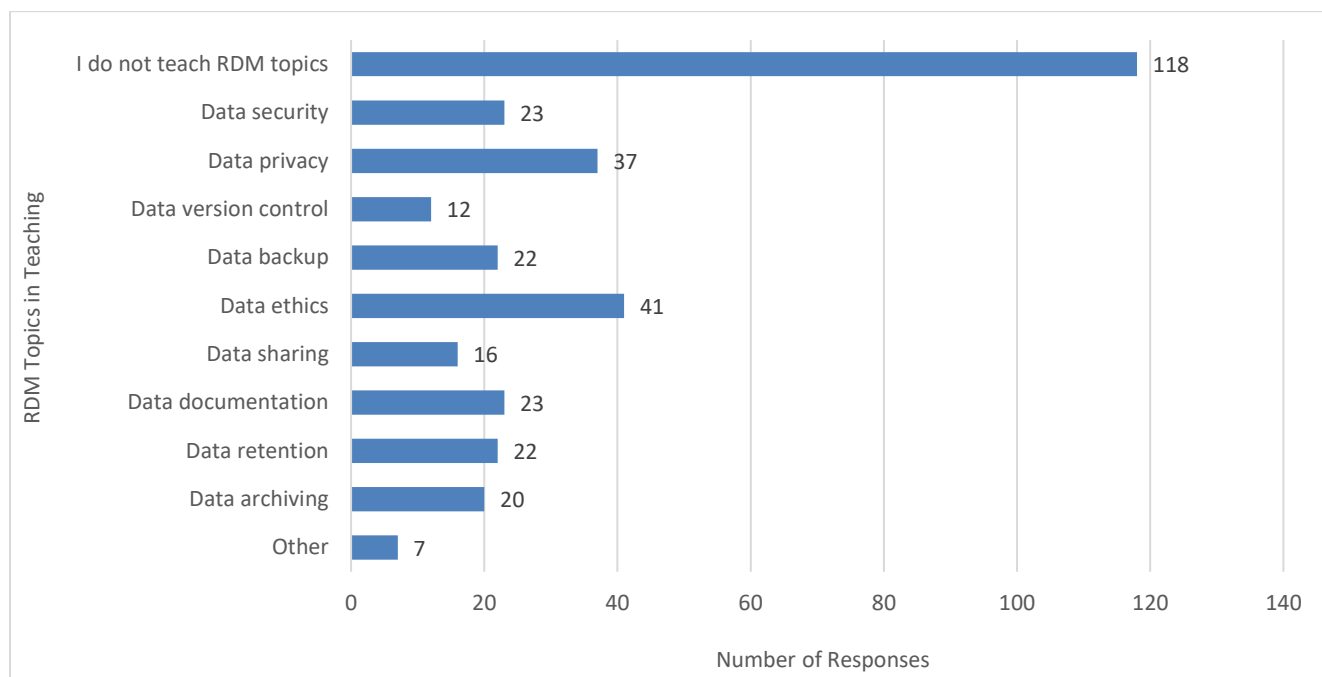


Figure 14. RDM topics in teaching. Survey participants were asked to select all the applicable topics related to RDM they include in their teaching practice (n=173).

The majority of respondents (118/173=68%) indicated that they do not teach RDM topics as illustrated in Figure 14.

Overall, the most popular RDM topics taught by respondents were data ethics and data privacy. When comparing disciplinary differences ([Appendix](#)), researchers in Arts, Engineering, Management, and Social Sciences indicated that they included all the topics listed in their teaching practice. While researchers in Science excluded data security and researchers in Medicine excluded version control from their teaching practice but included all the other topics listed. Researchers in Health Sciences did not teach version control and data documentation. Finally, researchers in Education taught the fewest number of RDM concepts only including security, privacy, ethics and retention in their teaching practice.

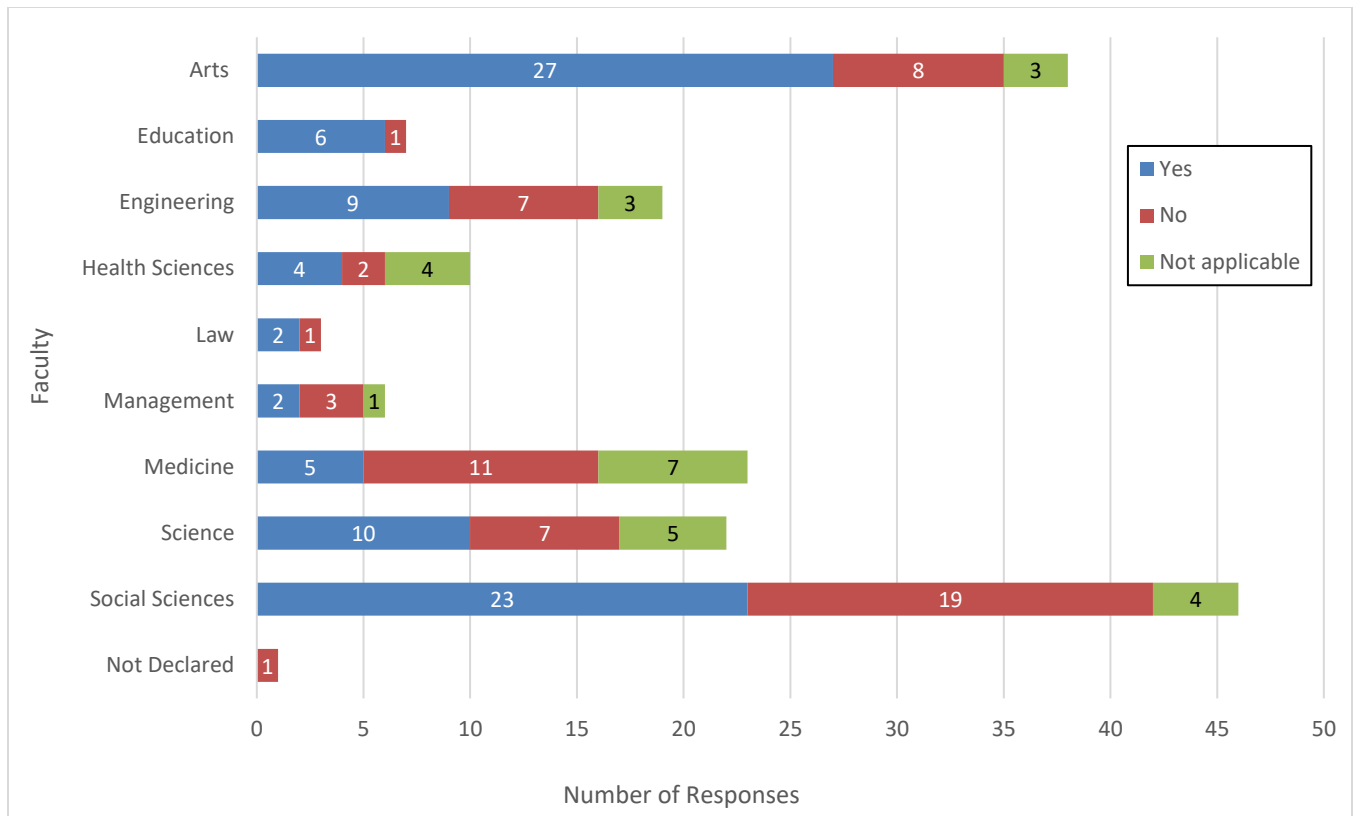


Figure 15. Use of own research data in teaching. Survey participants were asked if they use their own research data in their teaching practice (n=175).

Although the majority of respondents reported that they did not teach RDM topics (Figure 14), half of the respondents (88/175=50%) indicated that they used their own research data in their teaching practice (Figure 15), which corresponds with researchers indicating that sharing their data is beneficial to training the next generation of researchers (Figure 12).

Researchers in Arts were most likely to use their own data, while researchers in Medicine were least likely to use their own data. Taken together, these results suggest that researchers' teaching practices may include the use of data, either their own or from another source, but they do not cover how to manage that data. Therefore, there is potentially a lack of education and training in the essentials of research data management for the next generation of researchers.

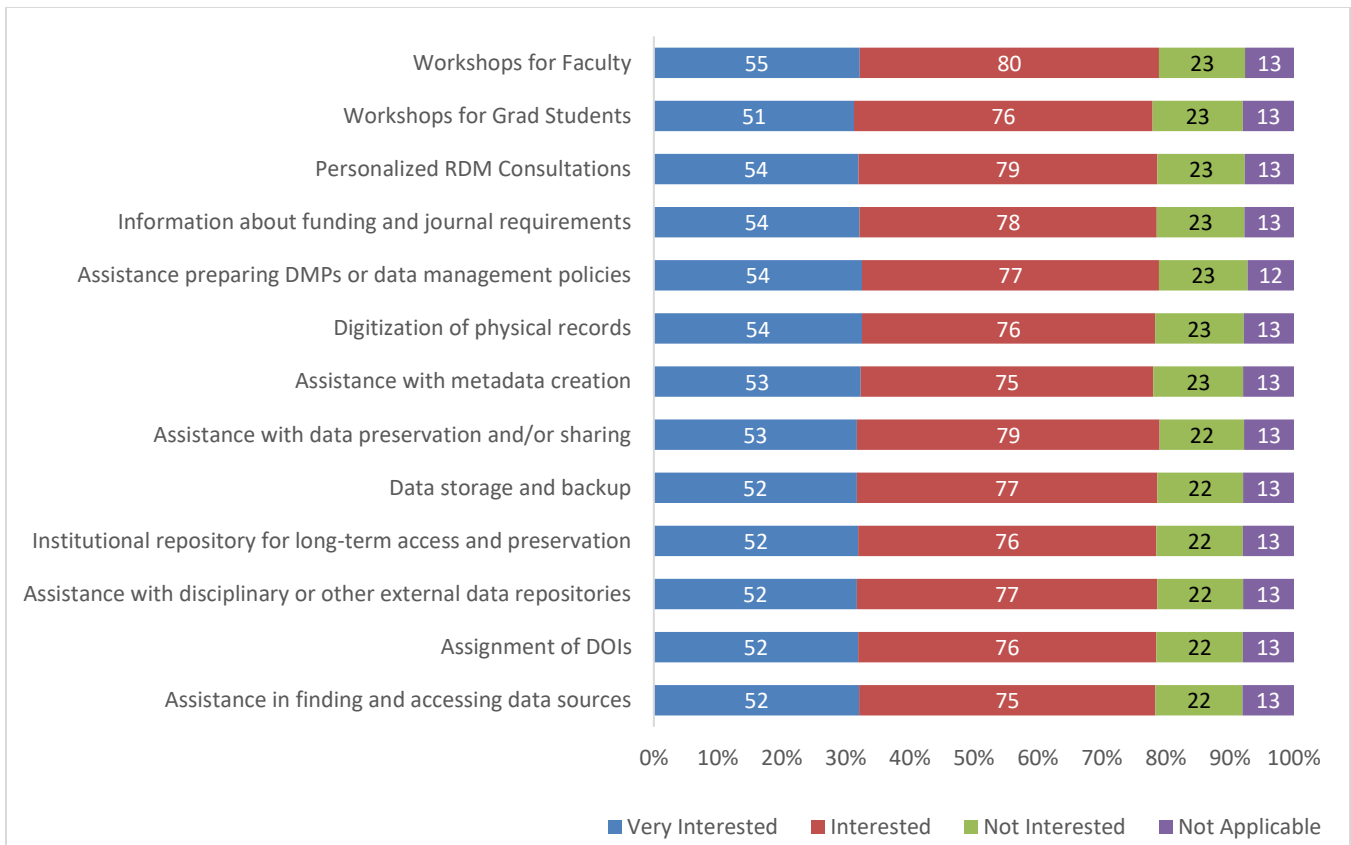


Figure 16. Interest in data services. Survey participants were asked to rate their interest in potential services if DMPs were required as part of grant applications (n=174).

Overall, respondents indicated they were interested in data services (Figure 16). However, respondents generally selected the same level of interest for each of the potential services listed and some only provided responses to services of interest while skipping over others. These results may reflect survey fatigue due to the long list of potential services in addition to a rather long survey questionnaire. Therefore, it is not possible to draw any significant conclusions regarding the level of interest in specific data services from this particular survey question.

However, gaps in researcher knowledge and readiness to meet RDM requirements have been identified in previous sections of this report. For example, researchers indicated they would need or prefer assistance and/or guided documentation to prepare DMPs (Figure 13). The identified gaps can be used to determine priorities for RDM service development. Furthermore, interest in the potential services proposed in the survey may increase when the Tri-Agency RDM policy comes into effect.

Research Data Management in the Digital Humanities

Digital Humanities (DH) or Digital Scholarship is a growing trend in research methodologies and there is an increasing interest from libraries to develop services and support for DH researchers. To explore RDM practices that are potentially unique to DH researchers, faculty members and postdoctoral fellows who were invited to participate in the Humanities and Social Sciences (HUSS) version of the survey were asked this additional question:

Digital Humanities, or Digital Scholarship, can be defined as the collection and use of digital research data (either through digitization of print resources, or using born-digital resources) combined with methodologies from traditional Humanities and Social Science scholarship.
Do you feel your research falls under this definition?

A total of 100 survey participants responded to this question (100/101=99%). Figure 17 shows that 46 respondents identified themselves as digital humanists/digital scholars, while 37 did not and 17 were not sure if their research fit this definition of digital humanities/digital scholarship.

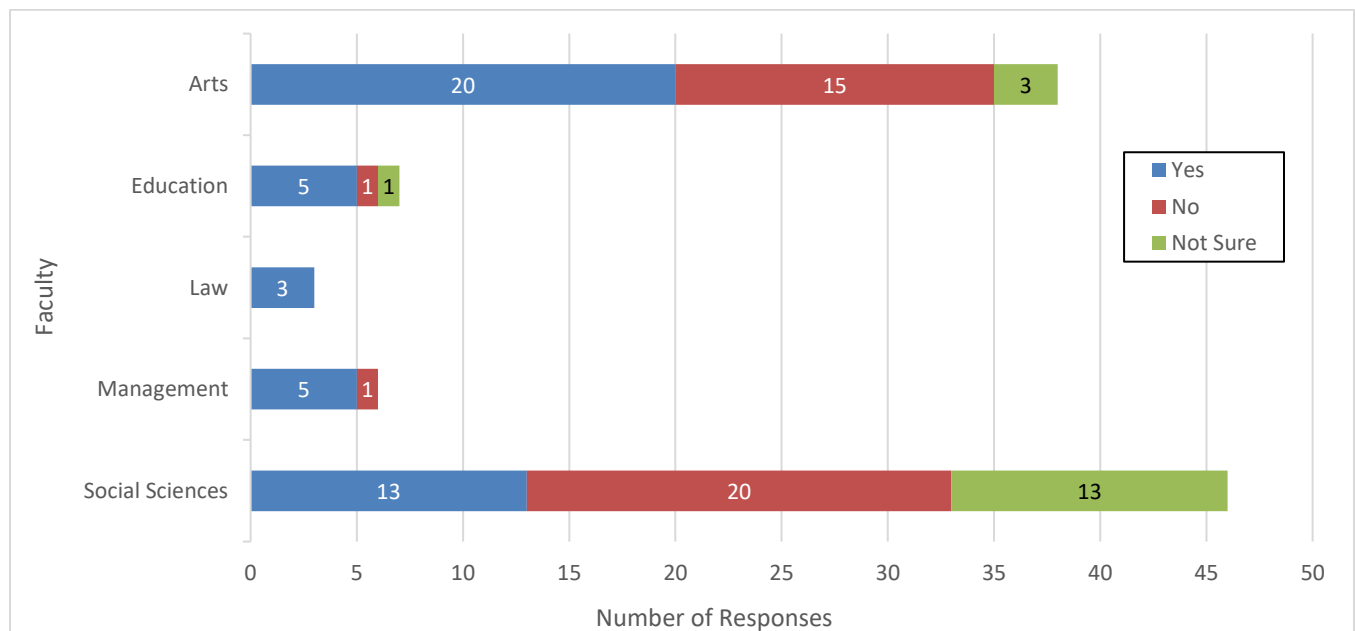


Figure 17 Digital Scholarship by Faculty. HUSS Survey respondents were asked if they identified themselves as digital scholars/digital humanists ($n=100$).

Interestingly, all three participants from the Faculty of Law identified themselves as digital scholars. Given the low response rate, the survey results are only representative of the individuals who participated and cannot be generalized to the larger population in the Faculty of Law.

Additionally, more than half of the researchers in Arts ($20/38=53\%$), Education ($5/7=71\%$), and Management ($5/6=83\%$) identified themselves as digital scholars; only 28% ($13/46$) of the researchers in Social Sciences identified themselves as digital scholars.

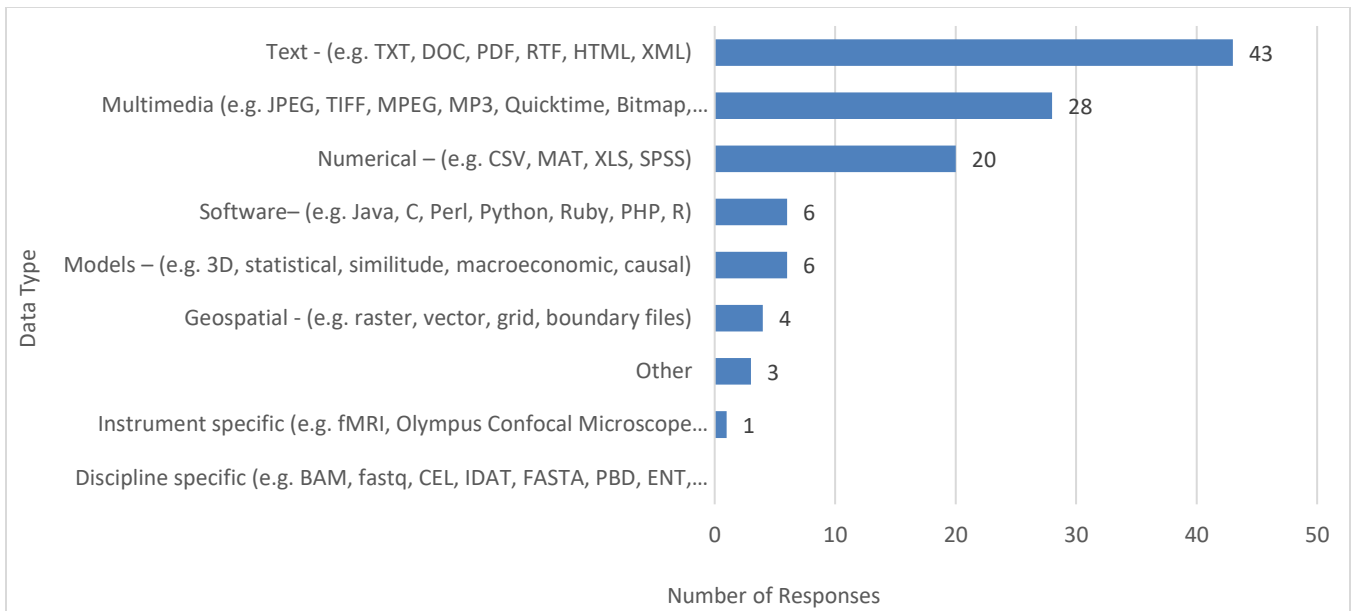


Figure 19 Type of research data generated or used in Digital Scholarship. Comparing the applicable types of research data generated or used in a typical research project by researchers who identified as a digital humanist/scholar (n=46).

The three most common types of research data generated or used in Digital Scholarship/Digital Humanities (Figure 19) were similar to those generated or used in the overall survey sample population (Figure 2): text, multimedia and numerical. However, multimedia data was cited as the second most frequently generated or used in Digital Scholarship, while in the overall survey sample population numerical data was the second most frequently generated or used data type. These results suggest that support for multimedia data may be needed more frequently among digital scholars/digital humanists compared to other researchers.

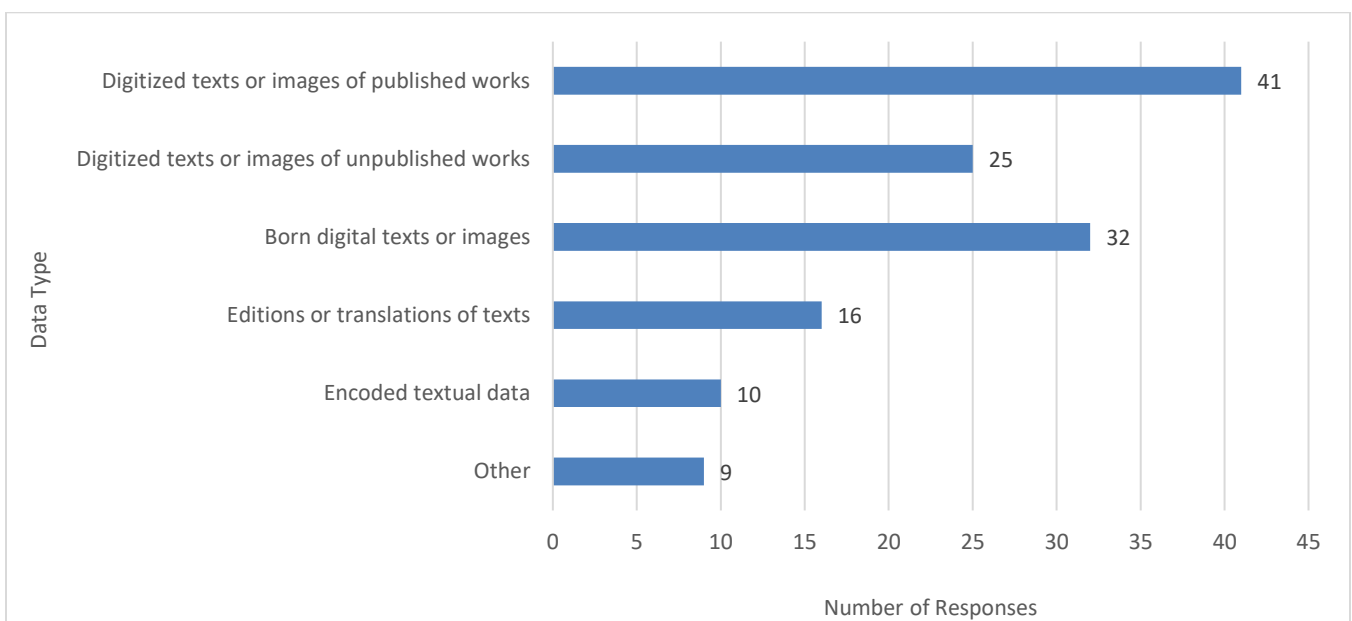


Figure 20 Type of data used in digital humanities/digital scholarship research. Survey participants who identified as digital scholars/digital humanists were asked to identify the type of data they use (n=46).

The 46 respondents who identified themselves as digital scholars/digital humanists were also asked to specify the type of data they used in their research in more detail. Figure 20 shows that the most frequently used type of data were digitized texts or images of published works, born digital texts or images and digitized texts or images of unpublished works. Further research and analysis are required to determine whether digitization of physical records is a data service that would be of interest to digital scholars/digital humanists.

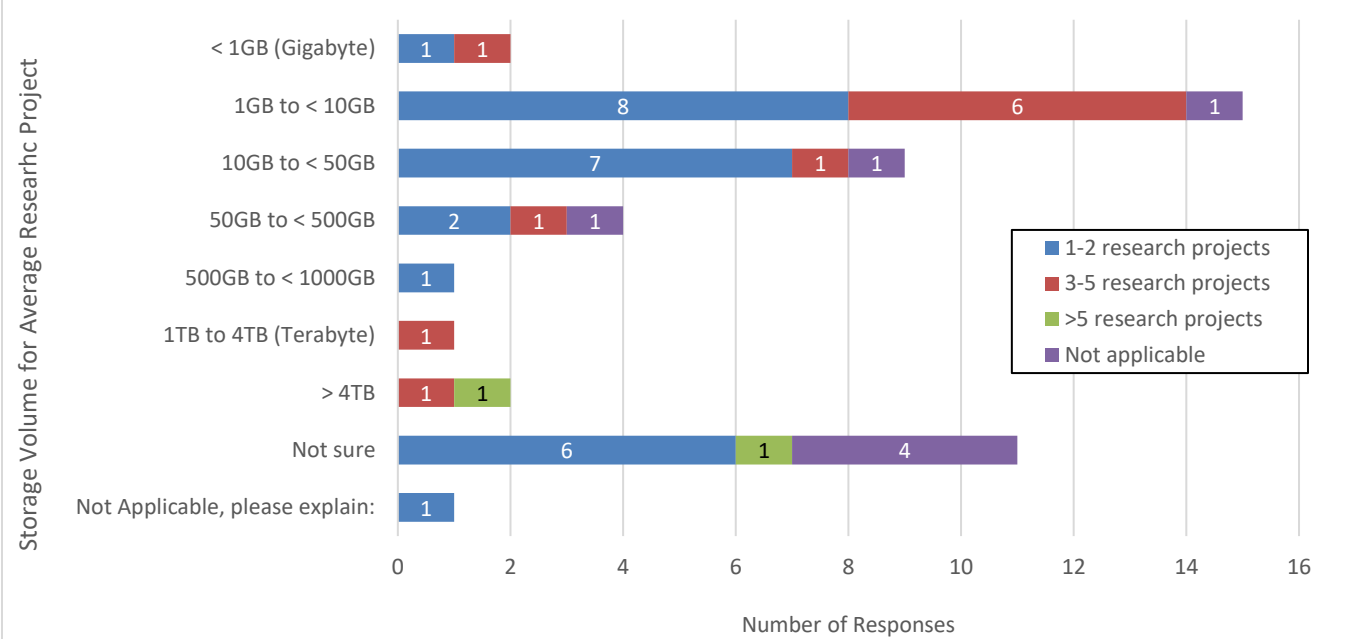


Figure 21. Amount of storage in relation to number of research projects in Digital Scholarship. Summary of the estimated amount of storage digital scholars/digital humanists use in an average research project as well as the number of research projects they lead in the past year (n=46).

Figure 21 illustrates the relationship between storage volume required for the average research project and the number of projects led by the digital scholar/digital humanist in the past year. At the time of the survey, the majority of DH research projects required less than 50 GB of storage (26/46=57%). Generally, digital scholars/digital humanists required less data storage compared to the overall survey sample population, where the majority of research projects required less than 1000 GB of storage (Figure 1). However, digital scholars/digital humanists were even more uncertain of how much storage they used in an average project: 24% (11/46) compared to 18% of the overall survey sample population (Figure 1).

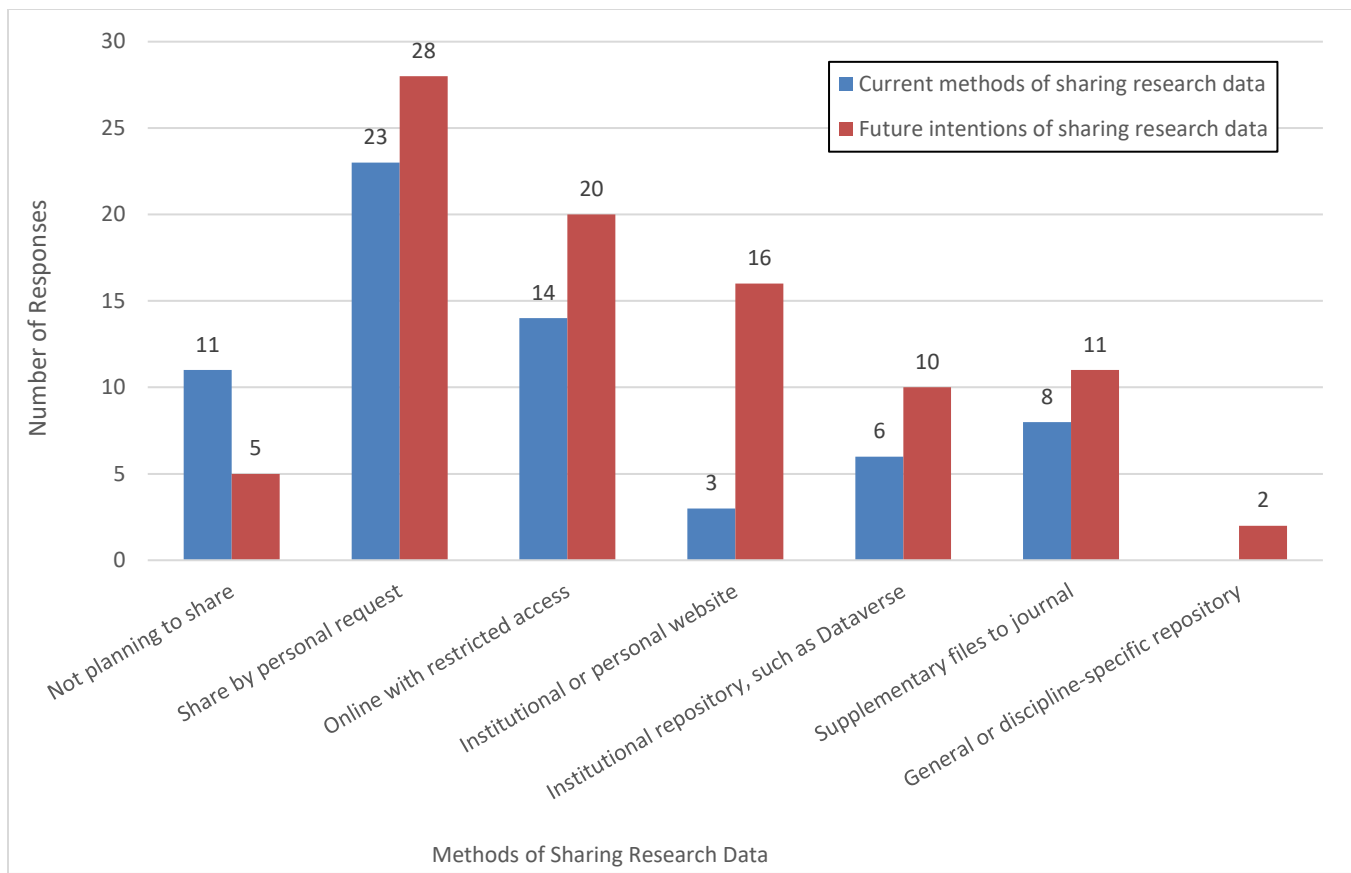


Figure 22. Comparison of current methods and future intentions of sharing research data in Digital Scholarship. Digital scholars/digital humanists (n=46) selected all the applicable methods of sharing research data they currently use and would consider using in the future.

Figure 22 compares current methods and future intentions of sharing research data among digital scholars/digital humanists at the time of the survey. The results indicated that trends in current practices and future considerations were similar to those seen in the overall survey sample population (Figure 8). Digital scholars/digital humanists indicated an increased interest in sharing their data in the future and that share by personal request was the most popular method of data sharing when the survey was conducted.

However, there were two notable differences in other data sharing methods used by digital scholars/digital humanists compared to the overall survey sample population. Firstly, the percentage of digital scholars/digital humanists who used the institutional data repository (Dataverse) as a current method of data sharing was larger than the overall survey sample population with 13% (6/46) of digital scholars/digital humanists using Dataverse compared to 7% of the overall survey sample population.

Secondly, the increase in interest in using an institutional or personal website to share data in the future was larger among digital scholars/digital humanists from 7% to 35% compared to the overall survey sample population (from 11% to 28%).

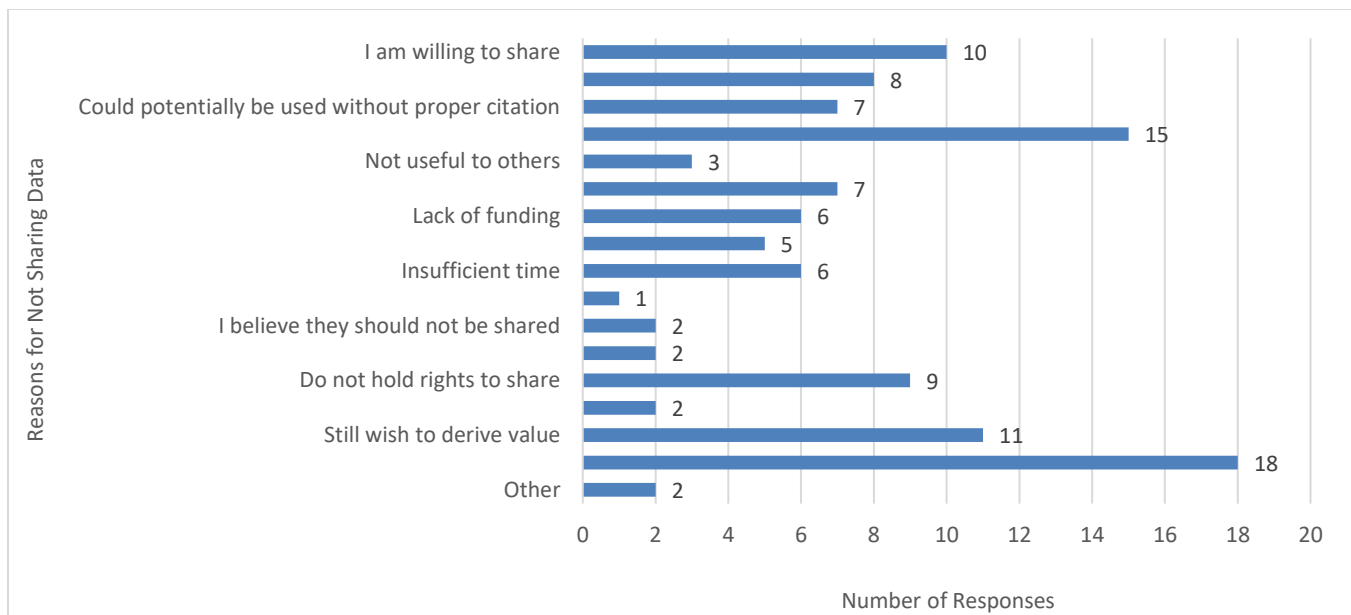


Figure 23. Reasons digital scholars/digital humanists would not share research data. Digital scholars/digital humanists indicate all applicable reasons why they would not be willing to share their research data and associated methods/tools (N=45).

Reasons why digital scholars/digital humanists would not be willing to share their research data and associated methods/tools (Figure 23) follow the same trends as the overall survey sample population (Figure 12). Incomplete or not finished data; privacy, legal or security issues; and the wish to derive value from their data were cited as the primary reasons they would not be willing to share. The percentage of digital scholars/digital humanists who were willing to share their data (10/46=22%) is similar to the overall survey sample population (25%).

Conclusions

At the time of the survey's dissemination (2016-2017), there was limited information about RDM practices or attitudes among uOttawa researchers. The Tri-Agencies had released their Statement of Principles on Digital Data Management and the research community anticipated a Tri-Agency policy on Research Data Management would soon follow. Librarians conducting this survey set out to address their knowledge gap by determining how uOttawa researchers in different disciplines were managing and sharing their research data and how the uOttawa Library might help to facilitate research data management activities in the future.

The key findings of the survey suggest that researchers, irrespective of discipline, indicate gaps in their RDM knowledge and may benefit from additional training and support to effectively manage their data and meet Tri-Agency RDM policy requirements. Specifically, researchers expressed interest in making use of a data repository associated with the University Ottawa, such as Dataverse, and in receiving guidance in the preparation of data management plans. The survey findings also imply that researchers could benefit from assistance with navigating RDM and research ethics policies as well as data documentation for reuse and reproducibility.

In addition, survey results point to a gap in infrastructure that facilitates collaboration, including international collaborations that may be subject to multiple RDM policies, and education and training in the essentials of research data management for students as the next generation of researchers.

Limitations of the Survey Results

Although the survey has revealed interesting insights and trends into the RDM practices and attitudes of uOttawa researchers, the low response rate and lack of statistical significance limit the interpretation of the results. Survey findings cannot be applied to the larger uOttawa academic community without further research, particularly in Law and Medicine where the survey sample populations only represent 2% of the target populations.

Additionally, the iterative process of developing and modifying the questionnaires led to errors and inconsistent terminology appearing in the survey that respondents may have found confusing. The question related to data storage for example, presented the stages of research data as: raw data, manipulated data (e.g. converted, curated, processes), and archived data (e.g. long-term storage or preservation). However, in a question that appeared later in the survey asking researchers how long they keep their data, the stages of research data were presented as: source material/raw data, intermediate/working data, and processed data ready for publication, which may include supporting information such as metadata and documentation.

Another challenge of conducting a multi-institutional study was the terminology used to describe institutional repositories and institutional Dataverses of the Scholars Portal Dataverse. At the time of the survey, institutional branding and support for users of the Scholars Portal Dataverse⁹ were still in development. As a result, there wasn't a clear consensus on how to describe Dataverse to local academic communities. At some Canadian institutions, Dataverse was presented as an institutional repository for data, while other institutions were promoting Dataverse as a data repository platform and still others were not using Dataverse at all. The survey presented Dataverse as an example of an institutional repository for data sharing, which was a bias towards institutions that were using Dataverse as an institutional data repository.

Furthermore, the RDM landscape at uOttawa has continued to evolve since the survey was conducted in 2016-2107. The survey reflected the priorities at the time: data storage, data sharing, DMPs and interest in RDM services. More nuanced or particular aspects of RDM were not included in the survey. For example, the survey did not address best practices in data backups and/or the prevention of data loss, awareness of RDM national and/or international RDM policies. As a result, the findings in this report provide a historical snapshot of the RDM landscape at uOttawa and may have limited practical use in shaping RDM support and services moving forward.

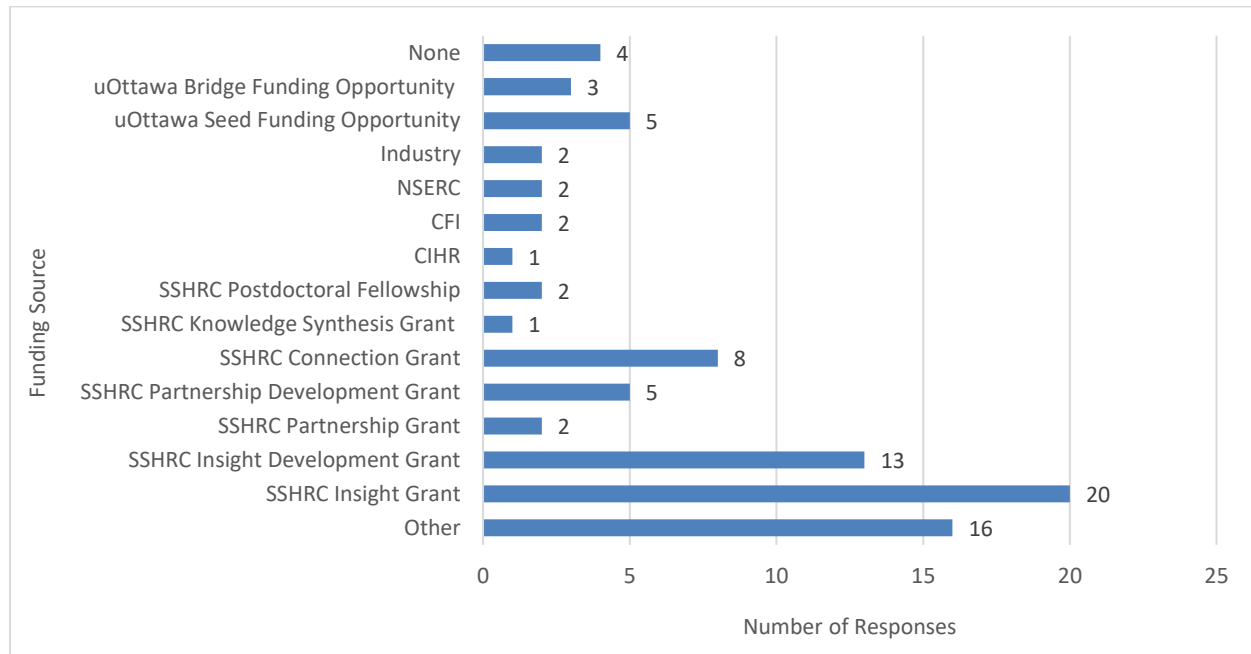
Future Directions

The survey findings provide information about researcher RDM practices at uOttawa that may serve to inform institutional policy, services, and infrastructure development that are aligned with funding agency requirements and effective data stewardship practices. Furthermore, this survey is part of a larger project with the Canadian RDM Survey Consortium, which aims to provide a national view of the RDM landscape in Canada by compiling survey data collected by participating members. The results of the Consortium's efforts will provide insights into trends in RDM practices across the country and will enable comparisons of unique practices and attitudes between disciplines and regions.

⁹ <https://dataverse.scholarsportal.info/>

Appendix

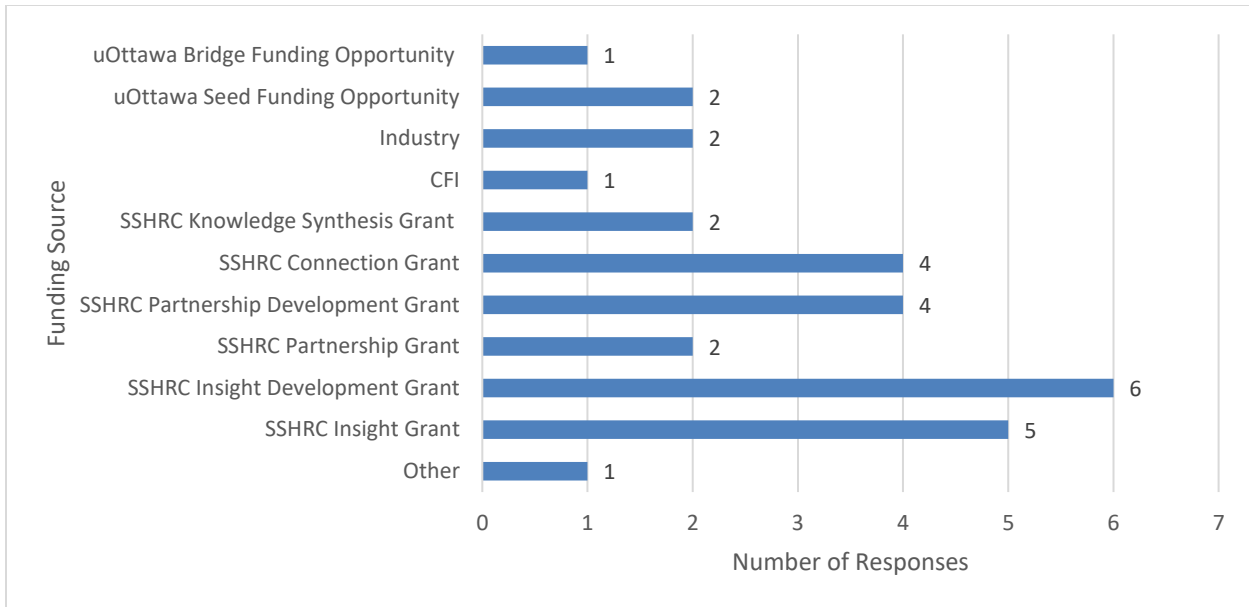
Survey Question: Which funding sources have you used within the past 5 years, or are planning to apply for in the next 5 years? Please exclude funding earmarked exclusively for operations and infrastructure. Select all that apply:



Funding sources in the Faculty of Arts.

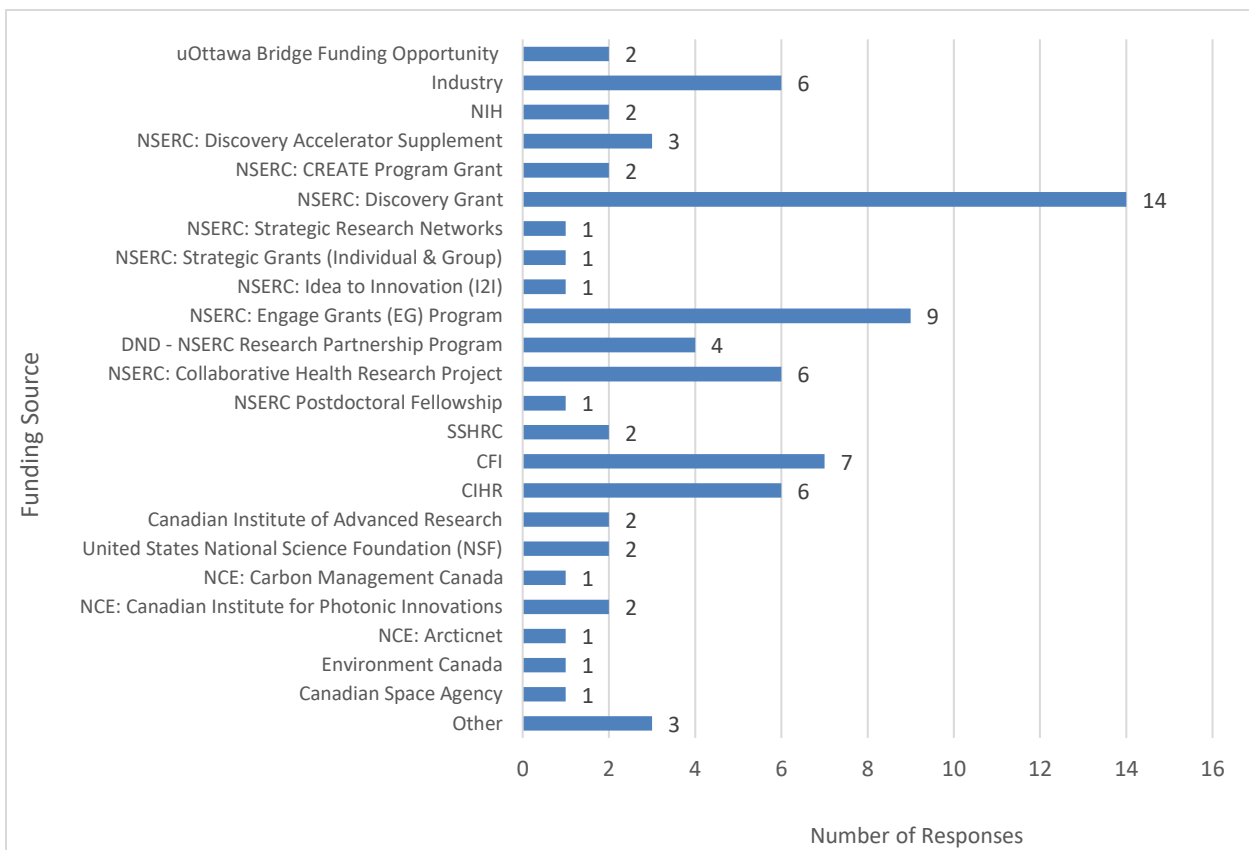
Other funding sources specified:

- Travel grants (APTPUO)
- ArcticNet, MEOPAR
- Educational Testing Services
- external funds (fellowships funded by other governments)
- Private donors and foundations outside this list
- grants for organizations such as Access Copyright Foundation and the Bibliographical Society of Canada
- Government contracts (IRCC), research funding from Pathways to Prosperity and the Centre for Research and Education on Women and Work at Carleton U.
- Canada Council Grants
- FQRSC Postdoc
- Bourse post-doctorale Fonds de recherche Société et culture du Québec
- City of Ottawa public funding campagne
- Financement externe BAnQ
- Ministère de l'Éducation
- Chaire de recherche de l'Université
- Fonds de développement de l'APTPUO



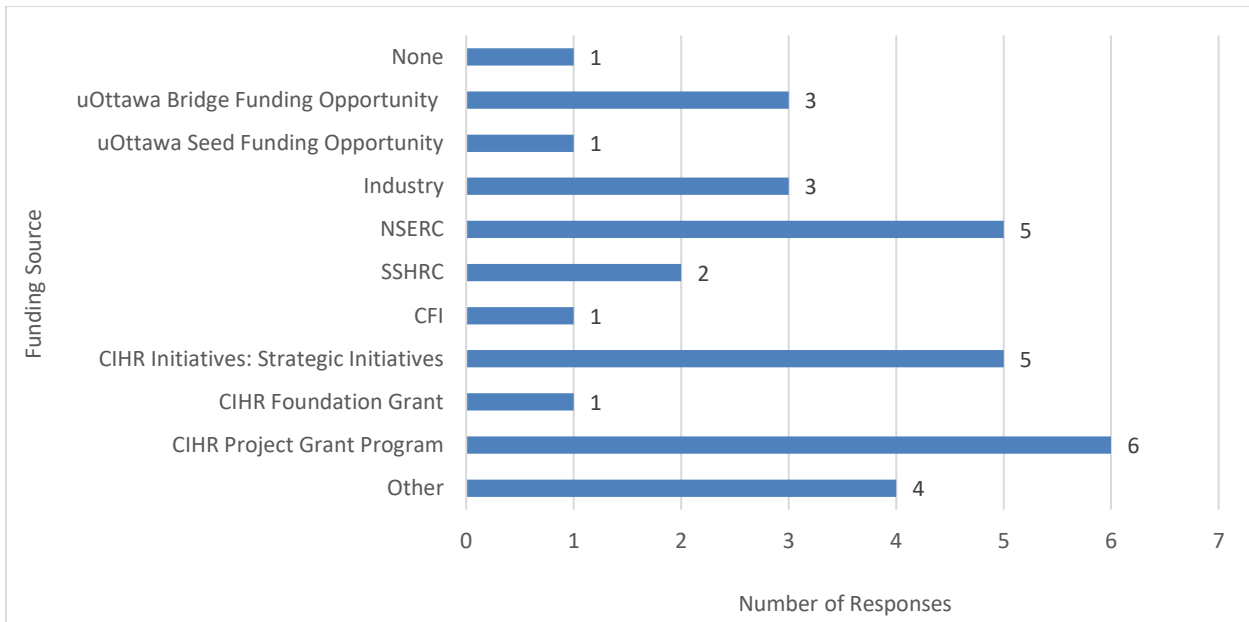
Funding sources in the Faculty of Education.

Other funding source specified: School Board funding.



Funding sources in the Faculty of Engineering.

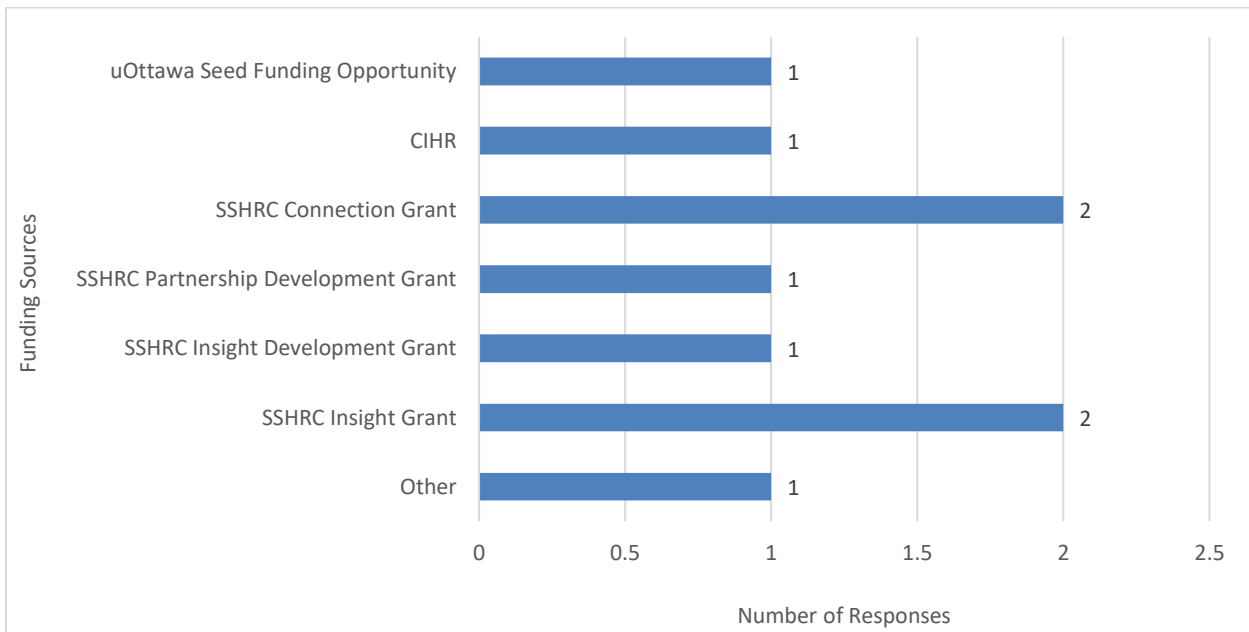
Other funding sources specified: NSERC CRD, Ontario Research Fund, Mitacs Postdoctoral Fellowship



Funding sources in the Faculty of Health Sciences.

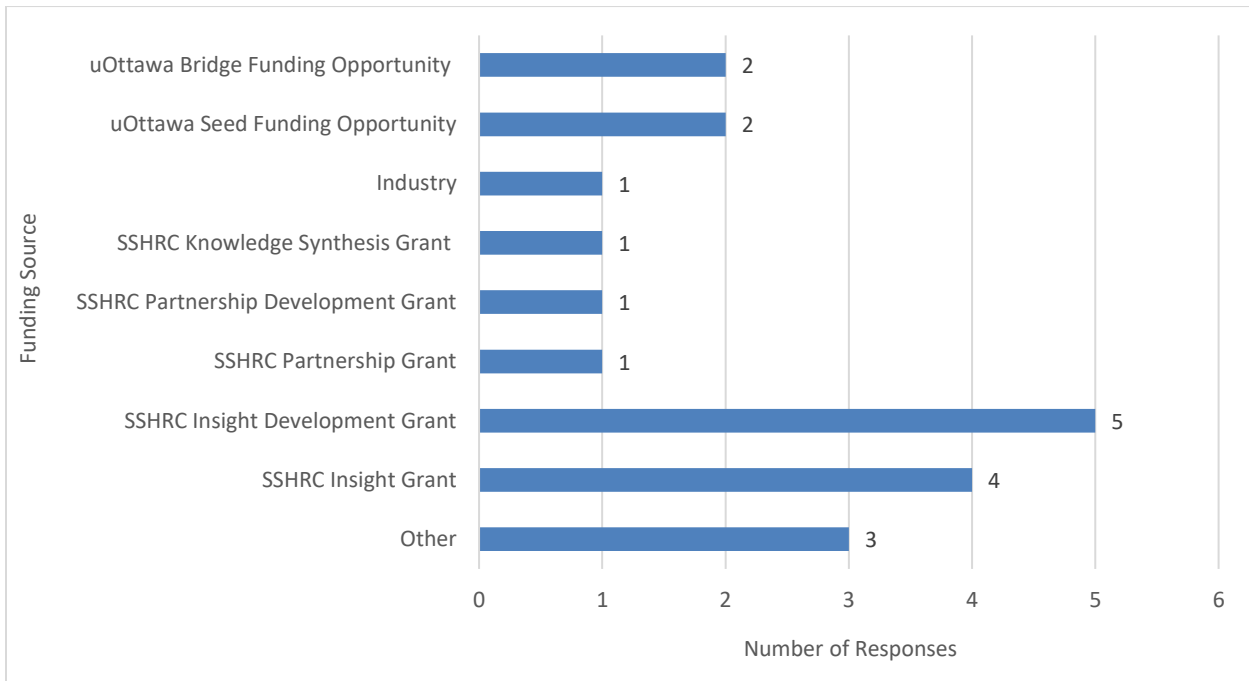
Other funding source specified:

- OCHSU, CHEO, IASP
- Consortium national de formation en santé, Action for Hearing Loss (International)
- Fondations ou associations (ex. diabète, diététique)
- Fondations professionnelles



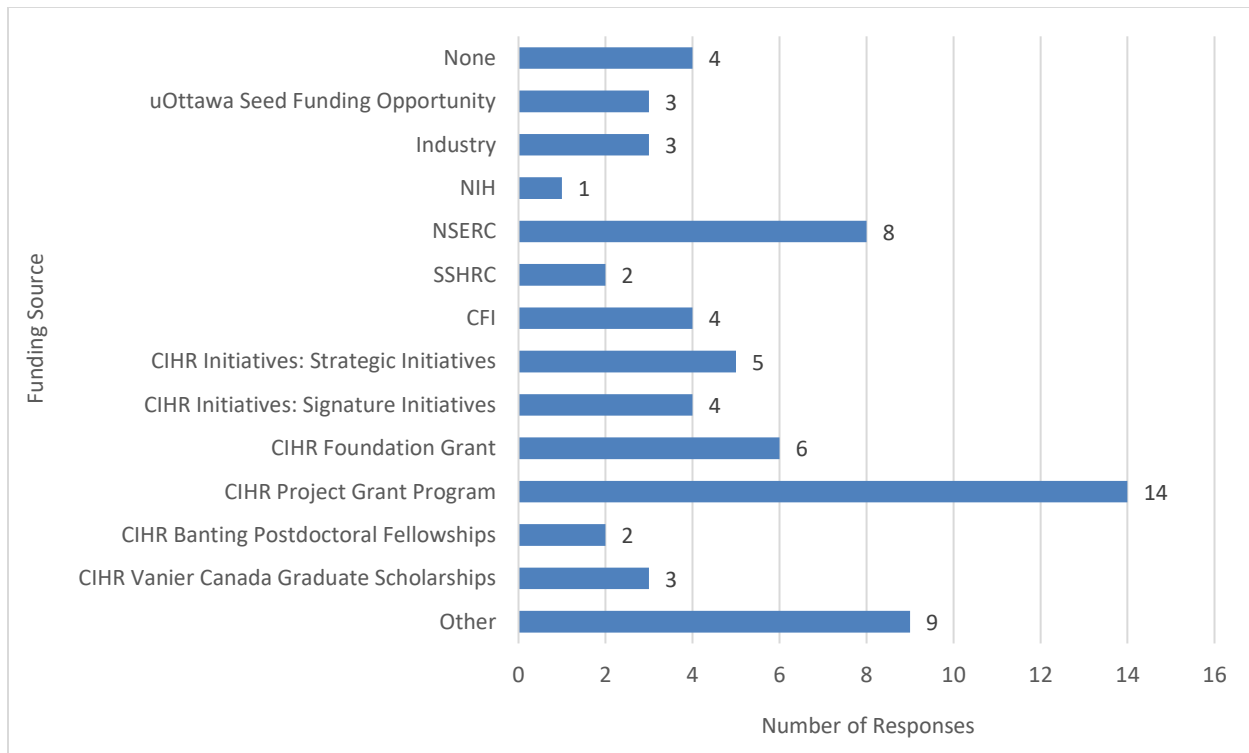
Funding sources in the Faculty of Law.

Other funding source specified: Law Foundation of Ontario.



Funding sources in the Faculty of Management.

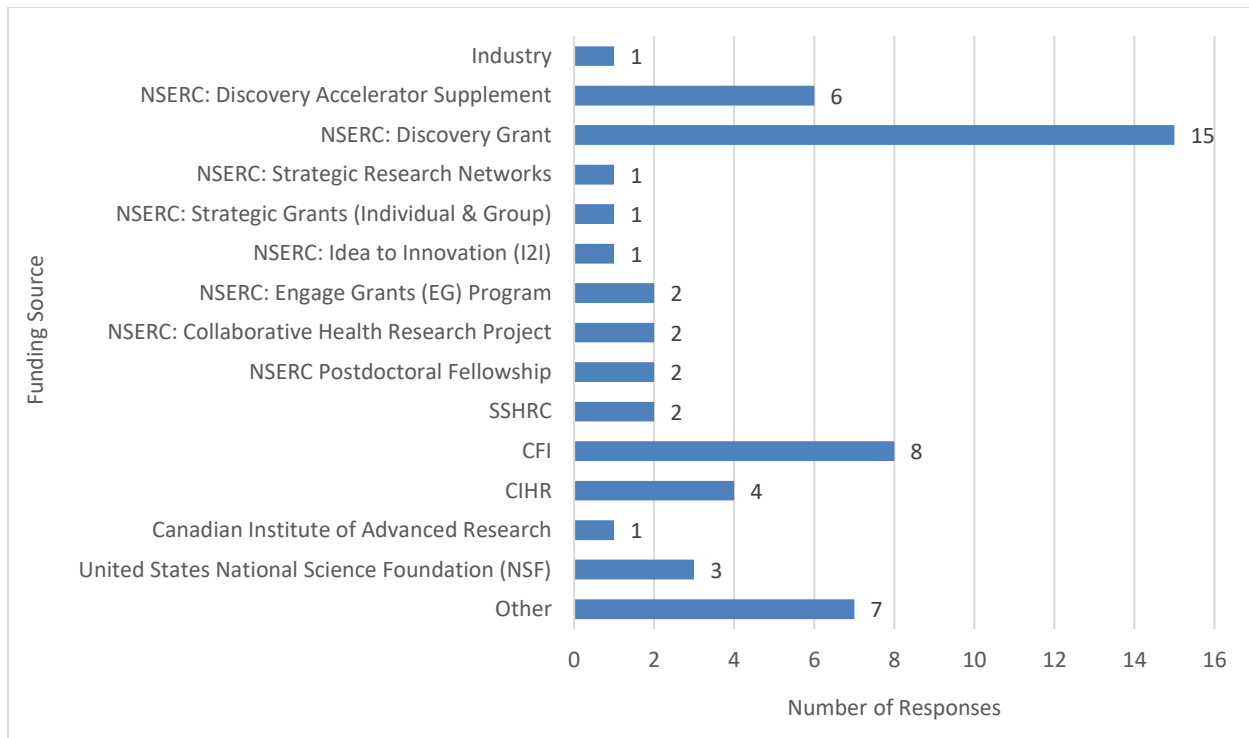
Other funding sources specified: Financement Telfer, Privées, International foundation grant (Germany).



Funding sources in the Faculty of Medicine.

Other funding sources specified:

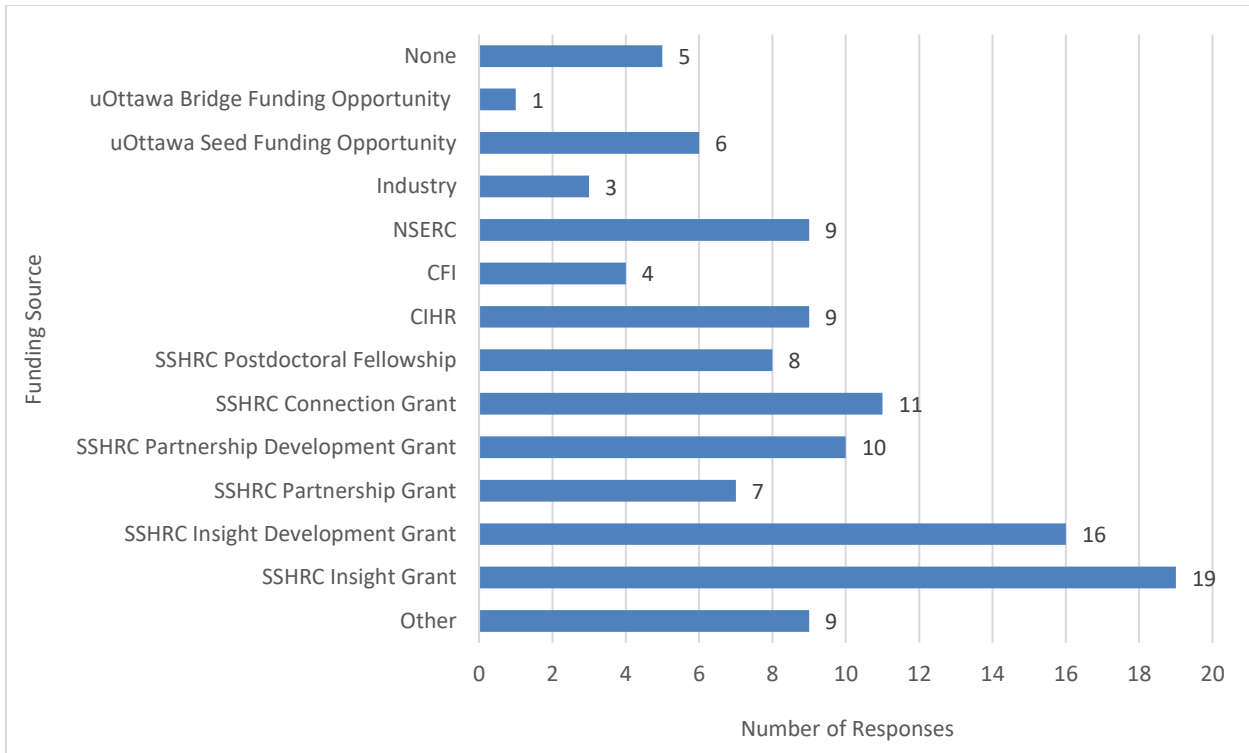
- Medical Council of Canada, Royal College of Physicians and Surgeons of Canada, TOHAMO
- Medical foundations; Private foundations
- AMS/Phoenix
- Chemical Management Plan, Health Canada
- International Development Research Centre project grant
- Cancer Research Society
- Heart and Stroke Foundation
- Royal College of Physicians and Surgeons of Canada Medical Education Research Grants
- Association Française contre les Myopathies (AFM)



Funding sources in the Faculty of Science.

Other funding sources specified:

- Contrats de recherche Ville d'Ottawa,
- US Government
- Canada Research Chairs Program
- Natural Resources Canada, Canadian Nuclear Safety Commission
- Funding from my supervisor.
- ERA: Early Researcher Award (Ontario)
- Gates Foundation

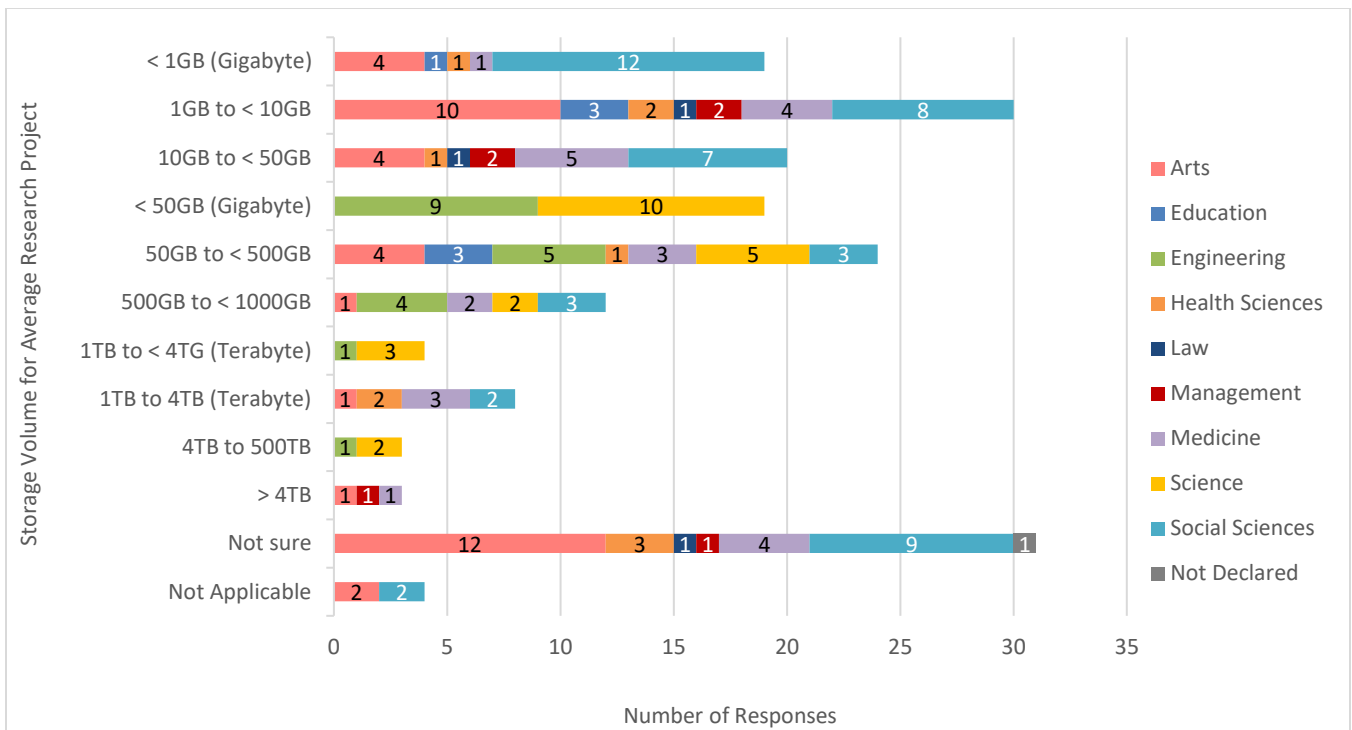


Funding sources in the Faculty of Social Sciences.

Other funding sources specified:

- Ontario Centres of Excellence
- French government/EU funding
- Alzheimer Society of Canada
- NSERC Equipment Grant
- CNFS
- Banting
- Fondation Chiang Ching-kuo, foundation indo-canadienne Shastri,
- Fond de developement professionnel U Ottawa
- CNFS UOttawa et CNFS Secrétariat National

Survey Question: How much data storage do you estimate you use in an average research project? Select one:

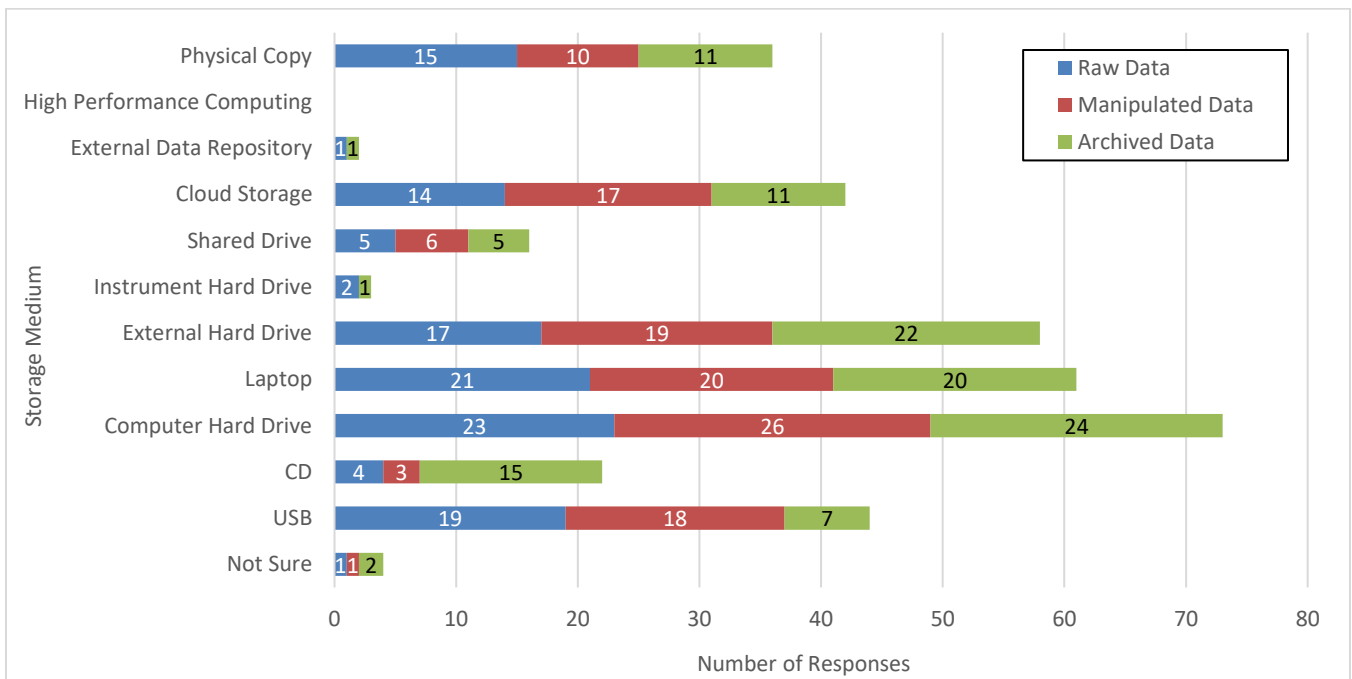


Amount of storage used in each Faculty.

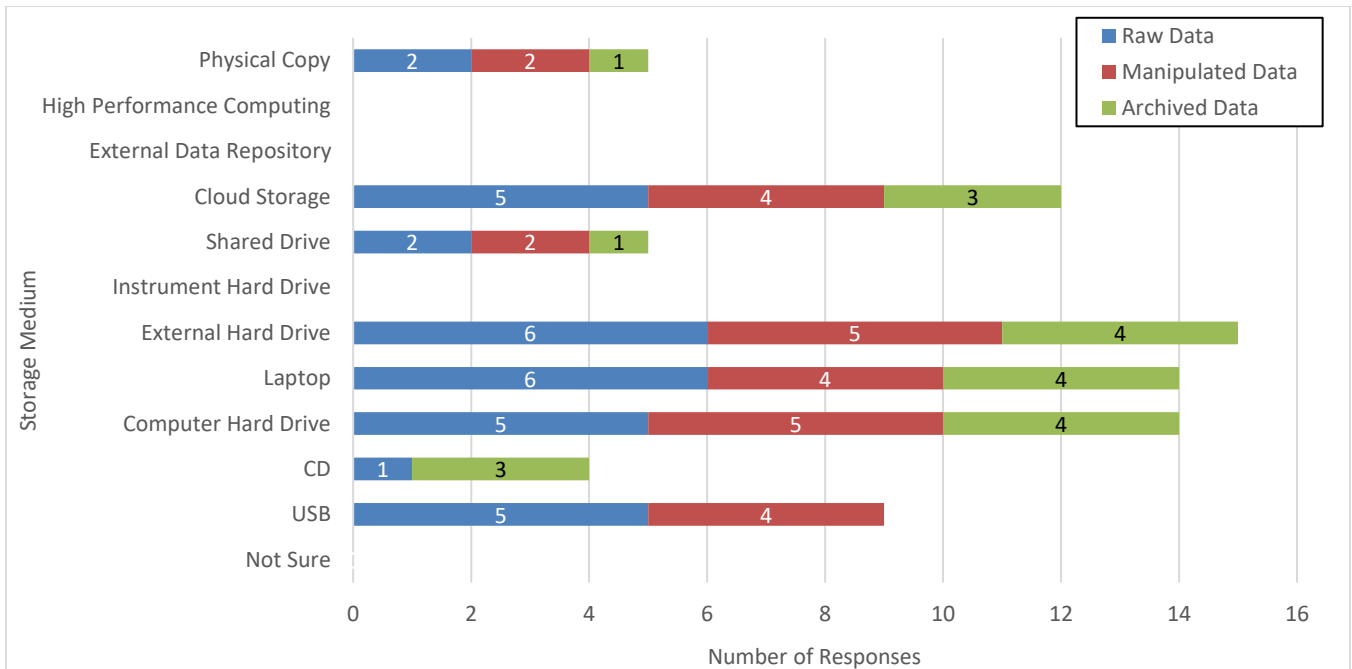
Survey Question: Which of the following best describes the type of research data you generate or use in a typical research project? Select all that apply:

	Text - (e.g. TXT, DOC, PDF, RTF, HTML, XML)	Numerical – (e.g. CSV, MAT, XLS, SPSS)	Multimedia (e.g. JPEG, TIFF, MPEG, MP3, Quicktime, Bitmap, Audio/Visual records)	Models – (e.g. 3D, statistical, similitude, macroeconomic, causal)	Software– (e.g. Java, C, Perl, Python, Ruby, PHP, R)	Instrument specific (e.g. fMRI, Olympus Confocal Microscope Data Format, FLIR Infrared Camera (SEQ))	Geospatial - (e.g. raster, vector, grid, boundary files)	Discipline specific (e.g. BAM, fastq, CEL, IDAT, FASTA, PBD, ENT, BRK, CIF, FITS, DICOM)	Other
Arts	37	13	21	3	2	1	3		4
Education	7	3	6		3				
Engineering	16	14	10	8	7	5	1		
Health Sciences	9	8	4	2	1	4			
Law	3	3		1					
Management	6	3	4	1	2				
Medicine	19	18	17	7	5	9	1	7	2
Science	15	17	13	7	9	7	3		3
Social Sciences	38	22	22	9	5	4	3		4
Not Declared	1	1	1			1			

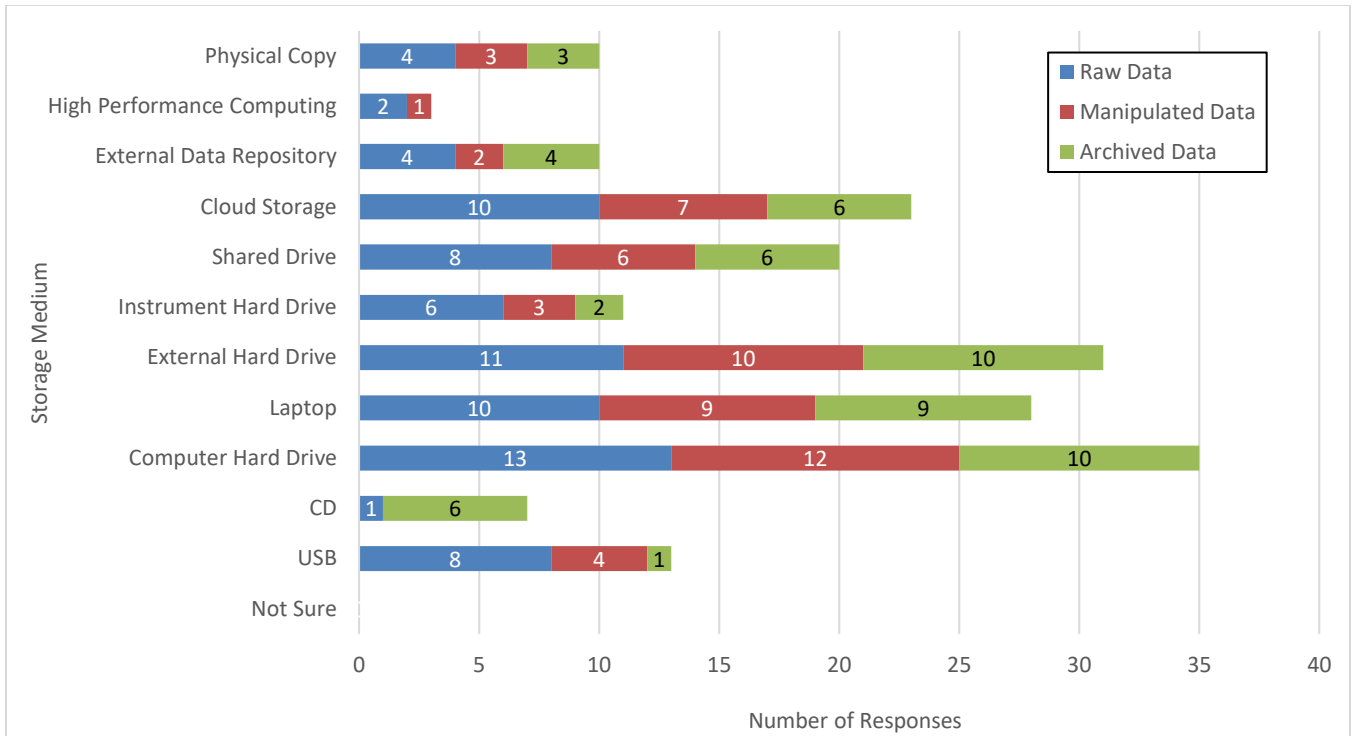
Survey Question: Please indicate where you store research data from your current project(s).
Select all that apply:



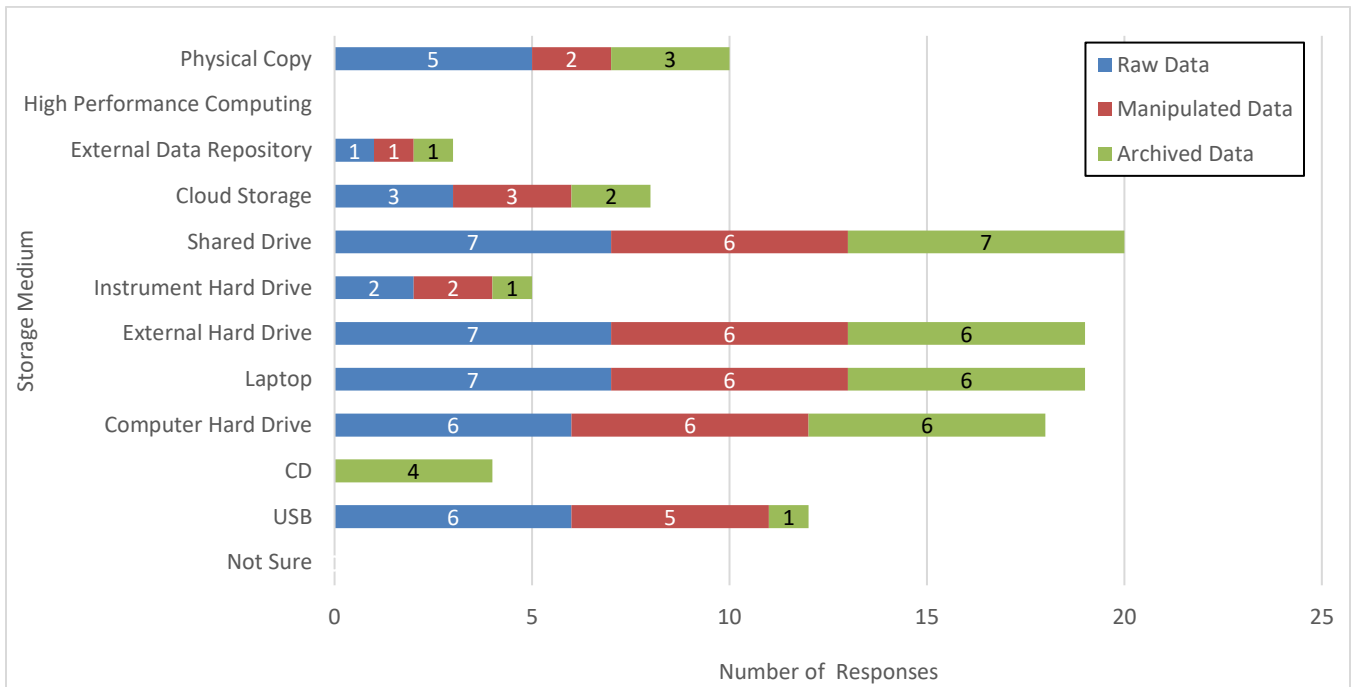
Data storage used in the Faculty of Arts.



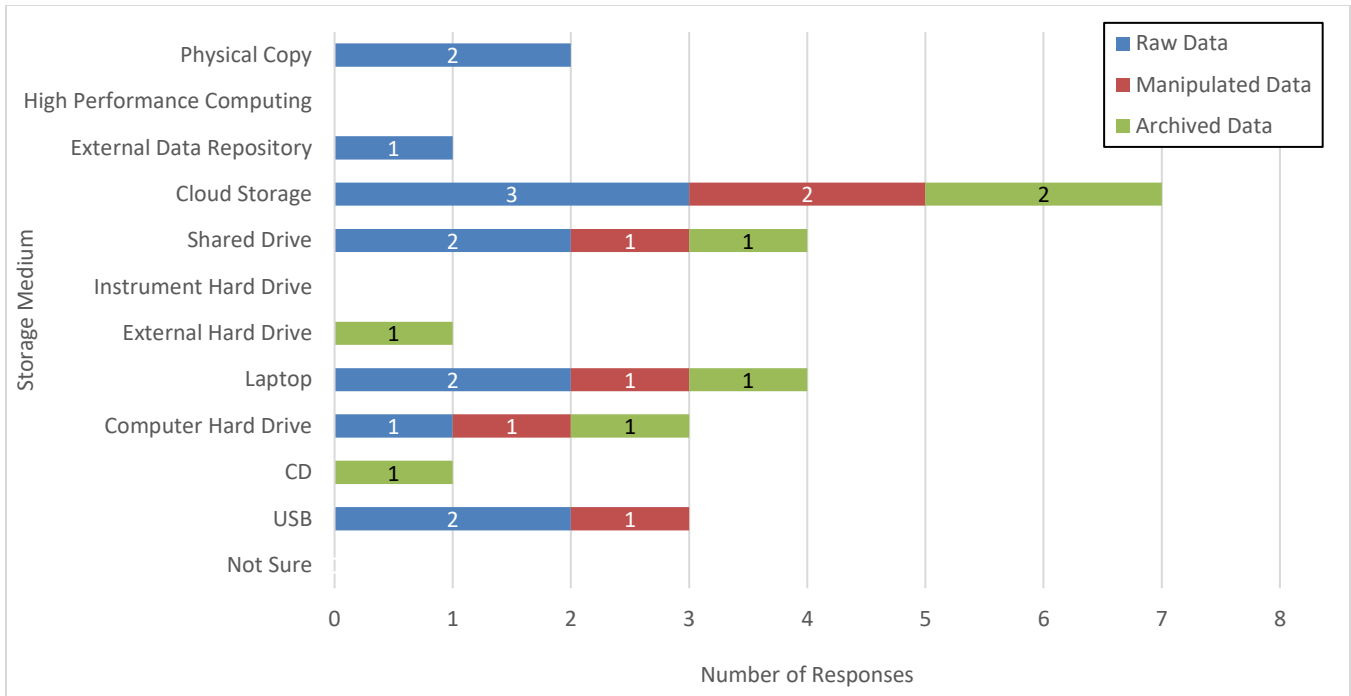
Data storage used in the Faculty of Education.



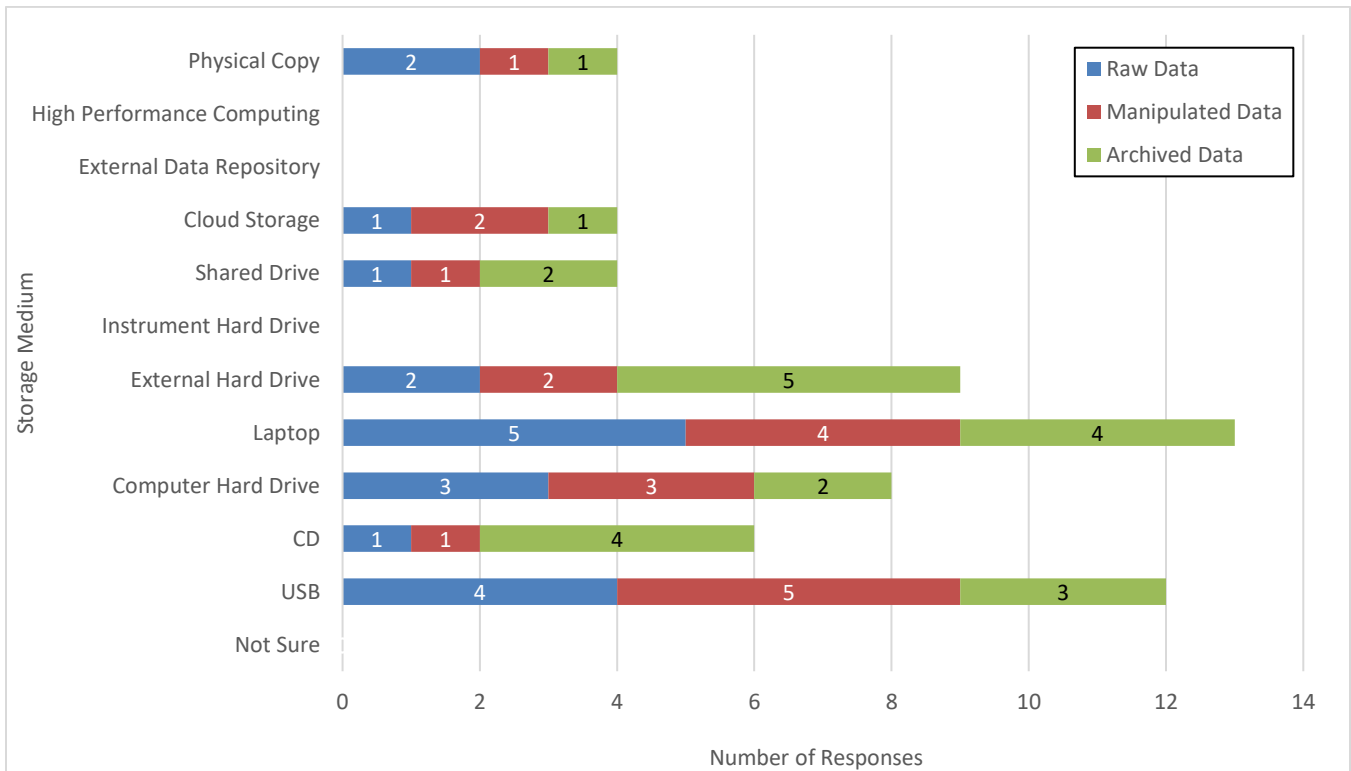
Data storage used in the Faculty of Engineering.



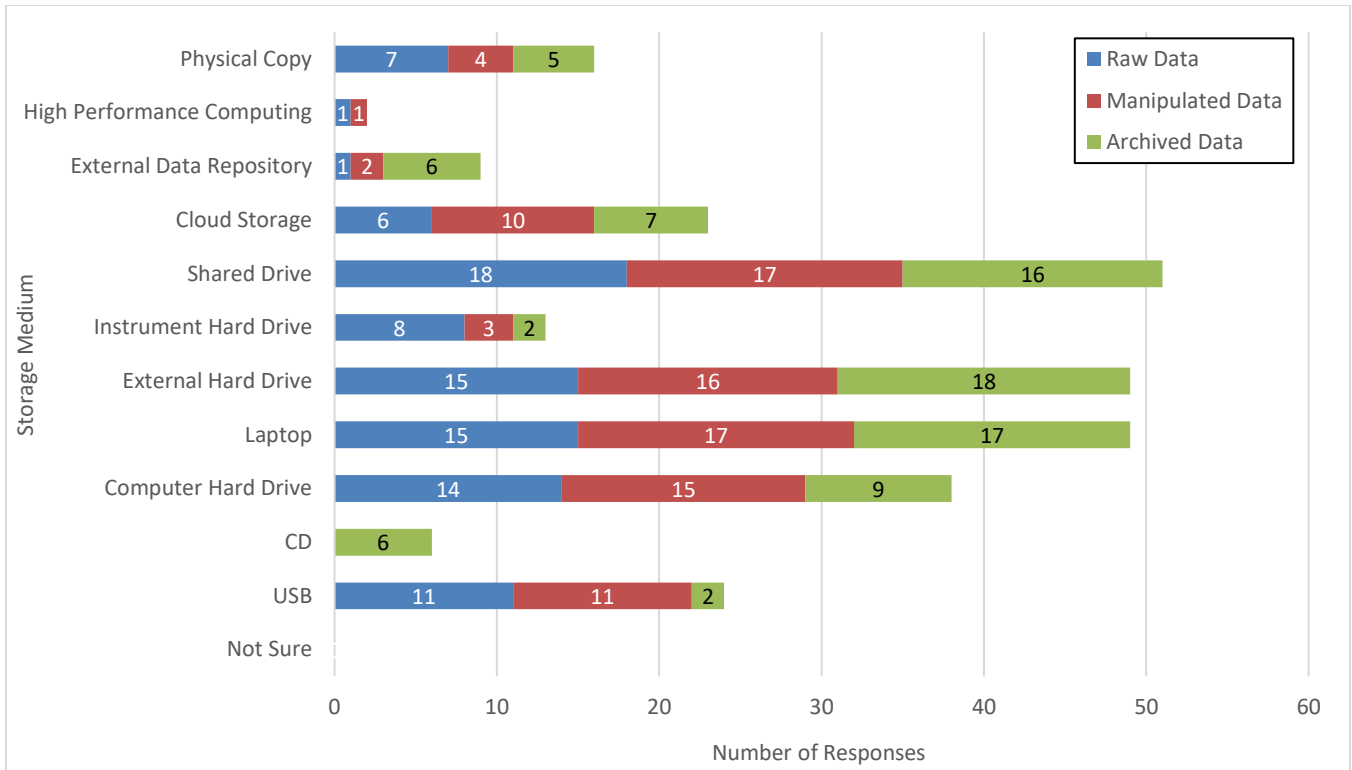
Data storage used in the Faculty of Health Sciences.



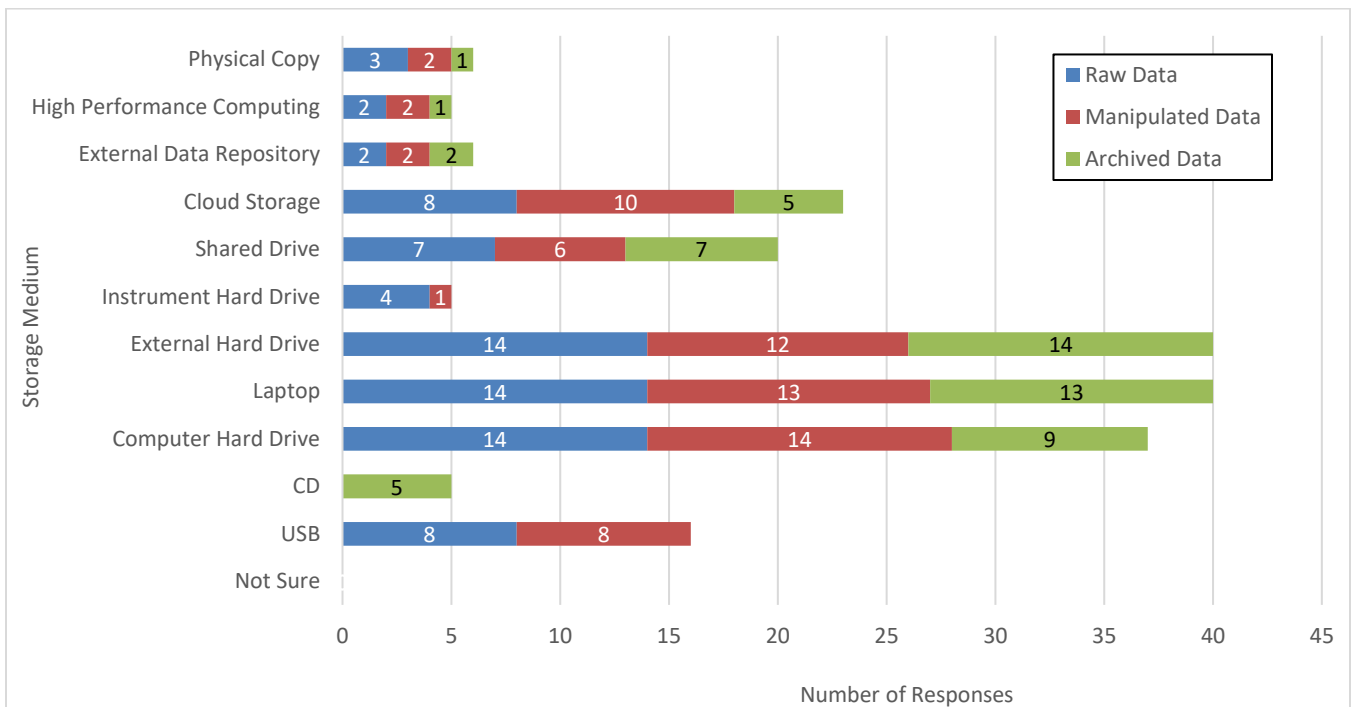
Data storage used in the Faculty of Law.



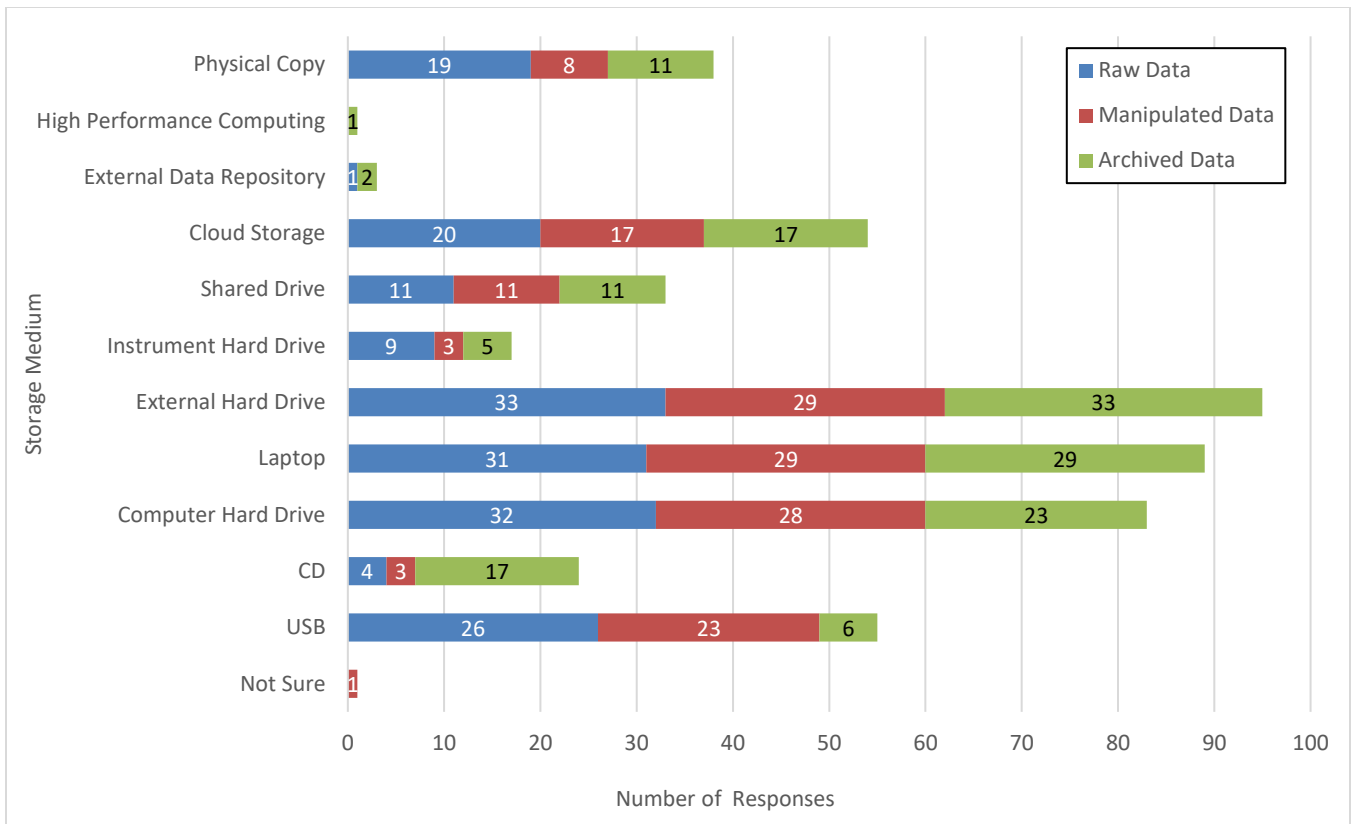
Data storage used in the Telfer School of Management.



Data storage used in the Faculty of Medicine.

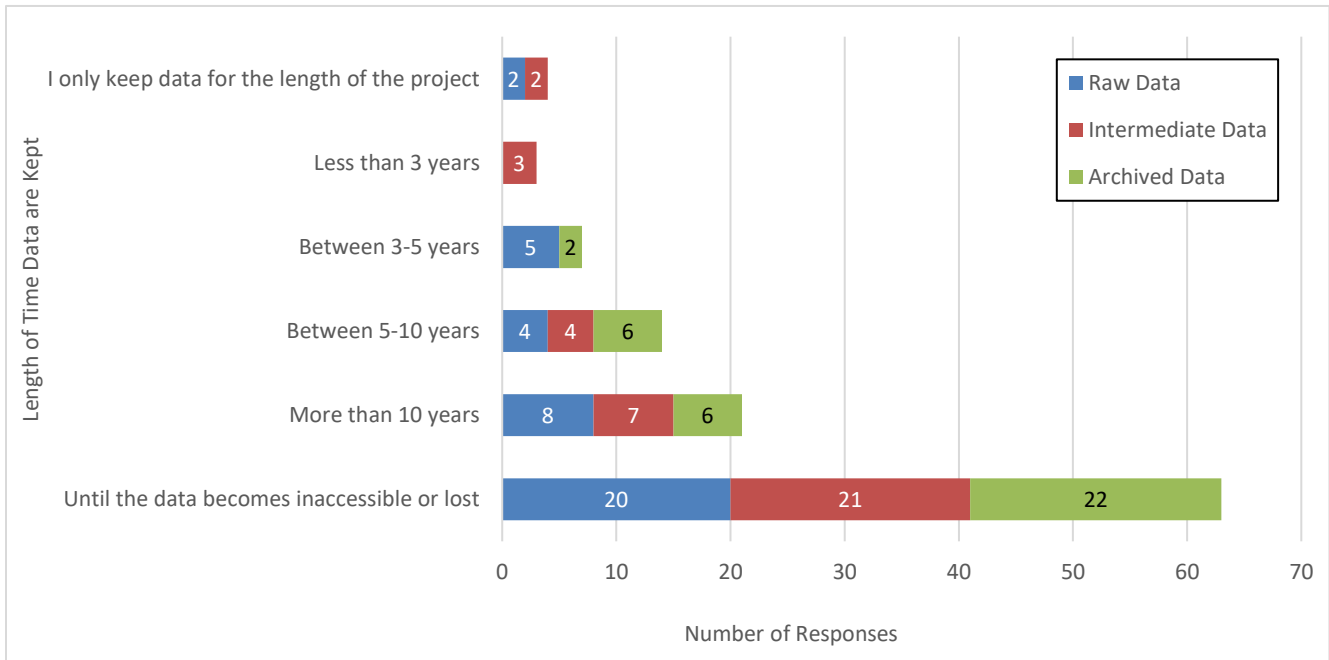


Data storage used in the Faculty of Science.



Data storage used in the Faculty of Social Sciences.

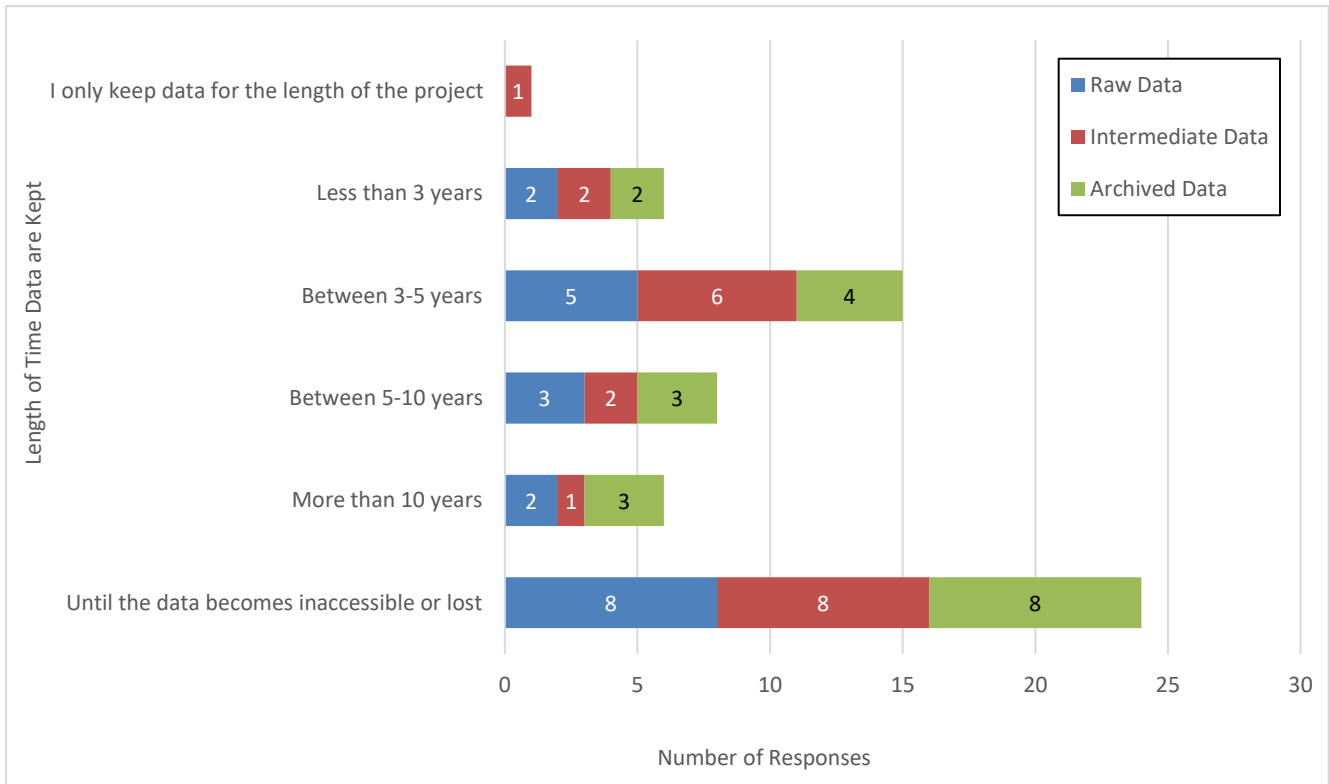
Survey Question: Use the chart below to indicate the length of time after project completion that you typically intentionally keep each type of research data. Project completion could include until publication, for example.



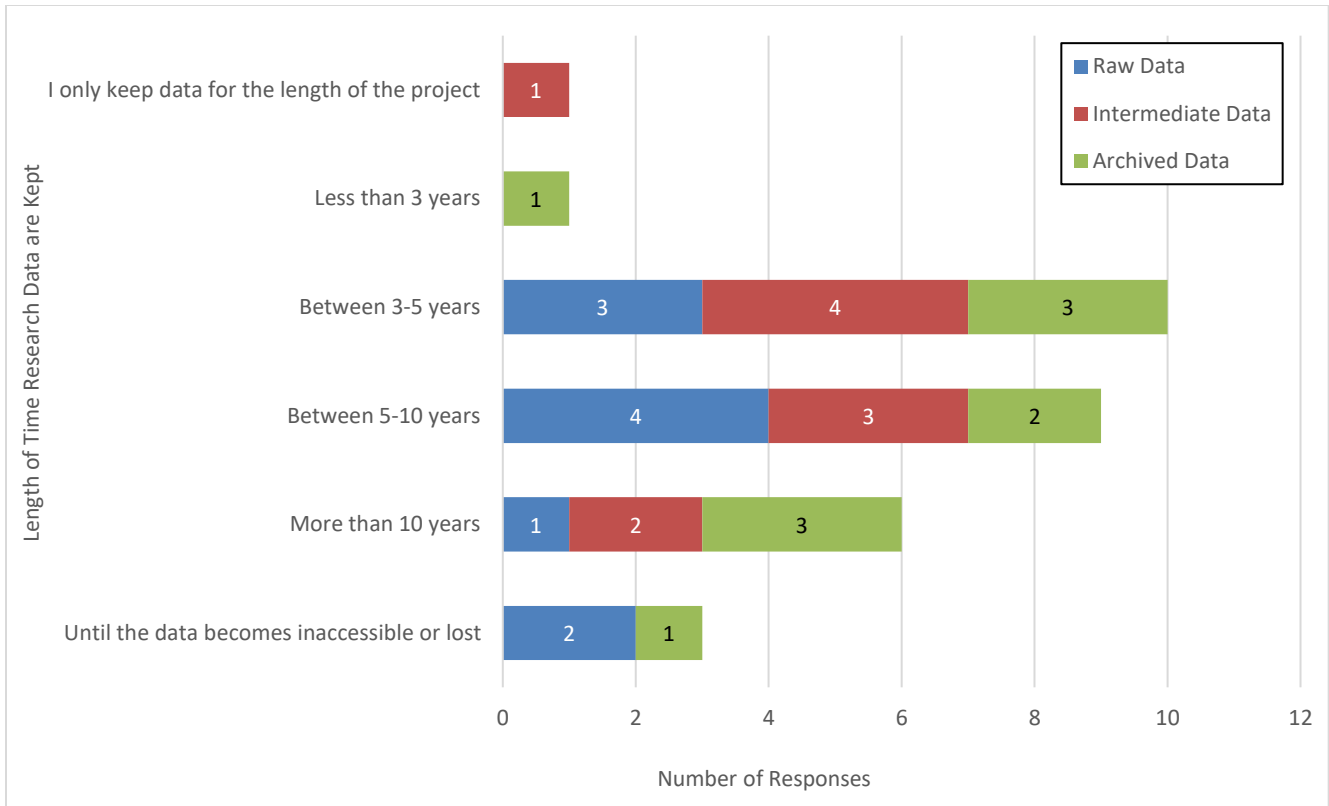
Length of Time Research Data are Kept in the Faculty of Arts.



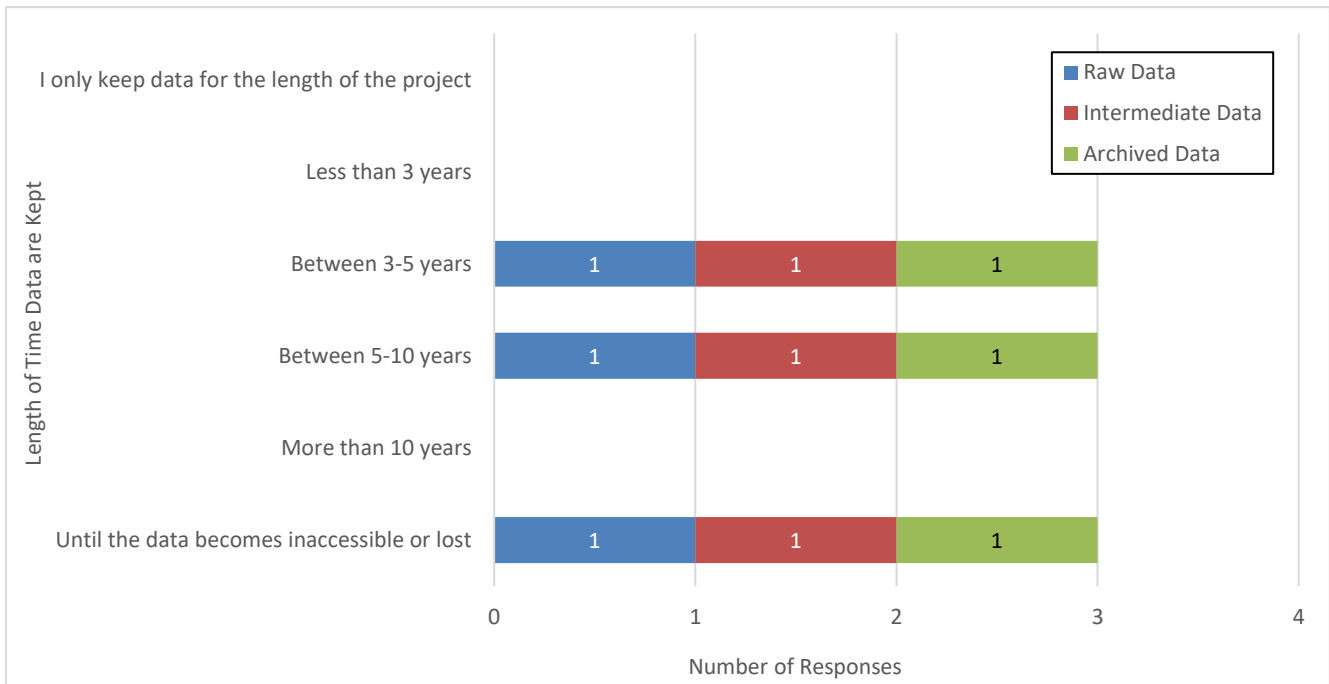
Length of Time Research Data are Kept in the Faculty of Education.



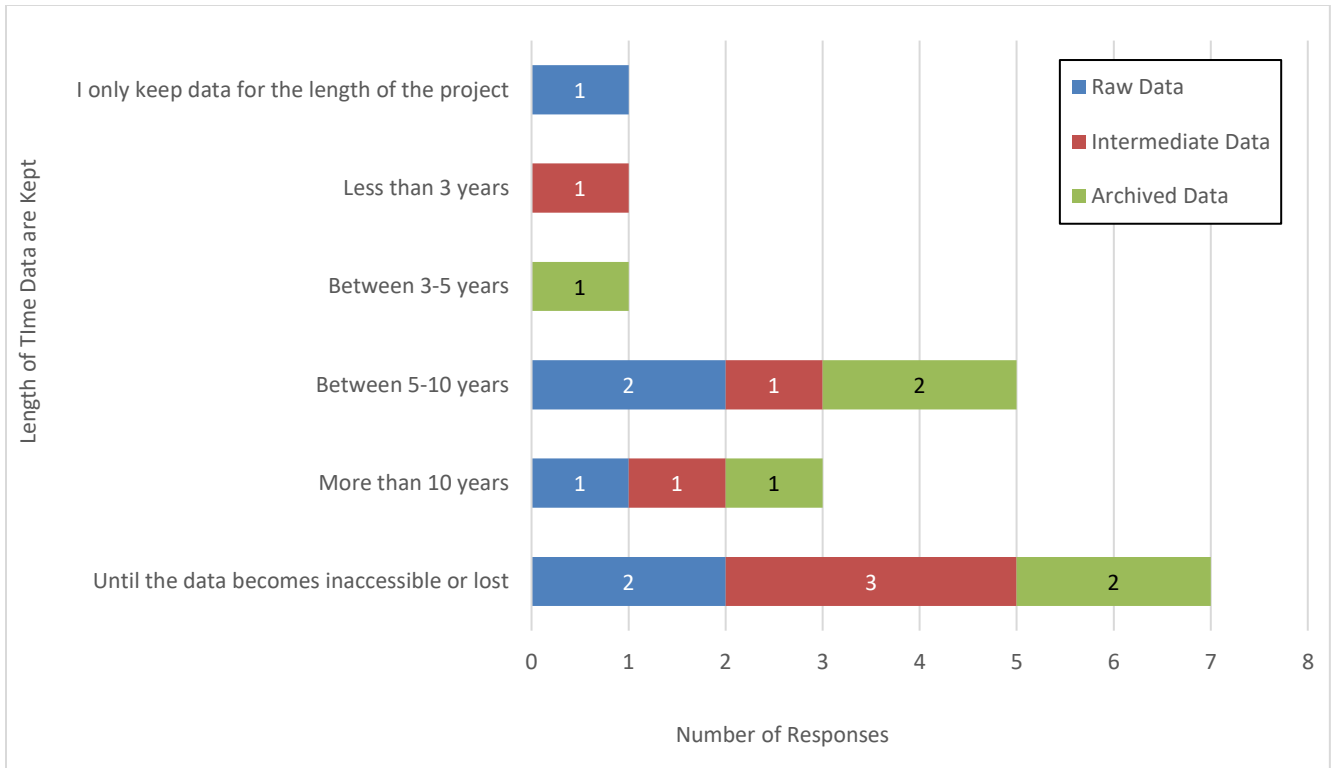
Length of Time Research Data are Kept in the Faculty of Engineering.



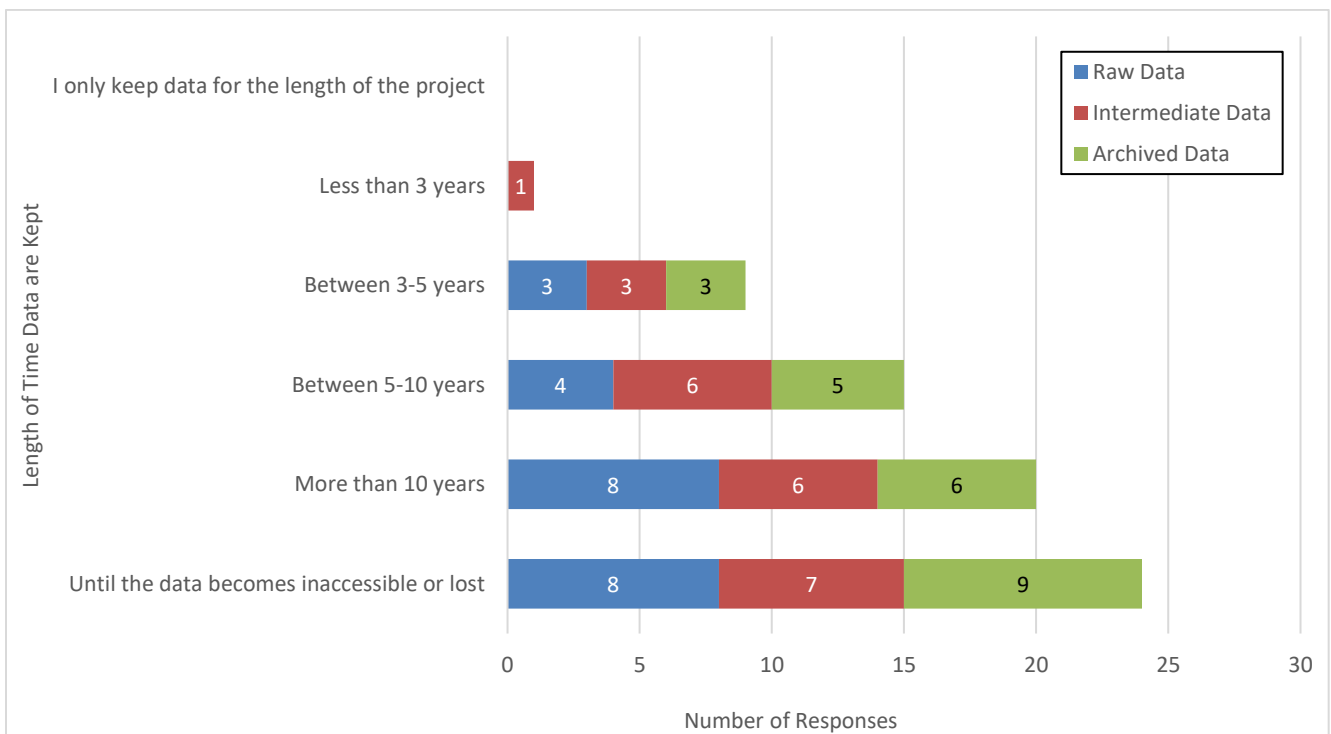
Length of Time Research Data are Kept in the Faculty of Health Sciences.



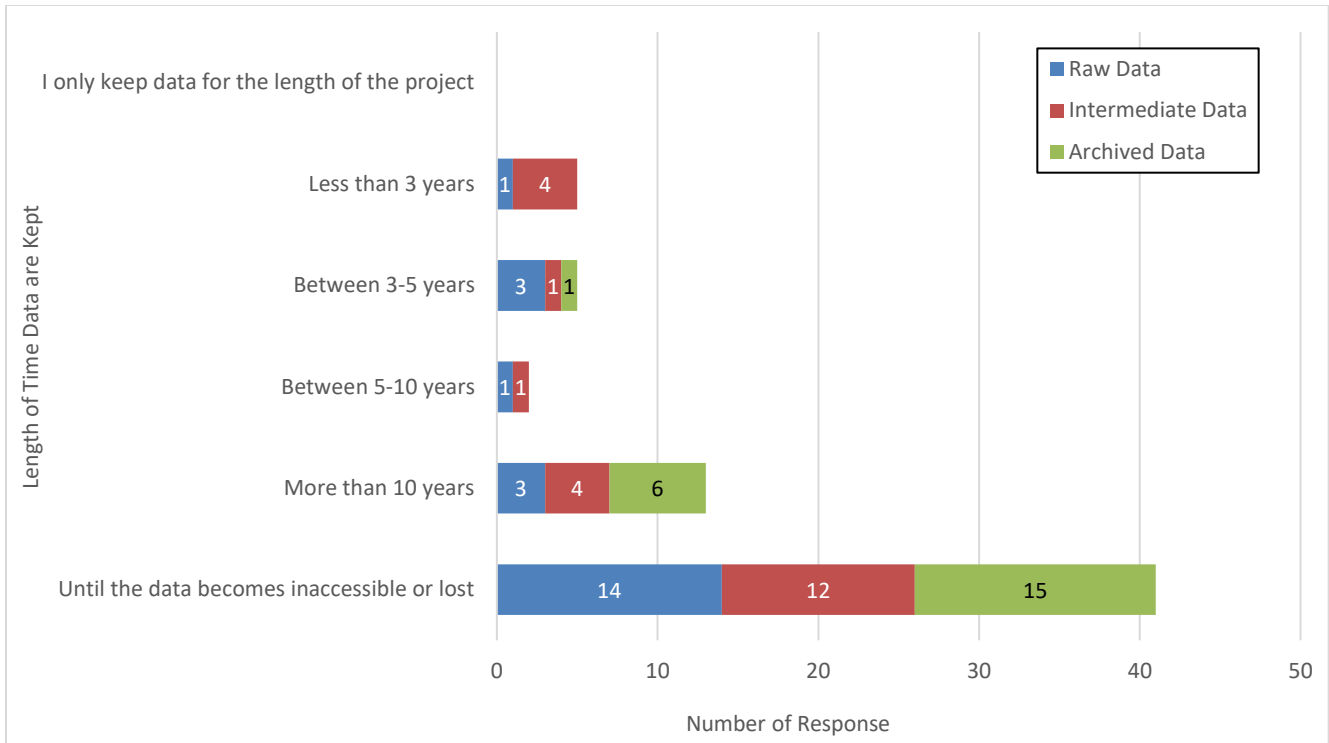
Length of Time Research Data are Kept in the Faculty of Law.



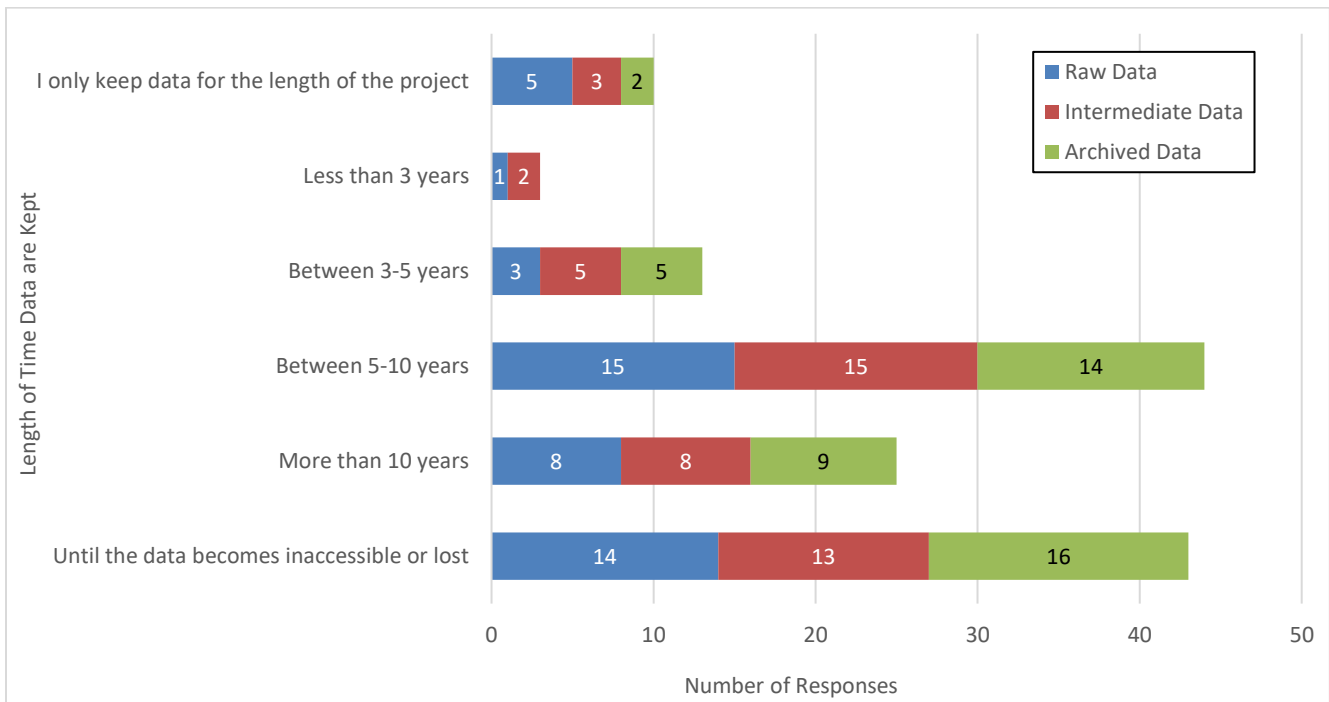
Length of Time Research Data are Kept in the Faculty of Management.



Length of Time Research Data are Kept in the Faculty of Medicine.



Length of Time Research Data are Kept in the Faculty of Science.



Length of Time Research Data are Kept in the Faculty of Social Sciences.

Survey Question: Which methods of sharing your research data do you currently use? Select all that apply. If you do not currently share your data, choose 'not currently sharing'.

	Not currently sharing	Personal request only	Online with restricted access	Institutional or personal website	Institutional repository, such as Dataverse	Supplementary files to journal	General or discipline-specific repository
Arts	12	20	8	4	4	4	1
Education	1	3	4	1		1	
Engineering	7	9	3	5		1	3
Health Sciences	2	6	3	1	3	2	
Law	1	1	1		1	1	
Management		5	1			1	
Medicine	3	15	6	2	2	13	6
Science	3	12	5	4		12	6
Social Sciences	20	21	8	2	3	5	1
Not Declared		1	1				

Survey Question: Hypothetically speaking, which methods of sharing your research data would you consider using in the future? Select all that apply. If you do not plan to share your data in the future choose 'not planning to share'.

	Not planning to share	Share by personal request	Online with restricted access	Institutional or personal website	Institutional repository, such as Dataverse	Supplementary files to journal	General or discipline-specific repository
Arts	5	23	15	14	7	7	2
Education		6	4	3	2	2	1
Engineering	2	13	6	7	4	4	6
Health Sciences	1	6	6	2	6	3	1
Law		1	1	1	2	1	
Management		5	2	2		1	
Medicine	1	11	9	7	11	12	9
Science		11	8	7	6	14	9
Social Sciences	9	28	15	7	6	9	2
Not Declared		1	1				

Survey Question: If your research data were not affected by restrictions or embargoes, with whom would you be willing to share them? Select all that apply:

	Nobody	Immediate collaborators	Researchers in my department	Researchers at uOttawa	Researchers in my field	Researchers outside my field	Anybody, including the general public
Arts	2	13	9	7	19	6	20
Education		5	3	3	3	1	3
Engineering	1	8	1	2	5	1	9
Health Sciences	1	8	4	3	6	1	1
Law		1	1		1	1	3
Management		6	4	3	4	2	
Medicine		17	12	8	12	9	5
Science		9	4	3	10	1	9
Social Sciences	1	28	12	8	22	9	9
Not Declared							

Survey Question: What benefits do you see to sharing your research data? Select all that apply. If you see no benefits, choose 'I see no benefits to sharing my data'.

	No benefits	Safeguards against misconduct	Training next generation researchers	Enables my data to be cited	Strengthens my academic portfolio	Increased my ability to obtain funding	Encourages collaborative scholarship	Encourages interdisciplinary research	Moves my field of research forward	Reduces redundant data collection	Supports open access to knowledge	Helps verify results	Help data integrity	Other
Arts	5	7	24	20	12	11	25	24	27	15	25	12	10	2
Education	1	1	3	2	1	1	4	3	2		2	3	2	1
Engineering	3	8	10	10	7	5	15	10	8	7	10	10	6	1
Health Sciences	1	1	5	4	1		6	6	7	6	4	4	2	1
Law		1	3	2	1		2	2	1	1	2		1	
Management	1		2	2	2	1	4	2	2	2	1	2	2	
Medicine	1	13	10	12	8	7	15	14	17	15	15	14	8	
Science		11	10	8	1		12	7	11	5	8	8	8	1
Social Sciences	8	21	27	20	6	7	26	17	17	13	19	15	12	1
Not Declared							1							

Survey Question: Data management plans typically address questions about research data types and formats: standards to be used for describing data; ethics and legal compliance; plans for preservation, access, sharing, and reuse; and responsibilities assigned and resources needed. 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? Select one:

	I would be able to draft a data management plan without assistance	I would need assistance and/or guided documentation	I would prefer to have assistance and/or guided documentation
Arts	7	17	13
Education		5	2
Engineering	4	12	3
Health Sciences	1	3	6
Law	1		2
Management		3	3
Medicine	3	11	9
Science	3	12	7
Social Sciences	7	19	19
Not Declared		1	

