Partnership
Canadian journal of library and information practice and research
Revue canadienne de la pratique et de la recherche en bibliothéconomie et sciences de l'information

Practices Before Policy: Research Data Management Behaviours in Canada
Les pratiques avant les politiques : comportements en matière de gestion des données de recherche au Canada

Melissa Cheung, Alexandra Cooper, Dylanne Dearborn, Elizabeth Hill, Erin Johnson, Marjorie Mitchell et Kristi Thompson

Volume 17, numéro 1, 2022

URI : https://id.erudit.org/iderudit/1090506ar
DOI : https://doi.org/10.21083/partnership.v17i1.6779

Aller au sommaire du numéro

Éditeur(s)

The Partnership: The Provincial and Territorial Library Associations of Canada

ISSN

1911-9593 (numérique)

Découvrir la revue

Citer cet article

Cheung, M., Cooper, A., Dearborn, D., Hill, E., Johnson, E., Mitchell, M. & Thompson, K. (2022). Practices Before Policy: Research Data Management Behaviours in Canada. *Partnership*, *17*(1), 1–80. https://doi.org/10.21083/partnership.v17i1.6779

Résumé de l'article

En prévision de la future Politique des trois organismes sur la gestion des données de recherche, un consortium de professionnels des bibliothèques universitaires canadiennes a interrogé des chercheurs sur leurs pratiques et leurs attitudes en matière de gestion des données de recherche (GDR) et sur leur intérêt pour les services de gestion des données. Des données recueillies de trois sondages ciblant les chercheurs en sciences et en génie, en sciences humaines et sociales ainsi qu'en sciences de la santé et en médecine ont été compilées pour créer un ensemble de données national. Cette étude est la première enquête à grande échelle sur les pratiques de GDR des chercheurs au Canada et est l'une des rares enquêtes récentes pluri-institutionnelles et multidisciplinaires sur ce sujet. Cet article présente les résultats du sondage afin d'évaluer l'état de préparation des chercheurs à répondre aux exigences de la politique de GDR, notamment en lien avec la préparation d'un plan de gestion des données (PGD) et le dépôt des données dans un dépôt numérique. Les résultats du sondage mettent aussi en évidence des tendances communes à l'ensemble du pays tout en révélant différentes pratiques et attitudes entre les disciplines. En se basant sur nos résultats, la plupart des chercheurs devront changer leurs comportements liés à la GDR afin de satisfaire aux exigences de la Politique des trois organismes sur la GDR. Les données recueillies fournissent des informations qui peuvent aider les institutions à prioriser le développement de services et d'infrastructures qui répondront aux besoins des chercheurs.

PARTNERSHIP
The Canadian Journal of Library and Information Practice and Research
Revue canadienne de la pratique et de la recherche en bibliothéconomie et sciences de l'information

vol. 17, no. 1 (2022)
Theory and Research (peer-reviewed)
DOI: https://doi.org/10.21083/partnership.v1i2.6779

Practices Before Policy: Research Data Management Behaviours in Canada

Les pratiques avant les politiques : comportements en matière de gestion des données de recherche au Canada

Melissa Cheung
Science & Engineering Research Librarian
University of Ottawa
melissa.cheung@uottawa.ca

Alexandra Cooper
Data Services Coordinator
Queen's University
coopera@queensu.ca

Dylanne Dearborn
Research Data Management Coordinator
University of Toronto
dylanne.dearborn@utoronto.ca

Elizabeth Hill
Data Librarian
Western University
ethill@uwo.ca

Erin Johnson
Acting Head - Discovery, Description and Metadata
Western University
ejohns83@uwo.ca

Marjorie Mitchell
Copyright, Scholarly Communications, and Research Data Management Librarian
University of British Columbia Okanagan Campus
marjorie.mitchell@ubc.ca

Kristi Thompson
Data Management Librarian
Western University
kthom67@uwo.ca

Abstract / Résumé

In anticipation of the then forthcoming Tri-Agency Research Data Management Policy, a consortium of professionals from Canadian university libraries surveyed researchers on their research data management (RDM) practices and attitudes, and their interest in data management services. Data collected from three surveys targeting researchers in science and engineering, humanities and social sciences, and health sciences and medicine were compiled to create a national dataset. The present study is the first large-scale survey investigating researcher RDM practices in Canada and one of the few recent multi-institutional and multidisciplinary surveys on this topic.

This article presents the results of the survey to assess researcher readiness to meet RDM policy requirements, namely the preparation of data management plans (DMPs) and data deposit in a digital repository. The survey results also highlight common trends across the country while revealing differences in practices and attitudes between disciplines. Based on our survey results, most researchers would have to change their RDM behaviors to meet Tri-Agency RDM policy requirements. The data we gathered provides insights that can help institutions prioritize service development and infrastructure that will meet researcher needs.

Data from the study are available in the Canadian RDM Survey Dataverse.

En prévision de la future Politique des trois organismes sur la gestion des données de recherche qui allait être adoptee, un consortium de professionnels des bibliothèques universitaires canadiennes a interrogé des chercheurs sur leurs pratiques et leurs attitudes en matière de gestion des données de recherche (GDR) et sur leur intérêt pour les services de gestion des données. Des données recueillies de trois sondages ciblant les chercheurs en sciences et en génie, en sciences humaines et sociales ainsi qu'en sciences de la santé et en médecine ont été compilées pour créer un ensemble de données national. Cette étude est la première enquête à grande échelle sur les

pratiques de GDR des chercheurs au Canada et est l'une des rares enquêtes récentes pluri-institutionnelles et multidisciplinaires sur ce sujet.

Cet article présente les résultats du sondage afin d'évaluer l'état de préparation des chercheurs à répondre aux exigences de la politique de GDR, notamment en lien avec la préparation d'un plan de gestion des données (PGD) et le dépôt des données dans un dépôt numérique. Les résultats du sondage mettent aussi en évidence des tendances communes à l'ensemble du pays tout en révélant différentes pratiques et attitudes entre les disciplines. En se basant sur nos résultats, la plupart des chercheurs devront changer leurs comportements liés à la GDR afin de satisfaire aux exigences de la Politique des trois organismes sur la GDR. Les données recueillies fournissent des informations qui peuvent aider les institutions à prioriser le développement de services et d'infrastructures qui répondront aux besoins des chercheurs.

Les données de cette enquête sont disponibles dans le Canadian RDM Survey Dataverse.

Keywords / Mots-clés

research data management (RDM), data management plans (DMP), data sharing, services, researcher practices

gestion des données de recherche (GDR), plans de gestion des données (PGD), partage des données, services, pratiques de recherche

Introduction

Institutions are facing new challenges as public funding agencies and publishers recognize the value of research data and encourage researchers to make their data available along with their publications. The success of these new policies and practices related to research data management (RDM), data sharing, and data preservation depends on researcher awareness and capacity to incorporate new practices into their workflows. In the US and the UK, funding mandates require the submission of a data management plan (DMP) as part of select grant applications and that research data be made available for reuse. In Canada, the Canadian Institutes of Health Research (CIHR), the Natural Sciences and Engineering Research Council of Canada (NSERC), and the Social Sciences and Humanities Research Council of Canada (SSHRC), collectively referred to as the Tri-Agencies, are the primary funding agencies. The Tri-Agency Statement of Principles on Digital Data Management, announced in 2015, outlines "the agencies' overarching expectations regarding research data management"

(Government of Canada, 2015). These principles were later formalized in the Tri-Agency Research Data Management Policy, released in March 2021 (Government of Canada, 2021). The incremental implementation of this policy will require researchers to submit a DMP for certain funding applications and to deposit their research data in a digital repository. Additionally, institutions must create an institutional RDM strategy outlining how they will support their researchers in meeting the Tri-Agencies' expectations.

A consortium of professionals from Canadian university libraries surveyed researchers between 2015 and 2019 to identify RDM practices and perceptions and to gauge interest in services in anticipation of the policy's release. The objectives of the survey were to inform the development of institutional services to help researchers meet RDM policy requirements. Data collected from the surveys will enable a deeper understanding of the Canadian RDM landscape, helping us to answer the questions:

- What RDM practices, perceptions, and interest in services do researchers have and do these vary by discipline?
- What do researcher practices tell us about how prepared they are to meet new RDM policy requirements?
- What do these practices and preferences tell us about opportunities for institutional support?

Literature Review

Building infrastructure and support for RDM is a "wicked problem" that is highly complex and involves a variety of stakeholders, each having a different perspective on, and definition of, the problem (Cox et al., 2016). It is difficult to develop clear solutions due to the lack of information on the scope of the problem, which may include long-standing dilemmas in other areas. A number of international studies surveying researchers on their RDM practices and their preferences for RDM support have been conducted. Major studies include a survey of National Science Foundation grant recipients by Tenopir et al. (2020), a study of researchers at five Illinois universities by Schumacher & VandeCreek (2015), a primarily U.S. snowball sample by Tenopir et al. (2015), and a national survey by Kim and Stanton (2015) of researchers in science, technology, engineering and math in the U.S. A number of single-institution studies of researchers have recently been carried out in the U.S. (Johnson & Steeves, 2019; Joo & Peters, 2020; Whitmire et al., 2015; Wiley & Burnette, 2019), Singapore (Majid et al., 2018), the Netherlands (Mancilla et al., 2019), and Australia (Kennan & Markauskaite, 2015; Krahe et al., 2020). However, RDM studies in the literature generally do not use a common

methodology (Goben & Griffin, 2019) and few survey multiple institutions using a common instrument.

A few studies on RDM behaviour in the Canadian context have been published. Among these, Argàez and Shearer (2008) completed a scan of Canadian and international research data management initiatives; Guindon (2014) conducted a small study of faculty at Concordia University; Perrier and Barnes (2018) published the results of a series of focus groups; and Abel et. al (2020) completed a survey to assess institutional capacity for RDM at Canadian institutions. The present study is the first large-scale survey covering researcher RDM practices in Canada and one of the few recent multi-institutional and multidisciplinary surveys on this topic. The aggregated results of our survey allow us to make direct comparisons between disciplines and researcher ranks while discovering commonalities and highlighting trends across the country.

Methodology

Survey Development and Deployment

An open partnership consisting of over 20 institutions[1], known as the Canadian RDM Survey Consortium, worked together to develop a survey instrument targeting three disciplinary areas: science and engineering (2015), humanities and social sciences (2016), and health and medical sciences (2017). Survey development was guided by previous RDM surveys or publications from the University of Minnesota (Johnston & Jeffryes, 2014), Purdue University (Carlson et al., 2011), Utah State University (Diekema et al., 2014), the University of Nottingham (Parsons et al., 2013), and the University of Colorado Boulder (Vice Chancellor for Research's Data Management Task Force, 2012).

The survey instrument had 25 standard questions organized into four sections: demographic and general contextual questions, working with research data, data sharing, and funding mandates and RDM services. The standard questions were modified for each disciplinary area with specific examples, choice of funding agencies, and other details. Any differences with specific questions will be noted where applicable.

[1] Participation in the Canadian RDM Survey Consortium was open to any Canadian post-secondary institution. Partner institutions were welcome to join at various stages of the research project and they were also free to end their participation at any time. The survey instrument was made publicly available under open access, and there may have been institutions that conducted the survey but chose not to share their data with the consortium. As a result, we do not have an exact number of participating institutions in the consortium.

Financial support for translating the survey instruments and guidance documents into French was provided by the Portage Network, CIHR, and SSHRC.

Survey participants included all ranks of faculty members, lecturers, and postdoctoral and research fellows. In addition, some institutions included graduate students at the master's and doctoral levels. Each institution followed its own ethics approval process and implemented the survey using its own survey software. Thirteen members of the consortium contributed data to this publication (see Table 1).

Table 1

National Dataset Survey Responses by Institution and Year Administered

Name of institution	2015	2016	2017	2018	2019	Total
Cape Breton University	--	--	--	10	--	10
École nationale d'administration publique	--	--	--	14	--	14
Institut national de la recherche scientifique	--	--	--	104	--	104
Université du Québec à Trois-Rivières	--	--	--	--	75	75
Université TELUQ	--	--	--	26	--	26
Queen's University*	396	395	213	--	--	1004
University of Ottawa	--	144	33	--	--	177
University of Toronto	95	--	181	--	--	276
University of Waterloo	46	--	17	--	--	63

Western University*	--	--	210	--	--	210
Trinity Western University	--	--	--	32	--	32
University of British Columbia*	94	66	89	--	--	249
University of Victoria*	--	--	418	--	--	418
Total	536	605	1256	186	75	**2658**

Note: Individual surveys were administered once at each institution. Institutions with multiple dates indicate that different disciplinary versions were administered in different years. Institutions marked with an asterisk (*) included graduate students (master's and PhD) in their surveys

Data Processing and Analysis Methods

The institutional datasets were standardized according to instructions provided by the Canadian RDM Survey Consortium. Each institution sent anonymized and cleaned datasets to the Data Manager, who compiled the files into a merged dataset. Analysis was carried out using SPSS and consisted of univariate frequencies and bivariate cross-tabulations. Associations between cross-tabulated variables were checked for significance using Pearson's chi-squared test of goodness of fit.

In the analysis, unless otherwise specified, Assistant Professors, Associate Professors, Full Professors, Clinical Colleagues, Adjunct Professors, Professors Emeriti, Lecturers, and Postdoctoral Fellows will be referred to as "Academic staff" when compared with "Graduate students" and "Other" respondents.

Limitations

As shown in Table 1, the surveys were administered over a period of several years and were conducted at different times at different institutions. This study took place after the Tri-Agency Statement of Principles on Digital Data Management (which offered a preview of the upcoming policy) was released in 2015 and before the formal Tri-Agency Research Data Management Policy was finalized in 2021. Therefore, each institution

collected survey data under a similar regulatory and funding environment and the results provide a snapshot of the Canadian RDM landscape during this time.

Participants that completed the survey were self-selected, which may lead to bias in the results. There were also some differences between different iterations of the survey instruments. One important difference is that the question related to teaching RDM topics was not included in the original science and engineering survey. Other differences in the survey instruments are detailed in the results section.

Institutions categorize subject areas into departments and faculties differently, so we created a standardized list of broad subject categories to merge the datasets. Grouping responses in broad subject categories, such as arts and humanities or science, limited our ability to compare disciplinary differences at a granular level, such as history or physics.

Graduate students were not originally part of the target population of the survey. Because of this, the wording of some questions may not have resonated with graduate students. For example, questions related to responsibilities of leading a project or preparing funding applications typically fall to a faculty member or principal investigator. As such, graduate students may have answered these questions from a perspective the survey was not designed to capture. Regardless, some schools chose to include graduate students as part of their survey populations and those results have been included in this study.

Results

The results present major findings from all survey respondents and highlight some differences by discipline or rank. Detailed disciplinary results tables can be found in Appendix A (RDM practices by discipline) and Appendix B (Interest in potential RDM services by discipline).

Demographics of survey participants

The national dataset used in the analysis for this study includes 2658 survey responses from 13 institutions. Regional distribution is shown in Table 2.

Table 2

National Dataset Responses by Region

Region	Percent (%)	N=
Atlantic and Quebec	8.6	229
Ontario	65.1	1730
West	26.3	699
Total	**100**	**2658**

An approximate target population and response rate could not be calculated given the variation in survey deployment at each participating institution.

All major fields of study (disciplines) were represented among the survey respondents as seen in Table 3. The participating institutions that also surveyed graduate students were located in Ontario and the Western provinces. The largest group of respondents self-identified as assistant, associate or full professors or clinical colleagues (49.7%, n=1190). Graduate students represented 32.5% (n=778) of the total survey respondents and between 20.5% and 43.7% of each disciplinary group included in the survey. The remaining respondents included postdoctoral fellows (5.9%, n=141), adjunct professors (3.6%, n=87), professors emeriti (3.3%, n=78), lecturers (2%, n=49), and other (2.9%, n=70).

Table 3

Respondent's discipline

Field of Study	Percent (%)	N=
Social Sciences	18.1	455
Arts/Humanities	15.2	383
Science	15.2	381

Engineering	11.9	299
Not declared	11.4	287
Medicine/Preclinical Sciences	11.2	281
Business/Management, Education, Law	9.4	237
Health Sciences	4.9	122
Interdisciplinary/Other	2.7	69
Total	**100**	**2514**
Not declared/not asked by administering institution		431

To understand how requirements from funding agencies may impact researchers, we asked respondents to indicate all funding sources they have used within the past five years or are planning to apply for in the next five years. A large number of academic staff reported that they have used, or plan to use, Tri-Agency funding (73.1%, n=1129). Of these respondents, 33.5% (n=508) selected SSHRC, 34.1% (n=518) selected NSERC, and 24.2% (n=367) selected CIHR.

Working with Research Data

We asked respondents to select all the types of research data they generated or used in a typical research project. The most frequently generated or used data types were text (65.4%, n=1738), numerical (47.5%, n=1262), and multimedia (36.2%, n=962). There were highly statistically significant differences in the type of research data generated or used across disciplines.

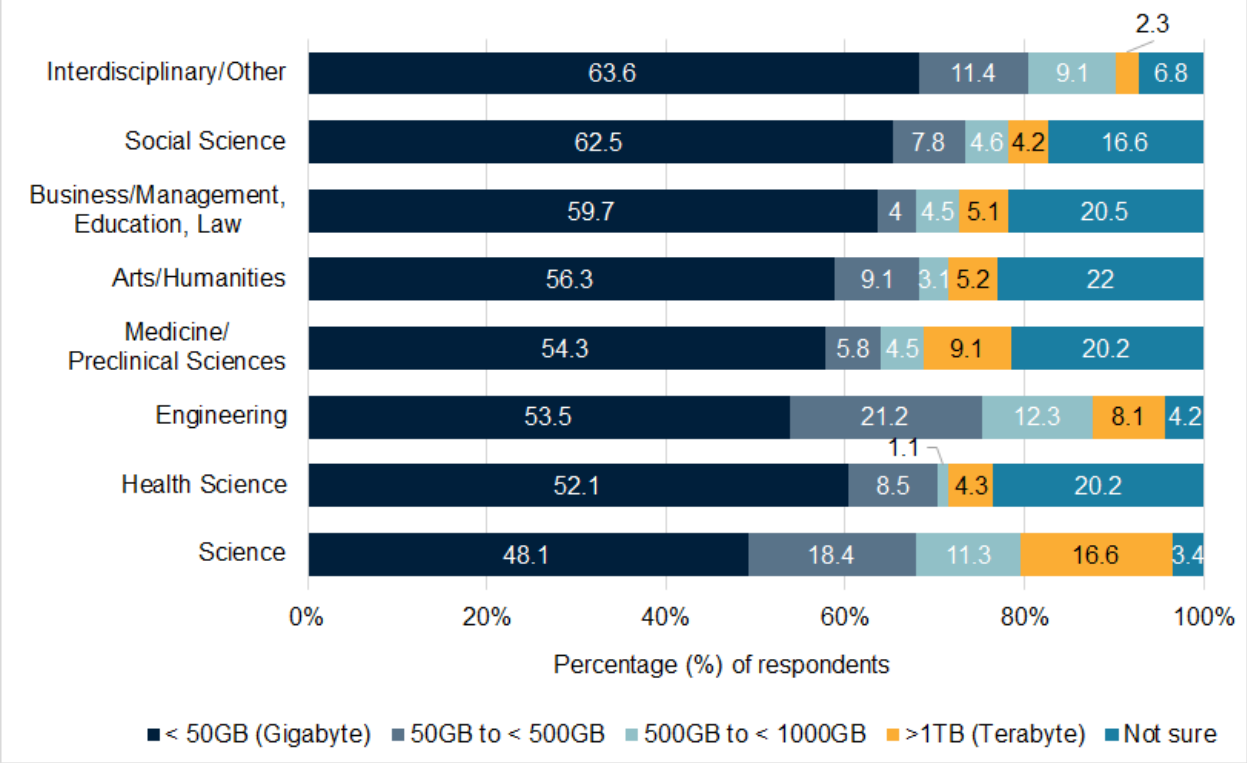
We also asked how many research projects respondents led in the past year and how much data storage they used in an average research project to assess approximately how much data researchers are generating. The results showed that 43.5% of respondents (n=852) led one to two research projects in the past year; 27% (n=529) led three to five projects; and 13.6% (n=267) led more than five projects. However, 2.3% (n=45) of respondents were not sure how many projects they led and 13.5% ("not applicable", n=265) did not lead any projects in the past year. Respondents in science led more than five research projects in a year (26.3%, n=83) more often than

respondents in other disciplines (4.4% – 16.4% lead more than five projects) with disciplinary differences being statistically significant ($X^2(28, N=1739)=144.93, p<.001$).

The majority of respondents (55.1%, n=1075) reported using less than 50GB of data storage in an average project while 7.9% (n=154) reported using more than 1TB and 15% (n=293) reported that they were not sure how much data storage an average project used. Respondents in science were more likely to indicate using more than 1TB of data storage (16.6%, n=53) than other disciplines. Statistically significant disciplinary differences also show that respondents in science, engineering, and interdisciplinary/other were more aware of how much data storage they used for an average research project compared to other disciplines ($X^2(35, N=1739)=252.26, p<.001$) (Figure 1).

Figure 1

Estimated amount of data storage used in an average research project by discipline



Cross-tabulation of Q8. “How much data storage do you estimate you use in an average

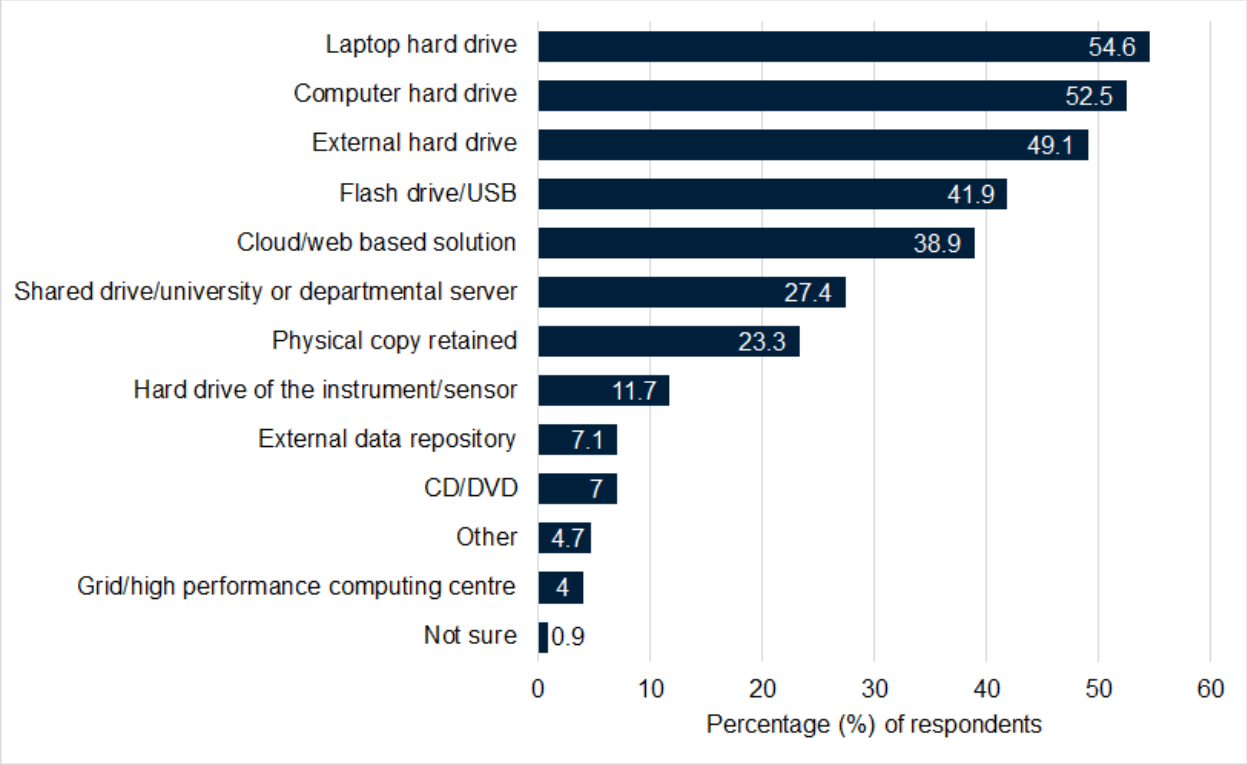
research project? Select one." and FIELDGR. "Respondent field of study (grouped)" [derived variable]." (n=1730)

When asked to indicate all the places where they were storing research data from their current project(s), the options most frequently selected by respondents were laptop hard drive (54.6%, n=1451), computer hard drive (52.5%, n=1345), and/or external hard drive (49.1%, n=1306) as shown in Figure 2. Flash drives or USBs were also frequently used with 41.9% (n=1113) of respondents selecting this response. There was a relatively low number of respondents reporting the use of institutional storage solutions such as a shared drive or university or departmental server (7.4%, n=728). We did not ask respondents to differentiate between short-term and long-term data storage; therefore, we are unable to determine if a respondent's choice in storage media is influenced by the stage of the research project. Furthermore, we cannot draw any conclusions about whether or not respondents were following best practices for data storage for their projects.

There was a statistically significant difference in graduate students' use of shared drives or university or departmental servers compared to other respondents. Only 17.6% (n=137) of graduate students selected this storage option compared to 35% (n=540) of academic staff and 42.9% (n=30) of respondents who identified as other ($X^2(2, N=2392)=80.90$, $p<.001$). These results may reflect differences in access to institutional infrastructure, with fewer graduate students having access to shared drives and/or university or departmental servers.

Figure 2

Where research data is stored for current project(s)



Q11. “Please indicate where you store research data from your current project(s). Select all that apply.” (n=2658)

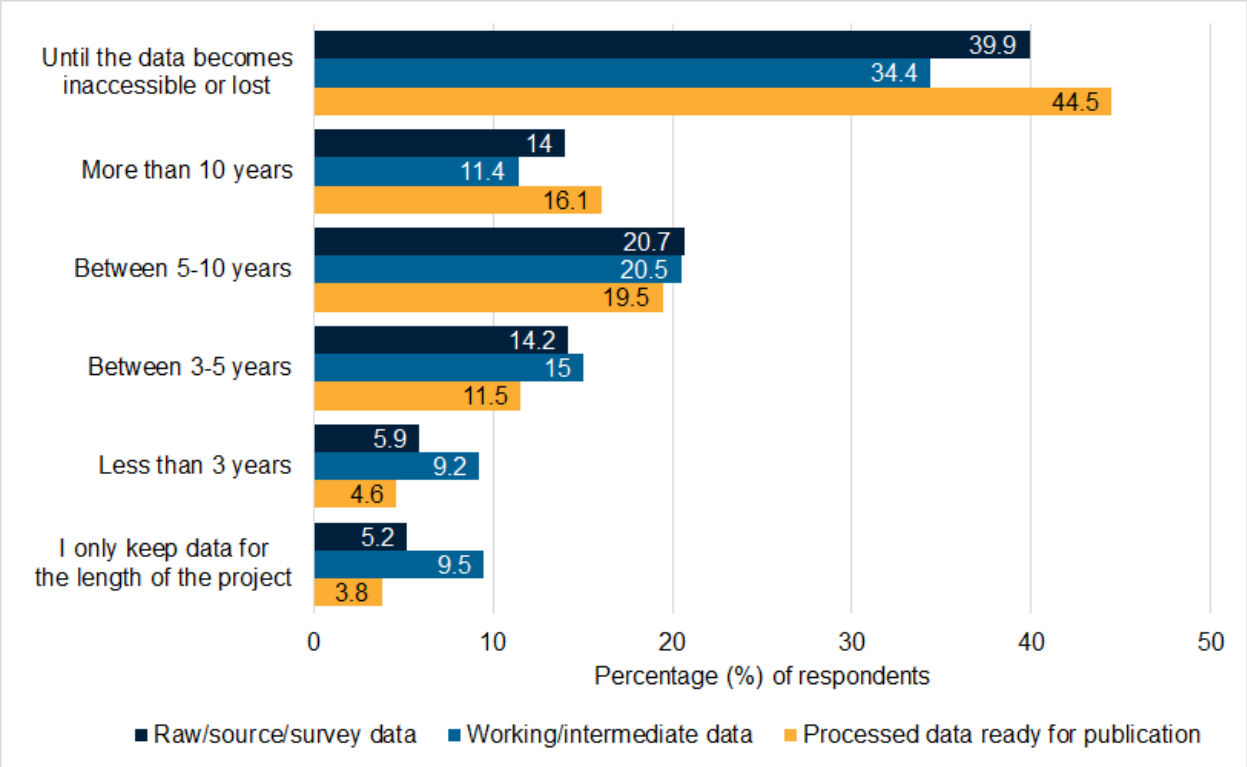
When comparing data storage options by discipline, respondents in science (26.5%, n=101) and engineering (24.7%, n=74) were more likely to report they use hard drives of instruments or sensors that generated the data (X^2 (10, N=2227) = 228.03; p<.001). Respondents in health sciences were more likely to keep physical copies of their data (41.8%, n=51, X^2 (10, N=2227) = 78.65; p<.001). These results reflect how research practices and the types of data generated vary across disciplines.

In addition to questions related to data storage, respondents were also asked how long they typically retain their data after project completion. Data were defined in stages as raw data (raw/source material/survey results), working data (or intermediate data), and processed data (ready for publication). As shown in Figure 3, respondents indicated that

they typically keep their research data until the data becomes inaccessible or lost for all stages: raw 39.9% (n=770), working 34.4% (n=657), and processed 44.5% (n=851). Notably, health sciences stood out as being more likely than other disciplines to report they retain their data for a set period of five to ten years after project completion.

Figure 3

Length of time research data are kept after project completion



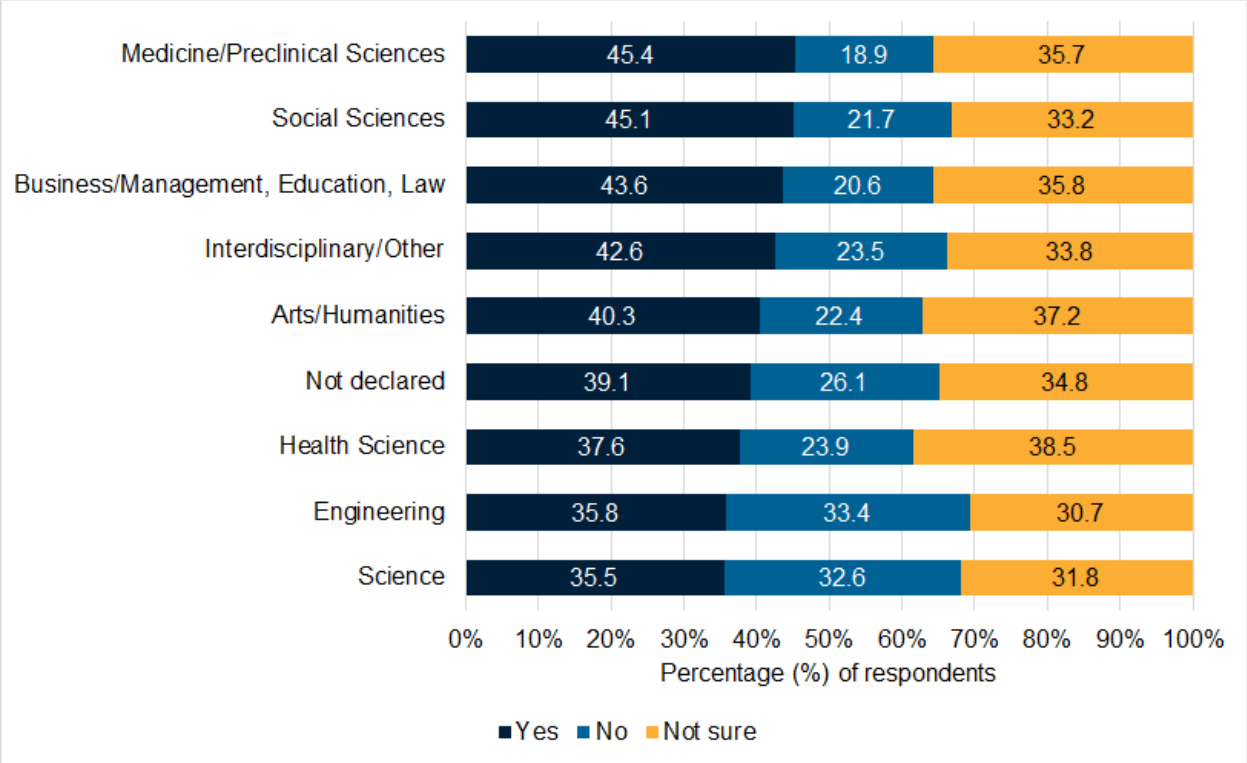
Q14. “Use the chart below to indicate the length of time after project completion that you typically intentionally keep each type of research data. Project completion could include until publication, for example.” (raw/source/survey data, n=310; working/intermediate data, n=1910; processed data ready for publication, n=1914)

Data documentation and description, such as variable and field definitions, codebooks, data dictionaries, metadata, and code for generating research results, are essential for data reuse. We asked respondents whether there is sufficient documentation for another person outside of their research team to understand and use their data, and we

also asked if the documentation was sufficient to replicate the methodologies that produced the data. Approximately 60% of respondents are not confident that their data had sufficient documentation for others to understand and use ("No" - 25.9%, n=604; "Not Sure" - 33.3%, n=775) and to replicate the results ("No" - 27.3%, n=530; "Not Sure" - 29.5%, n=572). The results were similar when comparing across disciplines and respondent rank as seen in Figure 4 and Figure 5.

Figure 4

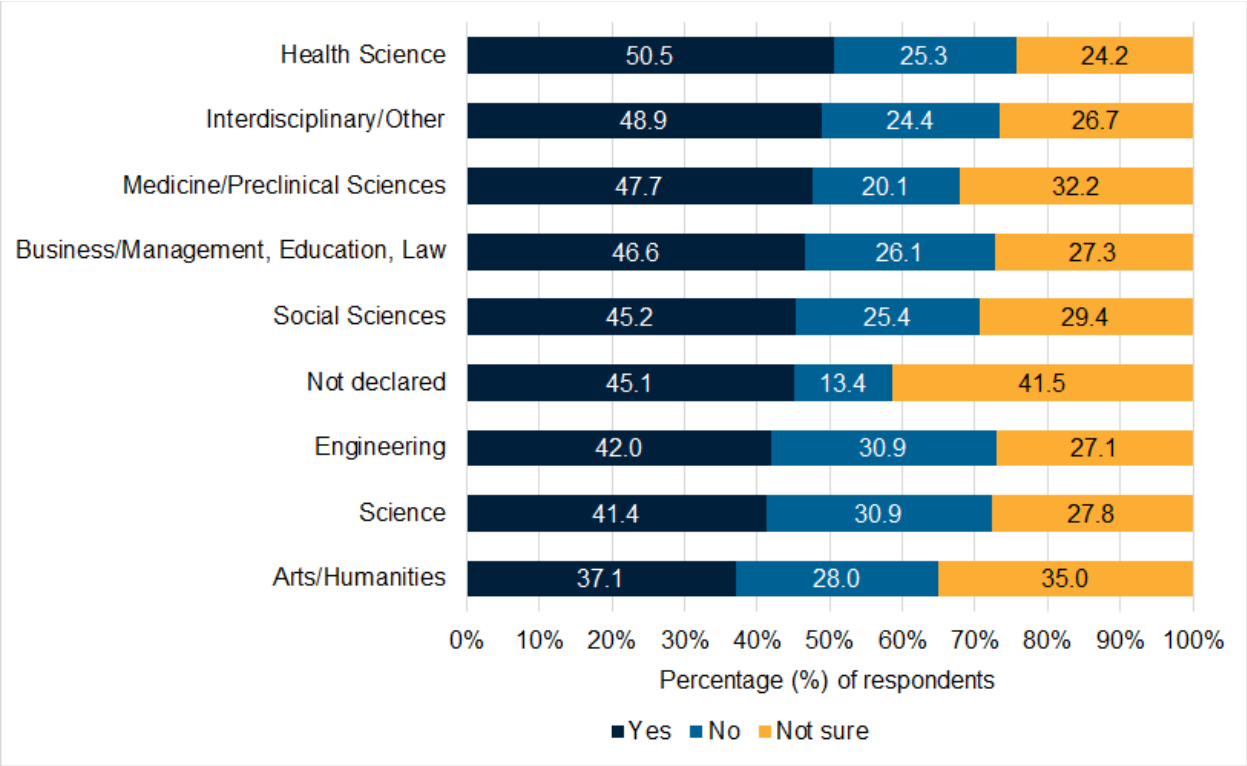
Sufficient documentation and description to understand and use the research data



Cross-tabulation of Q12. "Is there sufficient documentation and description (for example, variable and field definitions, codebooks, data dictionaries, metadata, scripts to run) for another person outside your research team to understand and use the research data?" and FIELDGR. "Respondent field of study (grouped)" [derived variable]. (n=2200)

Figure 5

Sufficient documentation and description for replication of methodologies that produced the data



Cross-tabulation of Q13. “Is there sufficient documentation and description (for example, variable and field definitions, codebooks, data dictionaries, metadata, scripts to run) retained in the same file, folder or document for another person outside your research team to replicate the methodologies that produced the data?” and FIELDGR. “Respondent field of study (grouped)” [derived variable]. (n=1812)

Data Deposit and Data Sharing

We gave respondents a list of methods of sharing research data and asked which methods they currently use and which they would consider using in the future. The most frequently selected data sharing method was sharing by personal request, with 41.8% of respondents (n=1108) reporting that they currently share their research data this way and 46% of respondents (n=803) reporting they would consider sharing their research

data this way in the future. However, some respondents may have believed that selecting “share by personal request only” excluded them from selecting any of the other responses to the question. Therefore, there may be more respondents who are or would consider sharing their data using other methods in addition to “share by personal request” than reported in the survey.

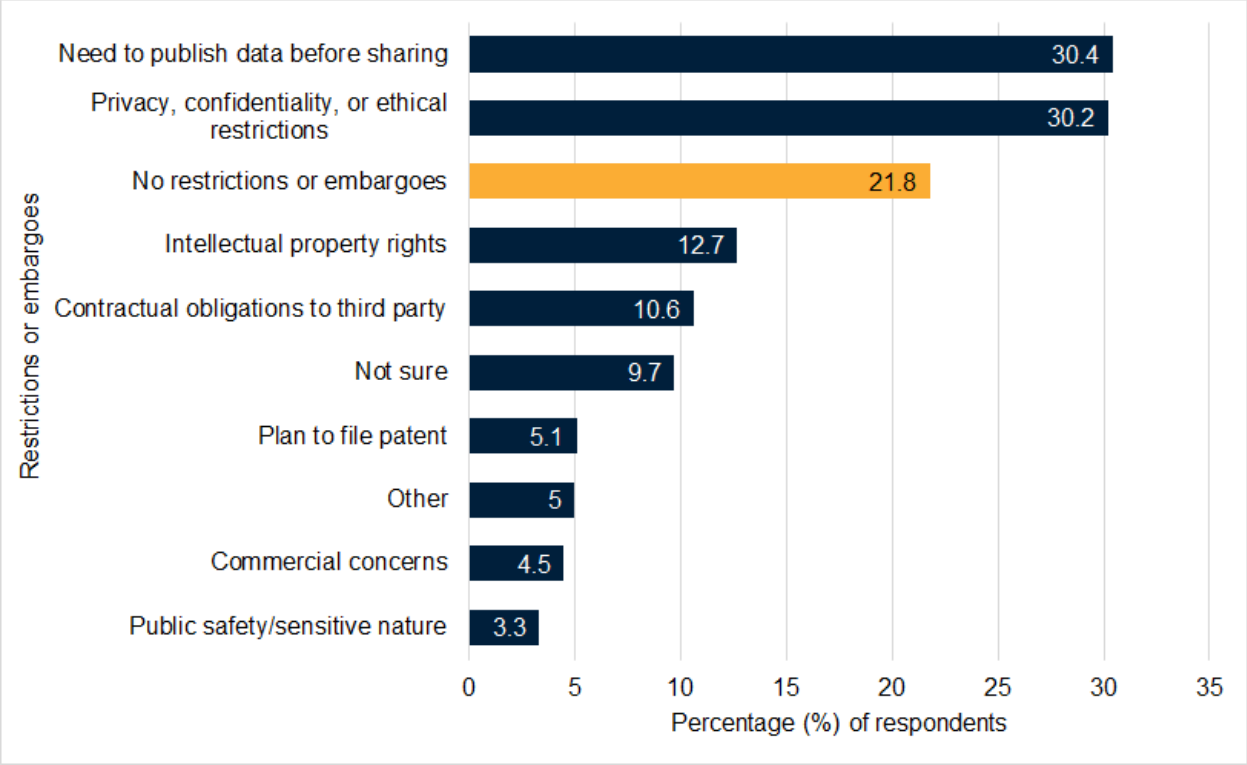
At the time of the survey, 28.4% of respondents (n=754) indicated they were not currently sharing their data and 13.7% of respondents (n=239) were not planning to share their data in the future. However, for each method of data sharing listed in the survey, more respondents were willing to consider using that method in the future than were using it at the time of the survey. This shows that some respondents who are not currently sharing their data are open to doing so in the future.

Data repositories were the least often selected method of currently sharing research data with only 4.9% of respondents (n=99) reporting that they currently deposit data in an institutional repository and 7.9% (n=198) indicating they deposit data in a general or discipline-specific repository. However, a larger percentage of respondents are willing to consider sharing data using these methods in the future: 20.2% (n=321) for institutional repositories and 14.2% (n=227) for general or discipline-specific repositories.

To understand the challenges and complexities related to data sharing, we asked respondents to identify all the applicable restrictions or embargoes on sharing their data. As shown in Figure 6, only 21.8% of respondents (n=554) reported that there are no restrictions or embargoes that prevent them from sharing their data. In comparison, 30.4% of respondents (n=786) reported they need to publish before they can share their data and 30.2% (n=779) indicated privacy, confidentiality, or ethics restrictions. Additionally, 9.7% of respondents (n=251) were not sure if they were allowed to share their data.

Figure 6

Restrictions or embargoes that may limit researchers' ability to share data with others

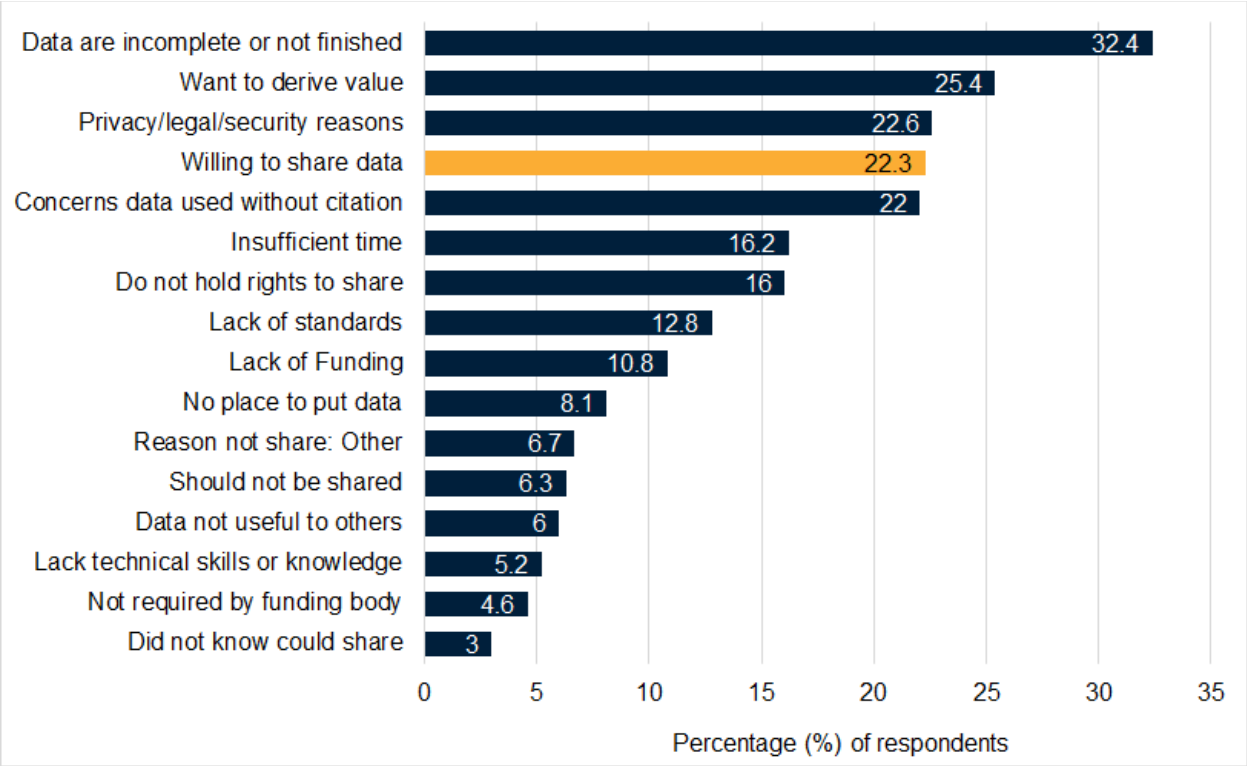


Q17. "Some research data cannot be shared because of legal or privacy restrictions or embargoes. Which of the following restrictions or embargoes may limit your ability to share your data with others? Select all that apply. If there are no restrictions or embargoes, choose 'there are no restrictions or embargoes on sharing my data with other parties'." (n=2438)

In addition to restrictions or embargoes, respondents reported a variety of reasons why they would not be willing to share their research data, as detailed in Figure 7. "Data are incomplete or not finished" was the most commonly selected reason to not share (32.4%, n=854), followed by the desire to still derive value from the data (25.4%, n=661). However, 22.3% (n=571) of respondents indicated they were willing to share their research data.

Figure 7

Reasons researchers are not willing to share their research data and associated methods/tools



Q19. “What, if any, are the reasons you would not be willing to share your research data and associated methods/tools? Select all that apply. If you are willing to share, choose ‘I am willing to share them’.” (n=2658)

Overall, 35.1% of respondents (n=639) were willing to share data with anyone, including members of the general public, 43.8% (n=798) were willing to share data with specified groups such as fellow researchers, 17.1% (n=312) would only share with immediate collaborators, and 4.0% (n=73) said they would not share data at all. This indicates most researchers are willing to share their research data with limitations, while only 29% of researchers (n=639) are willing to share their data openly.

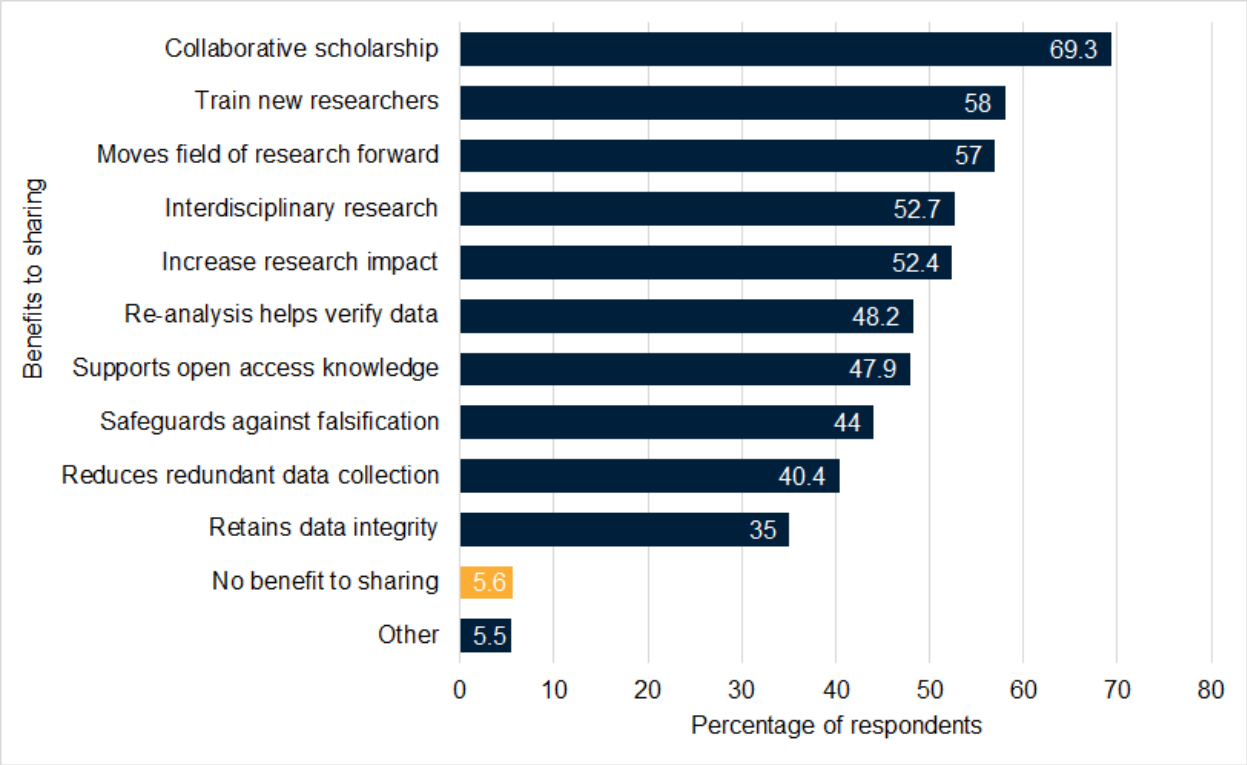
As seen in Figure 8, collaborative scholarship was the most commonly reported benefit for researchers willing to share their data. Respondents who were willing to share their

data with anybody, including the general public, were most likely to select support for open access to knowledge as a benefit to sharing their research data.

Of the respondents who indicated they were unwilling to share their research data, 59.3% (n=51) did not see any benefit to sharing. However, it is interesting to note that a small proportion of our survey respondents (17.4%, n=15) indicated they were unwilling to share their data but could see the benefit of sharing for collaborative scholarship.

Figure 8

Perceived benefits of sharing data



Q20. “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’.” (n=797)

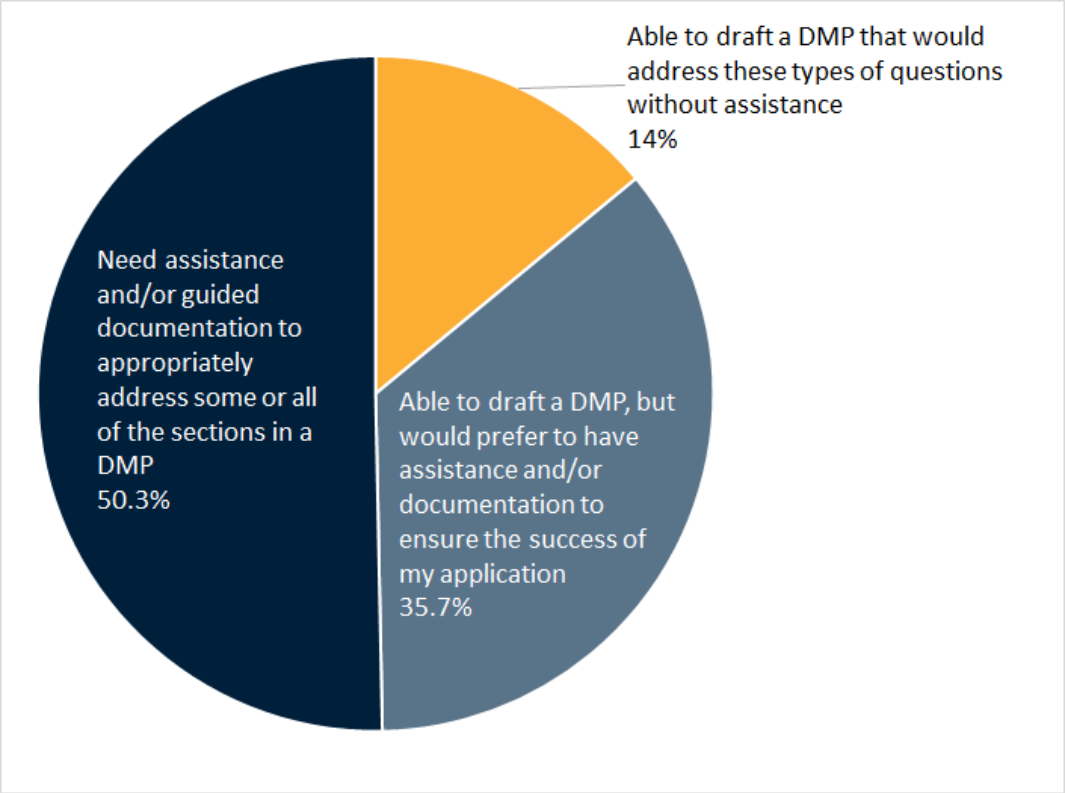
Ability to Draft Data Management Plans

We asked respondents about their ability to draft a DMP as part of a grant application. Very few respondents reported that they would be able to draft a DMP without

assistance, indicating a strong need for support services to assist researchers with data management planning (Figure 9).

Figure 9

Respondents' ability to draft a data management plan



Q22. "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." (n=2097)

Consistently, across all disciplines, respondents expressed a desire for assistance and/or guided documentation in creating DMPs. The responses from all disciplines

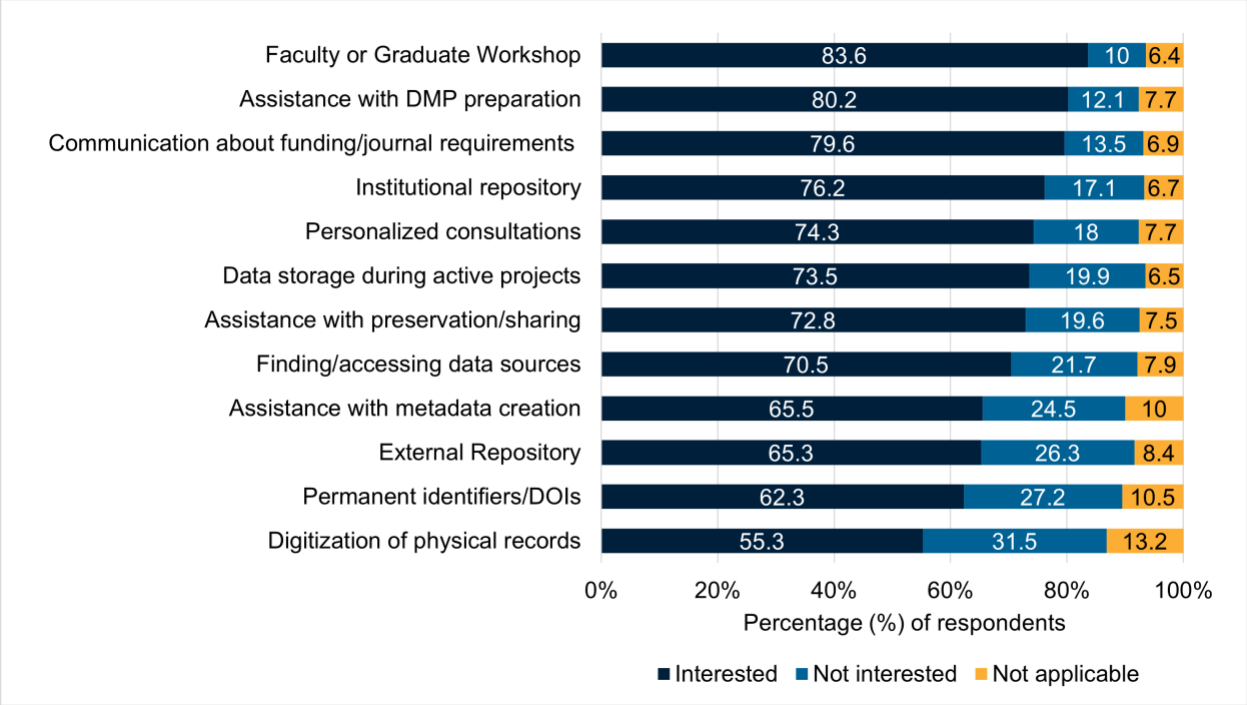
indicated high levels of need or preference for assistance, ranging from 82.2% in social sciences to 92.6% in the health sciences.

Interest in RDM Services

We asked respondents to rate their interest in proposed RDM services if DMPs were made part of grant applications from funding bodies such as SSHRC, CIHR, and NSERC. The majority of respondents expressed interest in all RDM services listed in the survey. Not all institutions surveyed graduate students, but respondent rank was a factor in interest in RDM workshops. In the survey, we asked about interest in workshops for faculty as well as interest in workshops for graduate students. To gauge interest in workshops, regardless of audience, we grouped the workshops on best practices in data management for faculty and for graduate students into one variable as illustrated in Figure 10. However, it should be noted that the service with the most interest among graduate students was RDM workshops for graduate students (84.7%, n=531; X^2 (4, N=2047) = 107.79; $p < .001$). In comparison, the services that garnered the most interest among academic staff were assistance with preparing DMPs to meet funding requirements.

Figure 10

Interest in RDM services



Q25. “If data management plans were made part of grant applications from funding bodies such as SSHRC, CIHR, and NSERC, how interested would you be in the following services? Please rate your interest in each service. If the service does not apply to your situation, choose ‘not applicable’.” (n= 2658)

There were also statistically significant differences in interest in all RDM services across disciplines. Respondents in health sciences expressed higher levels of interest in a greater number of services than other disciplines with the most frequently selected option being information about funding requirements and journal requirements regarding research data (91.5%, n=97; X^2 (16, N=1967) = 43.10; $p < .001$). Respondents in medicine and preclinical sciences also reported high levels of interest in this service (80.2%, n=182). In contrast, respondents in science were least likely to express interest in RDM services compared to other disciplines. The service most frequently selected by science respondents (77.4%, n=287) was an institutional repository for long-term access and preservation of research data. Respondents in the remaining disciplines reported highest levels of interest in assistance with DMP preparation.

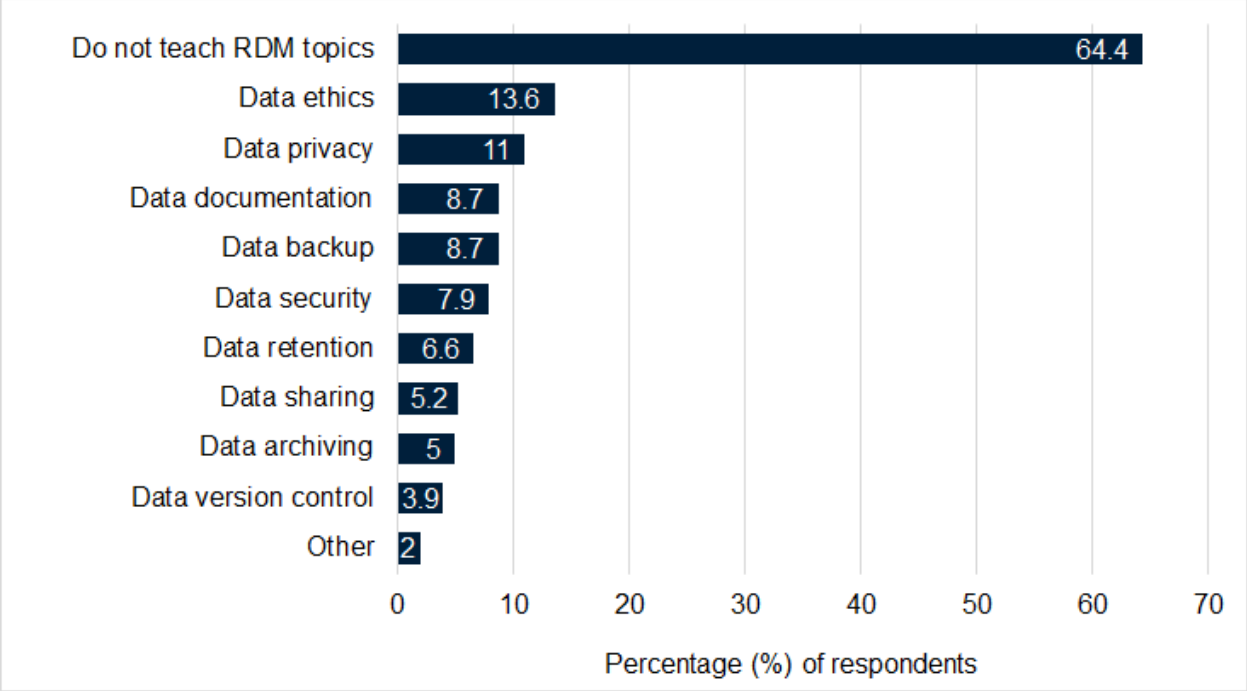
Teaching RDM

Respondents were asked which research data management topics they included in their teaching practice. As previously noted, this question was not asked in the early science and engineering survey, which hampered our ability to analyze and compare disciplinary differences in teaching RDM topics. The responses in Figure 11 are more representative of social sciences, arts and humanities, health sciences, and medicine disciplines.

Most respondents reported not teaching any RDM topics (64.4%; n=1002). Of those who did teach RDM topics, data ethics was highest reported (13.6%; n=211), followed by data privacy (11%; n=171), and data backup and data documentation (tied at 8.7%; n=135).

Figure 11

RDM topics taught in classes



Q23. “Do you include any of the following topics related to RDM in your teaching practice? Select all that apply. If you do not teach RDM topics, choose ‘I do not teach RDM topics’.” (n=1555)

Academic staff who covered RDM topics in their teaching practice indicated a higher level of interest in RDM workshops for graduate students (77.6%, n= 191) compared to those who do not teach RDM topics (57.6%, n= 389). These statistically significant results (X^2 (2, N=921) = 30.98, p <.000) could suggest that respondents who teach RDM topics see an inherent value in students learning these topics, or that further support is needed in educational efforts in RDM.

Discussion

Our survey study begins to address one of the key features Cox et al. (2016) attribute to the “wicked problem” of RDM: the lack of information about the scope of the problem. In particular, we focus on the lack of knowledge of researchers’ local RDM practices along with their perspective of the problem. Our findings identify researchers’ practices and reveal potential challenges they face in managing their research data, including the types of data, the number of research projects led in a year, and an approximation of how much data researchers are generating. These findings can inform the development of services and support to help researchers meet RDM policy requirements, such as drafting DMPs and preparing data to be deposited in a digital repository, with tailored approaches based on disciplinary differences in specific needs and priorities.

Our survey findings, which are consistent with previous studies (Tenopir et al., 2020; Whitmire et al., 2015), reveal that researchers’ RDM practices do not always correspond to accepted best practices and they may be putting their data at risk. Few researchers are taking advantage of institutional or departmental servers to store their research data. Instead, they are largely storing their data on personal computer and laptop hard drives. Additionally, graduate student use of institutional or departmental servers is particularly low, which could pose risks of data loss in thesis work or other research projects. The statistically significant difference between graduate students and academic staff in the use of shared drives and university or departmental servers may reflect differences in access to institutional infrastructure. Graduate students may not have access to shared drives and/or university or departmental servers. We suggest that these researchers would benefit from institutional services and infrastructure for appropriate data storage and retention strategies to mitigate data loss.

Although researchers reported they keep their data until it becomes inaccessible or lost, it is unclear whether they are taking active measures to prolong the data’s lifespan for

long-term storage and preservation, or simply ignoring it until it goes away. Our results suggest that they are not making deliberate decisions regarding what data to keep, for how long, and what data can or should be deleted, nor are they aware of the cost of retaining data long term. The exception is researchers in health sciences who reported an intentional decision to keep their data for a set period of five to ten years after project completion, which may be due to external requirements regarding the retention and destruction of confidential data. Writing a DMP would encourage more researchers to consider data retention periods and plan for potential reuse of their data after project completion. In addition to lack of planning for data preservation, our findings suggest that researchers are not producing adequate data documentation that will allow their data to be reused or their research findings to be reproduced. This is a concern because data deposit without sufficient data documentation does not fulfil the intent of open data. It is important to educate researchers on the importance of good data documentation and provide them with guidance on selecting appropriate data storage solutions along with data retention schedules and strategies.

One of the requirements of the Tri-Agency Research Data Management Policy is to deposit data in a digital repository, but very few researchers reported depositing their research data in an institutional, general, or discipline-specific repository, a finding consistent with previous studies (Johnson & Steeves, 2019; Joo & Peters, 2020; Majid et al., 2018). However, a larger percentage of researchers are interested in sharing their data in an institutional repository in the future. It is important to note that institutional data repositories were not as widely available at the time of survey deployment and there may have been some confusion over the differences between an institutional data repository (e.g. an institutional instance of Dataverse) and an institutional repository for open access papers. Even so, researchers expressed a high level of interest in their institutions providing an institutional repository for long-term access and preservation of research data. Given that most researchers do not have sufficient data documentation and do not currently make use of data repositories, our survey results reveal they may not be prepared to meet the data deposit requirement.

Although researchers expressed an openness to sharing their research data, there are instances when that may not be possible. As with previous studies (Joo & Peters, 2020; Majid et al., 2018; Tenopir et al., 2020), researchers reported a variety of restrictions, embargoes, or other reasons that prevent them from sharing their research data. However, it is encouraging that the majority of researchers see collaborative scholarship and open access to knowledge as benefits to sharing their research data. These findings suggest that there are opportunities to facilitate data deposit and sharing with proper planning, appropriate infrastructure, and education or training to mitigate the

risks and alleviate concerns of making research data widely available, which are consistent with Sieglitz et al.'s study published in 2020.

Our survey results indicate that more than half of researchers receive or intend to apply for funding from at least one of the Tri-Agencies, and they want help preparing DMPs to meet the RDM policy requirements. These results are unsurprising given that DMP requirements from funding agencies are often a "trigger of activity" that bring about researcher awareness of funders' RDM expectations (Cox et al., 2016, p.9; Mancilla et al., 2019). Our findings also show that a very low percentage of researchers would be able to draft a DMP without assistance or guided documentation. This is consistent with previous studies indicating that few researchers felt they could prepare a DMP that met all funder requirements (Akers & Doty, 2013; Nicholls et al., 2014; Kennan & Markauskaite, 2015; Majid et al., 2018; Johnson & Steeves, 2019; Tenopir et al., 2020). Krahe et al. (2020, p.111) noted that the vast majority of researchers consider having a DMP "good research practice,"but the most common reason they do not have one is a lack of understanding of "what a DMP is or what it should look like." Drafting a DMP represents a significant change in researcher practices and expectations of support as they face the upcoming requirement for including DMPs in funding applications. To create DMPs, researchers must be able to articulate how they work with and manage their data, and as previously discussed, our findings suggest they are not in the habit of doing that.

Similar to Akers & Doty's (2013) approach, we categorized our survey respondents into broad subjects based on their field of study to identify key differences in RDM practices. Our results reveal highly significant differences in the types and quantities of data generated and used across disciplines as well as data sharing practices. These differences should be considered when establishing services to facilitate the sharing and depositing of research data. Examining survey responses by discipline reveals typical RDM behaviours for different groups (Appendix A), allowing us to prioritize support for problem areas that each discipline is most concerned with and modify services and training to meet their respective needs.

When it comes to interest in RDM services, each discipline has different priorities (Appendix B). Preclinical sciences, medicine and the health sciences had the greatest number of researchers expressing interest in information about funding and journal requirements. Researchers in science prioritized an institutional repository for long-term access and preservation of research data, and researchers in the remaining disciplines were most interested in assistance with DMPs. Our findings were consistent with previous studies that reported researchers were interested in a variety of data services, including services that were not specific to RDM (Akers & Doty, 2013; Guindon, 2014;

Joo & Peters, 2020; Krahe et al., 2020; Majid et al., 2018). Researcher rank was also a determining factor in interest in RDM services. Graduate students were most interested in RDM workshops for graduate students, while academic staff were most interested in assistance with DMP preparation. These findings may reflect the different roles researchers have. Graduate students are more interested in learning about RDM in general. In contrast, academic staff are more likely to be responsible for preparing and submitting DMPs as part of grant applications. As RDM policies are developed and implemented, both researchers and service providers must continue to respond to changing guidelines.

Researchers are interested in increasing their knowledge of RDM through workshops on best practices. However, few of the researchers with teaching responsibilities include RDM topics in their teaching practice. Those who do include RDM topics in their teaching are more likely to be interested in a service offering RDM workshops for graduate students compared to those who do not. This paradoxical result—that those already teaching RDM topics themselves are more likely to want workshops offered to students—could demonstrate that these researchers see an inherent value in students learning about RDM, or may highlight a need for institutional teaching support in this area. As the importance of RDM becomes integral to funding requirements and expected research practices at institutions, there may be an increase in the inclusion of these concepts in teaching practices and an even greater interest in additional support for workshops on RDM for graduate students.

This study provides a snapshot of the Canadian RDM landscape prior to the release of the Tri-Agency Research Data Management Policy in 2021. The enactment of the policy creates an opportunity for follow-up studies to assess the impact that it has had in changing researcher RDM behaviour. Future research could include repeating the original survey to directly assess changes in researcher perceptions and practices over time. To gain further insights and standardize the survey instrument, questions related to RDM teaching and graduate student practices could be expanded.

Conclusion

Our survey findings indicate that many researchers need or want assistance with creating DMPs and want institutions to provide them with access to institutional data repositories to help them comply with the Tri-Agency Research Data Management Policy requirements. An obvious step is designing services that researchers have expressed interest in. Our findings also indicate that there are other areas where additional support and services are needed to foster best practices in RDM. Over time, as RDM practices continue to develop, services will continue to change.

Appropriate sharing of research data is a complex issue, and few researchers are prepared to navigate this complexity on their own. Additionally, researchers could use assistance with proper data management planning, storage strategies, and creating data documentation that makes their research data reusable and more widely accessible. It is encouraging that they expressed interest in all the potential RDM services suggested in our survey. It is important to note that each discipline is different and we should accommodate unique disciplinary RDM practices when designing services.

Researchers in Canada show an interest in support, education, and information on research data management practices that will allow them to comply with the Tri-Agency requirements. Canadian institutions can draw on our findings to build services that will provide the RDM support for effective data stewardship.

Acknowledgements

We would like to thank all members of the Canadian RDM Survey Consortium involved in the development of the survey. We are especially grateful to the institutional teams that contributed standardized data to the national dataset used in the analysis of this study.

Many thanks to the Portage Network, SSHRC, and CIHR for financial assistance to translate the survey.

Finally, thanks to Shahira Khair and Kathy Szigeti for their contributions to the analysis of the national dataset.

References

Abel, J., Cooper, A., Dearborn, D., Perry, C., Szwajcer, A., & Wang, M. (2020). Institutional research data management services capacity survey INSIGHTS report #1. RDM support within organizations: budget, structure, and strategies. Portage Network.

Akers, K. G., & Doty, J. (2013). Disciplinary differences in faculty research data management practices and perspectives. *International Journal of Digital Curation*, *8*(2), 5–26.

Argáez, D., & Shearer, K. (2008). Survey of Canadian and international data management initiative. CARL Data Management Working Group.

Carlson, J., Fosmire, M., Miller, C. C., & Nelson, M. S. (2011). Determining data information literacy needs: A study of students and research faculty. *Portal: Libraries and the Academy*, *11*(2), 629–657.

Cox, A. M., Pinfield, S., & Smith, J. (2016). Moving a brick building: UK libraries coping with research data management as a 'wicked' problem. *Journal of Librarianship and Information Science*, *48*(1), 3–17.

Diekema, A. R., Wesolek, A., & Walters, C. D. (2014). The NSF/NIH effect: Surveying the effect of data management requirements on faculty, sponsored programs, and institutional repositories. *The Journal of Academic Librarianship*, *40*(3–4), 322–331.

Goben, A., & Griffin, T. (2019). In aggregate: Trends, needs, and opportunities from research data management surveys. *College & Research Libraries*, *80*(7), 903.

Government of Canada. (2015, July 27). Tri-Agency Statement of Principles on Digital Data Management. Government of Canada.

Government of Canada. (2021). Tri-Agency Research Data Management Policy. Innovation, Science and Economic Development Canada.

Guindon, A. (2014). Research data management at Concordia University: A survey of current practices. *Feliciter*, *60*(2), 15–17.

Johnson, K., & Steeves, V. (2019). Research data management among life sciences faculty: Implications for library service. *Journal of EScience Librarianship*, *8*(1).

Johnston, L., & Jeffryes, J. (2014). Data management skills needed by structural engineering students: Case study at the University of Minnesota. *Journal of Professional Issues in Engineering Education and Practice*, *140*(2), 05013002.

Joo, S., & Peters, C. (2020). User needs assessment for research data services in a research university. *Journal of Librarianship and Information Science, 52*(3), 633–646.

Kennan, M. A., & Markauskaite, L. (2015). Research data management practices: A

snapshot in time. *International Journal of Digital Curation*, *10*(2), 69–95.

Kim, Y., & Stanton, J. M. (2015). Institutional and individual factors affecting scientists' data-sharing behaviors: A multilevel analysis. *Journal of the Association for Information Science and Technology*, *67*(4), 776–799.

Krahe, M. A., Toohey, J., Wolski, M., Scuffham, P. A., & Reilly, S. (2020). Research data management in practice: Results from a cross-sectional survey of health and medical researchers from an academic institution in Australia. *Health Information Management Journal, 49*(2–3), 108–116.

Majid S., Foo S., & Zhang X. (2018). Research data management by academics and researchers: Perceptions, knowledge and practices. In: Dobreva M., Hinze A., Žumer M. (eds). Maturity and Innovation in Digital Libraries. ICADL 2018. *Lecture Notes in Computer Science*, vol 11279. Springer.

Mancilla, H. A., Teperek, M., Dijck, J. van, Heijer, K. den, Eggermont, R., Plomp, E., Velden, Y. T. der, & Kurapati, S. (2019). On a quest for cultural change—Surveying research data management practices at Delft University of Technology. *LIBER Quarterly*, *29*(1), 1–27.

Nicholls, N. H., Samuel, S. M., Lalwani, L. N., Grochowski, P. F., & Green, J. A. (2014). Resources to support faculty writing data management plans: Lessons learned from an engineering pilot. *International Journal of Digital Curation*, *9*(1), 242–252.

Parsons, T., Grimshaw, S., & Williamson, L. (2013). Research data management survey. University of Nottingham.

Perrier, L., & Barnes, L. (2018). Developing research data management services and support for researchers: A mixed methods study. *Partnership: The Canadian Journal of Library and Information Practice and Research*, *13*(1).

Schumacher, J., & VandeCreek, D. (2015). Intellectual capital at risk: Data management practices and data loss by faculty members at five American universities. *International Journal of Digital Curation*, *10*(2), 96–109.

Tenopir, C., Dalton, E. D., Allard, S., Frame, M., Pjesivac, I., Birch, B., Pollock, D., & Dorsett, K. (2015). Changes in data sharing and data reuse practices and perceptions among scientists worldwide. *PLoS One*, *10*(8), e0134826.

Tenopir, C., Rice, N. M., Allard, S., Baird, L., Borycz, J., Christian, L., Grant, B., Olendorf, R., & Sandusky, R. J. (2020). Data sharing, management, use, and reuse: Practices and perceptions of scientists worldwide. *PLoS One*, *15*(3), e0229003.

Vice Chancellor for Research's Data Management Task Force. (2012). Research data management at the University of Colorado Boulder: Recommendations in support of fostering 21st century research excellence. University of Colorado Boulder.

Whitmire, A. L., Boock, M., & Sutton, S. C. (2015). Variability in academic research data management practices. *Program*, *49*(4), 382–407.

Wiley, C., & Burnette, M. (2019). Assessing data management support needs of bioengineering and biomedical research faculty. *Journal of EScience Librarianship*, *8*(1), e1132.

Appendix A

Legend:
Eng – Engineering
Sci – Science
A/H – Arts/Humanities
B/M/L/E – Business/Management, Law, Education
SS – Social Sciences
M/PS –Medicine/Preclinical Sciences
HS – Health Science
IO – Interdisciplinary/Other

Number of research projects led in past year by department

Question: How many research projects did you lead in the past year? Select one.

Variable: NUMPROJ

	Eng		Sci		A/H		B/M/L/E		SS		M/PS		HS		I/O	
1-2 projects	100	38.2%	102	32.4%	150	51.7%	78	43.6%	161	52.3%	109	44.9%	42	43.3%	13	28.9%
3-5 projects	68	26.0%	96	30.5%	52	17.9%	64	35.8%	71	23.1%	69	28.4%	38	39.2%	20	44.4%
>5 projects	43	16.4%	83	26.3%	20	6.9%	11	6.1%	37	12.0%	29	11.9%	8	8.2%	2	4.4%
Not sure	5	1.9%	5	1.6%	16	5.5%	3	1.7%	5	1.6%	5	2.1%	1	1.0%	1	2.2%
Not applicable	46	17.6%	29	9.2%	52	17.9%	23	12.8%	34	11.0%	31	12.8%	8	8.2%	9	20.0%

Estimated data storage for average project

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

Variable: SIZAVGALL

	Eng		Sci		A/H		B/M/L/E		SS		M/PS		HS		I/O	
< 50GB	139	53.5%	154	48.1%	161	56.3%	105	59.7%	192	62.5%	132	54.3%	49	52.1%	28	63.6%
50GB to < 500GB	55	21.2%	59	18.4%	26	9.1%	7	4.0%	24	7.8%	14	5.8%	8	8.5%	5	11.4%
500GB to < 1000GB	32	12.3%	36	11.3%	9	3.1%	8	4.5%	14	4.6%	11	4.5%	1	1.1%	4	9.1%
>1TB	21	8.1%	53	16.6%	15	5.2%	9	5.1%	13	4.2%	22	9.1%	4	4.3%	1	2.3%
Not sure	11	4.2%	11	3.4%	63	22.0%	36	20.5%	51	16.6%	49	20.2%	19	20.2%	3	6.8%
Not applicable	2	0.8%	7	2.2%	12	4.2%	11	6.3%	13	4.2%	15	6.2%	13	13.8%	3	6.8%

Derived variable based on SIZAVES, SIZAVGHUSSHS, and SIZAVGALL

Type of research data

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.

Variables: TYPEGEO, TYPEINS, TYPEMOD, TYPEMUL, TYPENUM, TYPESOF, TYPETEXT, TYPEOTH

	Eng		Sci		A/H		B/M/L/E		SS		M/PS		HS		I/O	
Geospatial	35	11.7%	58	15.2%	15	3.9%	2	0.8%	56	12.3%	13	4.6%	1	0.8%	6	8.7%
Instrument specific	86	28.8%	123	32.3%	12	3.1%	5	2.1%	31	6.8%	53	18.9%	20	16.4%	11	15.9%
Models	120	40.1%	128	33.6%	18	4.7%	31	13.1%	92	20.2%	41	14.6%	9	7.4%	16	23.2%
Multimedia	119	39.8%	157	41.2%	187	48.8%	71	30.0%	155	34.1%	100	35.6%	47	38.5%	28	40.6%
Numerical	192	64.2%	247	64.8%	78	20.4%	107	45.1%	246	54.1%	159	56.6%	72	59.0%	30	43.5%
Software	122	40.8%	141	37.0%	30	7.8%	29	12.2%	81	17.8%	40	14.2%	16	13.1%	15	21.7%
Text	181	60.5%	248	65.1%	311	81.2%	192	81.0%	317	69.7%	169	60.1%	91	74.6%	53	76.8%
Discipline specific	12	15.0%	19	11.5%	4	3.1%	2	2.8%	2	2.0%	37	13.9%	2	2.5%	2	6.5%
Other	28	9.8%	53	14.7%	24	7.0%	8	3.8%	26	6.1%	22	9.0%	10	10.4%	4	8.2%

Data storage media

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

Variables: STORUSB, STORCD, STORCOM, STORLAP, STOREHD, STORHDI, STORDSE, STORWEB, STOREXT, STORHPC, STORPHY, STORNSU, STOROTH

	Eng		Sci		A/H		B/M/L/E		SS		M/PS		HS		I/O	
Flash drive/USB	138	46.2%	148	38.8%	215	56.1%	112	47.3%	214	47.0%	101	35.9%	49	40.2%	36	52.2%
CD/DVD	24	8.0%	27	7.1%	37	9.7%	9	3.8%	37	8.1%	20	7.1%	3	2.5%	7	10.1%
Computer hard drive	197	70.6%	197	64.2%	182	47.5%	105	44.3%	267	58.7%	149	53.0%	55	45.1%	38	55.1%
Laptop hard drive	197	65.9%	229	60.1%	255	66.6%	143	60.3%	259	56.9%	123	43.8%	66	54.1%	37	53.6%
External hard drive	174	58.2%	235	61.7%	224	58.5%	104	43.9%	234	51.4%	101	35.9%	49	40.2%	37	53.6%
Hard drive of the instrument/sensor which generates the data	74	24.7%	101	26.5%	9	2.3%	3	1.3%	32	7.0%	31	11.0%	11	9.0%	6	8.7%
Shared drive/university or departmental server	90	30.1%	145	38.1%	83	21.7%	58	24.5%	116	25.5%	85	30.2%	52	42.6%	23	33.3%
Cloud/web-based solution	139	46.5%	157	41.2%	186	48.6%	109	46.0%	203	44.6%	78	27.8%	41	33.6%	25	36.2%
External data repository	26	8.7%	65	17.1%	17	4.4%	6	2.5%	26	5.7%	18	6.4%	2	1.6%	9	13.0%
Grid/high performance computing centre	12	4.0%	55	14.4%	2	0.5%	2	0.8%	7	1.5%	10	3.6%	0	0.0%	5	7.2%

Physical copy retained	36	12.0%	76	19.9%	134	35.0%	53	22.4%	128	28.1%	70	24.9%	51	41.8%	18	26.1%
Not sure	2	0.7%	1	0.3%	3	0.8%	3	1.3%	2	0.4%	10	3.6%	2	1.6%	0	0.0%
Other	13	4.3%	19	5.0%	17	4.4%	14	5.9%	20	4.4%	13	4.6%	8	6.6%	3	4.3%

Sufficient data documentation to understand the data

Survey question: Is there sufficient documentation and description (for example, variable and field definitions, codebooks, data dictionaries, metadata, scripts to run) for another person outside your research team to understand and use the research data?

Variables: DOCUND

	Eng		Sci		A/H		B/M/L/E		SS		M/PS		HS		I/O	
Yes	106	35.8%	135	35.5%	142	40.3%	95	43.6%	193	45.1%	113	45.4%	44	37.6%	29	42.6%
No	99	33.4%	124	32.6%	79	22.4%	45	20.6%	93	21.7%	47	18.9%	28	23.9%	16	23.5%
Not sure	91	30.7%	121	31.8%	131	37.2%	78	35.8%	142	33.2%	89	35.7%	45	38.5%	23	33.8%

Sufficient data documentation – to replicate methodologies

Survey question: Is there sufficient documentation and description (for example, variable and field definitions, codebooks, data dictionaries, metadata, scripts to run) retained in the same file, folder or document for another person outside your research team to replicate the methodologies that produced the data?

Variables: DOCREP

	Eng		Sci		A/H		B/M/L/E		SS		M/PS		HS		I/O	
Yes	110	42.0%	134	41.4%	106	37.1%	82	46.6%	137	45.2%	114	47.7%	48	50.5%	22	48.9%
No	81	30.9%	100	30.9%	80	28.0%	46	26.1%	77	25.4%	48	20.1%	24	25.3%	11	24.4%
Not sure	71	27.1%	90	27.8%	100	35.0%	48	27.3%	89	29.4%	77	32.2%	23	24.2%	12	26.7%

Length of time keep data – source/survey/raw data

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.

Variable names: KEEPSMR

	Eng		Sci		A/H		B/M/L/E		SS		M/PS		HS		I/O	
I only keep data for the length of the project	16	6.1%	9	2.8%	12	4.2%	15	8.7%	15	5.0%	22	9.2%	4	4.2%	0	0
Less than 3 years	24	9.2%	21	6.5%	10	3.5%	9	5.2%	17	5.7%	8	3.4%	2	2.1%	3	6.7%
Between 3-5 years	51	19.5%	46	14.2%	25	8.8%	29	16.8%	32	10.7%	24	10.1%	27	28.1%	7	15.6%
Between 5-10 years	34	13.0%	39	12.0%	55	19.4%	53	30.6%	69	23.0%	57	23.9%	48	50.0%	10	22.2%

More than 10 years	29	11.1%	41	12.7%	39	13.7%	18	10.4%	45	15.0%	57	23.9%	6	6.3%	5	11.1%
Until the data becomes inaccessible or lost	108	41.2%	168	51.9%	143	50.4%	49	28.3%	122	40.7%	70	29.4%	9	9.4%	20	44.4%

Length of time keep data – intermediate/working data

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.

Variable: KEEPINT

	Eng		Sci		A/H		B/M/L/E		SS		M/PS		HS		I/O	
I only keep data for the length of the project	31	11.8%	33	10.3%	16	5.7%	20	11.8%	21	7.0%	32	13.5%	8	8.4%	3	6.7%
Less than 3 years	34	13.0%	39	12.2%	17	6.1%	16	9.4%	21	7.0%	16	6.8%	3	3.2%	6	13.3%
Between 3-5 years	49	18.7%	43	13.4%	32	11.4%	27	15.9%	39	13.1%	31	13.1%	25	26.3%	12	26.7%
Between 5-10 years	31	11.8%	43	13.4%	60	21.4%	47	27.6%	70	23.5%	55	23.2%	49	51.6%	5	11.1%
More than 10 years	22	8.4%	33	10.3%	30	10.7%	14	8.2%	40	13.4%	40	16.9%	4	4.2%	4	8.9%
Until the data becomes inaccessible or lost	95	36.3%	129	40.3%	125	44.6%	46	27.1%	107	35.9%	63	26.6%	6	6.3%	15	33.3%

Length of time keep data – processed data ready for publication

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.

Variable: KEEPDAT

	Eng		Sci		A/H		B/M/L/E		SS		M/PS		HS		I/O	
I only keep data for the length of the project	12	4.6%	5	1.5%	6	2.1%	11	6.5%	7	2.4%	19	8.0%	3	3.2%	0	0.0%
Less than 3 years	17	6.5%	19	5.9%	7	2.5%	8	4.7%	8	2.7%	13	5.5%	2	2.1%	2	4.4%
Between 3-5 years	36	13.7%	31	9.6%	26	9.3%	22	12.9%	33	11.1%	20	8.4%	22	23.2%	6	13.3%
Between 5-10 years	39	14.9%	38	11.8%	46	16.4%	49	28.8%	68	23.0%	51	21.5%	44	46.3%	9	20.0%
More than 10 years	34	13.0%	54	16.7%	40	14.3%	21	12.4%	52	17.6%	50	21.1%	8	8.4%	7	15.6%
Until the data becomes inaccessible or lost	124	47.3%	176	54.5%	155	55.4%	59	34.7%	128	43.2%	84	35.4%	16	16.8%	21	46.7%

Methods of current data sharing

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.’

Variables: SHARENOT, SHAREREQ, SHAREONR, SHAREWEB, SHAREINR, SHAREJOU, SHAREDSP, SHAREOTH

	Eng		Sci		A/H		B/M/L/E		SS		M/PS		HS		I/O	
Not sharing	104	34.8%	67	17.6%	120	31.3%	78	32.9%	146	32.1%	92	32.7%	52	42.6%	20	29.0%
Personal request only	162	54.2%	219	57.5%	162	42.3%	89	37.6%	208	45.7%	106	37.7%	38	31.1%	35	50.7%
Online with restricted access	49	16.4%	68	17.8%	55	14.4%	42	17.7%	61	13.4%	35	12.5%	15	12.3%	7	10.1%
Institutional/personal website	35	11.7%	74	19.4%	51	13.3%	24	10.1%	50	11.0%	17	6.0%	7	5.7%	6	8.7%
Institutional repository	3	2.6%	33	14.9%	13	3.5%	7	3.1%	18	4.3%	8	2.9%	9	8.0%	2	3.6%
Supplementary materials to a journal publisher	38	12.7%	128	33.6%	23	6.0%	21	8.9%	53	11.6%	48	17.1%	5	4.1%	10	14.5%
Discipline-specific repository	21	7.0%	103	27.0%	10	2.6%	0		26	5.7%	23	8.2%	3	2.5%	10	14.5%
Other	1	1.2%	7	5.7%	11	3.7%	4	2.4%	13	3.9%	7	3.7%	7	11.1%	2	6.3%

Restrictions or embargoes

Survey question: Some research data cannot be shared because of legal or privacy restrictions or embargoes. Which of the following restrictions or embargoes may limit your ability to share your data with others? Select all that apply. If there are no restrictions or embargoes, choose 'there are no restrictions or embargoes on sharing my data with other parties.'

Variables: RESNOT, RESPUB, RESINP, RESPAT, RESCOM, RESTHP, RESPRI, RESSAF, RESUNS, RESOTH

	Eng		Sci		A/H		B/M/L/E		SS		M/PS		HS		I/O	
No restrictions	60	22.1%	133	37.4%	111	29.5%	52	22.9%	98	22.1%	34	12.4%	20	17.9%	11	19.0%
Need to publish before sharing	148	50.2%	160	43.4%	99	26.3%	48	21.1%	108	24.3%	95	34.5%	32	28.6%	26	38.8%
Intellectual property rights	72	24.4%	61	16.5%	37	9.8%	23	10.1%	36	8.1%	35	12.7%	5	4.5%	9	13.4%
Plan to file patent	45	15.3%	29	7.9%	1	0.3%	2	0.9%	1	0.2%	24	8.7%	1	0.9%	7	10.4%
Commercial concerns	32	10.8%	16	4.3%	9	2.4%	13	5.7%	10	2.3%	9	3.3%	5	4.5%	4	6.0%
Contractual obligations to third party	66	22.4%	49	13.3%	23	6.1%	23	10.1%	33	7.4%	22	8.0%	3	2.7%	12	17.9%
Privacy or ethical restrictions	57	19.3%	72	19.5%	92	24.5%	98	43.2%	188	42.3%	121	44.0%	55	49.1%	29	43.3%
Public safety/sensitive nature	11	3.7%	6	1.6%	7	1.9%	9	4.0%	23	5.2%	9	3.3%	6	5.4%	3	4.5%
Not sure	40	13.6%	28	7.6%	50	13.3%	17	7.5%	40	9.0%	31	11.3%	15	13.4%	6	9.0%
Other	12	4.1%	20	5.4%	27	7.2%	16	7.0%	27	6.1%	10	3.6%	8	7.1%	6	9.0%

Reasons for not sharing

Survey question: What, if any, are the reasons you would not be willing to share your research data and associated methods/tools/algorithms? Select all that apply. If you are willing to share, choose 'I am willing to share them.'

Variables: NOSWIL, NOSINC, NOSVAL, NOSIDK, NOSNOR, NOSFUN, NOSSBS, NOSDKC, NOSTIM, NOSSTD, NOSLFD, NOSPLC, NOSNOU, NOSPRI, NOSACK, NOSOTH

	Eng		Sci		A/H		B/M/L/E		SS		M/PS		HS		I/O	
Willing to share	67	24.0%	97	31.6%	94	24.5%	58	24.5%	110	24.3%	64	22.8%	28	23.0%	21	30.4%
Data are incomplete	134	44.8%	178	48.2%	147	38.4%	54	22.8%	145	32.1%	77	27.4%	37	30.3%	21	30.4%
Want to derive value	93	33.7%	121	32.9%	98	25.6%	60	25.3%	116	25.7%	59	21.0%	32	26.2%	14	23.3%
Lack technical skills	17	5.7%	17	4.5%	32	8.4%	10	4.2%	22	4.9%	22	7.8%	7	5.7%	0	0.0%
Do not hold rights	64	21.4%	67	17.6%	59	15.4%	48	20.3%	69	15.3%	41	14.6%	17	13.9%	13	18.8%
Not required by funding body	24	8.0%	19	5.0%	12	3.1%	9	3.8%	20	4.4%	8	2.8%	3	2.5%	2	2.9%
Should not be shared	18	6.0%	18	4.7%	23	6.0%	28	11.8%	34	7.5%	11	3.9%	11	9.0%	5	7.2%
Did not know could share	5	1.7%	12	3.1%	18	4.7%	6	2.5%	12	2.7%	9	3.2%	9	7.4%	2	2.9%
Insufficient time	85	28.4%	93	24.4%	68	17.8%	33	13.9%	57	12.6%	31	11.0%	15	12.3%	17	24.6%
Lack of standards	74	25.1%	80	21.7%	40	10.6%	23	10.1%	45	10.2%	20	7.3%	18	16.1%	10	14.9%
Lack of Funding	39	13.0%	53	13.9%	56	14.6%	22	9.3%	37	8.2%	30	10.7%	20	16.4%	9	13.0%
No place to put data	37	12.4%	44	11.5%	37	9.7%	18	7.6%	25	5.5%	24	8.5%	11	9.0%	7	10.1%
Not useful to others	23	8.3%	34	9.2%	33	8.6%	14	5.9%	25	5.5%	7	2.5%	9	7.4%	1	1.7%
Privacy/legal/security reasons	71	23.7%	61	16.0%	67	17.5%	77	32.5%	120	26.5%	78	27.8%	39	32.0%	22	31.9%

Concerns about citation	30	24.8%	78	33.3%	77	20.5%	49	20.7%	72	16.8%	61	21.7%	38	31.1%	14	24.6%
Other	15	5.4%	25	6.8%	25	6.5%	21	8.9%	45	10.0%	18	6.4%	13	10.7%	3	5.0%

With whom researchers are willing to share

Survey question: If research data were not affected by restrictions or embargoes, with whom would you be willing to share.

Variables: WHONON, WHOCOL, WHODEP, WHOINS, WHOFLD, WHOOUT, WHOALL

	Eng		**Sci**		**A/H**		**B/M/L/E**		**SS**		**M/PS**		**HS**		**I/O**	
Nobody	12	4.6%	1	0.3%	12	3.9%	9	4.7%	15	4.7%	13	4.8%	4	4.0%	3	6.7%
Immediate Collaborators	155	59.2%	175	54.2%	116	37.4%	92	48.4%	170	52.8%	135	49.5%	68	67.3%	22	48.9%
Same department	97	37.0%	98	30.3%	81	26.1%	65	34.2%	117	36.3%	90	33.0%	29	28.7%	14	31.1%
Same institution	81	30.9%	78	24.1%	74	23.9%	56	29.5%	92	28.6%	74	27.1%	21	20.8%	10	22.2%
Same field	128	48.9%	147	45.5%	146	47.1%	85	44.7%	157	48.8%	110	40.3%	42	41.6%	19	42.2%
Researchers outside field	36	13.7%	62	19.2%	62	20.0%	34	17.9%	72	22.4%	42	15.4%	12	11.9%	6	13.3%
Public	79	30.2%	141	43.7%	143	46.1%	56	29.5%	88	27.3%	62	22.7%	18	17.8%	16	35.6%

Ability to draft DMP

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.

Variable: DMP

	I would be able to draft a data management plan that would address these types of questions without assistance		I would be able to draft a data management plan but would prefer to have assistance and/or documentation to ensure the success of the application		I would need assistance and / or guided documentation to appropriately address some or all of the sections	
Eng	41	14.2%	103	35.8%	144	50.0%
Sci	62	16.7%	149	40.2%	160	43.1%
A/H	34	10.4%	89	27.2%	204	62.4%
B/M/L/E	23	11.5%	67	33.5%	110	55.0%
SS	71	17.8%	150	37.7%	177	44.5%
M/PS	24	10.5%	82	35.8%	123	53.7%
HS	8	7.4%	46	42.6%	54	50.0%
I/O	10	16.1%	24	38.7%	28	45.2%

Appendix B: Interest in potential RDM services by discipline

Question: If data management plans were made part of grant applications from funding bodies such as SSHRC, CIHR, and NSERC, how interested would you be in the following services? Please rate your interest in each service. If the service does not apply to your situation, choose 'not applicable.'

Engineering

Service	Level of interest	
Assistance with DMP preparation	236	82.52%
Communication/info about funding/journal requirements	234	82.39%
Institutional repository	230	80.70%
Personalized consultations	220	76.66%
Graduate student workshop	216	75.79%
Data storage during active projects	217	75.61%
Assistance with preservation/sharing	214	74.56%
Finding/accessing data sources	210	73.68%
Faculty workshop	193	67.48%
Permanent identifiers/DOIs	193	67.48%
External repository	186	64.81%
Assistance with metadata curation	185	64.46%
Digitization of physical records	154	53.47%

Sciences

Service	Level of interest	
Institutional repository	287	77.36%
Communication/info about funding/journal requirements	286	76.88%
Assistance with DMP preparation	280	75.07%
Data storage during active projects	271	72.27%
Graduate student workshop	251	67.65%
Personalized consultations	246	66.13%
Assistance with preservation/sharing	240	64.52%
Faculty workshop	237	63.37%
Permanent identifiers/DOIs	235	63.34%
Finding/accessing data sources	222	60.33%
Assistance with metadata curation	221	59.57%
External repository	220	59.30%
Digitization of physical records	180	48.52%

Arts and Humanities

Service	Level of interest	
Assistance with DMP preparation	258	79.88%
Communication/info about funding/journal requirements	253	78.33%
Data storage during active projects	245	76.09%
Assistance with preservation/sharing	233	73.04%
Personalized consultations	233	72.81%
Institutional repository	231	72.19%
External repository	229	71.56%
Finding/accessing data sources	228	71.25%
Assistance with metadata curation	216	68.14%
Faculty workshop	209	64.31%
Graduate student workshop	201	63.81%
Digitization of physical records	202	63.32%
Permanent identifiers/DOIs	189	59.06%

Business/Management, Law, and Education

Service	Level of interest	
Assistance with DMP preparation	162	84.38%
Personalized consultations	151	78.65%
Assistance with preservation/sharing	149	77.20%
Communication/info about funding/journal requirements	147	76.96%
Institutional repository	144	75.00%
Finding/accessing data sources	142	74.74%
Faculty workshop	139	71.65%
Assistance with metadata curation	133	70.00%
Graduate student workshop	132	69.84%
Data storage during active projects	132	69.84%
External repository	119	62.30%
Digitization of physical records	115	60.53%
Permanent identifiers/DOIs	116	60.42%

Social Sciences

Service	Level of interest	
Assistance with DMP preparation	312	80.83%
Communication/info about funding/journal requirements	301	78.18%
Institutional repository	289	74.29%
Personalized consultations	290	73.98%
Assistance with preservation/sharing	282	73.06%
Graduate student workshop	285	72.52%
Finding/accessing data sources	267	70.82%
Data storage during active projects	272	70.47%
External repository	252	65.63%
Assistance with metadata curation	250	64.43%
Faculty workshop	249	63.85%
Permanent identifiers/DOIs	227	59.74%
Digitization of physical records	205	54.23%

Medicine and Preclinical Sciences

Service	Level of interest	
Communication/info about funding/journal requirements	182	80.18%
Assistance with DMP preparation	176	77.19%
Personalized consultations	169	75.11%
Assistance with preservation/sharing	162	72.65%
Institutional repository	158	71.49%
Data storage during active projects	158	70.54%
Assistance with metadata curation	151	68.02%
Finding/accessing data sources	152	67.86%
Faculty workshop	149	65.64%
External repository	144	64.57%
Graduate student workshop	137	61.16%
Permanent identifiers/DOIs	134	60.63%
Digitization of physical records	120	53.81%

Health Sciences

Service	Level of interest	
Communication/info about funding/journal requirements	97	91.51%
Assistance with DMP preparation	95	87.96%
Institutional repository	93	87.74%
Assistance with preservation/sharing	89	85.58%
Finding/accessing data sources	89	83.96%
Personalized consultations	88	83.02%
Graduate student workshop	88	82.24%
Data storage during active projects	86	80.37%
Faculty workshop	84	80.00%
Assistance with metadata curation	79	75.24%
External repository	76	71.70%
Permanent identifiers/DOIs	75	70.75%
Digitization of physical records	71	66.98%

Interdisciplinary/Other

Service	Level of interest	
Institutional repository	54	84.38%
Personalized consultations	53	82.81%
Assistance with DMP preparation	51	80.95%
Communication/info about funding/journal requirements	51	79.69%
Graduate student workshop	48	76.19%
Finding/accessing data sources	48	76.19%
Data storage during active projects	47	75.81%
Digitization of physical records	47	74.60%
Assistance with preservation/sharing	47	74.60%
External repository	46	73.02%
Faculty workshop	44	68.75%
Assistance with metadata curation	40	63.49%
Permanent identifiers/DOIs	38	60.32%

Appendix C: Engineering and Science RDM survey instrument – generic version

The French version of this survey instrument is available in the Generic Resources for Canadian RDM Survey Dataverse.

In this survey

- a **'researcher'** can include any individual who worked on the research project that could legitimately claim academic authorship of the research project if the results of the research project were to be published in a scholarly work
- a **'research project'** can be defined as the research associated with investigating a hypothesis or group of hypotheses (and applicable set of predictions) aimed at answering a distinct or specific research question. A single research grant may support one research project or multiple research projects. In the context of this survey, a research project is associated with a distinct set of research data and is considered to be a subset of a research program, research activity or research area you may investigate
- **'research data'** can be defined as any data that are collected, observed, created or analyzed to produce research results. Research data could include:
    - Observational data such as sensor readings, telemetry, survey results, images;
    - Experimental data such as gene sequences, chromatograms, magnetic field;
    - Simulation data such as climate models, economic models, Monte Carlo calculations;
    - Derived or compiled data such as text and data mining, compiled database, 3D models which can come in many forms including text, numerical, multimedia, models, software, discipline specific, or instrument specific

Section 1: Working with Research Data

In this section you will be asked questions about your research data, including how you work with them, document them and store them.

Q1. How many research projects did you lead in the past year, for example, as a Principal Investigator or project lead? Select one:

- 1-2 research projects
- 3-5 research projects
- >5 research projects
- Not sure
- Not applicable

Q2. How much data storage do you estimate you use in an average research project? Select one:

- <50GB (Gigabyte)
- 50GB to <500GB
- 500GB to <1000GB
- 1TB to <4TG (Terabyte)
- 4TB to 500TB
- >500TB
- Not sure
- Not Applicable, please explain

Q3. Which of the following best describes the type of research data you generate or use in a typical research project? Select all that apply:

- Geospatial – (e.g. raster, vector, grid)
- Instrument specific – (e.g. Olympus Confocal Microscope Data Format, FLIR Infrared Camera (SEQ))
- Models – (e.g. 3D, statistical, similitude, macroeconomic, causal)
- Multimedia – (e.g. JPEG, TIFF, MPEG, Quicktime, Bitmap)
- Numerical – (e.g. CSV, MAT, XLS, SPSS)
- Software – (e.g. Java, C, Perl, Python, Ruby, PHP)
- Text – (e.g. TXT, DOC, PDF, RTF, HTML, XML)
- Other – (e.g. discipline specific such as CIF, FITS, DICOM) please specify

Q4. Please list any software used for analysis or manipulation of your research data, if applicable:

Q5. Please indicate where you store research data from your current project(s). Select all that apply:

- Flash drive/USB
- CD/DVD
- Computer hard drive (i.e. local hard drive)

- Laptop hard drive
- External hard drive
- Hard drive of the instrument/sensor which generates the data
- Shared drive/university or departmental server
- Cloud/web based solution (e.g. Dropbox, Google Drive, Amazon Cloud, Microsoft Cloud)
- External data repository (e.g. Protein Data Bank, Cambridge Structural Database, GitHub, Dryad, Figshare)
- Grid/high performance computing (HPC) centre
- Physical copy retained (in boxes, cabinets, etc.)
- Not sure
- Other, please specify

Q6. 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?

- Yes
- No
- Not sure

Q7. 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?

- Yes
- No
- Not sure

Q8. 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 or patent approval, for example.

	I only keep data for the length of the project	Less than 3 years	Between 3-5 years	Between 5-10 years	More than 10 years	Until the data becomes inaccessible or lost
Source material/ Raw data.						
Intermediate/ working data.						
Processed data ready for publication. Processed data						

may include supporting information such as spectra, synthesis methods.						

Section 2. Data Sharing

In this section you will be asked about your current practices and opinions on sharing your research data.

Q9. Which methods of sharing your research data do you currently use? Check all that apply. If you do not currently share your data, choose 'not currently sharing.'

- Not currently sharing
- Share by personal request only
- Share online with restricted access
- Upload online to an institutional or personal website
- Include as part of supplementary material files to a journal publisher
- Deposit in a general or discipline-specific repository, such as Protein Data Bank, Cambridge Structural Database, GitHub, Dryad, Figshare,

Q10. If you selected 'deposit in a general or discipline-specific repository, such as Protein Data Bank, Cambridge Structural Database, GitHub, Dryad, Figshare,' please specify:

Q11. Hypothetically speaking, which methods of sharing your research data would you consider using in the future? Check 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 only
- Share online with restricted access
- Upload online to an institutional or personal website
- Include as part of supplementary material files to a journal publisher
- Deposit in a general or discipline-specific repository, such as Protein Data Bank, Cambridge Structural Database, GitHub, Dryad, Figshare

Q12. If you selected 'deposit in a general or discipline-specific repository, such as Protein Data Bank, Cambridge Structural Database, GitHub, Dryad, Figshare,' please specify. Also, if you plan to use another data repository other than the ones listed above, please specify:

Q13. Some research data cannot be shared because of legal or privacy restrictions or embargoes. Which of the following restrictions or embargoes may limit your ability to share your data with others? Select all that apply. If there are no restrictions or

embargoes, choose ‘there are no restrictions or embargoes on sharing my data with other parties.’

- There are no restrictions or embargoes on sharing my data with other parties
- I need to publish my data before I can share them
- Sharing my data may jeopardize Intellectual Property rights
- I plan to file for a patent

- My data are commercially vulnerable
- I have a contractual obligation with a third party
- My data are subject to privacy or confidentiality restrictions (e.g. patient data)
- My data are a matter of public safety or of a sensitive nature
- I’m unsure if I am allowed to share my data
- Other, please specify

Q14. If your research data were not affected by such restrictions or embargoes, with whom would you be willing to share them? Select all that apply:

- Nobody
- My immediate collaborators
- Researchers in my department/institute/centre
- Researchers at [HOME UNIVERSITY]
- Researchers in my field
- Researchers outside my field
- Anybody, including the general public

Q15. What, if any, are the reasons you would not be willing to share your research data and associated methods/algorithms? Select all that apply. If you are willing to share, choose ‘I am willing to share them.’

- I am willing to share them
- They are incomplete or not finished
- I still wish to derive value from them
- I do not have the technical skills or knowledge
- I do not hold the rights to share them
- Funding body does not require sharing
- I believe they should not be shared
- I did not know I could share them
- Insufficient time
- Lack of standards to make them usable by others
- Lack of funding
- No place to put them
- They are not useful to others
- There are privacy, legal or security issues

- I'm concerned my data could be used without proper citation or acknowledgement
- Other, please specify

Q16. 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.'

- I see no benefits to sharing my data
- Data availability provides safeguards against misconduct, data fabrication and falsification
- Data sharing and/or replication studies help in the training of next generation researchers
- Data sharing enables my data to be cited and increases my research impact
- Data sharing encourages collaborative science
- Data sharing encourages interdisciplinary research
- Data sharing moves my field of research forward
- Data sharing reduces redundant data collection
- Data sharing supports open access to knowledge
- Re-analysis of data helps verify results
- Well-maintained data helps retain data integrity
- Other, please specify

Q17. Are you aware of any discipline-specific research 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.'

Section 3. Funding Mandates and Research Data Management Services

In this section you will be asked about your awareness of funding mandates for research data management, as well as your interest in potential research data management services.

Q18. 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 that would address these types of questions without assistance
- I would be able to draft a data management plan that would address these types of questions, but would prefer to have assistance and/or guided documentation to ensure the success of my application
- I would need assistance and/or guided documentation to appropriately address some or all of the sections

Q19. If data management plans were made part of grant applications from funding bodies such as NSERC, SSHRC, and CIHR, how interested would you be in the following services? Please rate your interest in each service. If the service does not apply to your situation, choose 'not applicable.'

	Very interested	Interested	Not interested	Not applicable
Workshops on best practices in data management for faculty.				
Workshops on best practices in data management for graduate students.				
Personalized consultation on data management practices for specific research groups or projects.				
Communication and information about funding requirements and journal requirements regarding research data.				
Assistance preparing data management plans to meet funding requirements, or assistance creating formal or documented data management policies.				
Digitization of physical records, such as lab notebooks.				
Assistance with documenting and describing data (i.e. metadata creation).				
Assistance with issues associated with data preservation and/or sharing (confidentiality, privacy, legal, intellectual property rights).				
Data storage and backup during active research projects.				
An institutional repository for long-term access and preservation of research data.				
Assistance with depositing research data in appropriate				

disciplinary or other external data repositories.				
Assignment of permanent digital object identifiers (DOIs) for datasets.				
Assistance in finding and accessing data sources.				

Q20. If there are other services you would like to see offered, please specify:

Section 4. Demographic & General Questions

[In this survey, the demographic questions were at the end of the survey. These questions were moved to the beginning of the survey for the Social Science and Humanities Survey.]

In this section you will be asked basic questions about your rank, departmental and research affiliations, and funding sources so we can put your answers into context.

Q21. Please indicate your rank at [HOME UNIVERSITY - rank will reflect institutional categories]:

- Graduate Student
- Postdoctoral Fellow
- Lecturer
- Adjunct Professor
- Assistant/Associate/Full Professor
- Professor Emeritus
- Other, please specify

Q22. Please, select your faculty, home institute or department at [HOME UNIVERSITY]. If your faculty, institute, or department is not listed, please select the one with which you are most closely affiliated: **[Draft list of subject areas - schools will list their own departments in these areas]**

- NSERC areas:
- Civil and mechanical engineering
- Chemical, biomedical and materials science engineering
- Electrical engineering
- Computing sciences
- Physics and astronomy
- Chemistry

- Earth sciences
- Evolution and ecology
- Cellular and molecular biology

Q23. Please indicate your cross-discipline affiliation outside of your home department, if applicable. Select all that apply: [Each institution enters their own list]

- I do not have a cross affiliation
- Prefer not to specify
- Other, please specify

Q24. Please indicate your research institute or centre, if applicable. Select all that apply: [Each institution enters their own list]

Other, please specify

Q25. 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:
[list to be based on each institution's funding sources but suggest basic list be included for each school's survey]

SSHRC
SSHRC Insight Grant
SSHRC Partnership Grant
CIHR
CFI
NSERC
Industry
None
Other, please specify:
_____

Section 5. Feedback, follow-up and additional information about the survey

In this final section you will be asked for your comments on the survey itself, as well as your willingness to be contacted for follow-up. This section also contains information about how the data collected in this survey will be used.

Please provide feedback on the survey content, or your thoughts on research data management practices. Any comments you have will be helpful.

We would be happy to follow up with any faculty member, postdoctoral fellow, graduate student or researcher who requires assistance from the library for current research data management practices or to help facilitate future collaborations. Are you willing to be contacted for follow up?

- Yes, please provide your email address

- No

This is the end of the survey. Thank you for taking the time to participate in the Research Data Management survey at [HOME UNIVERSITY].
You may withdraw from the survey at any time before hitting the submission button. Please note, once you submit your answers and close your browser you will no longer be able to withdraw your answers from the study. Anonymized survey submissions cannot be withdrawn. Any personal identifying information will be kept private, and all data will be anonymized in the event that results collected will be published. Please click the "Submit" button to submit your completed survey.

Appendix D: Humanities and Social Science RDM survey instrument – generic version

The French version of this survey instrument is available in the Generic Resources for Canadian RDM Survey Dataverse.

In this survey

- a 'researcher' can include any individual who worked on the research project that could legitimately claim academic authorship of the research project if the results of the research project were to be published in a scholarly work
- a 'research project' can be defined as the research associated with investigating a hypothesis or group of hypotheses (and applicable set of predictions) aimed at answering a distinct or specific research question. A single research grant may support one research project or multiple research projects. In the context of this survey, a research project is associated with a distinct set of research data and is considered to be a subset of a research program, research activity or research area you may investigate
- 'research data' can be defined as any data that are collected, observed, created or analyzed to produce research results. Research data could include:
    - Observational data such as sensor readings, telemetry, survey results, images
    - Simulation data such as climate models and economic models,
    - Derived or compiled data such as text and data mining, compiled database, 3D models which can come in many forms including text, numerical, multimedia, models, software, discipline specific or instrument specific

Section 1. Demographic & General Questions

In this section you will be asked basic questions about your rank, departmental and research affiliations, and funding sources so we can put your answers into context.

Q1. Please indicate your rank at [HOME UNIVERSITY - rank will reflect institutional categories]:

- Graduate Student
- Postdoctoral Fellow
- Lecturer
- Adjunct Professor
- Assistant/Associate/Full Professor
- Professor Emeritus
- Other, please specify

Q2. Please select your home department at [HOME UNIVERSITY]. If your department is not listed, please select the one with which you are most closely affiliated:
[Draft list of subject areas - schools will list their own departments in these areas]

- Anthropology/Archaeology
- Art
- Business/Management
- Communication Studies
- Diversity Studies
- Economics
- Education
- English Language and Literature
- Environmental Studies
- First Nations Studies
- Geography/Planning
- Global/Regional Studies
- History/Classics
- Law
- Library and Information Management
- Life Sciences
- Linguistics
- Modern Languages
- Performing Arts
- Philosophy/Religion
- Political Science
- Psychology
- Public Policy/International Relations
- Sociology
- Other, please specify

Q3. Please indicate your cross-discipline affiliation outside of your home department, if applicable.

Q4. Please indicate your research institute or centre, if applicable. Select all that apply:
[Each institution enters their own list]
Other, please specify

Q5. 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:
[list to be based on each institution's funding sources but suggest basic list be included for each school's survey]

- SSHRC
- SSHRC Insight Grant
- SSHRC Partnership Grant
- CIHR
- CFI
- NSERC
- Industry
- None
- Other, please specify

Q6. 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?

- Yes
- No
- Not Sure

Section 2: Working with Research Data

In this section you will be asked questions about your research data, including how you work with them, document them and store them.

Q7. How many research projects did you lead in the past year? Select one:

- 1-2 research projects
- 3-5 research projects
- >5 research projects
- Not sure
- Not applicable

Q8. How much data storage do you estimate you use in an average research project? Select one:

- < 1GB (Gigabyte)
- 1GB to < 10GB
- 10GB to < 50GB
- 50GB to < 500GB
- 500GB to < 1000GB
- 1TB to 4TB (Terabyte)
- >4TB
- Not sure
- Not Applicable, please explain

Q9. Which of the following best describes the type of research data you generate or use in a typical research project? Select all that apply:

- Geospatial – (e.g. raster, vector, grid)
- Instrument specific – (e.g. GPX, THING, LAS, ECW)
- Models – (e.g. 3D, statistical, similitude, macroeconomic, causal)
- Multimedia – (e.g. JPEG, TIFF, MPEG, MP3, Quicktime, Bitmap)
- Numerical – (e.g. CSV, MAT, XLS, SPSS)
- Software – (e.g. Java, C, Perl, Python, Ruby, PHP, R)
- Text – (e.g. TXT, DOC, PDF, RTF, HTML, XML, TEI)
- Other – (e.g. discipline specific such as CIF, FITS, DICOM) please specify

Q10. Please list any software and/or hardware used for the collection, analysis, or manipulation of your research data, if applicable:

Q11. Please indicate where you store research data from your current project(s). Select all that apply:

- Flash drive/USB
- CD/DVD
- Computer hard drive (i.e. local hard drive)
- Laptop hard drive
- External hard drive
- Hard drive of the instrument/sensor which generates the data
- Shared drive/university or departmental server
- Cloud/web based solution (e.g. Dropbox, Google Drive, Amazon Cloud, Microsoft Cloud)
- External data repository (e.g. Institutional Repository, GitHub, tDAR, CWRC, Artstor, Figshare, HathiTrust)
- Grid/high performance computing (HPC) centre
- Physical copy retained (in boxes, cabinets, etc.)
- Not sure
- Other, please specify

Q12. Is there sufficient documentation and description (for example, variable and field definitions, codebooks, data dictionaries, metadata, scripts to run) for another person outside your research team to *understand* and *use* the research data?

- Yes
- No
- Not sure

Q13. Is there sufficient documentation and description (for example, variable and field definitions, codebooks, data dictionaries, metadata, scripts to run) retained in the same file, folder or document for another person outside your research team to *replicate* the methodologies that produced the data?

- Yes
- No
- Not sure

Q14. 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.

	I only keep data for the length of the project	Less than 3 years	Between 3-5 years	Between 5-10 years	More than 10 years	Until the data becomes inaccessible or lost
Source material/ Raw data						
Intermediate/ working data						
Processed data ready for publication. Processed data may include supporting information such as spectra, synthesis methods						

Section 3. Data Sharing

In this section you will be asked about your current practices on sharing your research data.

Q15. 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

- Share by personal request only
- Share online with restricted access
- Upload online to an institutional or personal website
- Upload online to an institutional repository, such as Dataverse
- Include as part of supplementary material files to a journal publisher
- Deposit in a general or discipline-specific repository, such as GitHub, tDAR, CWRC, Artstor, Figshare, HathiTrust, please specify
- If you plan to use another data repository other than the ones listed above, please specify

Q16. 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
- Share online with restricted access
- Upload online to an institutional or personal website
- Upload online to an institutional repository, such as Dataverse
- Include as part of supplementary material files to a journal publisher
- Deposit in a general or discipline-specific repository, such as GitHub, tDAR, CWRC, Artstor, Figshare, HathiTrust. Please specify
- If you plan to use another data repository other than the ones listed above, please specify

Q17. Some research data cannot be shared because of legal or privacy restrictions or embargoes. Which of the following restrictions or embargoes may limit your ability to share your data with others? Select all that apply. If there are no restrictions or embargoes, choose 'there are no restrictions or embargoes on sharing my data with other parties.'

- There are no restrictions or embargoes on sharing my data with other parties
- I need to publish my data before I can share them
- Sharing my data may jeopardize Intellectual Property rights
- I plan to file for a patent
- My data cannot be shared because of commercial concerns
- I have a contractual obligation with a third party
- My data are subject to privacy, confidentiality, or ethics restrictions (e.g. survey data with personal information)
- My data are a matter of public safety or of a sensitive nature
- I'm unsure if I am allowed to share my data
- Other, please specify

Q18. If your research data were not affected by such restrictions or embargoes, with whom would you be willing to share them? Select all that apply:

- Nobody
- My immediate collaborators

- Researchers in my department/institute/centre
- Researchers at **[HOME UNIVERSITY]**
- Researchers in my field
- Researchers outside my field
- Anybody, including the general public

Q19. What, if any, are the reasons you would not be willing to share your research data and associated methods/tools? Select all that apply. If you are willing to share, choose 'I am willing to share them.'

- I am willing to share them
- They are incomplete or not finished
- I still wish to derive value from them
- I do not have the technical skills or knowledge
- I do not hold the rights to share them
- Funding body does not require sharing
- I believe they should not be shared
- I did not know I could share them
- Insufficient time
- Lack of standards to make them usable by others
- Lack of funding
- No place to put them
- They are not useful to others
- There are privacy, legal or security issues
- I'm concerned my data could be used without proper citation or acknowledgement
- Other, please specify

Q20. 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.'

- I see no benefits to sharing my data
- Data availability provides safeguards against misconduct, data fabrication and falsification
- Data sharing and/or replication studies help in the training of next generation researchers
- Data sharing enables my data to be cited and increases my research impact
- Data sharing encourages collaborative scholarship
- Data sharing encourages interdisciplinary research
- Data sharing moves my field of research forward
- Data sharing reduces redundant data collection
- Data sharing supports open access to knowledge
- Re-analysis of data helps verify results
- Well-maintained data helps retain data integrity
- Other, please specify

Q21. Are you aware of any discipline-specific research 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.’

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

In this section you will be asked about your awareness of funding mandates for RDM, as well as your interest in potential RDM services.

Q22. 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 that would address these types of questions without assistance
- I would be able to draft a data management plan that would address these types of questions, but would prefer to have assistance and/or guided documentation to ensure the success of my application
- I would need assistance and/or guided documentation to appropriately address some or all of the sections

Q23. Do you include any of the following topics related to RDM in your teaching practice? Select all that apply. If you do not teach RDM topics, choose 'I do not teach RDM topics.'

- I do not teach RDM topics
- Data security
- Data privacy
- Data version control
- Data backup
- Data ethics
- Data sharing
- Data documentation
- Data retention
- Data archiving
- Other, please specify

Q24. Do you use *your own* research data in your teaching practice?

- Yes
- No
- Not applicable

Q25. If data management plans were made part of grant applications from funding bodies such as SSHRC, CIHR, and NSERC, how interested would you be in the following services? Please rate your interest in each service. If the service does not apply to your situation, choose 'not applicable.'

	Very interested	Interested	Not interested	Not applicable
Workshops on best practices in data management for faculty.				
Workshops on best practices in data management for graduate students.				
Personalized consultation on data management practices for specific research groups or projects.				
Communication and information about funding requirements and journal requirements regarding research data.				
Assistance preparing data management plans to meet funding requirements, or assistance creating formal or documented data management policies.				
Digitization of physical records, such as lab notebooks.				
Assistance with documenting and describing data (i.e. metadata creation).				
Assistance with issues associated with data preservation and/or sharing (confidentiality, privacy, legal, intellectual property rights).				
Data storage and backup during active research projects.				
An institutional repository for long-term access and preservation of research data.				
Assistance with depositing research data in appropriate disciplinary or other external data repositories.				
Assignment of permanent digital object identifiers (DOIs) for datasets.				

Assistance in finding and accessing data sources.				

Q26. If there are other services you would like to see offered, please specify:

Section 5. Feedback, follow-up and additional information about the survey

In this final section you will be asked for your comments on the survey itself, as well as your willingness to be contacted for follow-up. This section also contains information about how the data collected in this survey will be used.

Please provide feedback on the survey content, or your thoughts on research data management practices. Any comments you have will be helpful.

We would be happy to follow up with any faculty member, postdoctoral fellow, graduate student or researcher who requires assistance from the library for current research data management practices or to help facilitate future collaborations. Are you willing to be contacted for follow up?

- Yes, please provide your email address
- No

This is the end of the survey.Thank you for taking the time to participate in the Research Data Management survey at [HOME UNIVERSITY]!

You may withdraw from the survey at any time before hitting the submission button. Please note, once you submit your answers and close your browser you will no longer be able to withdraw your answers from the study. Anonymized survey submissions cannot be withdrawn. Any personal identifying information will be kept private, and all data will be anonymized in the event that results collected will be published. Please click the "Submit" button to submit your completed survey.

Appendix E: Health and Medical Sciences RDM survey instrument – generic version

The French version of this survey instrument is available in the Generic Resources for Canadian RDM Survey Dataverse.

In this survey:

- a **'researcher'** can include any individual who worked on the research project that could legitimately claim academic authorship of the research project if the results of the research project were to be published in a scholarly work
- a **'research project'** can be defined as the research associated with investigating a hypothesis or group of hypotheses (and applicable set of predictions) aimed at answering a distinct or specific research question. A single research grant may support one research project or multiple research projects. In the context of this survey, a research project is associated with a distinct set of research data and is considered to be a subset of a research program, research activity or research area you may investigate
- **'research data'** can be defined as any data that are collected, observed, created or analyzed to produce research results. Research data could include:
    - Observational data such as sensor readings, telemetry, survey results, images
    - Simulation data such as climate models and economic models,
    - Derived or compiled data such as text and data mining, compiled database, 3D models which can come in many forms including text, numerical, multimedia, models, software, discipline specific, or instrument specific

Section 1. Demographic & General Questions

In this section you will be asked basic questions about your rank, departmental and research affiliations, and funding sources so we can put your answers into context.

Q1. Please indicate your rank at [HOME UNIVERSITY]
Customize rank to your home institution

- Graduate Student
- Medical Resident
- Postdoctoral Fellow
- Lecturer
- Adjunct Professor
- Assistant/Associate/Full Professor/Clinical Colleague
- Professor Emeritus
- Other, please specify

Q2. Please select your home department at [HOME UNIVERSITY]. If your department is not listed, please select the one with which you are most closely affiliated:

Q3. Please indicate your cross-discipline affiliation outside of your home department, if applicable.

Q4. Please indicate your research institute or centre, if applicable. Select all that apply:
Customize to your home institution

Q5. 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:

- CIHR
- SSHRC
- SSHRC Insight Grant
- SSHRC Partnership Grant
- CFI
- NSERC
- Industry
- Provincial funding, please specify
- None
- Other, please specify

Section 2: Working with Research Data

In this section you will be asked questions about your research data, including how you work with them, document them and store them.

Q7. How many research projects did you lead in the past year? Select one:

- 1-2 research projects
- 3-5 research projects
- >5 research projects

- Not sure
- Not applicable

Q8. How much data storage do you estimate you use in an average research project? Select one:

- < 1GB (Gigabyte)
- 1GB to < 10GB
- 10GB to < 50GB
- 50GB to < 500GB
- 500GB to < 1000GB
- 1TB to < 4TB (Terabyte)
- 4TB to 500TB
- >500TB
- Not Applicable, please explain

Q9. Which of the following best describes the type of research data you generate or use in a typical research project? Select all that apply:

- Geospatial – (e.g. raster, vector, grid, boundary files)
- Instrument specific (e.g. fMRI, LSM)
- Models – (e.g. 3D, statistical, similitude)
- Multimedia – (e.g. JPEG, TIFF, MPEG, MP3, Quicktime, Bitmap, Audio/Visual records)
- Numerical – (e.g. CSV, MAT, XLS, SPSS)
- Software – (e.g. Java, C, Perl, Python, Ruby, PHP, R)
- Text – (e.g. TXT, DOC, PDF, RTF, HTML, XML, Clinical, Client Reviews, Health Records)
- Discipline specific (e.g. BAM, FASTQ, CEL, IDAT, FASTA, PBD, BRK, DICOM)
- Other, please specify

Q10. Please list any software and/or hardware used for the collection, analysis, or manipulation of your research data, if applicable:

Q11. Please indicate where you store research data from your current project(s). Select all that apply:

- Flash drive/USB
- CD/DVD
- Computer hard drive (i.e. local hard drive)
- Laptop hard drive
- External hard drive
- Hard drive of the instrument/sensor which generates the data
- Shared drive/university or departmental server
- Cloud/web based solution (e.g. Dropbox, Google Drive, Amazon Cloud, Microsoft Cloud)
- External data repository (e.g. Dryad, Protein Data Bank, GenBank, PeptideAtlas, Cancer Imaging Archive)
- Grid/high performance computing (HPC) centre
- Physical/paper copy retained (in boxes, cabinets, etc.)
- Not sure
- Other, please specify

Q12. Is there sufficient documentation and description (for example, variable and field definitions, codebooks, data dictionaries, metadata, scripts to run) for another person outside your research team to *understand* and *use* the research data?

- Yes
- No
- Not sure

Q13. Is there sufficient documentation and description (for example, variable and field definitions, codebooks, data dictionaries, metadata, scripts to run) retained in the same file, folder or document for another person outside your research team to *replicate* the methodologies that produced the data?

- Yes
- No
- Not sure

Q14. 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.

	I only keep data for the length of the project	Less than 3 years	Between 3-5 years	Between 5-10 years	More than 10 years	Until the data becomes inaccessible or lost
Source material/Survey results/Raw data.						
Intermediate/ working data.						
Processed data ready for publication. Processed data may include supporting information such as metadata and documentation, spectra, or synthesis methods.						

Section 3. Data Sharing

In this section you will be asked about your current practices on sharing your research data.

Q15. 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
- Share by personal request only
- Share online with restricted access
- Upload online to an institutional or personal website
- Upload online to an institutional repository, such as Dataverse
- Include as part of supplementary material files to a journal publisher
- Deposit in a general or discipline-specific repository, such as Dryad, Protein Data Bank, GenBank, PeptideAtlas, Cancer Imaging Archive, please specify
- If you plan to use another data repository other than the ones listed above, please specify

Q16. 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
- Share online with restricted access
- Upload online to an institutional or personal website
- Upload online to an institutional repository, such as Dataverse
- Include as part of supplementary material files to a journal publisher
- Deposit in a general or discipline-specific repository, such as Dryad, Protein Data Bank, GenBank, PeptideAtlas, Cancer Imaging Archive. Please specify
- If you plan to use another data repository other than the ones listed above, please specify

Q17. Some research data cannot be shared because of legal or privacy restrictions or embargoes. Which of the following restrictions or embargoes may limit your ability to share your data with others? Select all that apply. If there are no restrictions or embargoes, choose 'there are no restrictions or embargoes on sharing my data with other parties.'

- There are no restrictions or embargoes on sharing my data with other parties
- I need to publish my data before I can share them
- Sharing my data may jeopardize Intellectual Property rights
- I plan to file for a patent
- My data cannot be shared because of commercial concerns
- I have a contractual obligation with a third party
- My data are subject to privacy, confidentiality, or ethics restrictions (e.g. survey data with personal/patient information)
- My data are a matter of public safety or of a sensitive nature
- I'm unsure if I am allowed to share my data
- Other, please specify

Q18. If your research data were not affected by such restrictions or embargoes as listed in Q17 above, with whom would you be willing to share them? Select all that apply:

- Nobody
- My immediate collaborators
- Researchers in my department/institute/centre

- Researchers at [HOME UNIVERSITY]
- Researchers in my field
- Researchers outside my field
- Anybody, including the general public

Q19. What, if any, are the reasons you would not be willing to share your research data and associated methods/tools/algorithms? Select all that apply. If you are willing to share, choose 'I am willing to share them.'

- I am willing to share them
- They are incomplete or not finished
- I still wish to derive value from them
- I do not have the technical skills or knowledge
- I do not hold the rights to share them
- Funding body does not require sharing
- I believe they should not be shared
- I did not know I could share them
- Insufficient time
- Lack of standards to make them usable by others
- Lack of funding
- No place to put them
- They are not useful to others
- There are privacy, legal or security issues
- I'm concerned my data could be used without proper citation or acknowledgement
- Other, please specify

Q20. 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.'

- I see no benefits to sharing my data
- Data availability provides safeguards against misconduct, data fabrication and falsification
- Data sharing and/or replication studies help in the training of next generation researchers
- Data sharing enables my data to be cited and increases my research impact
- Data sharing encourages collaborative scholarship
- Data sharing encourages interdisciplinary research
- Data sharing moves my field of research forward
- Data sharing reduces redundant data collection
- Data sharing supports open access to knowledge
- Re-analysis of data helps verify results
- Well-maintained data helps retain data integrity
- Other, please specify

Q21. Are you aware of any discipline-specific research 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.'

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

In this section you will be asked about your awareness of funding mandates for RDM,

as well as your interest in potential RDM services.

Q22. 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 that would address these types of questions without assistance
- I would be able to draft a data management plan that would address these types of questions, but would prefer to have assistance and/or guided documentation to ensure the success of my application
- I would need assistance and/or guided documentation to appropriately address some or all of the sections

Q23. Do you include any of the following topics related to RDM in your teaching practice? Select all that apply. If you do not teach RDM topics, choose 'I do not teach RDM topics.'

- I do not teach RDM topics
- Data security
- Data privacy
- Data version control
- Data backup
- Data ethics
- Data sharing
- Data documentation
- Data retention
- Data archiving
- Other, please specify

Q24. Do you use *your own* research data in your teaching practice?

- Yes
- No
- Not applicable

Q25. If data management plans were made part of grant applications from funding bodies such as CIHR, SSHRC, and NSERC, how interested would you be in the following services? Please rate your interest in each service. If the service does not apply to your situation, choose 'not applicable.'

	Very interested	Interested	Not interested	Not applicable
Workshops on best practices in data management for faculty.				
Workshops on best practices in data management for graduate students.				
Personalized consultation on data management practices for specific research groups or projects.				
Communication and information about funding requirements and journal requirements regarding research data.				
Assistance preparing data management plans to meet funding requirements, or assistance creating formal or documented data management policies.				
Digitization of physical records such as lab notebooks				
Assistance with documenting and describing data (i.e. metadata creation).				
Assistance with issues associated with data preservation and/or sharing (confidentiality, privacy, ethics, legal, intellectual property rights, data anonymization).				
Data storage and backup during active research projects.				
An institutional repository for long-term access and preservation of research data.				
Assistance with depositing research data in appropriate disciplinary or other external data repositories.				

Assignment of permanent digital object identifiers (DOIs) for datasets.				
Assistance in finding and accessing data sources.				

Q26. If there are other services you would like to see offered, please specify:

Section 5. Feedback, follow-up and additional information about the survey

In this final section you will be asked for your comments on the survey itself, as well as your willingness to be contacted for follow-up. This section also contains information about how the data collected in this survey will be used.

Please provide feedback on the survey content, or your thoughts on research data management practices. Any comments you have will be helpful.

We would be happy to follow up with any faculty member, postdoctoral fellow, graduate student, medical resident, or researcher who requires assistance from the library for current research data management practices or to help facilitate future collaborations. Are you willing to be contacted for follow up?

- Yes, please provide your email address
- No

This is the end of the survey. Thank you for taking the time to participate in the Research Data Management survey at [HOME UNIVERSITY]!

You may withdraw from the survey at any time before hitting the submission button. Please note, once you submit your answers and close your browser you will no longer be able to withdraw your answers from the study. Anonymized survey submissions cannot be withdrawn. Any personal identifying information will be kept private and all data will be anonymized in the event that results collected will be published. Please click the "Submit" button to submit your completed survey.