

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. 13, no. 1 (2018)

DOI: <http://doi.org/10.21083/partnership.v13i1.4115>

CC BY-NC-ND 4.0

Developing research data management services and support for researchers: A mixed methods study

Laure Perrier¹

Research Data Management Librarian

University of Toronto Libraries

l.perrier@utoronto.ca

 <https://orcid.org/0000-0001-9941-7129>

Leslie Barnes¹

Digital Scholarship Librarian

University of Toronto Libraries

leslie.barnes@utoronto.ca

¹ University of Toronto Libraries; 130 St. George Street; Toronto, Canada, M5S 1A5

Abstract

This mixed method study determined the essential tools and services required for research data management to aid academic researchers in fulfilling emerging funding agency and journal requirements. Focus groups were conducted and a rating exercise was designed to rank potential services. Faculty conducting research at the University of Toronto were recruited; 28 researchers participated in four focus groups from June–August 2016. Two investigators independently coded the transcripts from the focus groups and identified four themes: 1) seamless infrastructure, 2) data security, 3) developing skills and knowledge, and 4) anxiety about releasing data. Researchers require assistance with the secure storage of data and favour tools that are easy to use. Increasing knowledge of best practices in research data management is necessary and can be supported by the library using multiple strategies. These findings help our library identify and prioritize tools and services in order to allocate resources in support of research data management on campus.

Keywords

research data management; data sharing; focus groups; qualitative research

Introduction

Background and rationale

There is growing momentum in the research community both globally and nationally with funding agencies and journals moving towards increased transparency and requiring the development of data management plans, the preparation of data for sharing, and public access to raw data (Ioannidis et al., 2014; Shearer, 2015). In practice, this means that data must be selected, managed, and stored in a manner that ensures data integrity, protects sensitive data, and facilitates research replication. Research data management requires many skills, and academic researchers must determine how to meet these demands.

At present, Canada's federal granting agencies that promote and support research (known as the Tri-Agency) do not have a policy on digital data management that outlines obligations for researchers. In contrast, other countries such as the United Kingdom and the United States have established policies related to research data management at the government level or by major funding agencies, for example the Wellcome Trust in the United Kingdom (Shearer, 2015). In June 2016, Canada's Tri-Agency released a "Tri-Agency Statement of Principles on Digital Data Management" (Government of Canada, 2016). This document outlines expectations and responsibilities for activities such as the creation of data management plans and sharing research data that are similar to policies of other countries (Shearer, 2015). However no specific requirements or obligations were outlined, (e.g., that data management plans be submitted with grant applications). In May 2017, the Tri-Agency sought community feedback during face-to-face meetings at academic institutions and announced that a policy was forthcoming. The "Tri-Agency Statement of Principles on Digital Data Management" (Government of Canada, 2016) and the Tri-Agency community feedback sessions indicated some elements that will be included in the upcoming policy and provided an opportunity to consider how the University of Toronto Libraries could further support researchers.

While it is possible to anticipate needs and to establish services based on emerging policy requirements, a more direct and effective approach is to consult directly with researchers to determine what is most useful to aid their efforts in appropriately managing research data. Focus groups provide the opportunity to consult directly with end-users while also gathering information from a cross-section of people to collect multiple points of view at one time (Creswell, 2007). The dynamic exchange among participants can generate ideas in other members of the group and contribute to the quality of information, which would not be possible in one-to-one interviews (Creswell, 2007). The purpose of this study is to determine the essential tools and services researchers require for research data management to fulfill funding agency and journal requirements.

Literature Review

In this literature review, we first examine reviews of research data management services and studies of researcher attitudes and needs to identify common findings. We then examine the ways in which national circumstances have been found to affect the development of research data management services. Lastly, we suggest that our study expands on this field by adding a Canadian study that examines faculty attitudes in anticipation of emerging federal funding agency policies regarding data management.

Studies of research data management services have co-emerged with the development of research data management services in libraries and have charted both institutional and cross-institutional accounts. Using a variety of methodologies, researchers have evaluated faculty attitudes to research data management services (Peters & Dryden, 2011), modeled use cases for research data management services at an institution (Lage, Losoff, & Maness, 2011), and identified researcher needs for research data management services (Parham, Bodnar, & Fuchs, 2012). Cross-institutional studies have largely focused on the scope and breadth of research data management services offered by academic libraries (Soehner, Steeves, & Ward, 2010; Cox & Pinfield, 2013; Goldman, Kafel, & Martin, 2015; Richardson, Nolan-Brown, Loria, & Bradbury, 2012; Pinfield, Cox, & Smith, 2014; Corral, Kennan, & Afzal, 2013).

Research data management emerges as a complex practice that can occur at multiple stages in the research cycle. In a 2013 SPEC Kit on research data management, researchers identified numerous activities that can be considered research data management, including data management plan support, data management best practices, metadata support, organization, data citation, data sharing and access, and data storage and backup (Fearon, Gunia, Pralle, Lake, & Sallans, 2013). Research data management requires a diverse set of competencies.

Drivers for developing research data management services vary from institution to institution and in different national contexts. In their study of UK institutions, Pinfield et al. (2014) identify several motivators for research data management services: storage, security, preservation, compliance, quality, sharing, and jurisdiction. In the US, the 2010 announcement and 2011 enactment of data management plan requirements by the National Science Foundation (NSF) accelerated libraries' development of data services (Akers, Sferdean, Nicholls, & Green, 2014). A survey of NSF principal investigators at Cornell University, however, indicated that faculty were still uncertain about the NSF requirements (Steinhart, Chen, Arguillas, Dietrich, & Kramer, 2012).

While there have been several assessments of faculty needs in the area of research data management (Scaramozzino, Ramirez, & McGaughey, 2012; Bardyn, Resnick, & Camina, 2012; Parham et al., 2012), these studies focus on US and UK institutions. Though some reviews and surveys of research data management support include Canadian institutions (Soehner et al., 2010; Tenopir, Sandusky, Allard, & Birch, 2013; Sewerin et al., 2016a; Sewerin et al., 2016b), there is little qualitative literature about researchers' research data management needs in the Canadian context. Moreover, as Canadian universities anticipate the emergence of funding agency policies—as

indicated by the recent “Tri-Agency Statement of Principles on Digital Data Management” (Government of Canada, 2016)—it will be essential to understand faculty requirements.

Methods

Design

Focus groups were conducted to learn what services academic researchers believe are critical to support them in research data management. Participants were asked about their familiarity with writing data management plans and organizing, preserving, and sharing their research data. As well, they were asked about what potential support or services would be useful to them. The focus groups were planned with a post-positivist paradigm which assumes it is possible to capture true representations of the real world (Bhaskar, 1975; Collier, 1994; Patton, 2002; Pawson & Tilley, 1997; Miles & Huberman, 1994). This approach encourages insights that extend beyond the realm of measurable facts and discovery to generate in-depth, rich data that are based on the individuals' personal perspectives and experiences, leading to findings that can be transferred from samples to broader populations (Krueger, 1998).

Sampling and recruitment

University of Toronto has three campuses with a student (undergraduate and graduate) enrolment of 88,000 with 14,000 faculty receiving \$1.2 billion CDN in research dollars for 2015–2016. There are 44 libraries spanning the three campuses (University of Toronto, n.d.). University of Toronto faculty members (full- or part-time) currently conducting research were recruited for participation in the study. Several venues where we were recruiting had mixed groups, and we also included research coordinators who represented primary investigators, post-doctoral students, librarians conducting research, and IT professionals if they were involved with helping researchers manage their data. Recruitment for two focus groups took place by sending invitations to researchers within a Faculty (such as the Faculty of Arts & Science), and the participants for two other focus groups were recruited during a two-day conference held at the University of Toronto Mississauga campus. All focus groups were conducted on campus in a location convenient to participants. The moderators had professional interactions with some of the participants prior to conducting the focus groups.

Data collection

Three to five focus groups were planned with approximately five to ten participants per group. Recruitment efforts were organized to optimize including participants conducting research in a variety of disciplines. Once saturation of themes was identified, the focus groups were halted and no further data were collected. A moderator (one of the authors, either LP or LB) led each focus group. As well, a note taker was present at each focus group to record observations. Both moderators had previous experience in conducting qualitative research including focus groups, interviews, and usability studies. The purpose of the focus groups was explained to participants at the beginning of the

session, and they were asked to read and sign a consent form. Towards the end of the focus group session, participants were asked to rate potential services and tools using a Likert scale ranging from “very useful” to “not useful” (Appendix A). This quantitative exercise, based on services and tools that could be offered in a more immediate manner, was used to gauge prioritization amongst researchers. It was based on topics identified in the “Tri-Agency Statement of Principles on Digital Data Management” (Government of Canada 2016) and the emerging requirements of journals. Each session lasted approximately 60 minutes and was audio-recorded and transcribed. The sessions were based on a balance of gaining answers to a pre-planned agenda of questions from the interview guide (Appendix B) and hearing from each participant in their own words (Morgan, 1998; Burnard, 1991). To prepare, a mock focus group was conducted by a moderator (LP) and co-investigator (LB) with a group of five volunteers who were a mix of post-doctoral and PhD students studying in health sciences or earth sciences. This group understood research processes along with challenges related to research data management, making them an excellent proxy for our target audience. This session provided the opportunity to identify difficulties with questions (e.g., wording, clarity) and to streamline processes (e.g., equipment). Feedback was elicited in a debriefing session that was held immediately afterwards, and this information was used to improve the focus group sessions.

Data analysis

Transcripts were read and re-read to achieve immersion (Braun & Clark, 2006). Field notes and the focus group transcripts were both reviewed. A thematic approach was taken for the evaluation of the data; inductive content analysis and constant comparison were used to analyze the data (Hsieh & Shannon, 2005; Elo & Kyngäs, 2008). Two researchers (LP, LB) generated codes by reviewing the interview guide, reflecting on the information being sought after by the questions, and then independently coding the initial transcript. A meeting was held to compare codes that were generated and to discuss commonalities, refine wording, and harmonize inconsistencies. Modifications were done by consensus and included collapsing and adding codes to more accurately represent the data. From this meeting, a final set of codes was generated and used by the two investigators to independently code all further transcripts (Appendix C). A meeting was held after each transcript was received to review coding and resolve discrepancies. Once coding was completed, the data were reviewed so that groupings could be made within the codes and clustered into themes and sub-themes (Lincoln & Guba, 1985). Microsoft Excel (Microsoft Office Professional Plus 2013) was used to organize data by codes and to arrange by themes and sub-themes.

Rigour and quality

Lincoln and Guba’s (1985) framework was used to enhance the rigour and quality of the study. During the focus groups, probing questions were used to further clarify and to provide depth to participant’s contributions (Mays & Pope, 2000). Two investigators coded and analyzed the verbal data independently, discussing any discrepancies until agreement was reached. Quotes from focus group participants are provided for transparency as well as providing support for themes so that readers can judge whether

the findings reflect the perceptions of participants. Procedures were documented to generate an audit trail of coding and theme development; this process of triangulation ensured that findings arose from consensus amongst the investigators.

Results

Four focus groups were held between June and August 2016 (inclusive). The two investigators (LP, LB) independently coded all transcripts after they were transcribed. After the fourth focus group, the investigators agreed that no new information was being learned from participants; further recruitment was unnecessary.

Characteristics of participants

A total of 28 participants participated in four focus groups (Group 1: seven participants; Group 2: 9 participants; Group 3: five participants; Group 4: seven participants). Two focus groups were conducted in a meeting room within a building on the main campus of the University of Toronto. The other two focus groups were conducted during an educational event at the University of Toronto Mississauga campus. Table 1 summarizes the demographics and work profiles of the participants.

| | |
|---|----|
| Table 1 | |
| <i>Demographics and Work Profile of Study Participants (4 Focus Groups: N=28)</i> | |
| Age^a | |
| < 30 years | 2 |
| 30-39 years | 8 |
| 40-49 years | 9 |
| 50-59 years | 7 |
| 60-65 years | 1 |
| > 65 years | - |
| Gender^a | |
| Women | 17 |
| Men | 10 |
| Research Methods Used Most Often^b | |
| Qualitative | 12 |
| Quantitative | 10 |
| Mixed Methods | 10 |
| Years Conducting Research^a | |
| < 5 years | 1 |
| 5 – 10 years | 7 |
| 11 – 15 years | 6 |
| 16 – 25 years | 7 |
| > 25 years | 5 |

| Faculty or Division | |
|----------------------------------|----|
| Applied Science & Engineering | 1 |
| Architecture | - |
| Arts & Science | 16 |
| Dentistry | - |
| Education | - |
| Forestry | - |
| Information Science | - |
| Kinesiology & Physical Education | - |
| Law | - |
| Social Work | - |
| Management | - |
| Medicine | 9 |
| Music | - |
| Nursing | - |
| Pharmacy | - |
| Public Health | - |
| Other (Librarian) | 2 |
| Campus^b | |
| St. George (downtown) | 27 |
| Mississauga | 5 |
| Scarborough | 2 |

^a Three participants provided partial or no demographic information.

^b Some participants selected more than one answer.

Findings: Themes and sub-themes

Four major themes were identified: 1) seamless infrastructure; 2) data security; 3) developing skills and knowledge; and 4) anxiety about releasing data.

Theme: Seamless infrastructure

Researchers were aware that conscientious research data management is necessary for a variety of reasons including safety and security of data, long term preservation, compliance with research ethics board and funder requirements, and sharing. Careful and effective research data management is challenging, and researchers expressed interest in the University working with them to develop practical solutions that minimized the burden on their time.

Sub-theme: Safe storage of data

Technological obsolescence, cost, and uncertainty regarding the policies of commercial products were difficulties researchers encountered when selecting products or tools for data management. One researcher noted:

“And I would prefer not to use a third party . . . I would like to put something somewhere that it's persistent and that it's not Dropbox or Google Drive.” (*Focus Group 3, Participant 2*)

Many researchers spoke frequently about data storage and described using a wide range of storage solutions, from commercial products to non-proprietary databases they had built on their own. Researchers wanted the University to develop or purchase data storage infrastructure to streamline the process. Some wondered if vetted storage space could be “bought in bulk” so that the savings could be passed on to investigators, which would simultaneously allow them to maximize their grant funds and to ensure compliance with storage and security requirements. A participant stated:

“So if, at a massive scale, a University . . . were to arrange a certain space which is dedicated for data storage in the cloud—analogous to Amazon Glacier or something like that—then perhaps that could be a resource.” (*Focus Group 1, Participant 4*)

Similarly, researchers identified that if data storage was addressed at an institutional level, obsolescence of storage platforms and programs would not be a threat to their data as there could be a University-wide plan for updates. As well, criteria and standards required for ethical compliance could be assured.

Sub-theme: Hassle-free collaborations

One challenge with infrastructure involved the increasing numbers of global partnerships amongst researchers and the issues associated with managing the related data from their collaborative studies. When investigators at multiple institutions required access to data related to a single project, finding a solution was challenging due to restrictions, firewalls, and limitations set for those outside the institution. Even if participants met all requirements for maintaining safe and secure data, collaborating with individuals outside their own organization on active projects presented difficulties, especially when sharing data. This was highlighted in the following comments by participants:

“What do you do if your collaborators are from other universities and they don't want to put their work there, or they're not allowed to put their work there?” (*Focus Group 3, Participant 2*)

“I have a collaborator at a university in [name of place], where I work primarily, they're not considered to be part of UofT [University of Toronto] so there's this firewall between what I'm doing and what they're doing when in fact I would really love there not to be any wall and it would make our lives a lot easier, and more secure. . . .” (*Focus Group 3, Participant 6*)

Sub-theme: Ease-of-use

When participants spoke about the tools for data management, familiarity was an essential attribute. Tools that were uncomplicated and user-friendly were highly

favoured and considered necessary for researchers to incorporate into their research practice. This was described by one participant with the simple statement:

“[We want] things that are easy to use.” (*Focus Group 3, Participant 2*)

Dropbox (dropbox.com) was mentioned numerous times as an example of an effective tool that not only allowed data to be stored effortlessly, but also permitted easy sharing with collaborators. It was suggested that a Dropbox-like product housed under the University umbrella would be an optimal solution. This would address concerns with external provider policies and assure compliance related to appropriate storage, in particular security-related concerns. One researcher summed it up:

“If there were a kind of a hub, you know, or a portal—through the auspices of the University of Toronto—where one could store stuff as easily as using Dropbox.” (*Focus Group 4, Participant 2*)

Theme: Data security

Protection of data was taken very seriously by participants. For some researchers, this solely involved ensuring they could complete a study without data loss. For others, it also involved protecting human subjects and ensuring their privacy was not breached.

Sub-theme: Safeguarding participants

For the investigators who collected data involving human subjects, data security was an issue of significant concern. Questions were raised around being able to adequately protect human subjects. This included practical issues such as effectively de-identifying subjects within video or audio files and the associated work and technical expertise to do so. This was emphasized in the following statements:

“We had a lot of concerns in terms of privacy and confidentiality and to try and de-identify the people . . . if we’re going to be looking at having all data available those kinds of issues will be really, really big just because it is personal, identifiable information.” (*Focus Group 1, Participant 1*)

“I am committed to protecting [the] community and not directing hostile traffic to vulnerable people, but at the same time these are public documents, and I want to talk about them.” (*Focus Group 4, Participant 4*)

There was interest in having the University provide clarity and guidance on ways to anonymize this kind of data to satisfy legal and ethical requirements.

Sub-theme: At-risk data

There was an understanding that data should be backed up. For participants that had smaller datasets (contrasted to those with large datasets, (i.e., terabytes), tools provided by external companies were used to duplicate data. For instance, Gmail was frequently described as a back-up tool, with participants expressing uncertainty around

the policies related to this product and how this may affect their data. Typical comments included:

“ . . . because Gmail’s actually my alternate backup. Like, everything that I need were it to crash I send through Gmail.” (Focus Group 4, Participant 1)

“ . . . I don’t fully understand Google’s privacy policy . . . so I have been thinking about going through my Gmail and deleting everything that I have any sort of intellectual stake in, in my Gmail.” (Focus Group 4, Participant 5)

Participants felt that the research community would welcome clear guidance on appropriate backup tools.

Theme: Developing skills and knowledge

The research landscape is changing, and participants knew that expectations around sharing data were likely to change soon. Participants indicated a range of comfort levels with this prospect and raised concerns around addressing future requirements from funding bodies. Participants also expressed a universal interest in being exposed to best practices.

Sub-theme: Conflicting messages: sensitive vs. open data

One area of difficulty for researchers was understanding the perceived change from closely protecting sensitive data to sharing data in public repositories. This was seen as a discrepancy, and there was concern with how to resolve all obligations, from groups such as an Ethics Review Board to funders that may be asking for data to be shared publicly. This was demonstrated in the following remarks:

“We have an REB [Research Ethics Board] process which has been very much around ensuring privacy and confidentiality and safe, secure storage of data, and this is a pendulum swinging very much the other way.” (Focus Group 3, Participant 2)

“The REB is asking what we do with the data long term, you know, ‘When will you destroy it?’ . . . But that as well will have implications in terms of, like, keeping data, putting it into a repository, so it will need to have, I guess, alignment with what the REB will approve.” (Focus Group 2, Participant 6)

It was felt that clear direction needed to be offered if researchers were to successfully navigate through these requirements. The University was identified as needing to help them develop knowledge in this area by clearly outlining obligations and responsibilities.

Sub-theme: Reinforcing best practices

Participants believed that if funders imposed requirements in areas such as sharing data, documentation of data, or writing data management plans, there had to be a reliable standard or set of recommended methods. If best practices were outlined and

made available to researchers, this would provide dependable guidance when adapting to any new obligations. This is stressed in the participants' remarks:

"Having best practices around the safe collection/storage/transfer of data would be very useful." (Focus Group 2, Participant 2)

"What makes sense and what is common and what do a lot of other institutions or scholars use, so that I know that it can stay in that conversation, my work can stay in that arena." (Focus Group 3, Participant 4)

Participants did not readily agree upon the best method of delivering this information, and recommendations ranged from workshops to putting information on websites. Similarly, templates or examples of what would be considered ideal (e.g., data management plans) were identified as helping to contribute to the adoption of best practices.

Theme: Anxiety about releasing data

Participants understood that data sharing would likely be required by Canadian granting agencies in the future, and comfort levels with this fell on a wide spectrum. Wherever researchers fit on this spectrum, there was a clear understanding of how responsible research practices contribute to increasing the impact of the underlying data.

Sub-theme: How will other people use my data?

Participants raised concerns about who could potentially access and use their data. There was wariness around sharing data with unfamiliar individuals or groups. The following quotes encapsulate concerns expressed by the researchers:

"What seems rational is to have an agreement that if someone requests your data then you'll make it available to them, because then you know who it is who's asking, and you can explain to them, you know, any unusual features or what you have to do to understand this data." (Focus Group 1, Participant 3)

". . . I'm not sure it's been well thought through in terms of the nuances that are involved in making data—different kinds of data—available." (Focus Group 2, Participant 4)

Worries centered around the potential for misuse of data and how to manage this situation. Researchers were clear that if data were re-used by others, the dataset must be appropriately acknowledged and referenced. However, it was felt that properly referencing data was not well understood.

Sub-theme: Communicating context

Researchers described challenges with the usefulness of data when making them publicly available. Concerns raised around documentation included the time burden on researchers to make data independently understandable and challenges related to

providing sufficient contextual information (e.g., instrument calibration, weather conditions during data collection). The following comments illustrate these concerns:

“Other people who want to look at data—we can give them the data, but then there’s an issue, like, of what really was the condition here. . . .” (Focus Group 1, Participant 7)

“People are collecting data getting them to label . . . with reference to an ID, the date, the testing session, those kinds of details. I find it seems to slip through and which then can be hard later on if you’re trying to match back what that data relates to.” (Focus Group 2, Participant 9)

Findings: Rating Exercise

All 28 participants filled out the rating exercise to rank potential services or tools within research data management using a Likert scale ranging from “very useful” to “not useful.” Table 2 reports all responses and ratings for each item listed. Most items (10 out of 13) were identified as “very useful” or “useful” by the majority of participants (75% or more per item). The three remaining items had less support but still were still rated “very useful” or “useful” by over half of the participants: “How to Find Data from a Repository” (68%), “Consent Forms” (53%), and “Confidentiality” (53%).

Table 2

Rating Exercise

N = 28

| | | Rating, <i>n</i> (%) ^a | | | |
|---------------------------|---|-----------------------------------|-------------|---------------------------|------------|
| Item | | Very Useful or Useful | Average Use | Limited Use or Not Useful | No Opinion |
| Managing Data | Safe Storage of Data | 25 (89) | 3 (11) | 0 (0) | 0 (0) |
| | Data Management Plans | 24 (86) | 1 (4) | 2 (7) | 1 (4) |
| | Organizing Files | 23 (82) | 1 (4) | 2 (7) | 2 (7) |
| | Documenting Data | 22 (78) | 3 (11) | 2 (7) | 1 (4) |
| | Preserve Data for Long-Term Access | 22 (78) | 4 (14) | 1 (4) | 1 (4) |
| | Identify Data Repositories for Data Sharing | 21 (75) | 6 (21) | 0 (0) | 1 (4) |
| Research Support Services | Funding Agency Requirements | 25 (89) | 1 (4) | 0 (0) | 2 (7) |
| | Journal Requirements | 25 (89) | 1 (4) | 0 (0) | 2 (7) |
| | Intellectual Property | 23 (82) | 2 (7) | 1 (4) | 2 (7) |
| | Citing Data Sets | 21 (75) | 2 (7) | 3 (11) | 2 (7) |
| | How to Find Data from a Repository | 19 (68) | 4 (14) | 4 (14) | 1 (4) |
| | Consent Forms | 15 (53) | 4 (14) | 4 (14) | 5 (18) |
| | Confidentiality | 15 (53) | 4 (14) | 4 (14) | 5 (18) |

^a Percentages may not total 100 because of rounding.

Number of Participants

0-5 6-10 11-15 16-20 21-25

Discussion

The academic researchers in our study described their thoughts and preferences on the tools and services deemed important for research data management. Participants provided a multitude of strategies that could be considered to support compliance related to the management of research data. We had a group of researchers that came from a range of disciplines; however, our analysis revealed harmony on issues that were universal (e.g., data security). The results indicate academic libraries can provide services and support that would be welcomed by researchers, as well as identifying opportunities to partner with groups on campus to offer further assistance.

Libraries are well positioned to provide centralized services for research data management (Pinfield et al., 2014; Cox, Pinfield, & Smith, 2016). Researchers presented numerous questions around what they should be doing with regards to research data management, indicating a meaningful level of uncertainty. Some of this uncertainty stemmed from indications that funder requirements may soon include activities such as data sharing or writing data management plans. As well, participants highlighted data quality and data misuse as significant concerns. Training and education, query-answering services, and robust websites can be leveraged to address these topics and, in particular, focus on best practices. Although some issues are challenging (e.g., wariness around sharing data), it is possible to offer information (e.g., how to embargo data, how to cite data) or tools (e.g., templates for data dictionaries) that are helpful to researchers.

Challenges faced by researchers included the safety and storage of data, which involved confusion related to the policies associated with tools currently being used for backing up data. This confusion provided signs that data were potentially at risk. If tools specific to data management were offered at a University-level, participants indicated that making them seamless and uncomplicated needed to be a priority, ideally similar to tools that were already familiar (e.g., Dropbox). Some solutions for the safe storage of data could be addressed and managed solely by the library while others may require partnerships with groups on campus, such as information technology or the research office.

Participants indicated that most topics were considered “very useful” or “useful” on the rating exercise. Two of the three topics (confidentiality and consent forms) may have rated lower due to some of the researchers not conducting research with human subjects thus making these items irrelevant. “No opinion” was an option, but participants may have chosen “not useful” instead. The majority of topics generated interest amongst researchers, providing guidance for subject matter that can be incorporated into library services.

Limitations

Focus groups are characteristically limited in their ability to be generalized and extrapolated to a larger population. However, focus groups are well suited to providing experiences and perceptions (Kitzinger, 1995). Using a mixed method approach of

combining focus groups with a quantitative rating exercise provided a sound basis for the development of research data management tools and services at the institutional level.

Our focus groups included a broad spectrum of disciplines, and we speculated that we would have to conduct many sessions in order to generate themes and sub-themes due to heterogeneity. To address this, we set up individual focus groups to be homogeneous (e.g., individual groups had researchers from one discipline) and coded transcripts after each session. Despite participants conducting research in different areas, saturation was reached after a total of four focus groups. This may be due to the lack of demands being made on researchers related to their research data. For instance, researchers are currently not required to participate in writing data management plans for grant proposals submitted to major Canadian funding agencies. As a result, participants' responses reflect higher level matters such as interest in templates for data management plans or wanting to learn about best practices. When researchers are required to participate in stricter research data management practices, future studies may require more specificity and could assess researchers at specific career stages or examine a single discipline in relation to a particular issue (e.g., sharing sensitive data).

Conclusion

The results of our focus groups indicate that the library can play an important role in offering support to academic researchers as they face challenges related to research data management. Increasing knowledge of best practices is required, and this can be provided through a variety of offerings.

Our focus groups revealed areas where researchers were struggling and identified that assistance was required with data security and storage. Focus group participants indicated they favoured tools that were familiar (e.g., Dropbox). In addition to the provision of research data management tools and services, participants also indicated a strong need for guidance. Currently there are no explicit policies that would help direct actions on research data management at the University of Toronto. This presents an opportunity for work in this area to reach desired outcomes.

Focus groups accompanied with an exercise of rating key topics of importance provided clarification on areas needing attention. The valuable information collected in this study helped our library to identify and to prioritize areas for development, so we could strategically allocate resources in support of research data management on campus.

Contributors: LP conceived of the study. LP and LB conducted the focus groups and performed the coding and data analysis. LP and LB wrote the manuscript.

Funding: None

Declaration of Interest: None

Ethics approval: The Ethics Review Board of the University of Toronto, Toronto, Canada approved this study.

Acknowledgements: We would like to thank Dylanne Dearborn, Pam King, Andrew Nicholson, and Charlene Soobiah for taking notes during focus group sessions. We would like to acknowledge Dylanne Dearborn, Matt Gertler, Andrew Nicholson, Leanne Trimble, and Sarah Forbes for reviewing the protocol for the focus groups. We also thank Lily Yuxi Ren for collating information after the transcripts were coded and Andrés De los Ríos for assistance in anonymizing the transcripts.

Data sharing statement: Transcripts available from the Zenodo Repository: <http://doi.org/10.5281/zenodo.1187176>.

References

- Akers, K. G., Sferdean, F. C., Nicholls, N. H., & Green, J. A. (2014). [Building support for research data management: Biographies of eight research universities](#). *International Journal of Digital Curation*, 9(2), 171–191.
- Bardyn, T., Resnick, T., & Camina, S. (2012). [Translational researchers' perceptions of data management practices and data curation needs: Findings from a focus group in an academic health sciences library](#). *Journal of Web Librarianship*, 6(4), 274–287.
- Bhaskar, R. (1975). *A realist theory of science*. Leeds, England: Leeds Books.
- Braun, V., & Clark, V. (2006). [Using thematic analysis in psychology](#). *Qualitative Research in Psychology*, 3(2), 77–101.
- Burnard, P. (1991). [A method of analysing interview transcripts in qualitative research](#). *Nurse Education Today*, 11(6), 461–466.
- Collier, A. (1994). *Critical realism: An introduction to Roy Bhaskar's philosophy*. London, England: Verso.
- Corrall, S., Kennan, M. A., & Afzal, W. (2013). [Bibliometrics and research data management services: Emerging trends in library support for research](#). *Library Trends*, 61(3), 636–674.
- Cox, A. M., & Pinfield, S. (2013). [Research data management and libraries: Current activities and future priorities](#). *Journal of Librarianship and Information Science*, 46(4), 299–316.
- 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.

- Creswell, J. W. (2007). *Designing and conducting mixed methods research*. Thousand Oaks, CA: Sage Publications.
- Elo, S. & Kyngäs, H. (2008). [The qualitative content analysis process](#). *Journal of Advanced Nursing*, 62(1), 107–115.
- Fearon, D., Gunia, B., Pralle, B. E., Lake, S., & Sallans, A. L. (2013). [Research data management services](#). SPEC Kit 334. Washington, DC: Association of Research Libraries.
- Goldman, J., Kafel, D., & Martin E. R. (2015). [Assessment of data management services at New England region resource libraries](#). *Journal of eScience Librarianship*, 4(1), e1068.
- Government of Canada. (2016, December 21). [Tri-Agency Statement of Principles on Digital Data Management](#).
- Hsieh, H., & Shannon, S. E. (2005). [Three approaches to qualitative content analysis](#). *Qualitative Health Research*, 15(9), 1277–1288.
- Ioannidis, J. P., Greenland, S., Hlatky, M. A., Khoury, M. J., Macleod, M. R., Moher, D., Schulz, K. F., & Tibshirani, R. (2014). [Increasing value and reducing waste in research design, conduct, and analysis](#). *Lancet*, 383(9912), 166–175.
- Kitzinger, J. (1995). [Qualitative research: Introducing focus groups](#). *BMJ*, 311(7000), 299–302.
- Krueger, R. A. (1998). *Developing questions for focus groups*. Thousand Oaks, CA: Sage Publications.
- Lage, K., Losoff, B., & Maness, J. (2011). [Receptivity to library involvement in scientific data curation: A case study at the University of Colorado Boulder](#). *Libraries and the Academy*, 11(4), 915–937.
- Lincoln, Y., & Guba, E. (1985). *Naturalistic inquiry*. Newbury Park, CA: Sage Publications.
- Mays, N., & Pope, C. (2000). [Qualitative research in health care: Assessing quality in qualitative research](#). *BMJ*, 320(7226), 50–52.
- Miles, M. B., & Huberman, M. (1994). *Qualitative data analysis: An expanded sourcebook*. Thousand Oaks, CA: Sage Publications.
- Morgan, D. L. (1998). *The focus group guidebook*. Thousand Oaks, CA: Sage Publishing.

- Parham, S. W., Bodnar, J., & Fuchs, S. (2012). [Supporting tomorrow's research: Assessing faculty data curation needs at Georgia Tech](#). *College & Research Libraries News*, 73, 10–13.
- Patton, M. Q. (2002). *Qualitative research and evaluation methods*. Thousand Oaks, CA: Sage Publications.
- Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. Thousand Oaks, CA: Sage Publications.
- Peters, C. & Dryden, A. R. (2011). [Assessing the academic library's role in campus-wide research data management: A first step at the University of Houston](#). *Science & Technology Libraries*, 30(4), 387–403.
- Pinfield, S., Cox, A., & Smith, J. (2014). [Research data management and libraries: Relationships, activities, drivers and influences](#). *PLoS ONE*, 9(12), e114734.
- Richardson, J., Nolan-Brown, T., Loria, P. & Bradbury, S. J. (2012). [Library research support in Queensland: A survey](#). *Australian Academic & Research Libraries*, 43(4): 258–277.
- Scaramozzino, J. M., Ramirez, M. L., & McGaughey, K. J. (2012). [A study of faculty data curation behaviors and attitudes at a teaching-centered university](#). *College and Research Libraries*, 73(4): 349–365.
- Sewerin, C., Barsky, E., Dearborn, D., Henshilwood, A., Hwang, C., Keys, S., Mitchell, M., Spence, M., Szigeti, K., & Zaraiskay, T. (2016a, June). [From coast to coast: Canadian Collaboration in a changing RDM seascape](#). Paper presented at the 37th Annual IATUL Conference, Halifax, NS.
- Sewerin, C., Barsky, E., Dearborn, D., Henshilwood, A., Hwang, C., Keys, S., Mitchell, M., Spence, M., Szigeti, K., & Zaraiskaya, T. (2016b, June 2). [Research data management \(RDM\) needs of science and engineering researchers: A view from Canada](#).
- Shearer, K. (2015, April 7). [Comprehensive brief on research data management policies](#).
- Soehner, C., Steeves, C., & Ward, J. (2010, June). [E-science and data support services: A survey of ARL members](#). Paper presented at the 31st Annual IATUL Conference, West Lafayette, IN.
- Steinhart, G., Chen, E., Arguillas, F., Dietrich, D., & Kramer, S. (2012). [Prepared to plan? A snapshot of researcher readiness to address data management planning requirements](#). *Journal of eScience Librarianship*, 1(2), e1008.

Tenopir, C., Sandusky, R. J., Allard, S. & Birch, B. (2013). [Academic librarians and research data services: Preparation and attitudes](#). *International Federation of Library Associations and Institutions Journal (IFLA Journal)*, 39(1), 70–78.

University of Toronto (n.d.). [Quick Facts](#).

Appendix A. Rating Exercise

Below are a number of possible services or tools related to research data management. Please read each one and indicate to what extent you would find each item valuable.

| Managing Data | |
|---|--|
| Data management plans e.g. Offering templates and/or examples | |
| Organizing files e.g. Examples of how to create a system for quickly & reliably accessing your research records | |
| Documenting data (creating metadata) e.g. Guidelines on metadata standards | |
| Safe storage of data e.g. Identifying storage and back-up options for keeping your data secure | |
| Preserve data for long-term access e.g. Recommendations on file formats for long term readability and access | |
| Identify data repositories for data sharing e.g. Listing repositories by discipline | |
| Research Support Services | |
| Funding agency requirements e.g. Listing of major funding agency data management requirements | |
| Journal requirements e.g. Help interpreting journal data sharing policies | |
| Consent forms e.g. Information required for consent forms related to data sharing | |
| Confidentiality e.g. University of Toronto regulations and considerations related to research data | |
| Citing data sets e.g. Ensuring credit is given when data is cited | |
| How to find data from a repository e.g. Issues around accessing and using data | |
| Intellectual property e.g. University of Toronto regulations about ownership of research outputs | |
| Your Suggestions | |
| Other (please describe your suggestion): | |
| | |
| Other (please describe your suggestion): | |
| | |

Appendix B. Interview Guide

| |
|---|
| Welcome and introductions |
| <ul style="list-style-type: none"> ▪ Thank for participation ▪ Introduction: self and note-taker |
| Our topic is..... |
| <p>There is growing movement in the research community, both globally and nationally, with funding agencies and journals moving towards requirements that include writing research data management plans and sharing data in public repositories.</p> <p>Purpose of focus groups: To determine the essential tools and services required by researchers at UToronto for data management in fulfilling emerging funding agency and journal requirements</p> <p>The results will be used for.... The development of services and tools for researchers at the University of Toronto that are essential in doing things such,</p> <ul style="list-style-type: none"> • writing data management plans for grant applications • the documentation of data sets • organizing / managing data • preparing data for sharing in repositories or long-term access <p>What are we trying to accomplish?</p> <ul style="list-style-type: none"> • We are at the initial stages of development and want to hear your ideas and concepts <p>'Go around' – first name, research methods being used in a current study</p> |
| Guidelines |
| <p>Goal: To hear from participants based on their diverse experiences and viewpoints. The goal is not to come to any consensus, rather to hear a range of opinions.</p> <p>No right or wrong answers, only differing points of view</p> <p>Feel free to direct answers and comments towards each other, not just me.</p> <ul style="list-style-type: none"> ▪ Reminder of confidentiality and the basics of how the focus group will run. ▪ Remind participants that we are recording but no one will be identified in the final feedback report AND all information discussed is confidential. ▪ Ask that cell phones are turned off, not used. <p>The data for this project = everything discussed in the room.</p> <p>My role as moderator will be to guide the discussion.</p> |
| Discussion |
| <p>Organizing Your Research Data:</p> <p>Thinking back to a research project where you had to manage and track a reasonable amount of data,</p> |

| |
|---|
| <ol style="list-style-type: none"> 1. What strategies did you use to manage your data while you were conducting that study? [Prompts: team size, documentation, managing files, naming files] 2. When it comes to managing your research data, what provides challenges for you? What has contributed to making this a difficult process? 3. What facilitated or helped to make this an easier process? What worked well? |
| <ol style="list-style-type: none"> 4. What could the University of Toronto offer that would help you with organizing your data? (What things do you need the most help with? What kinds of supports or services could be offered that would be most useful?) 5. Can you describe any examples? |
| <p>Preserving Your Research Data: This involves keeping your data accessible and useful over the long-term (for yourself or for sharing in repositories). For example, saving files in non-proprietary formats such as using TXT files instead of Microsoft Word files.</p> |
| <p>Thinking about preserving your research data over the long term (say a study you did 10 or more years ago),</p> <ol style="list-style-type: none"> 1) What has helped you or what do you think would help you prepare your data so it could be accessed and used over the long term? 2) Can you describe any examples? |
| <ol style="list-style-type: none"> 3) Is there anything that the University of Toronto could offer that would help you with preserving your data? 4) Can you describe any examples? |
| <p>Sharing Your Research Data: As an example, some journals are requiring that you put your raw data into a public data repository before you can publish with them. A public data repository provides long term preservation and sharing of data.</p> |
| <ul style="list-style-type: none"> • What has helped you or what do you think would help prepare your data to be put into a public data repository? • Can you describe any examples? |
| <ul style="list-style-type: none"> • What could the University of Toronto offer that would help you with preparing your data for sharing? • Can you describe any examples? |
| <p>Data Management Plans: A data management plan is a formal document that outlines how you will handle your data during your research and after the project is completed. This is something that funding agencies are moving towards asking for in the future.</p> |
| <ol style="list-style-type: none"> 1. What has helped you or what do you think would help you prepare a data management plan? 2. Can you describe any examples? |

| |
|---|
| <p>3. Is there anything the University of Toronto could offer that would be helpful in developing a data management plan for your research projects?</p> <p>4. Can you describe any examples?</p> |
| Rating Exercise |
| <p>Presentation of listing of items:</p> <p>This is a listing of potential services or tools related to research data management that we've put together.</p> <p>We're going to give you 5 minutes to rate them on a scale from 'very useful' to 'not useful'. Please feel free to add addition features.</p> |
| Wrap-Up and Summary |
| <p>What is the best way to communicate to researchers about tools, services, and resources related to research data management at the University of Toronto?</p> <p>If you wanted information on any of the topics we talked about, how would you prefer to get this information?</p> |
| <p>Thinking of all the topics we discussed today: Suppose you had one minute to describe the most helpful things that could be offered to you related to managing your research data, what would you say?</p> <p>(What advice can you provide UToronto in creating resources that would work best for you and your colleagues related to managing your research data?)</p> |
| Have we missed anything? |
| Wrap-up |
| <p>Thank everyone for their participation.</p> |

Appendix C. Codes for Focus Groups

1. Infrastructure

(physical/technical and organizational structures and supports)

- a. data storage
- b. data security
- c. accessibility of data, e.g., ease of use of an online interface
- d. using, identifying, or accessing repositories
- e. policies/procedures
- f. organizational processes/arrangements
- g. communication

2. Content

- a. standards, e.g. metadata
- b. statistical information on publication and use, e.g., bibliometrics to track use
- c. ease / difficulty of finding data
- d. nature of the data, e.g., data life cycle: raw, processed, published
- e. value / importance of the data
- f. quality of data
- g. usefulness of data
- h. documentation

3. Capacity

- a. time
- b. money
- c. human resources
- d. personal comfort, e.g., with sharing data
- e. impact on career
- f. incentives, motivation
- g. knowledge