Research Data Management in Canada

Perceptions and Practices Across the Disciplines

Presented by Melissa Cheung, University of Ottawa and Alexandra Cooper, Queen’s University
21 July 2020
Background: The Survey Project

2015
Science and Engineering Survey

2016
Humanities and Social Sciences Survey

2017
Health Sciences and Medicine Survey

2018
Canadian RDM Survey Consortium

2019
National Dataset

2018
Consortium reaches 14 institutions

2020
Analysis of National Dataset
The Survey

- Common survey instrument consisted of 4 sections with questions that reflect the data lifecycle
- Each institution ran surveys individually using their own survey software
  - Customized the survey based on institutional characteristics and interest
  - Obtained ethics approval from their own research ethics board
Merging datasets

- Recode institutional datasets to match standard codebook
  - Regroup rank, funding variables
  - Regroup faculty/department into generic list of “field of study”
  - Remove text responses from analysis
Who participated?
Survey Results
Grad students made up 32.5% (n=778) of the total survey respondents and 20.5%-43.7% in each disciplinary group included in the survey.
Funding sources
Current RDM Practices

(reported at the time of the survey)

How are researchers working with and managing their data?
Estimated data storage size for a typical project

- < 50GB (Gigabyte): 55.1%
- 50GB to < 500GB: 11.6%
- 500GB to < 1000GB: 6.3%
- >1TB (Terabyte): 7.9%
- Not sure: 15%
- Not applicable: 4.1%
Where are researchers storing their data?

Note: respondents could select all that applied
Data kept after project

- I only keep data for the length of the project: 9.5% Source/Survey/Raw Data, 15% Intermediate/Working Data, 39.9% Processed data ready for publication
- Less than 3 years: 9.2% Source/Survey/Raw Data, 14.2% Intermediate/Working Data, 20.5% Processed data ready for publication
- Between 3-5 years: 11.5% Source/Survey/Raw Data, 19.5% Intermediate/Working Data, 20.7% Processed data ready for publication
- Between 5-10 years: 11.4% Source/Survey/Raw Data, 20.5% Intermediate/Working Data, 34.4% Processed data ready for publication
- More than 10 years: 16.1% Source/Survey/Raw Data, 19.5% Intermediate/Working Data, 44.5% Processed data ready for publication
- Until the data becomes inaccessible or lost: 34.4% Source/Survey/Raw Data, 44.5% Intermediate/Working Data, 16.1% Processed data ready for publication

Percentage (%) of respondents

- Percentage (%) of respondents

- Length of time data are kept

- Source/Survey/Raw Data

- Intermediate/Working Data

- Processed data ready for publication
Researcher Readiness for Tri-Agency RDM Policy

Would researchers be able to meet Tri-Agency RDM policy requirements for data management plans (DMPs) and to deposit data in digital repositories?
Audience poll: We asked researchers if they would be able to draft a DMP as part of a grant application. Guess which of the following statements most researchers said best describes their situation.

A) They do not need help drafting a DMP.
B) They prefer having help and/or guided documentation.
C) They need help and/or guided documentation to draft a DMP.
What if researchers were asked to draft a Data Management Plan (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 my application

I would need assistance and/or guided documentation to appropriately address some or all of the sections

No help: 14%
Prefer help: 35.7%
Need help: 50.3%

Percentage (%) of respondents
Data Sharing

- Other: 4.3% (Current), 3.4% (Future)
- Institutional repository: 4.9% (Current), 20.2% (Future)
- Discipline-specific repository: 7.9% (Current), 14.2% (Future)
- Institutional/personal website: 10.3% (Current), 25.2% (Future)
- Supplementary materials files to a journal publisher: 13.5% (Current), 26.8% (Future)
- Online with restricted access: 13.5% (Current), 30.3% (Future)
- Not sharing/Not planning to share: 28.4% (Current), 13.7% (Future)
- Personal request only: 41.8% (Current), 46% (Future)
Restrictions or embargoes on data sharing

- Public safety/sensitive nature: 3.3%
- Commercial concerns: 4.5%
- Other: 5%
- Plan to file patent: 5.1%
- Not sure: 9.7%
- Contractual obligations to third party: 10.6%
- Intellectual property rights: 12.7%
- No restrictions: 21.8%
- Privacy or ethical restrictions: 30.2%
- Need to publish before sharing: 30.4%
Restrictions or embargoes on data sharing

● Arts/Humanities and Science more likely to report no restrictions on data sharing
● Engineering and Science most likely to report need to publish first
● All other disciplines cited privacy, confidentiality, or ethics restrictions
● Arts/Humanities divided between need to publish first and privacy, confidentiality, or ethics restrictions
Audience poll: What do you think is the most common reason why researchers would not be willing to share their data?

A) Researchers believe their data should not be shared
B) The data are incomplete or not finished
C) Researchers still wish to derive value from the data
D) Funding body does not require sharing
Reasons for not sharing research data and associated methods/tools

1. Did not know could share: 3%
2. Not required by funding body: 4.6%
3. Lack technical skills: 5.2%
4. Not useful to others: 6%
5. Should not be shared: 6.3%
6. Reason not share: Other: 6.7%
7. No place to put data: 8.1%
8. Lack of Funding: 10.8%
9. Lack of standards: 12.8%
10. Do not hold rights: 16%
11. Insufficient time: 16.2%
12. Concerns about citation: 22%
13. Willing to share: 22.3%
14. Privacy/legal/security reasons: 22.6%
15. Want to derive value: 25.4%
16. Data are incomplete: 32.4%

Percentage (%) of respondents
Would researchers be able to meet Tri-Agency RDM policy requirements for data management plans (DMPs) and to deposit data in digital repositories?

Answer: NOPE

(Not based on their RDM practices as reported, anyway)
RDM Services

What supports do researchers want or need to create DMPs for grant applications?

What supports do researchers want or need to deposit their data?
Interest in RDM services

- Faculty or Graduate Workshop: 83.6% Interested, 10% Not Interested, 6.4% Not Applicable
- Assistance with DMP preparation: 80.2% Interested, 12.1% Not Interested, 7.7% Not Applicable
- Communication about funding/journal requirements: 79.6% Interested, 13.5% Not Interested, 6.9% Not Applicable
- Institutional repository: 76.2% Interested, 17.1% Not Interested, 6.7% Not Applicable
- Personalized consultations: 74.3% Interested, 18% Not Interested, 7.7% Not Applicable
- Data storage during active projects: 73.5% Interested, 19.9% Not Interested, 6.5% Not Applicable
- Assistance with preservation/sharing: 72.8% Interested, 19.6% Not Interested, 7.5% Not Applicable
- Finding/accessing data sources: 70.5% Interested, 21.7% Not Interested, 7.9% Not Applicable
- Assistance with metadata creation: 65.5% Interested, 24.5% Not Interested, 10% Not Applicable
- External Repository: 65.3% Interested, 26.3% Not Interested, 8.4% Not Applicable
- Permanent identifiers/DOIs: 62.3% Interested, 27.2% Not Interested, 10.5% Not Applicable
- Digitization of physical records: 55.3% Interested, 31.5% Not Interested, 13.2% Not Applicable
## Interest in services: differences by discipline

<table>
<thead>
<tr>
<th>Field of study</th>
<th>Most popular RDM services</th>
<th>Interested</th>
</tr>
</thead>
<tbody>
<tr>
<td>Engineering</td>
<td>Assistance with DMP preparation</td>
<td>82.52%</td>
</tr>
<tr>
<td>Arts/Humanities</td>
<td>Assistance with DMP preparation</td>
<td>79.88%</td>
</tr>
<tr>
<td>Business/Management, Education, Law</td>
<td>Assistance with DMP preparation</td>
<td>84.38%</td>
</tr>
<tr>
<td>Social Sciences</td>
<td>Assistance with DMP preparation</td>
<td>80.83%</td>
</tr>
<tr>
<td>Medicine/Preclinical Sciences</td>
<td>Communication/info about funding/journal requirements</td>
<td>80.18%</td>
</tr>
<tr>
<td>Health Science</td>
<td>Communication/info about funding/journal requirements</td>
<td>91.51%</td>
</tr>
<tr>
<td>Science</td>
<td>Institutional repository</td>
<td>77.36%</td>
</tr>
<tr>
<td>Interdisciplinary/Other</td>
<td>Institutional repository</td>
<td>84.38%</td>
</tr>
</tbody>
</table>
Some Implications

How can the survey results help us with developing RDM services?
Gaps between practices and policy requirements

- Clarification on policies related to data sharing and deposit in digital repositories
  - Privacy, confidentiality and ethics
  - Funding and journal policies
  - Which data should be retained and/or shared

- High demand for assistance with DMPs and institutional repositories for data preservation
But wait, there’s more!

- Plans to publish our findings
- National Dataset will be made publicly available
- Learn more about the project on the [Canadian RDM Survey Consortium’s page](#) of Portage’s website
Thank you!

Questions?