# 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



### 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

Variable name	FUNCIHR			
Variable label	Funding source: CIHR			
Variable description	Funding source respondent has used in last 5 years or intends to use in next 5 years: CIHR			
Survey question	Which funding sources have you used within the past 5 years, or are planning to apply for in the next 5 years? Please exclude funding earmarked exclusively for operations and infrastructure. Select all that apply:			
Value label	Value description Cou	unt Percen		
0	Not Chosen 21	64 81.4%		
1	Chosen 45	55 17.1%		
99	Not Declared 1	3 0.5%		
999	Not asked by administering institution 24	6 1.0%		
Variable name	FUNSSHRC			
Variable label	Funding source: SSHRC			
Variable description	Funding source respondent has used in last 5 years or intends to use in next 5 years	ears: SSHRC		
Survey question	Which funding sources have you used within the past 5 years, or are planning to the next 5 years? Please exclude funding earmarked exclusively for operations a infrastructure. Select all that apply:			
Value label		unt Percen		
-		36 72.8%		
		83 25.7%		
99	Not Declared 1	-		
999	Not asked by administering institution 2	6 1.0%		
Variable name	FUNCFI			
Variable label	Funding source: CFI			
Variable description Survey question	Which funding sources have you used within the past 5 years, or are planning to apply for in the next 5 years? Please exclude funding earmarked exclusively for operations and			
Value label	infrastructure. Select all that apply: Value description Con	unt Percen		
	Not Chosen 22			
		27 8.5%		
		1 0.0%		
	Hot becaled			
	Not asked by administering institution 21	19 8.2%		
99 999 Variable name	Not asked by administering institution 21 FUNNSERC	19 8.2%		
99 999		19 8.2%		
99 999 Variable name	FUNNSERC			
99 999 Variable name Variable label Variable description Survey question	FUNNSERC Funding source: NSERC	ears: NSERC o apply for in		
99 999 Variable name Variable label Variable description	FUNNSERC Funding source: NSERC Funding source respondent has used in last 5 years or intends to use in next 5 y Which funding sources have you used within the past 5 years, or are planning to the next 5 years? Please exclude funding earmarked exclusively for operations a	ears: NSERC o apply for in and		
99 999 Variable name Variable label Variable description Survey question Value label	FUNNSERC Funding source: NSERC Funding source respondent has used in last 5 years or intends to use in next 5 ye Which funding source respondent has used within the past 5 years, or are planning to the next 5 years? Please exclude funding earmarked exclusively for operations a infrastructure. Select all that apply.	ears: NSERC o apply for in and unt Percen		
99 999 Variable name Variable label Variable description Survey question Value label 0	FUNNSERC Funding source: NSERC Funding source: respondent has used in last 5 years or intends to use in next 5 ye Which funding sources have you used within the past 5 years, or are planning to the next 5 years? Please exclude funding earmarked exclusively for operations a infrastructure. Select all that apply: Value description Not Chosen 19	ears: NSERC o apply for in and unt Percen		
99 999 Variable name Variable label Variable description Survey question Value label 0 1	FUNNSERC Funding source: NSERC Funding source: respondent has used in last 5 years or intends to use in next 5 ye Which funding sources have you used within the past 5 years, or are planning to the next 5 years? Please exclude funding earmarked exclusively for operations a infrastructure. Select all that apply: Value description Not Chosen 19	ears: NSERC o apply for in and unt Percen 87 74.8% 32 23.8%		
99 999 Variable name Variable label Variable description Survey question Value label 0 1 9 9	FUNNSERC Funding source: NSERC Funding source: NSERC Funding source: respondent has used in last 5 years or intends to use in next 5 years, or are planning to the next 5 years? Please exclude funding earmarked exclusively for operations a infrastructure. Select all that apply. Value description Not Chosen 19 Ghosen 63	ears: NSERC o apply for in and unt Percen 187 74.8% 32 23.8% 3 0.5%		

Northwest Territories

Yukon

Alberta

Manitoba

Nunavut

5-0

3

Ontario

#### Who participated?

Sha Columb

Quebec

N.L.

N.B.P.E.I.

## Survey Results

#### Respondent Field of Study



#### 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



#### Where are researchers storing their data?



#### Data kept after project



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



#### Data Sharing



#### Restrictions or embargoes on data sharing



### 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

Did not know could share 3 Not required by funding body 4.6 Lack technical skills 5.2 Not useful to others 6 Should not be shared 6.3 Reason not share: Other 6.7 No place to put data 8.1 Lack of Funding 10.8 Lack of standards 12.8 Do not hold rights 16 Insufficient time 16.2 **Concerns** about citation 22 Willing to share Privacy/legal/security reasons Want to derive value Data are incomplete

0



# Back to our research question

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)

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

## **RDM Services**

What supports do researchers want or need to deposit their data?

#### Interest in RDM services

Faculty or Graduate Workshop Assistance with DMP preparation Communication about funding/journal requirements Institutional repository Personalized consultation: Data storage during active projects Assistance with preservation/sharing Finding/accessing data sources Assistance with metadata creation **External Repository** Permanent identifiers/DOIs Digitization of physical records

0%	20% 40 <sup>9</sup> Percer	% 60% ntage (%) of respondents	80%	1009
s	55-3		31.5	13.2
is line	62.3		27.2	10.5
у	65.3		26.3	8.4
n 📔	65.5		24.5	10
s	70.5		21.7	7.9
g 📕	72.8		19.6	7.5
s IIII	73.5		19.9	6.5
IS STATE	74.3		18	7.7
у	76.2		17.1	6.7
s l	79-	6	13.5	6.9
n [	80.	.2	12.1	1 7.7
р	8	3.6	1	l <b>o</b> 6.4

#### Interest in services: differences by discipline

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

## 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 <u>Consortium's page</u> of Portage's website



# Thank you! Questions?