LEARNING HOW TO SUPPORT RDM ACROSS CANADA: 
A MULTI-INSTITUTIONAL SURVEY OF HEALTH AND MEDICAL RESEARCHERS' RDM PRACTICES AND ATTITUDES

Francine Berish¹  
Melissa Cheung⁵  
Alexandra Cooper¹  
Dylanne Dearborn⁶  
K-Lee Fraser³  
Matthew Gertler²  
Vincent Gray⁷  
Mayu Ishida⁴  
Marjorie Mitchell⁴  
May Yan²  

Queen’s University¹  
Ryerson University²  
University of Alberta³  
University of British Columbia⁴  
University of Ottawa⁵  
University of Toronto⁶  
Western University⁷
The changing RDM landscape (the short version)

- The Canadian landscape
  - CIHR, NSERC, SSHRC
  - Parallels to Europe/UK
- Tri-Agencies adopt Statement of Principles on Digital Data Management
- Portage Network
Establishment of Portage (CARL initiative)

2015

Tri-Agency Policy on Open Access Publications

DMP Assistant Launched

2016

RDM Survey Clearinghouse created

Portage RDM Survey

2017

Tri-Agency Statement of Principles on Data Management

2018

Consortium reaches 14
Motivation for survey(s)

- Understand our researchers’ practices and needs
- Plan services to support research data management
- Identify differences across disciplines
- Collaborative creation of a national dataset
Evolution of the survey project

- 2015 - engineering and science version was created
- 2016 - modified for social sciences and humanities
- 2017 - modified for health and medical sciences
- Translation support provided by CIHR and SSHRC
- Discussions/result sharing with CIHR, SSHRC and NSERC
Benefits of consortia - Logistics

- Collaborative nature - many hands make light work
- Shared documents to work from
  - e.g. survey instruments, survey question exports, informed consent, ethics documents, codebooks, data merging instructions
- Leveraging knowledge and expertise of the group
- Increase in credibility for survey efforts

www.digitalbevaring.dk
Benefits of consortia - Results

- More data!
- Greater understanding of disciplinary trends
- Can compare institutional differences
- National snapshot
- Can be used by other groups interested in RDM in Canada
Collaborative & inclusive: mechanisms

- Operate cohesively as a group - decisions are made together

- Mechanisms for collaboration
  - Group calls for updates, decisions and results dissemination
  - Google Drive space for working documents and resources
  - Listserv for discussion questions and new ideas
Collaborative & inclusive: administration, outreach

- **Administration**
  - Identified contacts for new schools to assist with conducting the survey, applying for ethics approval, and data wrangling
  - Institutional contacts list and institutional progress spreadsheet
  - Clearinghouse space on Portage website (managed by Research Intelligence Expert Group)

- **Outreach to other institutions**
  - Finalizing instructions
  - Will do an outreach push when documents available on the clearinghouse
  - New institutions are still joining in the meantime
Working Across Time Zones

- Pacific Time Zone
- Mountain Time Zone
- Central Time Zone
- Eastern Time Zone
- Atlantic Time Zone
- Newfoundland Time Zone
Methods
Data privacy & survey instrument

● Each institution ran this survey individually on different survey software
  ○ Customized the survey based on their on local interest and institutional differences
    ■ Differences include Program names, Ranks, etc.
    ■ Mapping departments (NLH categories) & exceptions (researcher degrees of freedom)

● Different survey software used: Fluidsurveys, Qualtrics, SurveyMonkey
  ○ Privacy restrictions varies by province (British Columbia on strict end)
  ○ Institutional considerations varied
Merging datasets

- Edit institutional dataset to match codebook
- Regroup rank, funding variables
- Regroup faculty/department into generic list:
  - Preclinical Sciences
  - Medicine
  - Health Sciences
  - Interdisciplinary/Other
- Remove text responses from analysis

<table>
<thead>
<tr>
<th>Codebook</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Variable Name</strong></td>
</tr>
<tr>
<td><strong>Variable Label</strong></td>
</tr>
<tr>
<td><strong>Variable Description</strong></td>
</tr>
<tr>
<td><strong>Value Label</strong></td>
</tr>
<tr>
<td><strong>Value Description</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Variable Label</th>
<th>Variable Description</th>
<th>Value Label</th>
<th>Value Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTID</td>
<td>ID assigned by administering institution</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LANG</td>
<td>Language in which survey was completed</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RANK</td>
<td>C2. Rank of respondent</td>
<td>The respondent’s rank at his/her institution. For consistency, each survey response was recoded to a controlled vocabulary.</td>
<td>1</td>
<td>English</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>French</td>
</tr>
<tr>
<td>FIELD</td>
<td>C2. Respondent’s field of study</td>
<td>The field of research or study of the respondent. Institutions asked for specific fields of study, which were mapped to broader fields for consistency.</td>
<td>1</td>
<td>Preclinical Sciences</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>Medicine</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>Health Sciences</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>Interdisciplinary/Other</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td>Not declared</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>6</td>
<td>Assistant/Associate/Full Professor/Clinical Colleague</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>7</td>
<td>Professor Emeritus</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>8</td>
<td>Other</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>9</td>
<td>Not Declared</td>
</tr>
</tbody>
</table>
Survey questions reflect research data lifecycle:

- Data practices throughout an active project
- Practices and perceptions around sharing and archiving data
- What services and support would researchers like to receive related to data management
Preliminary Results
Survey responses by institution

- Ryerson University: 23 respondents
- University of Ottawa: 33 respondents
- University of Alberta: 62 respondents
- Western University: 66 respondents
- University of Toronto: 67 respondents
- University of British Columbia: 89 respondents
- McGill University: 146 respondents
- Queen's University: 152 respondents

Number of respondents (N=638)
Respondents by field and rank

- Lecturer/Professor/Clinical Colleague:
  - Medicine: 197
  - Health Sciences: 81
  - Preclinical Sciences: 22
- Graduate Student:
  - Medicine: 52
  - Health Sciences: 35
  - Preclinical Sciences: 12
- Postdoctoral Fellow:
  - Medicine: 10
  - Health Sciences: 25
  - Preclinical Sciences: 2
- Other:
  - Medicine: 6
  - Health Sciences: 8
  - Preclinical Sciences: 19

N=607
Research Data Management Planning
Able to draft a Data Management Plan without assistance

11%
Ability to draft data management plan (DMP)

89% (n=513)
Need assistance or would like to have assistance some form of assistance.

11% (n=63)
Would be able to draft a DMP
Ability to draft data management plan (DMP)

- **Lecturer/Professor/Adjunct Professor/Clinical Colleague**
  - Would be able to draft a DMP: 32.5%
  - Would be able to draft a DMP but would prefer to have assistance and/or documentation: 24.3%
  - Need assistance and/or documentation to write DMP: 8.5%

- **Graduate Student**
  - Would be able to draft a DMP: 10.2%
  - Would be able to draft a DMP but would prefer to have assistance and/or documentation: 8.9%

- **Postdoctoral Fellow**
  - Would be able to draft a DMP: 5.4%
  - Would be able to draft a DMP but would prefer to have assistance and/or documentation: 3.6%

- **Other**
  - Would be able to draft a DMP: 1.9%
  - Would be able to draft a DMP but would prefer to have assistance and/or documentation: 2.3%
CURRENT DATA MANAGEMENT PRACTICES
Amount of Data Created

Not sure 20%

(N=572) (n=115)
Amount of Data Created

< 1GB
22.6%

(n=129)
Amount of Data Created

< 10GB
42.3%
(n=242)
Amount of Data Created

< 50GB
51.2%

(n=357)
Amount of Data Created

> 50GB
28.7%

(n=167)
# Data storage

<table>
<thead>
<tr>
<th>Large Number:</th>
<th>Computer Hard Drive</th>
<th>Laptop Hard Drive</th>
<th>Flash Drive</th>
<th>External Hard Drive</th>
<th>Cloud/web based solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Some:</td>
<td>Cloud/Web based</td>
<td>Shared Drive/Servers</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Few:</td>
<td>Repository</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Number of data storage locations

Selected Weaker Locations such as Hard Drives
Limited ability to share & interest in institutional repositories
Reasons for not sharing

- Other: 4.1%
- Not useful to others: 4.1%
- Not required by funding body: 5.2%
- Lack technical skills: 5.8%
- Should not be shared: 5.9%
- Did not know could share: 5.9%
- No place to put data: 11.9%
- Lack of standards: 14.6%
- Lack of Funding: 15.4%
- Insufficient time: 17.2%
- Do not hold rights: 17.2%
- Want to derive value: 26.1%
- Willing to share: 27.3%
- Concerns about citation/acknowledgement: 28.3%
- Data are incomplete: 29%
- Privacy/legal/security reasons: 34.6%
Current and future sharing methods

- **Other**:
  - Current: 3.7%
  - Future: 5.5%

- **Discipline-specific repository**:
  - Current: 13.6%
  - Future: 15.3%

- **Not sharing/not planning to share**:
  - Current: 33.8%
  - Future: 6.7%

- **Institutional repository**:
  - Current: 22.1%
  - Future: 6.3%

- **Institutional/personal website**:
  - Current: 24.3%
  - Future: 7.4%

- **Supplementary materials to journal**:
  - Current: 30.8%
  - Future: 18.1%

- **Online with restricted access**:
  - Current: 34.5%
  - Future: 14.5%

- **Personal request only**:
  - Current: 44.4%
  - Future: 45.8%
Data sharing benefits

- Collaborative scholarship: 64.4%
- Interdisciplinary research: 55.4%
- Train new researchers: 55.0%
- Moves field of research forward: 54.8%
- Supports open access knowledge: 52.2%
- Re-analysis helps verify data: 49.1%
- Safeguards against falsification: 47.2%
- Increase research impact: 46.7%
- Reduces redundant data collection: 45.5%
- Retains data integrity: 42.6%
- None: 5.3%
- Other: 2.2%
Planning services for our researchers

- Digitization of physical records: 70.3%
- Permanent identifiers/DOIs: 74.5%
- External Repository: 78.3%
- Assistance with metadata creation: 81.4%
- Data storage during active projects: 82.3%
- Finding/accessing data sources: 82.5%
- Assistance with preservation/sharing: 83.3%
- Institutional repository: 83.4%
- Personalized consultations: 83.9%
- Graduate student workshop: 85.3%
- Faculty workshop: 86.7%
- Assistance with DMP preparation: 89.3%
- Communication - funding/journal requirements: 89.5%
Conclusions
69% (n=441) of respondents received funding from at least one of the Federal agencies.

Funder Used

- 63% (n=397) CIHR
- 13% (n=76) SSHRC
- 23% (n=141) NSERC
Canadian Federal Tri-Agencies - ability to draft DMP

<table>
<thead>
<tr>
<th>Agency</th>
<th>Would not need help</th>
<th>Would need or want help</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIHR</td>
<td>11.3%</td>
<td>87.2%</td>
</tr>
<tr>
<td>SSHRC</td>
<td>3.9%</td>
<td>96.1%</td>
</tr>
<tr>
<td>NSERC</td>
<td>14.2%</td>
<td>85.8%</td>
</tr>
</tbody>
</table>
Considerations when preparing RDM service

- There are diverse data types, sizes and practices when looking at data management throughout an active project
- There is a willingness to share data
- There is a high level of interest in services around RDM
Next steps

- Portage Clearinghouse:
  - Data amalgamation documentation to combine datasets to create national dataset
- Sharing results with Tri-Agencies (CIHR/SSHRC/NSERC)
- Expand consortium membership
References


Slide 5, 14, 40, 43 - Digital Bevaring images - https://digitalbevaring.dk/om-sitet/about-us/

Canadian RDM Survey Consortium
Dylanne Dearborn, University of Toronto, Chair