Research Data Management at the University of Alberta:
Preparing for new funder policies

James Doiron
Research Data Management Services Coordinator
University of Alberta Libraries

UofA Research Funding Fair
November 22, 2018
My background

- RDM Services Coordinator, UofA Libraries
- Educational background in social sciences (Psychology & Criminology)
- Cross disciplinary & multi-institutional applied research data management experience

RDM Initiatives:

- CARL Portage Network: Data Management Planning & RDM Training Expert Groups; Dataverse North Working Group; Dataverse Training, Working Group (Chair)
- Statistic Canada Data Liberation Initiative (DLI) External Advisory Committee
- Confederation of Open Access Repositories (COAR) RDM Interest Group
What is Research Data Management?

RDM = managing data throughout all phases of the research lifecycle…through active phases and beyond

RDM Incentives

‘Carrots’

★ Competitive funding advantage
★ Increase efficiencies
★ Quality assurance → data validity
★ Increase likelihood of data preservation
★ Increases research visibility & impact
★ Facilitate data discovery & repurposing
★ Data citations
★ Increase opportunities for collaboration

‘Sticks’

➢ Helps ensure compliance w/ funding agencies’ policies
➢ Risk management → minimizes security & confidentiality risks
➢ Helps to meet institutional requirements
The Canadian Research Data Management landscape...at times it can feel like a maze!

- Metadata & Documentation
- Funder Requirements
- Data Management Planning
- Data Security
- Data Access
- Data Collection
- Data Preservation
- Data Confidentiality
- Institutional Requirements
- Data Repurposing
- Data Quality

(change ahead)

(never came from comfort zones)

(be prepared)
Tri-Agency Statement - 2016

- In June 2016, The Tri-Agencies (SSHRC, CIHR, NSERC) released a Statement of Principles on Digital Data Management
- Statement aligns with other national endeavours relating to RDM, including the Government of Canada’s Action Plan on Open Government (2014)

The statement outlines:

1. **Expectations** → DMP’s, standards, collection/storage, metadata, preservation, retention, and sharing
2. **The responsibilities of all stakeholders** to meet those expectations
Data Management Planning:

“...is necessary at all stages of the research project lifecycle, from design and inception to completion.”

“...should be developed using standardized tools”

Metadata:

“All research data should be accompanied by metadata”

Preservation:

“Research data resulting from agency funding should be preserved in a publicly accessible, secure and curated repository or other platform for discovery by others”
Tri-Agency RDM Policy (Draft) - 2018

- In the Spring of 2018, the Tri-Agencies released a draft of the “Tri-Agency Research Data Management Policy”

- Incremental implementation of a final policy is expected to begin in later 2019

The draft policy includes suggested requirements related to three primary areas:

- Institutional Strategy
- DMPs
- Deposit
Institutional Strategy:

“Each institution administering tri-agency funds is required to create an institutional research data management strategy”

Data Management Plans:

“...specific funding opportunities may require DMPs to be submitted to the appropriate agency at time of application

Data Deposit:

“Grant recipients are required to deposit into a recognized digital repository all digital research data, metadata and code that directly support the research conclusions in journal publications, pre-prints, and other research outputs that arise from agency-supported research
The Canadian Picture - Many Stakeholders
Portage Network

Portage is a national RDM network launched by CARL in 2015 which coalesces initiatives to build capacity and coordinate RDM activities in Canada

1. **Networks of Expertise:**
   - Pan-Canadian RDM expertise
   - Provide access to resources, tools, and experts in the area of RDM

2. **Infrastructure Platforms**
   - Working with library consortia, institutions, and other infrastructure partners to assemble essential RDM infrastructure and service components

*Portage provides bi-lingual expertise and resources*
Portage Network

Portage is a national RDM network launched by CARL in 2015 which coalesces initiatives to build capacity and coordinate RDM activities in Canada.

1. **Networks of Expertise**
   - Pan-Canadian RDM expertise
   - Provide access to resources, tools, and experts in the area

   **Networks of Expertise by the numbers:**
   - 6 Expert Groups
   - 6 Working Groups
   - >115 members
   - >40 institutions/organizations

   *UofA representation on 3 EGs & 3 WGs!

2. **Infrastructure Platforms**
   - Working with library consortia, institutions, and other infrastructure partners to assemble essential RDM infrastructure

   **Infrastructure Platforms:**
   - DMP Assistant
   - Federated Research Data Repository (FRDR)

*Portage provides bi-lingual expertise and resources*
UofA:
Institutional shared research data stewardship

Libraries

Researchers

Faculties

Individuals, Groups and Services

Graduate Studies

IT

Research Ethics Office

Research Services Office
UofA Libraries: Delivering RDM supports and services

Preparing for the Tri-Agency RDM Policy
Responses to Tri-Agency Policy

Institutional RDM Strategies

Data Management Planning

Data Repository Options

DMP Assistant: National, online, bilingual, Data Management Planning Tool

The Dataverse Project
Pillar 1: Institutional RDM Strategies

1. **Raise awareness**: Policy requirements & RDM best practices
2. **Assess institutional RDM readiness**
4. **Define an RDM roadmap**: to help with capacity-building over the medium term, to ensure research excellence, efficiency, and transparency.
Pillar 2: Data Management Planning

Portage Expert Groups:
- Data Management Planning
- RDM Training

https://www.library.ualberta.ca/research-support/data-management
Pillar 2: Data Management Planning

- The Portage DMP Assistant is a national & freely available ‘gold standard’ platform for supporting researchers in developing data management plans.

- The Tri-Agencies draft RDM policy states:

  “Grant applicants are encouraged to use standardized tools to develop their DMPs, such as the Portage Network’s DMP Assistant.”

University of Alberta Libraries provides all of the necessary infrastructure, development expertise and hosting services for supporting the DMP Assistant nationally.

https://www.library.ualberta.ca/research-support/data-management
Pillar 2: Data Management Planning

DMP Assistant Features:

Freely available & web-based

Bi-lingual

Customizable templates

Walks through key RDM categories & questions

Guidance & examples provided

Living RDM document → can be revised during active research

Can have multiple DMPs

Shared access - read only, editors, & co-owner

Export function

DMP sharing
DMP Assistant is a bilingual tool for preparing data management plans (DMPs). The tool follows best practices in data stewardship and walks researchers step-by-step through key questions about data management.

1. Sign up with DMP Assistant
2. Sign in and select a template under Organizations. The Portage template is the default.
3. Answer the questions that are relevant to your work. Guidance and examples are provided.
4. Revisit the tool throughout your research to review or revise your answers.

Sign in
If you have an existing account with DMP Assistant or previous version of DMP Builder.

Sign up
New to DMP Assistant? Sign up today.

Please note that we are currently working on single sign-in authentication. For now, please create a new DMP Assistant account. You will have the option to link your DMP Assistant account to your campus ID when that feature becomes available.
Upon signing in, researchers can either access an existing DMP that they have.....

.....or create a new one
Each DMP has study level information associated with it.
Researchers can choose to answer questions within any given section at any time.

---

**Tips**

Not all questions will apply to all research projects. Researchers are encouraged to answer the questions relevant to their work.

Researchers should revisit the tool throughout their research to review or complete their responses.

---

<table>
<thead>
<tr>
<th>Plan details</th>
<th>University of Alberta Data Management Questions</th>
<th>Share</th>
<th>Export</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Collection</td>
<td>6 questions, 0 answered</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Documentation and Metadata</td>
<td>3 questions, 0 answered</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Storage and Backup</td>
<td>2 questions, 0 answered</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preservation</td>
<td>3 questions, 0 answered</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Sharing and Reuse</td>
<td>4 questions, 0 answered</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Responsibilities and Resources</td>
<td>4 questions, 0 answered</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
What types of data will you collect, create, acquire and/or record?

- B: Images, audio, video, text, tabular data, modeling data, spatial data, instrumentation data.

- A: Unspecified.

Guidance: University of Alberta Guidance

Examples: Images, audio, video, text, tabular data, modeling data, spatial data, instrumentation data.

Save

Not answered yet

What file formats will your data be collected in? Will these formats allow for data re-use, sharing and long-term access to the data?

- B: Proprietary file formats which require specialized software or hardware to use are not recommended, but may be necessary for certain data collection or analysis methods. Using open file formats or industry-standard formats (e.g., those which are widely used by a given community) is preferred whenever possible.

- A: Unspecified.

Guidance: University of Alberta Guidance

Proprietary file formats which require specialized software or hardware to use are not recommended, but may be necessary for certain data collection or analysis methods. Using open file formats or industry-standard formats (e.g., those which are widely used by a given community) is preferred whenever possible.

Save

Not answered yet

Plan details
University of Alberta Data Management Questions
Share
Export

Data Collection
(8 questions, 0 answered)

Documentation and Metadata
(3 questions, 0 answered)

Storage and Backup
(2 questions, 0 answered)

Preservation
(3 questions, 0 answered)

Not answered yet

Which data are selected for preservation and access will depend on potential reuse value, whether there are obligations to retain or destroy data, and the resources required to properly curate the data and ensure that it remains usable in the future. In some circumstances, it may be feasible to preserve all versions of the data (e.g., raw, processed, analyzed, final), but in others, it may be preferable to only keep only selected data (e.g., transcripts instead of audio interviews).

Guidance: University of Alberta Guidance

Which data are selected for preservation and access will depend on potential reuse value, whether there are obligations to retain or destroy data, and the resources required to properly curate the data and ensure that it remains usable in the future. In some circumstances, it may be feasible to preserve all versions of the data (e.g., raw, processed, analyzed, final), but in others, it may be preferable to only keep only selected data (e.g., transcripts instead of audio interviews).

Save

Not answered yet
Share note with collaborators

Add collaborator

Email

Permissions:
- Co-owner
- Edit
- Read only

Add collaborator

From here you can download your plan in various formats. This may be useful if you need to submit your plan as part of a grant application.

Select what format you wish to use and click to ‘Export’.

Format
- pdf

Export

Settings (Using default PDF formatting values)
Pillar 3: Data Deposit into a Repository

- Institutional RDM Strategies
- Data Management Planning
- Data Repository Options
Pillar 3: Data Deposit into a Repository

Dataverse Positives:

- Open source software
- Mature platform
- Widely adopted
- Community support
- User-friendly
- Brandable

Portage Training Materials in development

UofA Dataverse by the numbers:
- ~400 datasets
- ~4000 files
- >23,000 downloads

http://dataverse.org/
Pillar 3: Data Deposit into a Repository

Dataverse Features:

- Digital Object Identifier (DOI) assigned
- User controlled data access
- Many types of data
- Built in data citations
- Usage metrics
- Brandable

**Portage Training Materials in development**

Persistent identifier assigned when upload data

Automatically registered with DataCite
https://www.datacite.org/

From fully open to restricted access

Tabular (CSV, SPSS, R, etc)
Documentation (pdf, doc, text)
Geospatial data
Multimedia (audio-visual)

Madueke, Ijeoma Sylvia, 2018, "NIGERIAN LITERATURE IN FRENCH TRANSLATION (NILIFT)",
https://doi.org/10.7939/DVN/CHNOHA

Including a customizable guestbook

High level as well as in-depth web based training modules
The Dataverse Network

http://dataverse.org/
**Pillar 3: Data Deposit into a Repository**

Federated Research Data Repository

**Main Features:**
- ‘Big Data’ upload/download
- National Discovery Platform
- National Repository Option
- Preservation Processing capacity

Currently in Limited Production

**Strategic Relevance**
- Fits into broader Portage vision for Preservation
- Strategic partnership between Portage & Compute Canada (ARC)
- Aim is to fill key gaps in the RDM ecosystem

https://portagenetwork.ca/frdr-dfdr
FRDR DEMO SITE AVAILABLE
https://demo.frdr.ca/repo/?locale=en
UofA Libraries RDM & Data Services - Come See Us!

https://www.library.ualberta.ca/research-support/data-management

Research Data Management

Organize my data
How to prepare a data management plan & protect your data

Share my data
Why you would share, data sharing tools, metadata & how to get credit

Access existing data
How to locate data and access data tools, terms & citations

Archive my data
How to find a repository, prepare your data & submit your data

Consult us at data@ualberta.ca

Data Help Request
Use this form to contact the University of Alberta Data Team with any data-related questions. Information on the U of A Libraries’ data collections and services is available at http://guides.library.ualberta.ca/data

* Required

What best describes your request? *
- Help with finding statistics/data
- Help with GIS/spatial/mapping data
- Help with research data management
- Help with LDC Text Corpora
- Other

Please provide your contact name and preferred e-mail address or phone number. *

Your Faculty/Department, if applicable?

Brief description of the help or information you require. *

Best day for an appointment if you would like a data consult.
Thank-you

James Doiron
Research Data Management Services Coordinator
Digital Initiatives, University of Alberta Libraries
jdoiron@ualberta.ca