Pillar 2: Data Management Plans

James Doiron
RDM Services Coordinator, UAlberta Library
Academic Director, UAlberta Research Data Centre
Co-Chair, Portage DMP Expert Group

Putting the Tri-Agency Policy into Practice
October 25-28, 2021
Session overview:

1. **DMPs** - What they are, their importance & benefits

1. **Content of DMPs** - General categories & guidance

1. **DMP Assistant** - How to access and use the tool; key features overview

1. **DMP Expert Group resources** - General guidance, DMP exemplars; upcoming resources

1. **Questions & Discussion**
What is Research Data Management?

RDM spans across all phases of the research lifecycle - managing data through active phases and beyond.

Source: “University of Virginia Library LibGuide: Research Data Management”
What is a data management plan (DMP)?

A DMP:

- Is a formal document which clearly articulates the strategies and tools you will implement to effectively manage your data.
- Speaks to the management of data both during the active phases of your research and after the completion of the research project.

The objective of a DMP is to address issues related to data management prior to starting your research project!
Why are DMPs important?

A data management plan is important to the research process as it can help you to:

- **set out consistent strategies** prior to starting your research for how data will be managed throughout its entire lifecycle

- **identify the strengths & weaknesses** in your current practices and make decisions on how to integrate effective data management practices into your process

- **prepare data for future reuse, preservation and sharing**

- **reduce the overall cost of research** by increasing project efficiencies
A DMP provides information across key research lifecycle categories:
When should I start creating my DMP?

A DMP should be developed as early in the research process as possible!

Every research project should ideally **begin** with the creation of a DMP, as it can be used to guide the research process.
DMP guidance resources

Creating an effective DMP (English/French)
DMP General guidelines

- Begin by providing a description of your research project, its focus, and purpose.
- Avoid extensive use of discipline specific jargon - your DMP should be easily understood by anyone!
- Provide clarification for any acronyms used.
- Do not leave sections or questions blank.
- Provide rationale for decisions made - help others understand why you have made a decision.
- Your DMP is a living document - update it as needed!
DMP Sections: Data Collection

Include descriptions of how you will collect data, including from where and in what format(s).

Describe any software and/or platforms that will be used for data collection.

Clearly explain how you will both store and transfer data.

Provide an estimate of the amount of data you will collect (e.g., MBs/GBs/TBs).

Explain how you will organize your data, including details relating both to file naming and versioning.
DMP Sections: Documentation & Metadata

- Choose a metadata standard suited to your discipline and/or chosen data repository, or provide rationale for creating your own.
- Describe what information will be needed for others to understand or reuse your data.
- Describe how you will consistently capture documentation throughout the project.
DMP Sections: Storage & Backup

State a data backup schedule, automatic being most ideal.

Provide an estimate of storage space needed during the active phases of your research - remember to take into account file versioning, backups, and data growth!

Describe how collaborators or research team will be able to access, modify, contribute, and work with your data.
DMP Sections: Preservation

Not all data that you create necessarily needs to be preserved - consider such things as the value of your data, funding requirements, etc., and decide which, if any, should be preserved. Consult with experts in the Libraries as needed!

Consider optimal file formats (preferably non-proprietary) for supporting long-term preservation. Optimally preserved data are easily accessible and use by anyone, without requiring proprietary software to do so.
DMP Sections: Sharing & Reuse

Consider the appropriate sharing of your data, including any funding or confidentiality requirements.

Consult with colleagues or librarians to choose an appropriate data repository or search re3data.org to find one.

Choose a repository that assigns permanent identifiers to datasets (e.g., DOI) to enhance discoverability, accessibility, and citability.

Explain what uses can be made of your data through licenses like Creative Commons.

If applicable, describe how you will ensure file integrity, anonymization and de-identification.
DMP Sections: Responsibilities & Resources

- Identify data stewardship roles and responsibilities of project members and other organizations during and after the project.

- Estimate and describe any required resources and costs for data management and long-term access to your data.
DMP Sections: Ethics & Legal Compliance

Describe if there are any legal, ethical, and intellectual property issues when managing and sharing your data.

Explain how you will comply with any applicable privacy legislation and laws, including funding and institutional requirements.

Describe how you will ensure your data are securely managed after the project is completed including the secure management of sensitive data and in accordance with any ethical obligations.
THE PORTAGE NETWORK

is dedicated to the shared stewardship of research data in Canada through:

- Developing a national research data culture
- Fostering a community of practice for research data
- Building national research data services and infrastructure
Portage DMP Assistant

- Hosted by the University of Alberta Library and supported by the Portage Network.
- A national, open, bilingual data management planning (DMP) tool to help researchers better manage their data throughout the lifespan of a project.
- Develops a DMP by prompting researchers to answer a number of key data management questions, supported by best-practice guidance and examples.

https://assistant.portagenetwork.ca/
Portage DMP Assistant

Researchers can:

- Create project plans, using a generic template or an institution-, discipline-, or methodology-specific template;
- See guidance appropriate to the particular template they are using;
- Collaborate with multiple researchers on a plan;
- Connect to local guidance and support for data management at their academic institutions.

https://assistant.portagenetwork.ca/
Portage DMP Assistant

Institutions can:

- Customize the default Portage template for their institution;
- Create institution-specific templates and guidance;
- View information about DMP Assistant usage at their institution.

https://assistant.portagenetwork.ca/
Web-based, bilingual tool
Guidance & Examples
Can have multiple DMPs
Shared access - co-owner, editor, or read-only
Living document - modifiable when needed
Exportable (e.g., pdf, text, MS Word)
Freely available to end users
Endorsed by the Tri-Agency
Supports cross institutional research collaboration

DMP Assistant Key Features
Notice: Signed out successfully.

Welcome to DMP Assistant.
DMP Assistant has been developed by the Portage Network to help you write data management plans.

Getting started:
- Digital Curation Centre
- UC3: University of California Curation Center
- UK funder requirements for Data Management Plans
- US funder requirements for Data Management Plans
- DCC Checklist for a Data Management Plan
- DMP Assistant equivalent in France
- France funder requirements for Data Management Plans
Upon signing in, you can either access an existing DMP

Or create a new one
You can ‘copy’ or ‘clone’ an existing DMP for similar projects or funder requirements.
Select a template to use when creating a new plan:
DMP Templates

There are 13 discipline and methodology-specific templates...so far!

- Advanced Research Computing
- Arts-Based Research
- Interdisciplinary Health
- Software/Technology Development
- Mixed Methods (Surveys & Qualitative Research)
- Qualitative Health Sciences Research
- CRDCN: Accessing Data from Research Data Centres
- CRDCN: Research Data Centres and External Analysis
- History and the Humanities
- Neuroimaging in the Neurosciences
- Open Science Workflows
- Studying Molecular Interactions
- Systematic Reviews
- Water Quality Research

https://portagenetwork.ca/tools-and-resources/training-resources/
Each DMP has study level information associated with it. You can choose from a range of visibility settings - from public to private.
You can choose to answer questions within any given section at any time...

<table>
<thead>
<tr>
<th>Section</th>
<th>Questions</th>
<th>Answered</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Collection (6 / 6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Documentation and Metadata (3 / 3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Storage and Backup (2 / 2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preservation (3 / 3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Sharing and Reuse (4 / 4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Responsibilities and Resources (4 / 4)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Total questions answered: 22/22
Dedicated space as well as guidance customized guidance is provided for each question...

**Documentation and Metadata (3 / 3)**

What documentation will be needed for the data to be read and interpreted correctly in the future? This includes study-level documentation, data-level description, and any other contextual information required to make the data usable by other researchers.

Survey data will be collected within REDCap, a secure electronic data capture and management software hosted by the Women & Children's Health Research Institute (WCHRI) at the University of Alberta. REDCap features include the ability to develop and export a data dictionary which will outline all codes and variables within the survey. Key documentation related variables will be automatically populated within the survey data including time and date stamps and other key information to support data management and analytic activities.

All qualitative interviews will include summary information including: data collector, location of interview, and the date that the interview was conducted. Additionally, qualitative interviews will have accompanying field notes containing key contextual information and metadata.

File naming documentation will be developed and implemented. Components of file names will include as needed: file version (raw, edit, master, analytic), date (i.e., dd/mm/yyyy), and any applicable key contextual information, (e.g., geographical location, interviewer initials or code).

Your documentation may include study-level information about:

- who created/collected the data
- when it was created
- any relevant study documents
- conditions of use
- contextual details about data collection methods and procedural documentation about how data files are stored, structured, and modified.

A complete description of the data files may include:

- naming and labelling conventions
- explanations of codes and variables
- any information or files require to reproduce derived data.

More information about data documentation is available at the UK Data Archive.
You can easily add collaborators to your DMP, with different permissions settings...

Portage DMP Exemplar - "Mixed Methods"

Set plan visibility
Public or organisational visibility is intended for finished plans. You must answer at least 80% of the questions to enable these options. Note: test plans are set to private visibility by default.
- Private: visible to me, specified collaborators and administrators at my organisation
- Organisation: anyone at my organisation can view
- Public: anyone can view

Manage collaborators
Invite specific people to read, edit, or administer your plan. Invitees will receive an email notification that they have access to this plan.

<table>
<thead>
<tr>
<th>Email address</th>
<th>Permissions</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="mailto:mitchelldoiron@hotmail.com">mitchelldoiron@hotmail.com</a></td>
<td>Owner</td>
</tr>
<tr>
<td><a href="mailto:jdoiron@ualberta.ca">jdoiron@ualberta.ca</a></td>
<td>Co-owner</td>
</tr>
</tbody>
</table>

Invite collaborators

Submit
A range of options are available both for formatting and downloading your DMP...
The Portage DMP Expert Group helps to develop guidelines, best practices, and content in support of the DMP Assistant. Upcoming outputs include:

- **Tutorial Videos**
  - Introduction to DMPs
  - Introduction to DMP Assistant

- **Updated Institutional Administrator Guide**
  - Institutional Administrator Training

- **More DMP Exemplars**

- **And More!**

<table>
<thead>
<tr>
<th>Discipline and/or methodology focus</th>
<th>Title &amp; Link</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Digital Humanities</strong></td>
<td>Belgians and French in the Prairies</td>
</tr>
<tr>
<td></td>
<td>Soundscape Study</td>
</tr>
<tr>
<td></td>
<td>Women’s Print History Project (1750-1830)</td>
</tr>
<tr>
<td></td>
<td>Historical Canadian Census Data</td>
</tr>
<tr>
<td><strong>Mixed Methods</strong> (surveys &amp; qualitative interviews)</td>
<td>Mixed Methods Fictional Exemplar</td>
</tr>
<tr>
<td><strong>Natural Sciences</strong></td>
<td>Ecohydrology Research Group</td>
</tr>
<tr>
<td></td>
<td>Computational Reproducibility in High-Performance Computing</td>
</tr>
<tr>
<td><strong>Social Sciences</strong></td>
<td>People, Places, Policies &amp; Prospects: Affordable Rental Housing for Those in Greatest Need</td>
</tr>
<tr>
<td></td>
<td>Usage of Academic Profile Websites</td>
</tr>
<tr>
<td>Discipline and/or methodology focus</td>
<td>Digital Humanities</td>
</tr>
<tr>
<td>-------------------------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td><strong>Digital Humanities</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Mixed Methods</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Natural Sciences</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Social Sciences</strong></td>
<td></td>
</tr>
</tbody>
</table>

### Data Management Plan Exemplar #3: Mixed Methods

**Fictional Exemplar**

**Data Collection**

- **What types of data will you collect, create, acquire and/or record?**
  - We will be collecting surveys which will then be exported into tabular format.
  - We will also be conducting both semi-structured interviews and focus groups that will produce both digital audio and text (transcripts) based data.
- **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?**
  - Our file formats will exist both in non-proprietary and proprietary formats. The non-proprietary formats will ensure that these data are able to be used by anyone wishing to do so once they are deposited and made openly available.
  - Surveys will exist in .csv (non-proprietary), .xls (.xlsx for Microsoft Excel) and SPSS (both proprietary) formats. For more information regarding SPSS see: SPSS Wikipedia [https://en.wikipedia.org/wiki/SPSS](https://en.wikipedia.org/wiki/SPSS).
  - Interviews & focus groups data will exist in .mp3 (non-proprietary), .txt, .docx and .pdf (both proprietary) formats. For more information regarding NVivo see: NVivo Wikipedia [https://en.wikipedia.org/wiki/NVivo](https://en.wikipedia.org/wiki/NVivo).
- **Any survey data deposited for sharing and long-term access will be in .csv format so that anyone can use them without requiring proprietary software.**
- The final de-identified versions of the interviews and focus groups transcripts will be exported into a basic non-proprietary text format for deposit, long-term preservation and access.
- **If data are collected using laptops or mobile devices, please explain how you will securely store and transfer the data.**
  - Laptops are not being used for any data collection, though encrypted digital voice recorders (DVRs) will be used to collect both interviews and transcripts. Interviews and focus group digital audio files will not be stored on the DVRs, only collected and then securely transferred to the project's cloud based virtual research environment (VRE) via a secure FTP (File Transfer Protocol).

---

Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0)
Portage Network | portage@carleton.ca | portagenetwork.ca
If you are interested in learning more about using the DMP Assistant, you can access and view the below webinar!

Support Your Research with DMP Assistant 2.0!

Presenter: Robyn Nicholson, DMP Coordinator, Digital Research Alliance of Canada
The DMP Assistant Steering Committee provides essential guidance, support and direction including on the development of **new proposed features** such as:

- **API Integration** with other systems (e.g., grants, ethics)
- **Reviewing & Evaluating** DMPs
- Building a **DMP Repository** with DMP Assistant integration
- **DMP Versioning**
- Leveraging the power & potential of **PIIDs**
- Improving machine-actionability > **maDMPs**

Questions & Discussion

James Doiron
RDM Services Coordinator, UofA Library
Academic Director, UofA Research Data Centre
Co-Chair, Portage DMP Expert Group

jdoiron@ualberta.ca