INTRODUCTION TO RESEARCH DATA MANAGEMENT

Tools and resources to get started

Scholarship & Dal Series
November 2018

Presenters:
Erin MacPherson, MLIS, P.Ag
Sarah Stevenson, MLIS, B.Sc.

This session will help you:

• Understand the importance and utility of management plans for research data/materials of research.

• Be able to built a data management/materials of research plan for a current project using the Portage DMP Assistant.

• Gain awareness of the Dataverse and FRDR data sharing platforms.
Research Data Management

A broad term used to describe the structure, organization, maintenance, and overall stewardship of research data.

These elements are given consideration at each stage of research from the beginning of a project to its conclusion and can be revisited through the project.

Research Data Lifecycle

http://data.library.virginia.edu/data-management/lifecycle/
What are Research Data?

Research data may be defined as the factual records (e.g. microarray, numerical and textual records, images and sounds, etc.) used as primary sources for research, and that are commonly accepted in the research community as necessary to validate research findings.

(Canadian Association of Research Libraries, n.d.)

Examples

- Documents (text, MS Word), spreadsheets
- Can include online surveys, transcripts, etc
- Scanned documents
- Such as lab and field notebooks
- Digital media
- Audio, video and other recording files
- Scanned photos and digital images
- Transcribed test responses
- Database contents
- Digital models, algorithms, scripts
- Contents of an application (input, output, logfiles for analysis software, simulation software)
- Documented methodologies and workflows
- Records of standard operating procedures and protocols

(University of Bristol, 2012)
Let’s watch a video…

How to avoid a data management nightmare, by Karen Hanson, Kevin Read & Alisa Surkis, New York University Health Sciences Library. Creative Commons Licence CC-BY (Attribution, reuse allowed)
Thoughts?

- Variables
- Workflows
- Naming conventions
- Backups

Why manage research data?

Many benefits...

- SHARE it. (data re-use by you or by others) – allows others to benefit from your data, increase your citation rate, reduce duplication.
- SAFE-KEEPING. (protect from theft, loss, device failure, natural disaster).
- Prepare for CHANGE. (technological, administrative, personnel).
- To be ORGANIZED. (making accessing and using data easier).
- Tri-Agency FUNDING guidelines.
- Journal submission & publication requirements
Tri-Agency Statement of Principles

In June 2016, NSERC, SSHRC and CiHR adopted the *Tri-Agency Statement of Principles on Digital Data Management*. The statement covers many things, including:

- digital research data management and expectations,
- responsibilities of researchers, research communities, institutions and funders in meeting these expectations.

As publicly-funded organizations, they advocate for access to research results, including research data.

Why?

- Avoid duplication
- Advance knowledge
- Maximize research benefits to Canadians
- Promote Canadian researchers and their accomplishments (Government of Canada, 2015).

Some journals have policies on data sharing

http://www.journalofanimalecology.org/view/.../authorGuideline.html
Data Archiving

Data are important products of the scientific enterprise, and they should be preserved and usable for decades in the future. The British Ecological Society thus requires, as a condition for publication, that all data (or, for theoretical papers, mathematical and computer models) supporting the results in papers published in its journals will be archived in an appropriate public archive, such as Dryad, Treebase, NERC data centre, CentraIR, Figshare, or another archive of the author’s choice that provides comparable access and guarantee of preservation. Authors may elect to have the data made publicly available at time of publication or, if the technology of the archive allows, may opt to embargo access to the data for a period of up to a year after publication.

Exceptions, including longer embargoes, may be granted only in exceptional circumstances at the discretion of the editor, especially for sensitive information such as confidential social data or the location of endangered species.

For further information we have a Q&A and you can see a list of suitable repositories here.

http://www.esa.org/esa/science/data-sharing/resources-and-tools/
What is a DMP and why should I care?

A Data Management Plan, or DMP, is a formal document that outlines the steps taken before, during, and after your research project to ensure that the data will be useable, discoverable, and accessible beyond the purpose for which it was originally collected.

Benefits of a DMP:

• Start thinking about your data at the outset of a project
• Identifies strategies and services to support future use
• Cuts down on the duplication of efforts
• Satisfies funder requirements

Components of a DMP

1. Data Collection
2. Documentation & Metadata
3. Storage and Backup
4. Preservation
5. Sharing and Reuse
6. Responsibilities & Resources
7. Ethics & Legal Compliance

Note: The DMP Assistant or your institution may have a slightly different template, but these are the general components of a DMP.
The DMP Assistant will help you with all of this!

Access the DMP Assistant tool via:

Portage website: https://assistant.portagenetwork.ca/

OR

Your institution's website: http://dal.ca.libguides.com/rdm/plan

Launched in 2015 by CARL, the Portage Network has two interlinked mandates:

- A library-based network of expertise on research data management
- National platforms for planning, preserving, and discovering research data
12/10/2018

Note: For Librarians and those who may have RDM responsibilities - you can customize the DMP Assistant for your institution. There is a guide available on the Portage website to help:
https://portagenetwork.ca/working-with-portage/dmp-assistant-custom-guide/

Since I already have an account, we are brought directly to the “My Plans” section upon logging in. You can see some of the sample plans I’ve created. This section lists the name, template used, who owns the plan, if it has been shared, last edit date. You can also edit existing plans, share, export and delete them.
I’m going to choose Dalhousie’s branded version with the standard portage template. Institutions can customize template questions and add guidance. Researchers need to only answer questions that are applicable to them.

These fields can all be edited - you probably don’t want to call your plan “My plan”. You can also save each section as you go, and come back later.
These are the questions that you can fill out for your DMP. You only need to answer questions that are relevant to your project. Guidance is provided, and your institution can also add additional guidance.

Once you fill out the basic plan details, you are brought to the Question section. Guidance to the questions is provided on the right. You can also add links, tables and format the text. On the right, you also have the option to provide notes for any collaborators. You can also save each question as you answer it and return to the plan later.
You can add collaborators and add permissions. There are three permission levels. Collaborators will have to register for the DMP Assistant if they don’t already have an account.

If you need to make changes, you can always go back and edit plan details or the questions anytime.
Click on the “Settings” option for more great features before you export your document. You can select which sections to export (including plan details and questions). For some formats, you can also choose your font type, size and margins.
**Data deposit**

**DATAVERSE**

- A free open source web application
- Developed by the Institute of Quantitative Social Science at Harvard University
- Designed specifically for research data
- Platform to contribute, share, preserve and find data.
- Dataverse is used all over the world. In Canada there are several installations available, including Dalhousie and UNB in Atlantic Canada.

**FRDR**

FRDR = Federated Research Data Repository (DFDR en français)

A scalable, federated platform for digital research data management and the discovery of Canadian research data

*Discovery, Storage, Curation, Preservation*
Repositories

Find repositories to share, preserve your data and more:

- **re3data.org** – a global registry of subject specific research data repositories.  
  http://www.re3data.org/

- Some popular repositories:
  - **Figshare**: Figshare offers 20GB of free, private space for research. hosted using Amazon Web Services.  
    https://figshare.com/features
  - **ICPSR** (Interuniversity Consortium for Political and Social Research):  
    https://www.icpsr.umich.edu/icpsrweb/
  - **Dryad**: primarily lifesciences  
    http://datadryad.org/
  - **Dataverse**: Institutional based (Dalhousie) but Harvard & Scholars Portal  
    Dataverse is open to the world.
  - **GitHub**:  
    https://github.com/
  - **ComputeCanada**: various types of computing solutions:  
    https://www.computecanada.ca/

RDM Infrastructure and tools

**Start here:**

- **Research Data Management Libguide**: A “mini website” to help you get started on RDM planning.
- **Dal Data Management Services website**: https://libraries.dal.ca/services/data-management-services.html

**DMP Assistant**: a free tool to Canadian researchers to help you develop a good data management plan for your project. Developed by CARL Portage Network, this tool will walk you step-by-step through a series of questions to create a data management plan.

**Dataverse**: An open source, self-serve research data repository software. Share, organize and preserve your data.

**DalSpace**: institutional repository that collects, preserves and distributes digital content produced by members of Dalhousie University. Suitable for documentation related to research datasets, such as journal articles and theses. It can also be used to preserve final datasets (those that will not require changes or editing).

**Federated Research Data Repository**: is for Canadian researchers to deposit and share research data. FRDR is a good repository to deposit very large files and datasets. It also supports automatic preservation using Archivematica.
References


Photo Credits

- Slide 18: Portage logo: https://portagenetwork.ca/
- Slide 30: Dataverse logo: https://dataverse.org/
- FRDR logo: https://www.frdr.ca/repo/