The “need-to-knows” about Research Data Management

McMaster Library SLDC Workshop - 01-Dec, 2016

V. Jadon  <vivek@mcmaster.ca>
J. Brodeur <brodeujj@mcmaster.ca>
Outline

✧ What is Research Data Management (RDM)?
✧ The research data lifecycle
✧ RDM support at various scales
✧ RDM tools and resources
If you shared your data another researcher or collaborator, would they be able to:

a. interpret and understand it?

b. use it in new analyses?

Would someone (including you) be able to find, interpret and use your data 20 years from now?
Research Data Management (RDM) is...

... the active organization & maintenance of data

... the application of best practices to ensure data security, accessibility, usability, and integrity

... a set of activities resulting in self-describing data sets that can be discovered and reused
Applying RDM best practices will benefit...

Researchers and their collaborators
- Improves research efficiency and productivity
- Provides extra credit for research work
- Increases research impact
- May help to meet funding requirements

Research Communities
- Accelerates discovery
- Enables validation and verification

Funders, governments and the public
- Improves return on investment
- Increases research transparency
- Data as a public good
Publishers & Scientific Organizations

✧ Data sharing policies
✧ Recommended data repositories
✧ Publisher-supported data repositories
Tri-Agency Statement of Principles on Digital Data Management

1. Preamble

The Canadian Institutes of Health Research (CIHR), the Natural Sciences and Engineering Research Council of Canada (NSERC), and the Social Sciences and Humanities Research Council of Canada (SSHRC) (the agencies) are federal granting agencies that promote and support research, research training, knowledge transfer and innovation within Canada.

As publicly funded organizations, the agencies are strong advocates for making the results of the research they fund as accessible as possible. In promoting access to research results, they aspire to advance knowledge, avoid research duplication and encourage reuse, maximize research benefits to Canadians and showcase the accomplishments of Canadian researchers. These aspirations align with the Government of Canada’s commitment to open science, as described in Seizing Canada’s Moment: Moving Forward in Science, Technology and Innovation (2014).

Research data include observations about the world that are used as primary sources to support scientific and technical inquiry, scholarship and research-creation, and as evidence in the research process.¹ Research data are gathered through a variety of methods, including experimentation, analysis, sampling and repurposing of existing data. They are increasingly produced or translated into digital formats. When properly managed and responsibly shared, these digital resources enable researchers to ask new questions, pursue novel research programs, test alternative hypotheses, deploy innovative methodologies and collaborate across geographic and disciplinary boundaries. The ability to store, access, reuse and build upon digital research data has become critical to the advancement of science and scholarship, supports innovative solutions to economic and social challenges, and holds tremendous potential for Canada’s productivity, competitiveness and quality of life.
Tri-Agency Statement of Principles on Digital Data Management: Expectations

- Data management planning
- Constraints and obligations
- Adherence to standards
- Collection and storage
- Metadata
- Preservation, retention and sharing
- Timeliness
- Acknowledgement and citation
- Efficient and cost-effective

Tri-Agency Statement of Principles on Digital Data Management
http://www.science.gc.ca/default.asp?lang=En&n=83F7624E-1
Tri-Agency Statement of Principles on Digital Data Management: Responsibilities

Researchers
- incorporating best practices
- developing DMPs
- adhering to policies and standards

Research Institutions
- supporting best practices
- providing access to resources
- creating guidance and policies

Research Communities
- developing & promoting standards
- fostering excellence
- selecting repositories

Research Funders
- developing policies & guidance
- promoting data management
- providing peer reviewers
RDM in the Research Data Lifecycle: Things to Consider
RDM in the Data Lifecycle

- Plan
- Create/Collect
- Preserves
- Analyse
- Share
Planning

What are the stipulations in institutional, funder or publisher data policies to be followed?

What resources do you require to manage your data?

Who is responsible for data management and long-term stewardship?
Creating and Collecting

How much data will you collect or create? Where will you store it and back it up?

Is your created or collected data in a suitable format for sharing and long-term preservation?

What documentation and metadata should accompany the data?
Analyzing

- Plan
- Create/Collect
- Share
- Preserve
- Analyse

How will you manage any ethical or privacy issues before analysing the data?

How will you securely store (potentially large and cleaned) data pre- and post-analysis?

Who will have access to this data for analysis?
Preserving

Plan

Share

Create/Collect

Preserve

Analyse

What data should be retained and preserved?

Where will you preserve your data? Who will have access to this preserved data?

For how long you are going to preserve your research data?
What data will you make publicly available for research and reuse?

What resources are in place for sharing your data among multiple sites?

How will you manage restrictions e.g. license, privacy issues etc. associated with your data before sharing?
Supporting RDM needs at various scales
RDM capacity development at various scales

Institutional

Provincial / National

International

RDM @McMaster

Scholars Portal

DataCite

Portage

MacDATA Institute

RHPACS

RDA

SHARE

A Service of the Ontario Council of University Libraries

Association of Research Libraries
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

https://portagenetwork.ca/
Portage Expert Groups

- Data Management Planning Expert Group
- Preservation Expert Group
- Discovery Expert Group
- Research Intelligence Expert Group
- Training Expert Group
- Curation Expert Group
Institutional shared-stewardship

Libraries → Researchers

Graduate Studies → Individuals, Groups and Services

Ethics

Research Services → IT
What are researchers’ data management obligations?

What challenges do researchers face in managing their data?

How can the library help researchers address their data management needs?
Data Management Resources

1. Planning for RDM
...it all starts with a (good) plan

A research Data Management Plan (DMP) should:
✧ Describe how you will manage data through all stages of your research
✧ Communicate a strategy for creating share-worthy data products

A good DMP will:
✧ Be completed at the time of study design
✧ Ensure compliance with policies / obligations
✧ Document and organize research activities
✧ Help identify support requirements
✧ (Likely) evolve with your study...
Portage DMP Assistant

✧ A web-based, bilingual data management planning tool.
✧ Available to all researchers in Canada.
✧ A guide for best practices in data stewardship.
✧ Exportable data management plans.

https://www.youtube.com/watch?v=zgLaJpJfehQ

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

Login: mcm.rdm.demo@gmail.com
Pass: demopass
Data Management Resources

II. In-project data management
Things to consider...

How will you manage digital data in your research?
✧ What types of data will you collect and how much?
✧ How will you **organize** and **secure** your data?

How will you share data between collaborators?
✧ How will you **describe** your data so that others understand it?
✧ How will you control **access** to this data?
Open Science Framework
https://osf.io/

✧ Free, open source web application
✧ Connects and supports the research workflow
✧ Integrates with existing services (e.g. Google Drive, Dropbox, etc.)
✧ Supports collaborative workflows
✧ Facilitates data publishing to selected data repositories

GETTING STARTED WITH THE OPEN SCIENCE FRAMEWORK (OSF)

https://youtu.be/2TV21gOzfhw

Workshop demo: https://osf.io/muq5f
Data Management Resources

III. Preserving and sharing data products
Things to consider...

How will your data products be stored in the long-term?
✧ How to ensure that it remains integral and secure?
✧ Who will assume long-term responsibility for your data?

How will others access your data products?
✧ What data should be shared?
✧ How will you manage legal, commercial & ethical constraints?

How to maximize credit for sharing your data?
✧ How to ensure that your data is findable, accessible and citable?
Scholars Portal Dataverse

✧ A data repository for researchers at Ontario's universities.
✧ An online platform to share, preserve, cite, explore and analyze research data.
✧ Allows researchers to control how they share their data.
✧ Supports data DOI registration through Datacite Canada.

https://www.youtube.com/watch?v=UDFGqRY61fQ

http://dataverse.scholarsportal.info
Thank You.

For more information:

RDM @McMaster
library.mcmaster.ca/rdm
rdmgmt@mcmaster.ca

portagenetwork.ca/

McMaster Library’s RDM webpage and primary contact
CARL Portage Network page: Access to RDM information and DMP Assistant
Other Links

https://portagenetwork.ca/ - CARL Portage Network page: RDM information and access to DMP Assistant


http://library.mcmaster.ca/rdm - RDM@McMaster homepage

http://www.re3data.org/ - Registry of research data repositories

http://dataverse.scholarsportal.info/dvn/ - Scholars Portal Dataverse