The need-to-knows of Research Data Management

Pending Tri-Agency policies and resources to manage your research materials

Jay Brodeur
McMaster University Library
CRESS Research in Focus
06-Feb, 2018
Research data are contents that are used as primary sources to support research, scholarship, artistic activity or research-creation, and that are used as evidence in the research process and commonly accepted in the research community as necessary to validate research findings and results.

Canadian Tri-Agency

Research Data Management is the active organization & maintenance of data throughout the research cycle to ensure its security, accessibility, usability, and integrity.
The research data lifecycle

Applying RDM best practices will benefit...

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

Researchers and their collaborators

- Accelerates discovery
- Enables validation and verification

Research Communities

- Improves return on investment
- Increases research transparency
- Data as a public good

Funders, governments and the public
Requirements and Expectations

Funding Agencies and Governments

Publishers and Research Communities

Institutions
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: **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: **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 Communities**
- developing & promoting standards
- fostering excellence
- selecting repositories

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

**Research Funders**
- developing policies & guidance
- promoting data management
- providing peer reviewers
Tri-Agency Policy Development Timeline

✧ Capitalizing on Big Data: Toward a Policy Framework for Advancing Digital Scholarship in Canada [2013]
✧ Canada's Action Plan on Open Government [2014-2016; 2017]
✧ Tri-Agency Statement of Principles on Digital Data Management [2016]
Draft Tri-Agency RDM Policy

✧ Tri-Agency draft data management policy — expected April, 2018
✧ 6-month consultation period; feedback will inform policy

Proposed policy — 3 possible requirements:
1. **Institutions**: Institutional Strategy
2. **Researchers**: Data Management Plans
3. **Researchers**: Data Deposit\(^{a,b,c,d,e,f,g}\)

✧ Phased and incremental implementation

[Link to M. Lucas’ slides — Portage Network Event](30-Jan, 2018)
Draft Tri-Agency RDM Policy

✧ Tri-Agency draft data management policy — expected April, 2018
✧ 6-month consultation period; feedback will inform policy

Proposed policy — 3 possible requirements:

1. **Institutions**: Institutional Strategy
2. **Researchers**: Data Management Plans
3. **Researchers**: Data Deposit\(^{a,b,c,d,e,f,g}\)

✧ Phased and incremental implementation

[Link to M. Lucas’ slides — Portage Network Event](#) (30-Jan, 2018)
Institutional shared-stewardship
RDM capacity development at various scales

Institutional
- McMaster University Library
- RDM
  @McMaster
- MREB
- MacDATA Institute
- RDC

Provincial / National
- Scholars Portal
  A Service of the Ontario Council of University Libraries
- DataCite
- Portage
- Research Data Canada – Données de recherche Canada
- RDC
- DRC
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

I. 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/
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, secure, and backup** your data?
  ○ Are there ethical or commercial restrictions?

How will you provide access to collaborators?
✧ How will you **describe** your data so that others understand it?
✧ How will you control **access** to this data?
✧ How will you manage data versions?
copies of your data
copies of your data

copies are on-hand (easily accessible)

- 1 “production” (working) copy
- 1 “production backup” copy
copies of your data

copies are on-hand (easily accessible)

- 1 “production” (working) copy
- 1 “production backup” copy

copy is in another location (“off-site”), with a trusted service provide
“Production” copy → Where you work with the data

- PC, laptop, mobile device, etc.

“Production backup” copy → Easily accessible (+ versioning?) backup

- External hard drive with backup software
- MacDrive (seafile): [https://macdrive.mcmaster.ca](https://macdrive.mcmaster.ca)
- Dropbox, Google Drive, etc.
Off-site “Archived” Backup

Providers / Services:

- Campus / Consortium-hosted (RHPSCS)
- Remote, Commercial
  - Backblaze, Iron Mountain, JustCloud, etc.
Why Distance is Important

Gustavus Adolphus College

March 29, 1998

F3 tornado
Considerations for “Archived” Backup

- Security (physical and electronic)
- Automation
- Availability (and time to recover)
- Versioning
- Integrity-checking and error correction
- Data storage (locational) requirements
- Cost
Open Science Framework

- 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

https://osf.io/

https://youtu.be/2TV21gOzfhw

GETTING STARTED WITH THE OPEN SCIENCE FRAMEWORK (OSF)

Workshop demo: https://osf.io/mug5f
Data Management Resources

III. Preserving and sharing data products

- Plan
- Create/Collect
- Analyse
- Preserve
- Share
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 (if any) can/should be shared? Who should have access?
✧ How will you manage legal, commercial & ethical constraints?

How to maximize credit for sharing your data?
✧ In which repository should you deposit your data?
✧ How to ensure that your data is *FAIR* *(findable, accessible, interoperable and reusable)*?
The FAIR Guiding Principles

F1: (meta)data have a globally unique and eternally persistent identifier
F2: data are described with rich metadata
F3: metadata clearly and explicitly includes the ID of the data it defines
F4: (meta)data are registered and indexed in a searchable resource

A1: (meta)data retrievable by their ID using a standardized protocol
A1.1: protocol is open, free and universally implementable
A1.2: protocol allows for AuthT/ AuthZ where needed
A2: metadata is always accessible

R1: meta(data) richly described with accurate and relevant attributes
R2: (meta)data released with a clear and accessible data usage license
R3: (meta)data associated with detailed provenance
R4: (meta)data meet domain-relevant community standards

I1: (meta)data use a formal, accessible, shared, broadly applicable language for knowledge rep.
I2: (meta)data use vocabularies that follow FAIR principles
I3: (meta)data include qualified references to other (meta)data

re3data.org

Registry of research data repositories
Federated Research Data Repository (FRDR)

- Compute Canada + Canadian Association of Research Libraries (CARL)
- National data repository
- Framework for federating existing and future data repositories

https://www.frdr.ca
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

McMaster Library’s RDM webpage and primary contact

portagenetwork.ca/

CARL Portage Network page: Access to RDM information and DMP Assistant