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

Courtesy: Leadership Council for Digital Research Infrastructure (LCDRI) Data Management Position Paper, 2017-08-31

Applying RDM best practices will benefit...

Researchers and their collaborators

Research Communities

Funders, governments and the public

- ♦ Improves research efficiency and productivity
- ♦ Provides extra credit for research work
- ♦ Increases research impact
- ♦ May help to meet funding requirements
- ♦ Accelerates discovery
- ♦ Enables validation and verification

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





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 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 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 Open Access Policy on Publications [2015]
- Tri-Agency Statement of Principles on Digital Data Management [2016]
- Draft Tri-Agency Research Data Management Policy [2017-2018]

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
 <u>Link to M. Lucas' slides Portage Network Event</u> (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
 <u>Link to M. Lucas' slides Portage Network Event</u> (30-Jan, 2018)

Institutional shared-stewardship







McMaster

University



RDM capacity development at various scales

Institutional







Provincial / National













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





- 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

2

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): <u>https://macdrive.mcmaster.ca</u>
- Dropbox, Google Drive, etc.

Off-site "Archived" Backup



Providers / Services:

- Campus / Consortium-hosted (RHPCS)
- 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





<u>Open Science Framework</u>

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)

Workshop demo: <u>https://osf.io/muq5f</u>

https://youtu.be/2TV21gOzfhw

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 (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





Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... & Bouwman, J. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific data, 3. *doi:10.1038/sdata.2016.18*



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

	Search Dades	Seatt.
Search Results to Granges Memorie	< The Muster'	everyal based. Tax
hitly inners +	Dieles Millertade 20 18 18	田王市井市
Raffine Reserve Deligined Texanology Ref Stream Appliest Group(sines Stream) (K) Texano (New York Stream) (K) Texano (New York Stream) (K) Next Stream View Of Stream (K) Next Stream View Stream) (K) Next Stream View Stream) Next Stream View Stream (K) Next Stream View Stream) Next Stream View Stream (K) Next	Notify the set of the	India additional and a second
Marine, Million A. (%) Aran. M. Allar (1) Wellich A. (marine) Merine, M. A. (2) Laborato, F. (2)	Incomparison mapping and produce of the anise of particular sectors and an anisotropy of the sectors of the sector of particular sectors and particular sectors and the sector of the se	NAT TRANSPORT
Notice Advantage Moderne (Anisate (17) National Research Control, Organic (1) National Research Control, of Canada (1) Moderne Deveratio (1)	Integrating patients engenites and to perspected mapping to the step of \$2 patients of the constant of the Capital Marga Devices, Nanada the Thermite (1994), 100000, 100000, 100000,	NE INCLUTIO L CONTINUE Las Present Sec.7 2016

http://dataverse.scholarsportal.info



Thank You.

For more information:



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

McMaster Library's RDM webpage and primary contact



SERVICES PARTAGÉS POUR LES DONNÉES DE RECHERCHE SHARED STEWARDSHIP OF RESEARCH DATA

portagenetwork.ca/

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