

RESEARCH DATA MANAGEMENT 101

HEATHER GANSHORN, Librarian
Libraries & Cultural Resources
Heather.Ganshorn@ucalgary.ca

Research data management guide:
<https://libguides.ucalgary.ca/researchdatamanagement>

Grad Success Week – May 1, 2024

OUTLINE

- Introductions
- What is research data management?
- What do I need to do to comply with funding requirements?
- Data Management Planning with the DMP Assistant

WHAT BRINGS YOU HERE TODAY?

- What are you hoping to get out of this session?
- What are your current data management practices?



Hanson, K., A. Surkis & K. Yacobucci. Data Sharing and Management Snafu in 3 Short Acts. <https://youtu.be/N2zK3sAtr-4>



UNIVERSITY OF
CALGARY

Lab **NEXT**

WHAT ARE RESEARCH DATA?

- “Any information that has been collected, observed, generated or created to validate original research findings” (University of Leeds)

- Can include:

Images • Email, chat logs • Sound or video recordings • Presentations • Document data • Oral history & interviews • News stories • Bibliography • Measurements & statistics • 3D models, virtual environments, simulations • Software, source code • Transcripts • Procedures & workflows

RESEARCH DATA MANAGEMENT

- Application of standard practices towards the creation and analysis of data for current and future purposes

Why Research Data Management?

- Maintain data integrity (versions), avoid accidental data loss
- Meet requirements
- Easier collaboration
- Share your data, get more citations
 - Colavizza, G., et al. (2020). The citation advantage of linking publications to research data. PloS one, 15(4), e0230416.
<https://doi.org/10.1371/journal.pone.0230416>
- Increase reproducibility & verifiability

Research Data Management

Before

- Using others' data
 - Finding data
 - Copyright
 - Data licensing
- Funding applications
- Data management plan
- Ethics

During

- Documentation & metadata
- Storage & data protection
- Managing sensitive data
- Collaboration

After

- Sharing data
- Data repositories
- Preservation
- Data citation

DATA MANAGEMENT PLANS (DMPs)

- Short, living document
- Describe how you will ***organize, store, and share*** your research data through the lifetime of the research
- Does not contain data

DMP POLICY IN CANADA:

2021: Tri-Agency Research Data Management Policy

- **FAQ:** http://www.science.gc.ca/eic/site/063.nsf/eng/h_97609.html
- **Institutions:** will need to develop RDM policies by March 1, 2023
- **Individual Researchers:**
 - Will need to submit Data Management Plans for Tri-Agency funding starting in spring 2022 (phased-in approach)
 - Will have to meet requirements for data deposit/sharing (not yet specified, will be phased in).
 - CIHR-funded researchers may have to comply with Section 3.2 of the [Tri-Agency Open Access Policy](#)

Indigenous Data Sovereignty

- [OCAP Principles](#) for First Nations data
- [CARE principles](#) – general guidance on centring data sovereignty when working with Indigenous data

WORLD WIDE

- In the US, Data management policies and requirements for DMP have been introduced by NIH (2020), NSF (2011), and more to come for federally funded agencies.
- In the UK, “Common Principles on Data Policy” issued by Research Council UK (2011), and most UK funders now require a DMP
- In the EU, a pilot action on data management requirements within the Horizon 2020 framework. The approach is “as open as possible, as closed as necessary.”

FAIR Principles

- <https://www.go-fair.org/fair-principles/>
- Data should be:
 - Findable (persistent identifiers, human and machine-readable metadata)
 - Accessible (easily accessed online)
 - Interoperable – with standard applications for analysis, storage, processing of data
 - Reusable – data should be well described so it can be replicated/used in different settings

DMP Assistant – for creating DMPs

- Designed with the Tri Agency policies in mind
- Create an account for free and get started on your plan at <https://assistant.portagenetwork.ca/>
- I'm available to consult on writing your DMP
- Digital Research Alliance of Canada (formerly Portage Network) has exemplars and templates; more coming soon (<https://alliancecan.ca/en/services/research-data-management>)

CREATING A DMP

[Home](#)[Public DMPs](#)[DMP Templates](#)[Help](#)

Language ▼

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](#)

[Sign in](#)[Create account](#)

* Email

* Password

DMP Assistant (Canadian, bilingual), <https://assistant.portagenetwork.ca>

EXAMPLE SCENARIO

Louise is a new PhD student in the School of Public Health and Preventive Medicine. Her PhD topic relates to policy interventions to prevent the outbreak of infectious diseases like bird flu. She is interested in this topic because of her work as a policy analyst with AHS and her background in volunteering in developing countries and sees completing the PhD as a good way to further her policy career as well as her interests in social development.

Louise's research will involve several field interviews with health workers and policy makers in Canada, Vietnam, Indonesia and China. She has an iPhone and thought that she would use this to make audio recordings of the interviews, which she will later analyze (possibly using NVivo).

Louise also wants to access the policy documents of government agencies and health service providers (including hospitals) in Calgary and other jurisdictions in Canada and overseas. She thinks she will do content analysis on these, probably also using NVivo. Some agencies freely provide these documents on their websites, while other agencies have internal documents that are not readily available to the general public, which she may have to approach the organizations for directly. Louise wants to test her hypothesis that a speedy response from policymakers can reduce the spread of infectious diseases. This will require doing some cross-analysis of her findings from the policy documentation and interviews along with the World Health Organization's Cumulative number of confirmed human cases of avian influenza A(H5N1) dataset, which is available for download from the WHO website as a series of PDFs published monthly.

In doing her literature review there are several industry publications and academic journals that Louise has identified as potential places in which she might try to publish later. There are also some big international conferences coming up, and her supervisor has encouraged her to consider presenting her results at these.

DATA COLLECTION

1. What types of data will you collect, create, link to, acquire and/or record?
2. 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?
3. What conventions and procedures will you use to structure, name and version-control your files to help you and others better understand how your data are organized?





TIP: FILE TYPES

If possible, DO NOT use proprietary file formats

If you must use a proprietary format, also include an open version (ex. rtf with docx)

TIP: FILE ORGANIZATION/STRUCTURE

- Set up your structure BEFORE beginning the project; modify as needed and based on group discussion

Name	Date modified	Type
 Interview_Data	2020-05-04 2:38 PM	File folder
 NVivo_files	2020-05-04 2:38 PM	File folder
 Policy_Documents	2020-05-04 2:30 PM	File folder
 WHO_dataset	2020-05-04 2:38 PM	File folder

TIP: FILE NAMING

Be Descriptive:

- Unique identifier (ie. Project Name or Grant # in folder name)
- Project or research data name
- Conditions (Lab instrument, Solvent, Temperature, etc.)
- Run of experiment (sequential)
- Date (in file properties too)
- Version #

TIP: FILE NAMING CONVENTIONS

Be Consistent

YYYYMMDD
MMDDYYYY
YYMMDD
MMDDYY
DDMM

MAINTAIN ORDER

TimeDate
DateProjectID
TimeProjectID

**INCLUDE THE SAME
INFORMATION**

Sample00001234
Sample01234
Sample1234

**INCLUDE THE SAME
INFORMATION**

TIP: FILE NAMING CONVENTIONS

Files without a naming convention:

- Test_data_2013
- Project_Data
- Design for project.doc
- Lab_work_Eric
- Second_test
- Meeting Notes Oct 23

Files with a naming convention:

- 20130503_DOEProject_DesignDocument_Smith_v2-01.docx
- 20130709_DOEProject_MasterData_Jones_v1-00.xlsx
- 20130825_DOEProject_Ex1Test1_Data_Gonzalez_v3-03.xlsx
- 20130825_DOEProject_Ex1Test1_Documentation_Gonzalez_v3-03.xlsx
- 20131002_DOEProject_Ex1Test2_Data_Gonzalez_v1-01.xlsx
- 20141023_DOEProject_ProjectMeetingNotes_Kramer_v1-00.docx

DOCUMENTATION AND METADATA

1. What documentation will be needed for the data to be read and interpreted correctly in the future?
2. How will you make sure that documentation is created or captured consistently throughout your project?
3. If you are using a metadata standard and/or tools to document and describe your data, please list here.

METADATA

- Data about your data
- The main purpose of metadata is to facilitate the discovery or search of your data through relevant criteria
- Everything you'd need to recreate your data

TIP: METADATA

- Describe your data as thoroughly as possible
- Use standards as appropriate, ex.
 - DDI - Archiving and Social Science
 - Darwin Core – Biology
 - DIF - Scientific data sets
 - CSDGM - Geographic data
 - TEI - Humanities, social sciences and linguistics
- For data deposit, most data repositories will walk you through minimum metadata, but the more you can add, the better.

WHEN METADATA AREN'T STANDARDIZED

- What's wrong with this (fictional) spreadsheet of ladybugs collected within Alberta?

1	Species	Collection location	Date of collection
2	Convergent lady beetle	51.043540740947094, -114.13092022832767	01-Jul-21
3	<i>Anisosticta bitriangularis</i>	Peter Lougheed Provincial Park	2021-06-30
4	Polkadot ladybeetle	University of Calgary	Aug-21
5	<i>Psyllobora vigintimaculata</i>	Calgary, AB	Aug. 1/21
6	Transverse lady beetle	Drumheller	26 July, 2021
7			

DATA DOCUMENTATION TOOLS

- [README template](#)
- Data documentation initiative: <https://www.ddialliance.org>
- DDI + Excel integration:
<https://www.colectica.com/software/colecticaforexcel/>

STORAGE AND BACKUP

1. What are the anticipated storage requirements for your project, in terms of storage space (in megabytes, gigabytes, terabytes, etc.) and the length of time you will be storing it?
2. How and where will your data be stored and backed up during your research project?
3. How will the research team and other collaborators access, modify, and contribute data throughout the project?

STORAGE RESOURCES AT UCALGARY

- [Research Computing Service](#) for general data storage
- [Secure Computing Platform](#) for sensitive, level 3 or 4, data
- [Other computing resources offered through IT](#)
- [UCalgary Information Security Classification standard](#)

Data Security Classification Levels

1. Public data

- public announcements, telephone directory, published research data, open data

2. Internal Use

- Info where disclosure or loss would be inconvenient but it is not likely to result in financial loss or reputational damage
- Unpublished research data, anonymized or de-identified human subject data

3. Confidential

- Restricted use information where disclosure or loss would cause harm or have to be reported to a regulating organization
- Employment applications, personnel files, intellectual property, passwords

4. Restricted

- Information that is confidential and subject to specific privacy and security safeguards. Problems would have to disclosed to all affected individuals
- Credit card info., health information linked to an identifiable individual, identifiable human subject research data

https://www.ucalgary.ca/it_files/ea/Information%20Security%20Classification%20Standard.pdf

DATA LOSS HORROR STORIES – TOY STORY 2



Link to video: https://youtu.be/8dhp_20j0Ys

TIP: BACKUP

- 3-2-1 Backup strategy:
 - 3 total copies of your data, on at least 2 different mediums, and at least 1 copy offsite
- Keep your anti-virus software up-to-date
- Keep your passwords secure
- Consider using encryption
- Do not store master data copies on personal computers or laptops
- Be mindful of ethics constraints

PRESERVATION

1. Where will you deposit your data for long-term preservation and access at the end of your research project?
2. Indicate how you will ensure your data is preservation ready.
Consider preservation-friendly file formats, ensuring file integrity, anonymization and de-identification, inclusion of supporting documentation.

STUDY: The Availability of Research Data Declines Rapidly with Article Age

- Study on biology data, looked at 516 articles published 1991-2011
- Dataset availability: decreases by 17% per year
- Working email address: decreases by 7% per year

Vines et al., The Availability of Research Data Declines Rapidly with Article Age. 2013. Current Biology, Vol 24(1).

TIP: FILE FORMATS FOR PRESERVATION

IMAGES

- TIFF, PNG, JPG

SPREADSHEETS

- CSV, TSV

VIDEO

- MPG, MOV, AVI

SOUND

- MP3, FLAC

TEXT

- TXT, CSV, PDF/A, ASCII, UTF-8

EBOOKS

- EPUB, PDF/A

DATABASES

- XML, CSV, SQL

RETENTION RULES FOR RESEARCH DATA

- [UCalgary Master Records Retention Schedule](#)
 - [Research Involving Human Subjects](#)
 - [Research Data – Not Involving Human Subjects](#)

Often research data/records must be retained for five years after end of project. May be longer in some cases. Check with the Principal Investigator

SHARING AND REUSE

1. What data will you be sharing and in what form? (e.g. raw, processed, analyzed, final).
2. Have you considered what type of end-user license to include with your data?
3. What steps will be taken to help the research community know that your data exists?

TIP: SHARING CONCERNS

You must ensure the confidentiality of your subjects. Watch for:

- Personally Identifiable Information
- Protected Health Information
- Sensitive Information
- Information that seems generic, but could still be used to narrow information down to an individual

TIP: SHARING CONCERNS

If you can't scrub your data?

- Embargoes
- Technological Access Restrictions
- Data Use Agreements

SHARING: DATA REPOSITORIES

Designed for data

- Secured-stored
- Authoritative copy of your data
- Discoverable through Google, etc.
- Easily citable
- Open access
- Some are free to use, some are not

TIP: DATA REPOSITORIES

- UCalgary PRISM Data: <https://borealisdata.ca/dataverse/calgary>
- Other disciplinary and general data repositories available, see here for good list (scroll down to Disciplinary Repositories):
<https://libguides.ucalgary.ca/c.php?g=395022&p=5066153>
- For a robust, discipline-agnostic repository, I recommend the [Open Science Framework](#)

RESPONSIBILITIES AND RESOURCES

1. Identify who will be responsible for managing this project's data during and after the project and the major data management tasks for which they will be responsible.
2. How will responsibilities for managing data activities be handled if substantive changes happen in the personnel overseeing the project's data, including a change of Principal Investigator?
3. What resources will you require to implement your data management plan? What do you estimate the overall cost for data management to be?

ETHICS AND LEGAL COMPLIANCE

1. If your research project includes sensitive data, how will you ensure that it is securely managed and accessible only to approved members of the project?
2. If applicable, what strategies will you undertake to address secondary uses of sensitive data?
3. How will you manage legal, ethical, and intellectual property issues?
 - See [Creative Commons](#) or [Open Data Commons](#) for choosing a licence

GET HELP WITH DATA MANAGEMENT

Data Management Planning, general research data management info

Library/Jennifer Abel/Heather Ganshorn
research.data@ucalgary.ca

<https://libguides.ucalgary.ca/researchdatamanagement>

Data Storage & Computing During Your Project

Research Computing Services

<https://it.ucalgary.ca/research-computing-services>

Data Sharing/Repositories

Library –

PRISM Data (smaller datasets, 3GB per file)

<https://dataverse.scholarsportal.info/dataverse/calgary>

Federated Research Data Repository (national-level, for large datasets)

<https://www.frdr-dfdr.ca/repo/?locale=en>

QUESTIONS?

Email: research.data@ucalgary.ca
heather.ganshorn@ucalgary.ca

Library Guide:
<https://libguides.ucalgary.ca/researchdatamanagement>