Learning outcomes

By the end of this session you should be able to:

1. *Understand why research data management plans are important*
2. *Be able to identify the benefits and challenges of data management*
3. *Know what resources UVic Libraries has to support you*
Outline

• What is a data management plan?
• Why is it important?
• Components of a data management plan
• UVic Library supports
Data management plans

• Describe how data are collected, formatted, preserved and shared, as well as how existing datasets will be used and what new data will be created
Why? Research Data Management Plans (and the data itself) can…

- Satisfy grant & journal requirements
- Satisfy research ethics board requirements
- Ensure data access and longevity
- Increase research efficiency
- Promote collaboration & maximize transparency
- Promote inquiry and innovation
- Raise researcher(s) profile(s) and increase impact of research
- Provide greater resources for education and training
Background - USA

From Developing data services: a tale from two Oregon universities -
http://www.slideshare.net/amandawhitmire/20140618-rml-rendezvousfinal
Canadian context – major players

Science.gc.ca

RDC DRC
Research Data Canada - Données de Recherche Canada

Canadian Institutes of Health Research
Natural Sciences and Engineering Research Council of Canada
Social Sciences and Humanities Research Council of Canada

compute Canada
calcul Canada
3. Expectations

Data management planning is necessary at all stages of the research project lifecycle from design and inception to completion.

4. Responsibilities

Responsibilities of researchers include:

“Incorporating data management best practices into their research, and developing data management plans to guide the responsible collection, formatting, preservation and sharing of their data throughout the entire lifecycle of a research project and beyond….”

Seven questions:
3. What do you see as the biggest challenges to effective data management and the development of data standards in Canada?

4. What is the current capacity within post-secondary institutions to support research data curation?
Issues still to be hammered out

1. Capacity within post-secondary institutions to support research data curation…
   “Researchers do not have the research funding to support data curation….Likewise, institutions do not have funding to support data curation…”

2. Share data throughout the entire lifecycle of a research project and beyond…
   “what data must be preserved and for how long?”

3. Identify and encourage use of specific repositories and platforms
   “not clear what a researcher’s or an institution’s responsibility is…”

Comments from D. Michael Miller (Uvic), Assoc. Vice-President Research; Co-chair Research Computing Advisory Committee
“The purpose of this statement of principles is not to create one national RDM policy, nor to replace existing institutional policies, but rather to try to ensure that there is a common core to the principles governing RDM among Canada's universities.”

The principles

1. The importance of data for research
2. National and international collaboration
3. Access
4. Ethical, legal, and privacy issues
5. Privileged use
6. Recognition of intellectual contributions
7. A public trust
8. Data management plans – follow int’l standards and community best practices
9. Metadata and discoverability
10. Multilingual access
Research data lifecycle

• Managing your research data occurs at each stage of your research project.
Video

A data management horror story

by Karen Hanson, Alisa Surkis and Karen Yacobucci (NYU Health Sciences Libraries)

https://youtu.be/N2zK3sAtr-4
Reflect on your research

• If you were asked to share your data with another researcher would they be able to make sense of your data?
• If you needed to locate your data files from 5 years ago, how easy would they be to find and use?
The Data Management Plan: Common Misconceptions

• Does *not* require that *all* data must be shared
  – Sensitive information/patient privacy
  – Intellectual property rights and commercial value

• Sharing can take many forms

• Funders recognize that different disciplines have different “cultures” of data sharing

• Sharing “at no more than incremental costs and within a reasonable time.”
Portage (data management planning assistant)

Seven sections:
1. Data collection
2. Documentation & metadata
3. Storage & backup
4. Preservation
5. Sharing & re-use
6. Responsibilities & resources
7. Ethics & legal compliance

With several questions to answer in each section

https://portagenetwork.ca/
1. Data collection

• What kinds of data do you collect/generate in your research?
• What file formats will you use?
• What conventions will you use to structure your data files?
• What constitutes a (distinct) dataset?
  – Location
  – Occurrences
  – Time period?
2. Documentation & metadata

**Descriptive:** title, author, keywords

**Administrative:** information needed to use the data, e.g. software requirements, copyright

**Structural:** how different data sets relate to one another, e.g. file formats, information about the relationship between data sets in a database
Disciplinary Metadata

While data curators, and increasingly researchers, know that good metadata is key for research data access and re-use, figuring out precisely what metadata to capture and how to capture it is a complex task. Fortunately, many academic disciplines have supported initiatives to formalise the metadata specifications the community deems to be required for data re-use. This page provides links to information about these disciplinary metadata standards, including profiles, tools to implement the standards, and use cases of data repositories currently implementing them.

For those disciplines that have not yet settled on a metadata standard, and for those repositories that work with data across disciplines, the General Research Data section links to information about broader metadata standards that have been adapted to suit the needs of research data.

Please note that a community-maintained version of this directory® has been set up under the auspices of the Research Data Alliance.

Search by Discipline

Biology

Earth Science

General Research Data

Physical Science

Social Science & Humanities

More at: http://www.dcc.ac.uk/drupal/resources/metadata-standards
3. Storage & backup

• Anticipated storage requirements
• Length of time for storage
• Where?
  – 3-2-1 rule:
    • 3 copies
    • 2 different media
    • 1 backup offsite

• Security
  – Passwords
  – Data encryption
  – Room secure
  – Network firewalls, etc…
4. Preservation

• For archiving/preservation, convert to or export to non-proprietary formats (e.g.)
  – ASCII: .txt
  – Comma-separated values: .csv
  – Tab-delimited: .tsv or .tab
  – Images: .tiff

UK Data Service:
https://www.ukdataservice.ac.uk/manage-data/format/recommended-formats
Preservation – file naming guidelines

File names/titles:
• No spaces or special characters
• Use _ as a delimiter
• Use descriptive file names
• Include dates (int’l standard YYYY_MM_DD or YYYYMMDD)
• Include version numbers
• Be consistent!

File and variable naming guidelines:
• https://www.ukdataservice.ac.uk/manage-data/document/
• https://www.ukdataservice.ac.uk/manage-data/document/data-level/tabular
5. Sharing and re-use

• What data will you be sharing and in what form?
  – Raw
  – Processed
  – Analyzed
  – Final

• What type of end-user license (if any) do you need for your data?
  – Creative Commons?
  – Open Data?

• What steps will you take to let the research community know your data exists?
Sharing – Locating research data repositories

- Browse by content type
- Browse by subject
Sharing and re-use

• Obtain a DOI (digital object identifier) for data
• Why?
• Provides a persistent link to facilitate data visibility
• Enable tracking of citations to provide proof of research impact

Sharing and re-use

Why cite data?
– Facilitate re-use and verification of data
– Enable impact of data to be tracked
– Create scholarly structure that recognizes and rewards data producers

When citing data, include:
– Author/creator
– Date created
– Title
– Institution/organization
– Identifier (DOI or handle)

Author. Date. Title of dataset. Institution/Organization. DOI

https://www.datacite.org/services/cite-your-data.html

*No official format for citing data; some journals and conferences have established data citation guidelines
Scholarly journals

Access to Archives of Scientific Psychology® Data

The Archives of Scientific Psychology® Review Committee is responsible for approving all research applications requesting access to the datasets stored at the Interuniversity Consortium for Political and Social Research (ICPSR) in connection with articles published in Archives of Scientific Psychology (hereafter, Archives).

Authors whose manuscripts are accepted for publication in Archives deposit the data that underlie the analyses presented in their manuscripts to ICPSR. However, only those variables used in the analyses reported in the published article are initially made available to qualified next-users. Contact the Archives Review Committee to confirm the scope of the available data.

An Open Methodology, Collaborative Data Sharing, Open Access Journal.

Editorial: Archives of Scientific Psychology: A New Journal for a New Era

http://openpsychologydata.metajnl.com/about/

Recommended Repositories

The following repositories meet our peer-review requirements and are recommended for the archiving of JOPD datasets. Please contact us if you would like to use another repository or recommend that we add it to our list.

International repositories
- JOPD Dataverse
- Dryad
- Figshare
- OpenfMRI
- Zenodo

National repositories
- DANS (Netherlands)
- Gesis (Germany)
- FORS (Switzerland)
- Odum (United States)
- SND (Sweden)
- TARKI (Hungary)

Institutional repositories
- UCL Discovery
6. Responsibilities & resources

• Who will be responsible for managing data during and after the research?
• Who will be responsible if personnel changes happen?
Ethics & legal compliance

- How will you ensure sensitive data are securely managed and accessible only to approved members of the project?
- What strategies for secondary use of sensitive data?
How can UVic Libraries help?

• Portage (DMP assistant)
• Dataverse
• Research Data guide
  http://libguides.uvic.ca/rdmp
Acknowledgements


• *UBC Library DataGuide*, (version 3.1, October 2015) by Eugene Barsky, UBC Research Data Librarian

• My colleagues at UVic:
  – Kathleen Matthews and Katy Nelson (Sci & Eng)
  – Daniel Brendle-Moczuk (Geog and Data)
  – Rebecca Raworth (Island Med Program)
  – Lisa Goddard (AUL, Digital Scholarship & Strategy)
Thank you!

Jessica Mussell, BA, MLIS: jmussell@uvic.ca