Introduction to Research Data Management Planning
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
Background - Canada

2004: Declaration on Access to Research Data from Public Funding (OECD)

2013: Consultation document: Toward a Policy for Advancing Digital Scholarship in Canada

2014: Portage data management planning assistant (CARL)

2015: Draft Tri-Agency Statement of Principles on Digital Data Management

March 2016: Research Data Management in Canadian Universities: A Statement of Principles (RDC)

2016: Tri-Agency Statement of Principles on Digital Data Management
Canadian context – major players
Data Management

Tri-Agency Statement of Principles on Digital Data Management

High-quality data management is fundamental to research excellence. The ability to store, access, reuse and build upon digital data is critical to the advancement of research, supports innovative solutions to economic and social challenges, and holds tremendous potential for improvements in quality of life nationally and internationally.

The three federal research funding agencies – 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) – are committed to fostering a robust environment for data stewardship in Canada and internationally, and to helping ensure that Canada is well positioned to contribute to and capitalize on data-intensive science and scholarship. To achieve this, and while recognizing the diversity of data practices and needs within the Canadian research community, the agencies are in the process of reviewing and enhancing their data management requirements for agency-supported researchers.

As a step in this process, the agencies have developed the Tri-Agency Statement of Principles on Digital Data Management (the Statement). This builds on the 2013 consultation document, Toward a Policy Framework for Advancing Digital Scholarship in Canada, information provided by a Comprehensive Brief on Research Data Management Policies, and advice received through stakeholder engagement. The agencies sought feedback from the research community in summer 2015 on a draft version of the Statement.

The Statement of Principles on Digital Data Management outlines the agencies’ overarching expectations with regard to digital research data management, and the responsibilities of researchers, research communities, institutions and funders in meeting these expectations. It complements and builds upon existing agency policies, and will serve as a guide to assist the research community in preparing for, and contributing to the development of, Tri-Agency data management requirements. The HTML version is accessible here, and the PDF version is accessible here.

As the context for research data management evolves, the agencies – in consultation with the stakeholders for research data management in Canada – will review and revise the Statement as appropriate.

Over the coming months, the agencies will engage the research community to seek advice on realizing the principles expressed in the Statement, particularly with a view to developing new data management requirements.

“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.”
Issues still to be sorted 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 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 Misconception

• *All* data must be shared
  – Sensitive information/patient privacy
  – Intellectual property rights and commercial value
  – Sharing can take many forms
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/)
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: Locally-hosted data repository

http://dvn.library.ubc.ca/dvn/dataverses/UVIC_MAIN/faces/site/BrowseDataversesPage.xhtml?groupId=3
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 data journals

Examples include:

• *Journal of Open Archaeology Data* - published by Ubiquity (2011)
• *Geoscience Data Journal* - published by Wiley (2012)
• *Scientific Data* - published by Nature (2013)
• *Biodiversity Data Journal* – published by Pensoft (2013)
• *Archives of Scientific Psychology* - published by APA (2013)

Citation counts for tracking impact of your data
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?

1. UVic librarians
2. Portage (DMP assistant) – create your data management plan
3. Dataverse – store and share your data sets and get a DOI for your data
4. Research Data Management guide – lots of data management information can be found here
Acknowledgements


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

• Uvic Libraries’ Research Data Task Group:
  – Daniel Brendle-Moczuk (Geog & Data)
  – Aditi Gupta (Engineering)
  – Kathleen Matthews (Science)
  – Jessica Mussell (Psyc & Public Admin)
  – Rebecca Raworth (Health Sciences)
  – Lisa Goddard, Group Chair (AUL, Digital Scholarship & Strategy)
  – CLIR Data Curation Postdoctoral Fellow, Jacqueline Quinless (joining us Jan 2017)