UVIC LIBRARIES

Research Data Management Workshop

Presented by Kathleen Matthews, Science Librarian

January 27, 2016

http://libguides.uvic.ca/rdmp
Workshop Objectives… Why Now?


- Emergence of *Data Journals*, DOIs, Data Repositories, Metadata standards.
Data Sharing and Management Snafu in 3 Short Acts
http://youtu.be/N2zK3sAtr-4
Overview of Best Practices in Research Data Management

- Adapted from Best Practices for Research Data Management PPT, original prepared in December 2014, by Eugene Barsky, Sally Taylor and Jennifer Abel, UBC Library

- Excerpts from Research Data Management Guide, Version 2.8, by Eugene Barsky, UBC Research Data Librarian
Take a Moment....to think about your research.

What kinds of data do you generate?

- Where is your data stored and how is it organized?
- If you were asked to share your data with another researcher would they be able to make sense of your work?
- If you needed to locate your data files from 5 years ago, how easy would they be to find and use?
Research Data Management Topics

- Workshop Modules:
  - RDM Basics & Data Management Plans
  - Metadata and organizing your data
  - Data storage and security
  - Data sharing and re-use

- Goal: To help researchers better manage their research data.

Image - https://www.flickr.com/photos/34547181@N00/
Module 1: Research Data and Data Management Plans

- Research data can be a great many things, from DNA samples to interview transcripts to photographs.

- What kinds of data do you generate in your research? What file formats?

Image - https://www.flickr.com/photos/gsfc
Why Data Management?

Proper data management can make it easier for you to:

● Keep track of files & manage versions

● Compile information at project end

● Reproduce your work

● Share your data

● Satisfy research ethics board, grant & journal requirements

Image - https://www.flickr.com/photos/kenfagerdotcom/
Research Data Lifecycle

- Managing your research data occurs at each stage of your research project.

- A look at the life cycle of research data - http://www.data-archive.ac.uk/create-manage/life-cycle

- GOOGLE search "Data Life Cycle" images. This is one result set!

- DataGuide, page 4

Image - https://www.flickr.com/photos/gspragin/
Data Management Plan

● You need a plan…

● Data Management Plan

● A data management plan is a document that helps you manage your research data by planning out what you will do with your data during and after your research (DataOne, 2012).

● Federal requirements emerging…
Federal Mandates – US Data Management Plan Requirements

1985: National Research Council
2003: NIH Data Sharing Policy
2011: NSF DMP requirement
2013: NSF bi-sketch change

1999: OMB Circular A-110 revisions
2008: NIH Public Access Policy
2012: NEH, Office of Digital Humanities DMP requirement
2013: OSTP memo on public access to results of federally funded data

From Developing data services: a tale from two Oregon universities -
http://www.slideshare.net/amandawhitmire/20140618-rml-rendezvousfinal
In Canada

- Data management plans are an expectation.


“This statement of principles outlines the Agencies’ overarching expectations regarding research data management, and the responsibilities of researchers, research communities, research institutions and research funders in meeting these expectations.”

Image - https://www.flickr.com/photos/twosevenoneonineighthreesevenatenzerosix/
Typical Data Management Plan

A typical plan includes information about:

- Types of data produced
- Metadata or documentation standards
- Data security and encryption
- Data storage
- Intellectual property rights
- Data sharing
- Data archiving

Image - https://www.flickr.com/photos/cross_stitch_ninja/
Online Tools

- In Canada - DMP Assistant: https://portagenetwork.ca
- In US (for NSF, NIH, NOAA, etc) - Data Management Planning Tool: https://dmptool.org
- UK Data Archive Checklist http://data-archive.ac.uk/create-manage/planning-for-sharing/data-management-checklist

Image - https://www.flickr.com/photos/derpunk/
Module 2: Metadata and Data Organization

- Metadata is often described as “data about data”

Image - https://www.flickr.com/photos/centralasian/
Metadata

**Descriptive**: content and context of your data at both the dataset and item level.
Examples: title, author, keywords

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

**Structural**: how different data sets relate to one another
Examples: Information about the relationship between data sets in a database, file formats

Image - [https://www.flickr.com/photos/wakingtiger](https://www.flickr.com/photos/wakingtiger)
Metadata Standards

- Many disciplines have metadata standards: e.g. Dublin Core, Darwin Core, DDI

- List of standards in your field - http://www.dcc.ac.uk/resources/metadata-standards

Image - https://www.flickr.com/photos/pamilne/
Best practices in data documentation

- At the very least you should document your data in a `readme.txt` file stored alongside your data:
  - context of data collection (the goal of your research)
  - data collection methods (protocols, sampling, instruments, coverage…)
  - structure of files
readme.txt file….

- sources used
- quality assurance (data validation, checking)
- data modifications
- confidentiality and permissions
- names of labels and variables
- explanations of codes and classifications
- DataGuide, page 8 & 9
More Best Practices in data documentation

- Don’t wait to document your data! If you wait until the end of your project you might lose valuable information!
- Avoid jargon & symbols … Use keywords
- Define terms and acronyms
- State limitations
- Use descriptive titles
- Be specific and quantify
File Naming Conventions

- Keep file names under 32 characters
- Avoid spaces & special characters. No periods.
- Use descriptive file names. Use underscores.
- Include dates (international standard YYYY_MM_DD or YYYYMMDD)
- Include version numbers
- Be consistent! DataGuide, page 9 - Downloadable instructions http://researchdata.library.ubc.ca/organize/
Module 3: Data Storage and Security

- Essential parts of your plan
- A best practice is to have **three copies** stored in **at least two locations**
- DataGuide, page 11-13

Image: https://www.flickr.com/photos/hi-phi/14699924741
Data Storage and Security

- Where do you store your data?
- Do you secure it?
- Attach metadata alongside stored data
- Data can be lost for any number of reasons

Image - https://www.flickr.com/photos/ian-s/
File Formats

- Format recommendations:
  - Non-proprietary
  - Open, with documented standards
  - Used by your community
  - Encoded using standard character encoding
  - Uncompressed

- Downloadable instructions –
  [http://researchdata.library.ubc.ca/format/](http://researchdata.library.ubc.ca/format/)

- DataGuide, page 11-12

Image - [https://www.flickr.com/photos/chiselwright/](https://www.flickr.com/photos/chiselwright/)
Data Security

- Network security: who has access to the network? Are there firewalls?

- Physical security: who has access to the computers? Who can access physical files? How is data transported?

- Computer security: anti-virus software? power surges? passwords and firewalls? data encryption? data storage secure?

Image - https://www.flickr.com/photos/toxi/
Module 4: Data Sharing and Re-use

Why share data?

● Transparency and integrity

● Promote innovation and collaboration

● Required by funding agencies or journals:
  o PLOS - http://goo.gl/mP0JBS
  o JDAP - http://datadryad.org/pages/jdap

Image - https://www.flickr.com/photos/developmentseed/
Sharing Research Data can…

- Satisfy journal & grant requirements
- Promote scholarly rigor & increase research efficiency
- Make research more open & raise researcher profiles
- Promote collaboration & maximize transparency
- Promote inquiry and innovation
- Increase economic & social impact of research
- Provide greater resources for education and training
And…. If that isn’t incentive enough…

Grantees are required to deposit their data in publically accessible repositories by….

- Canadian Social Sciences and Humanities Research Council (SSHRC)
- Canadian Institutes of Health Research (CIHR)
Challenges to Sharing Data

- Privacy or ethical issues
  - Compliance with institutional regulations
  - Anonymize data

- Intellectual property issues (e.g. third-party data, co-authors)

- Practical issues (e.g. metadata, technology)

Image - https://www.flickr.com/photos/blpmnt/
Sharing data in an ethical manner

- evaluate the anonymity of your data
- obtain a confidential review (repository may look it over)
- comply with institutional regulations (research ethics, HREB)
- comply with other regulations (ORS, BREB, HIPAA)
- have informed consent for data sharing
- restrict use of confidential data
How to Share Data

● Find a home for your data:
  o Subject specific repository or archive
  o Institutional repository or archive
  o Journal website
  o Project website

● License your data (e.g. Creative Commons or Open Data Commons)

● Provide suggested data citation

● DataGuide, page 15

Image - https://www.flickr.com/photos/arenamontanus/
Repositories for Data Deposit & Sharing

● To find a data repository for your discipline
  ● re3data.org

● Consider:
  ● Who may want access?
  ● Discipline specific repository?
  ● Access policies?
  ● Preservation plan?
  ● What kind of data is accepted?
  ● Metadata standards?
  ● Fees?

Image - https://www.flickr.com/photos/jwyg/
Licencing Data

- To protect copyright & allow access
- Ensure you have permission to license the data
- Creative Commons and Open Data Commons Licenses
  - Creative Commons zero license CC0 – like a waiver
- Digital Curation Centre guide to licensing research data at http://www.dcc.ac.uk/resources/how-guides/license-research-data
- DataGuide - page 16
Citing data

- How to cite data?

- No official format for citing data, many journals and conferences have established data citation rules.

- Generally, it is a good idea to include the following information:
  - Author/creator
  - Date created
  - Title
  - Publisher
  - Identifier (e.g. DOI or handle)

- DataCite Canada – DOIs for data sets
How can we help?

- Host and manage your data in Dataverse
  [UVic’s Research Data Repository hosted at UBC]
  http://dvn.library.ubc.ca/dvn/

- Digital preservation & access for your data within UVicSpace

Image - https://www.flickr.com/photos/98216330@N00/
How can we help?

- Help you with **metadata** (find right standard, etc)
- Help you with **DMPs** - DMP Assistant at [https://portagenetwork.ca](https://portagenetwork.ca)
- Research Data Subject Guide [http://libguides.uvic.ca/rdmp](http://libguides.uvic.ca/rdmp)

THANK YOU… And Now…

Image - [https://www.flickr.com/photos/danielygo/](https://www.flickr.com/photos/danielygo/)
Hands-on with DMPbuilder

Intro to Data Management Planning Tools

- Types of data produced
- Metadata or documentation standards
- Data security and encryption
- Data storage
- Intellectual property rights
- Data sharing
- Data archiving
Online DMP Tools

● In Canada - DMP Assistant: https://portagenetwork.ca

● In US (for NSF, NIH, NOAA, etc) - Data Management Planning Tool: https://dmptool.org

● UK Data Archive Checklist http://data-archive.ac.uk/create-manage/planning-for-sharing/data-management-checklist
Demo of DMPbuilder

- Online Data Management Planning Tool for Canadian Researchers
  [https://dmp.library.ualberta.ca/](https://dmp.library.ualberta.ca/)

- Hands-on Activity - Register with the DMPtool & explore