Best Practices for Research Data Management (RDM) - Using Dataverse for Research Data

November 2016

Eugene Barsky, UBC
eugene.barsky@ubc.ca
http://researchdata.library.ubc.ca/

Image by https://www.flickr.com/photos/40032755@N06/
Outline

- General overview of RDM
- Tri-Agency approach to RDM
- Best Practices for RDM
- Focus on Dataverse

You will also need UBC Library DataGuide (version 4.1, July 2016) – a quick guide we have developed to get you started on Research Data Management.
Data

Information

Presentation

Knowledge

Image - http://epicgraphic.com/metaphors/
Data rich

Race teams at the U.S. Grand Prix collected more than 243 TB of data in 2015, a few terabytes more data than there are in the Library of Congress.

Image - https://www.flickr.com/photos/caterhamf1/
Data rich

- Feed of about 150 key parameters, which is transmitted in real time with a 4G connection

- 5 GB of raw data per boat per day

Image - https://www.flickr.com/photos/markdoliner/
Data poor...

- APGAR test

Image - https://www.flickr.com/photos/hippie/
Image - http://goo.gl/H96iv0
What issues are raised in this video?

Data Sharing and Management Snafu in 3 Short Acts

http://youtu.be/N2zK3sAtr-4
Why Data Management

Proper data management can make it easier for you to:

- Find your files
- Keep track of different versions of your data
- Organize and compile information at the end of a project
- Reproduce your work (if required for a journal or patent)
- Pass on your work to another researcher
- Share your work
- Satisfy grant requirements
- Satisfy research ethics board and institutional requirements
- Satisfy journal requirements

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

- Let’s take a look -
  http://www.data-archive.ac.uk/create-manage/life-cycle

Image - https://www.flickr.com/photos/gspragin/
Federal Mandates in the U.S.

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...

- **Tri-Agency Statement of Principles on Digital Data Management** - June 2016

And you know about the **Tri-Agency Open Access policy**, right? - Feb 2015

Image - [https://www.flickr.com/photos/twosevenoneonenineeightthreesevenatenzerosix/](https://www.flickr.com/photos/twosevenoneonenineeightthreesevenatenzerosix/)
Tri-Agency expectations for RDM

Researchers:

- Incorporate RDM **best practices** (in their discipline)

- Develop **Data Management Plans** (DMPs)

- Follow institutional **policies** and standards
Tri-Agency expectations for RDM

Funders

- Develop **policy** and requirements that facilitate responsible data management
- Provide clear guidance for fulfill RDM requirements
- **Promote** the importance of excellent RDM
- Provide **peer-reviewers** with guidance for applications assessment

Image - [https://www.flickr.com/photos/sonson/](https://www.flickr.com/photos/sonson/)
Tri-Agency expectations for RDM

Institutions

- Provide researchers access to repositories that securely preserve, curate and provide access to research data
- Provide researchers with guidance to properly manage their data, including DMPs
- Promote the importance of research data management to researchers, students and staff
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://assistant.portagenetwork.ca/](https://assistant.portagenetwork.ca/)  
  (CAF authentication coming soon)

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

Image - [https://www.flickr.com/photos/derpunk/](https://www.flickr.com/photos/derpunk/)
Metadata and Documentation

● Metadata is often described as “data about data”

● It helps answer the questions:
  who?
  what?
  where?
  when?
  why?

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

Image: https://i.ytimg.com/vi/MsKDq_SIJOA/maxresdefault.jpg
File Naming

- Downloadable instructions – [http://researchdata.library.ubc.ca/organize/](http://researchdata.library.ubc.ca/organize/)
- DataGuide, page 9

[Image - http://library.umassmed.edu/necdmc/modules](http://library.umassmed.edu/necdmc/modules)
**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. DDI, Dublin Core, Darwin Core

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

Image - https://www.flickr.com/photos/pamilne/
Data Documentation

- At the very least you should document your data in a `readme.txt` file stored alongside your data
- DataGuide, page 8

Image - https://www.flickr.com/photos/karola/
Data Storage and Security

- Where do you store your data?
- Do you secure it?
Data Storage

3-2-1 backup rule:

- Have at least **three** copies of your data
- Store the copies on **two** different media
- Keep **one** backup copy offsite

- DataGuide, page 11

Image: https://www.flickr.com/photos/hi-phi/14699924741
What UBC provides

- [WestGrid Data Storage](https://www.flickr.com/photos/pascalcharest/) network
- UBC’s [Storage Grid](https://www.flickr.com/photos/pascalcharest/)
- UBC’s [Workspace 2.0](https://www.flickr.com/photos/pascalcharest/)
- Talk to your IT admins!
File Formats

- Downloadable instructions – http://researchdata.library.ubc.ca/format/

Try to stay with:
- Non-proprietary
- Open, with documented standards
- Used by your community
- Encoded using standard character encoding
- Uncompressed

- DataGuide, page 11-12

Image - 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**  Is anti-virus software up to date? Are you protected against power surges? Do you use passwords and firewalls? Is data encrypted? Is data storage secure?

- **UBC IT encryption service**
Data Sharing and Re-use

Why share data?

- Transparency and integrity
- Promote innovation and collaboration
- Required by funding agencies or journals:
  - PLoS - [http://goo.gl/mP0JBS](http://goo.gl/mP0JBS)
  - JDAP - [http://datadryad.org/pages/jdap](http://datadryad.org/pages/jdap)

Image - [https://www.flickr.com/photos/developmentseed/](https://www.flickr.com/photos/developmentseed/)
Challenges to Sharing Data

● Privacy or ethical issues
  ○ Compliance with institutional (e.g. BREB) or other regulations
  ○ De-identify data
  ○ checklist from UBC ARC

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

Image - https://www.flickr.com/photos/blprnt/
How to Share Data

- **UBC Abacus Dataverse**
  - License
  - Data Citation
  - DOIs
  - Version control
  - Integrity checks
  - Preservation
  - Granular access for sensitive data

- **Other repositories** - [re3data.org](https://re3data.org)

Image - [https://www.flickr.com/photos/jwyg/](https://www.flickr.com/photos/jwyg/)
Dataverse - Live presentation

20-30 min of live presentation, driving [UBC Abacus Dataverse](#):

- Create and change records
- Metadata, types of metadata, standards
- Uploading files, large files, zipping
- Version control
- Tabular analysis in your browser
- Granular access to datasets: public, institutional, groups
- UNFs for data analysis
- OAI for discoverability
How can we help?

- Host and let you **manage** your data in [UBC Abacus Dataverse](https://assistant.portagenetwork.ca/)

- Help you with **DMPs** - [https://assistant.portagenetwork.ca/](https://assistant.portagenetwork.ca/)

Image - [https://www.flickr.com/photos/danielygo/](https://www.flickr.com/photos/danielygo/)
Questions?

Image - https://www.flickr.com/photos/debord/