Research Data Management Plan(ning)

• Why?

• Background

• Why now?

• Research data life cycle

• Components of a data management plan
Research Data Management Plans (and the data itself) can...

- satisfy grant & journal requirements
- ensure data access and longevity
- endorse scholarly rigor
- increase research efficiency
- raise researcher(s) profile(s)
- promote collaboration & maximize transparency
- promote inquiry and innovation
- increase impact of research
- provide greater resources for education and training

WHY?
Background
USA

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

US National Science Foundation (January 2011)

- Data Sharing Policy
  ...expected to share with other researchers...the primary data, samples...

- Data Management Plan requirements
  ...Proposals...must include a “Data Management Plan”
3. Expectations:

**Data management planning** is necessary at all stages of the research project lifecycle from **design** and **inception** to **completion**.
3. Expectations: (con’t)

Data management plans: key element of the data management planning process

How data is:
- collected
- formatted
- preserved
- shared

and what new data will be created…
4. Responsibilities of researcher:
develop data management plans to guide the...
- collection
- formatting
- preservation
- sharing of their data
throughout the entire lifecycle of a research project…
…and beyond…
Seven questions:

3) …biggest challenges to effective data management…

4) …capacity within post-secondary institutions to support research data curation…
UVic “responses” by: D. Michael Miller, Assoc. Vice-President Research; Co-chair Research Computing Advisory Committee

4) ...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...”

Dr. Miller points to Canadian Association of Research Libraries (CARL) response
4) …share data throughout the entire lifecycle of a research project and beyond…

(2.1) “what data must be preserved and for how long?”

4) …Identify and encourage use of specific repositories and platforms

(2.3) “not clear what a researcher’s or an institution’s responsibility is…”
Research Data Lifecycle

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

Data Sharing and Management Snafu in 3 Short Acts
http://youtu.be/N2zK3sAtr-4
major players in Canada

Science.gc.ca

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

Canadian Association of Research Libraries-CARL
DMP Assistant is a bilingual tool for preparing data management plans (DMPs). The tool follows best practices in data stewardship and walks researchers step-by-step through key questions about data management.
Data Management Plan(ning) tool

Seven sections:

- Data Collection
- Documentation and Metadata
- Storage and Backup
- Preservation
- Sharing and Reuse
- Responsibilities and Resources
- Ethics and Legal Compliance

with several questions to answer in each section
Responsibilities and Resources
(not the first section of DMP Portage but perhaps most important…)

• Who will be responsible for managing data during and after the research?

• Who will be responsible if personnel changes happen?
Research Data

Research data are many things, from interview transcripts, observations/occurrences, photographs, samples to survey responses, etc.

- What kinds of data do you collect / generate in your research?
- What file formats?
- What conventions will you use to structure your data files?

- What constitutes a (distinct) “dataset”?  
  - location
  - occurrences
  - time-period
  - etc?
Research Data

types of data:

• audio

• imagery

• numeric

• geo-spatial

• textual

• etc!
Research Data

for archiving/preservation, (convert or export your data to…)

non-proprietary file formats:

• ASCII .txt

• comma-separated values: .csv

• JavaScript Object Notation: .json; .geojson

• OpenDocument: .odf

• tab-delimited: .tsv; .tab

• .tiff

• XML: .xml
Research Data:
for archiving/preservation,
(convert or export your data to…)

recommended File formats / types

https://www.ukdataservice.ac.uk/manage-data/format/recommended-formats

Research data

File names / titles

- No spaces & special characters
- Use _ as delimiter
- Use descriptive file names
- Include dates (international standard YYYY_MM_DD or YYYYMMDD)
- Include version numbers
- Be consistent!
Documentation and Metadata

File(s) and Variable(s) name guidelines

https://www.ukdataservice.ac.uk/manage-data/document/

https://www.ukdataservice.ac.uk/manage-data/document/data-level/tabular

https://www.dataone.org/best-practices/assign-descriptive-file-names
Documentation and Metadata

Data about the data...

answers the questions:
• who?
• what?
• where?
• when?
• why?
• how?
Documentation and Metadata

Document:
• classification systems
• data capture and collection methods
• units of measure
• data coding (including syntax files)
• format and file type of the data
• research methodology
• analysis methods
• Who did what when ?!
• etc!
Documentation and Metadata

some (major) Metadata standards
Documentation and Metadata

Metadata standards for ecology

**ABCD - Access to Biological Collection Data**
A standard for the access to and exchange of primary biodiversity data, including specimens and observations.

**Darwin Core**
A body of standards, including a glossary of terms (in other contexts these might be called properties, elements, fields, columns, attributes, or concepts) intended to facilitate the sharing of information about biological diversity by providing reference definitions, examples, and commentaries.

**EML - Ecological Metadata Language**
Ecological Metadata Language (EML) is a metadata specification particularly developed for the ecology discipline.

More at [http://www.dcc.ac.uk/drupal/resources/metadata-standards](http://www.dcc.ac.uk/drupal/resources/metadata-standards)
Storage and Backup

• anticipated storage requirements

• length of time for storage

• Where?
  • “3 – 2 – 1 rule”
    • 3 copies
    • 2 different media
    • 1 backup offsite

• versioning (different updated versions)
Storage and Backup

Security

- Anti-virus software
- Passwords
- Data encryption

- Room secure

- Who has access to the network? Firewalls

- Power surges

- How is data transported?
Preservation

- preservation ready data
  - preservation-friendly file formats (non-proprietary)

- Where for long-term preservation?

- include supporting documentation!
Sharing and re-use

• what data will you be sharing and in what form?
  • raw
  • processed
  • analyzed
  • final

• what type of end-user license for your data?
  • Creative Commons
  • Open Data

• steps to let the research community know your data exists
Sharing and re-use

Obtain a DOI (Digital Object Identifier) for data

• Why a DOI for data?
  • provide persistent link for articles, presentations, websites, etc to facilitate data visibility
  • enable tracking of data citations to provide proof of research impact

Sharing and re-use

• Why cite data?
  • facilitate reuse and verification of data
  • enable impact of data to be tracked
  • create scholarly structure that recognises and rewards data producers

https://www.datacite.org/services/cite-your-data.html
Sharing and re-use

Cite your data:

- Include:
  - Author / Creator
  - Date created
  - Title
  - Institution / Organisation
  - Identifier (DOI or handle)

Author. Date. *Title of dataset*. Institution/Organisation. DOI

(No official format for citing data; some journals and conferences have established data citation guidelines)
Responsibilities and Resources (again!)

(Who will be responsible for managing data during and after the research? Who will be responsible if personnel changes happen?)

Now that you’ve considered numerous questions...

• What resources will you need for your data management?
  • hardware
  • software
  • personnel (data manager)
  • Physical space
  • etc?

• What do you estimate the overall cost for data management to be?
  https://www.ukdataservice.ac.uk/manage-data/plan/costing
Ethics and Legal compliance

• How will you ensure sensitive data is securely managed and accessible only to approved members of the project?

• What strategies for secondary use of sensitive data?
Best Practices

on-line guides

https://www.ukdataservice.ac.uk/manage-data

http://www.data-archive.ac.uk/media/2894/managingsharing.pdf

https://www.dataone.org/best-practices

http://www.icpsr.umich.edu/files/ICPSR/access/dataprep.pdf
How can we help?

- select metadata standard
- metadata itself
- DMP Assistant at [https://portagenetwork.ca](https://portagenetwork.ca)
- Research Data Subject Guide [http://libguides.uvic.ca/rdmp](http://libguides.uvic.ca/rdmp)

kmatthew@uvic.ca
danielbm@uvic.ca
How can we help?

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

- Digital preservation & access for theses / dissertations via UVicSpace