Library Research Data Management Seminar

BIOL 565

Thursday, September 12, 2018

Shahira Khair
Data Curation Librarian
skhair@uvic.ca

Kathleen Matthews
Biology & Forest Biology Subject Librarian & Sciences Research Data Librarian
kmatthew@uvic.ca
Learning Outcomes

By the end of this workshop, participants will have learned about:

1. The phases of the research data lifecycle
2. The importance of good research data management
3. Data management plans and the DMP Assistant
4. Services and tools offered by the UVic Libraries to support students and researchers in managing their data
Online Guide

http://libguides.uvic.ca/researchdata
What are research data?
Research Data

Data that are used as primary sources to support technical or scientific enquiry, research, scholarship, or artistic activity, and that are used as evidence in the research process and/or are commonly accepted in the research community as necessary to validate research findings and results.

All other digital and non-digital content have the potential of becoming research data. Research data may be experimental data, observational data, operational data, third party data, public sector data, monitoring data, processed data, or repurposed data.

https://dictionary.casrai.org/Research_data
Exercise 1:

In groups of 3 or 4, create a list of the different types of research data that you have encountered in your research or studies and provide some context surrounding the use of each data type.
Types of data

- Numeric
- Text
- Image
- Audio
- Video
- Geospatial
- Physical
- Etc. !

File formats

- xls, csv, tab, sav, accdb
- rtf, doc, txt, pdf, odf, xml, html
- jpeg, tiff, raw
- mp3, wav, flac
- mp4
- shp, mdb
- ??
Research data lifecycle
The lifecycle of research data often exceeds the time span of the research project responsible for creating it.

It is important that data are well managed over the course of a research project in terms of their organization, description, and preservation, so that they can be accessed and reused over time.
Research Data Lifecycle

**Creating Data**
- Design research
- Plan data management (formats, storage, etc.)
- Plan consent for sharing
- Locate existing data
- Collect data (experiment, observe, measure, simulate)
- Capture and create metadata

**Processing Data**
- Enter data, digitise, transcribe, translate
- Check, validate, clean data
- Anonymise data where necessary
- Describe data
- Manage and store data

**Analysing Data**
- Interpret data
- Define data
- Produce research outputs
- Author publications
- Prepare data for preservation

**Giving Access to Data**
- Distribute data
- Share data
- Control access
- Establish copyright
- Promote data

**Preserving Data**
- Migrate data to new format
- Migrate data to suitable medium
- Back-up and store data
- Create metadata and documentation
- Archive data

**Reusing Data**
- Follow-up research
- New research
- Undertake research reviews
- Localise findings
- Teach and learn

University of Victoria
The importance of good data management

“We back up our data on sticky notes because sticky notes never crash.”
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 requirements
- Satisfy journal requirements

Research Data Management Data Guide v.5.1 (Eugene Barsky, University of British Columbia)
Tri-Agency Expectations & Draft Policy

Statement of Principles for Digital Data Management (2016)

- Data management planning
- Adherence to standards
- Collection and storage
- Providing high quality metadata
- Retention, preserving and sharing
- Citation and acknowledgement
- Following legal and ethical obligations
Tri-Agency Expectations

Responsibilities of researchers include:

• incorporating data management best practices into their research;

• 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;

• acknowledging and citing datasets that contribute to their research; and
Exercise 2:

Reflect on your own research...

- If you were asked to share your data with another researcher, would they be able to make sense of your data? *Why or why not?*

- If you needed to locate your data files from 5 years ago, how easily would you be able to find and use them? *Why or why not?*
Library Support
Support from the Library

- Create a data management plan
- Find data sources for your research needs
- Find and implement relevant metadata standards
- Store data in a suitable repository
  - UVic Dataverse
- Share and promote data with DOIs

http://libguides.uvic.ca/researchdata
Data management plans
Data Management Plans

DMPs are formal documents that describes how you plan to collect, document, store and share your data throughout the lifecycle of a research project.

** It is a living document! **
Data Collection

• The type and volume of data you will collect, the file formats and software you will use.

  – 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?
File naming and version control

- Recommended conventions:
  - Develop a systematic convention
  - Keep file names short but meaningful
  - Avoid non-alphanumeric characters
  - Denote dates in the "YYYYMMDD" format

- The most important rule is be consistent.
Data Documentation

• How will you describe and structure the data to make it independently understandable?

  – What documentation will be needed for the data to be read and interpreted correctly in the future?
  – How will you make sure that documentation is created or captured consistently throughout your project?
  – If you are using a metadata standard and/or tools to document and describe your data, please list here.
Metadata

- Metadata is often thought of as “data about data”

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

- **Descriptive**: content and context of your data at both the dataset and item level. E.g. measurement, taxonomic, names, spatial, temporal

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

- **Structural**: describes how different data sets relate to one another. E.g.: data sets in a database, file formats
Metadata Standards

- Metadata should follow defined standards in make it findable, accessible and reusable.
- Many disciplines have their own metadata standards.

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

• What are your strategies and solutions for keeping your data securely accessible?

  – What are the anticipated storage requirements for your project, in terms of storage space (in megabytes, gigabytes, terabytes, etc.) and the length of time you will be storing it?
  – How and where will your data be stored and backed up during your research project?
  – How will the research team and other collaborators access, modify, and contribute data throughout the project?
Storage and Backup

• Accidental data loss can be disastrous for any research project.

• **Follow the 3-2-1 rule:** Save three copies of your data, on two different storage mediums, and one copy off site.

• [UVic University Systems](#) offers a range of storage services:
  – Personal and Departmental File Storage
  – Automated Computer Backup
  – File Archiving
  – Research Storage
Ethics and Legal Compliance

• If you have sensitive data, how will it be protected?
• What strategies do you have for the secondary use of sensitive data, if applicable?
• How will you manage legal, ethical, and intellectual property issues?
Preservation

- Where will the data be deposited for long-term preservation and access?
  
  - How you will ensure your data is preservation ready?
    
    Consider preservation-friendly file formats, ensuring file integrity, anonymization and de-identification, inclusion of supporting documentation.

  - Where will you deposit your data for long-term preservation and access at the end of your research project?
Open File Formats

• Open-File Formats can be used by anyone.
  – increase your ability to open and read your files in the future
  – make your data usable and accessible to more researchers immediately

• Recommended File Formats:
  Databases: XML, CSV
  E-Books: EPUB
  Images: JPG, PNG, PDF, TIFF, BMP
  Sound: MP3, FLAC
  Text: TXT, CSV, PDF/A, ASCII, UTF-8
  Video: MPG, MOV, AVI
  Spreadsheets: CSV
  Medical Images: DICOM
Sharing and Reuse

• What data will you share, and in what form? How will your peers know that your data exists?
  – What data will you be sharing and in what form? (e.g. raw, processed, analyzed, final).
  – Have you considered what type of end-user license to include with your data?
  – What steps will be taken to help the research community know that your data exists?
Why Share?

- Improves transparency and data integrity
- Promotes innovation and collaboration
- Increasingly required by funding agencies and publishers
- Allows for recognition and citation of work

Find a repository:

- **Dataverse** is our multi-disciplinary repository supporting the storage, licensing and sharing of research data files.
- **UVic Space** is our institutional open-access repository for digital scholarly works.
- **re3data.org** is a directory of external, domain-based repositories to help researchers find a repository that meet any discipline-specific needs.
Challenges to sharing

- Privacy or ethical issues

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

- Practical issues (e.g. metadata, technology)
Responsibilities and Resources

• Who will be responsible for the management of the data, and how will those responsibilities be delegated?
• Do you have sufficient funding and resources to implement your plan, ensuring secure and effective use of your data?
The DMP Assistant

• A free online tool for preparing data management plans.
• Offers guidance and walks researchers through key questions.

https://assistant.portagenetwork.ca
Exercise 3:

In groups of 3 or 4, read through the case study and answer your assigned DMP questions.