Research Data Management

A Brief Introduction

Leslie Barnes, Digital Scholarship Librarian and Dylanne Dearborn, Research Data Librarian (Science and Engineering)
SciNet, 9 Sep 2015

Outline

- What is RDM?
- Why is RDM important?
- Overview of RDM issues and topics
- Questions

What is Research Data Management (RDM)?

- Keeping your data organized makes it easier to work with
  - Describing and documenting data
  - Knowing roles and responsibilities
  - Having an accurate inventory
  - Adhering to policies

Why is RDM Important

- Helps you understand what you have
- Helps you preserve and/or share your data
- Helps you adhere to funder policies
- Helps you adhere to publisher policies
- Helps you adhere to REB policies
- For your own sanity and the well-being of your colleagues

Funding Policies - Canada

Preservation

- SSHRC, 2 years
- CIHR, 5 years

Sharing

- CIHR, must "deposit bioinformatics, atomic, and molecular coordinate data into the appropriate public database"

Management

- Data management supports other operations
- Requirements in Canada TBD, but just released:
  - "Draft Tri-Agency Statement of Principles on Digital Data Management"

Publisher Policies

"[…] a condition of publication in a Nature journal is that **authors are required to make materials, data and associated protocols promptly available to readers without undue qualifications**. Any restrictions on the availability of materials or information must be disclosed to the editors at the time of submission."

nature publishing group npg

Research Ethics Board (REB), UofT

- Research involving human participants must be reviewed and approved by the REB
- The REB requires you to describe:
  - the data to be collected and where/how it will be analyzed
  - how the data will be protected
  - how long the data will be retained
  - where the data will be stored
  - if data will be shared (and will there be restrictions on access)

Understanding What you Have

- Data management is intended to support data throughout its entire lifecycle, enabling discovery, understandability, and citation
- With proper data management, you can:
    - find what you've created
    - know what version you're looking at
    - know who is responsible for storing, backing up, and maintaining data
    - work easily in collaborative groups (everyone knows what's going on!)
    - understand what the data means in the future

Organization and documentation

- Have defined roles, if you're working in a group
  - Know who is responsible for the data and who is going to "keep" it
- Keep an inventory of your data
- Describe your data to make sure you know what it is and how it was made
- Include any programs, scripts, or software used to create or needed to view the data

Metadata

- Commonly called "data about data"
- A more structured way of describing data
- Often used in repositories and other collections
- Supports discoverability and findability
- Makes data meaningful over time
- Lots of standards: ask around in your field or consult a librarian

Confidential and Sensitive data

- Personally identifiable data is deemed "confidential"
- Other data that, if disclosed, would be negative is "sensitive" (for example, coordinates for the nest of an endangered bird)
- Regulations for how to govern this data will by the REB board

Retention, Preservation, and Sharing

- Retaining is just keeping data
- Preservation is keeping data in *meaningful* ways over the long-term (i.e. properly described and able to be viewed and used)
- Sharing data means making data understandable to potential users and putting in in an accessible space
- RDM supports all of these activities

Data Management Plans

- A short document to state your RDM plans
- Created at the start of a project and helps to anticipate challenges and design a strategy for your research project
- Online tools have templated questions that reflect criteria to meet funding regulations
- You can draft your own - make sure you address all required areas

Questions and Contact

leslie.barnes@utoronto.ca

dylanne.dearborn@utoronto.ca