DES HORIZONS ÉVOLUTIFS
La gestion de vos données de recherche

SHIFTING HORIZONS
Managing your research data
Spend less time managing your research data and more time on your research

For Health Sciences and Medicine Researchers

Laure Perrier
University of Toronto Libraries
WHAT WE’LL COVER

- RDM landscape (10 mins)
- 4 Simple Things (40 mins with Q&A)
- Small-group exercise (30 mins)
- Wrap up (10 mins with Q&A)
LEARNING OBJECTIVES

To provide participants with,

- An overview of current research data management requirements
- Strategies for implementing best practices in managing research data
WHY IS IT IMPORTANT?

Funding agencies
WHY IS IT IMPORTANT?

Funding agencies

Journals
FUNDING AGENCIES: TRI-AGENCY

CIHR
NSERC
SSHRC
FUNDING AGENCIES: TRI-AGENCY

CIHR

Bioinformatics, atomic, and molecular coordinate data
FUNDING AGENCIES: TRI-AGENCY

CIHR

Bioinformatics, atomic, and molecular coordinate data

All researchers: retain original data for a minimum of 5 years at end of grant
FUNDING AGENCIES: TRI-AGENCY

NSERC
None
FUNDING AGENCIES: TRI-AGENCY

SSHRC
Data preserved and made available within 2 years of completion of research project
TRI-AGENCY: DRAFT RDM POLICY

Data Management Plans
Data Management Plans

Share research data
TRI-AGENCY: DRAFT RDM POLICY

Data Management Plans

Share research data

University: Create an institutional strategy
TRI-AGENCY: DRAFT RDM POLICY

Timeline: Summer / Fall 2019
TRI-AGENCY: DRAFT RDM POLICY

Timeline: Summer / Fall 2019

Implementation:
- phased
- incremental
International Committee of Medical Journal Editors:

Require trials to have a data sharing statement (July 2018)
International Committee of Medical Journal Editors:

Require trials to have a data sharing statement (July 2018)

PLOS One; BMJ
PLOS journals require authors to make all data underlying the findings described in their manuscript fully available without restriction at the time of publication. When specific legal or ethical requirements prohibit public sharing of a dataset, authors must indicate how researchers may obtain access to the data. Refusal to share data and related metadata and methods in accordance with this policy will be grounds for rejection.
What is data?
WHAT IS DATA?

✗ No
Published Table in journal: Aggregate data

✓ Yes
Raw data file: individual responses

Table 1:

<table>
<thead>
<tr>
<th>Statistic</th>
<th>N</th>
<th>Mean</th>
<th>St. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>rating</td>
<td>30</td>
<td>64.633</td>
<td>12.173</td>
<td>40</td>
<td>85</td>
</tr>
<tr>
<td>complaints</td>
<td>30</td>
<td>66.600</td>
<td>13.315</td>
<td>37</td>
<td>90</td>
</tr>
<tr>
<td>privileges</td>
<td>30</td>
<td>53.133</td>
<td>12.235</td>
<td>30</td>
<td>83</td>
</tr>
<tr>
<td>learning</td>
<td>30</td>
<td>56.367</td>
<td>11.737</td>
<td>34</td>
<td>75</td>
</tr>
<tr>
<td>raises</td>
<td>30</td>
<td>64.633</td>
<td>10.397</td>
<td>43</td>
<td>88</td>
</tr>
<tr>
<td>critical</td>
<td>30</td>
<td>74.767</td>
<td>9.895</td>
<td>49</td>
<td>92</td>
</tr>
<tr>
<td>advance</td>
<td>30</td>
<td>42.933</td>
<td>10.289</td>
<td>25</td>
<td>72</td>
</tr>
<tr>
<td>high.rating</td>
<td>30</td>
<td>0.333</td>
<td>0.479</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
WHAT IS DATA?

× No
Published Table in journal: Aggregate data

✓ Yes
Raw data file: individual responses

Table 1:

<table>
<thead>
<tr>
<th>Statistic</th>
<th>N</th>
<th>Mean</th>
<th>St. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>rating</td>
<td>30</td>
<td>64.633</td>
<td>12.173</td>
<td>40</td>
<td>85</td>
</tr>
<tr>
<td>complaints</td>
<td>30</td>
<td>66.600</td>
<td>13.315</td>
<td>37</td>
<td>90</td>
</tr>
<tr>
<td>privileges</td>
<td>30</td>
<td>53.133</td>
<td>12.235</td>
<td>30</td>
<td>83</td>
</tr>
<tr>
<td>learning</td>
<td>30</td>
<td>56.307</td>
<td>11.737</td>
<td>34</td>
<td>75</td>
</tr>
<tr>
<td>raises</td>
<td>30</td>
<td>64.633</td>
<td>10.397</td>
<td>43</td>
<td>88</td>
</tr>
<tr>
<td>critical</td>
<td>30</td>
<td>74.767</td>
<td>9.895</td>
<td>49</td>
<td>92</td>
</tr>
<tr>
<td>advance</td>
<td>30</td>
<td>42.933</td>
<td>10.289</td>
<td>25</td>
<td>72</td>
</tr>
<tr>
<td>high.rating</td>
<td>30</td>
<td>0.333</td>
<td>0.479</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
WHAT IS DATA?

Any material that will allow your results to be reproduced
WHAT IS DATA?

Any material that will allow your results to be reproduced

Includes notes and details
WHAT IS DATA?

Any material that will allow your results to be reproduced
Includes notes and details
Others should be able to understand
4 SIMPLE THINGS

Make a Plan
DATA MANAGEMENT PLANS

Type, format, volume of data
DATA MANAGEMENT PLANS

Type, format, volume of data

Storage, backup, access
DATA MANAGEMENT PLANS

Type, format, volume of data

Storage, backup, access

Ethical, legal issues
DATA MANAGEMENT PLANS

Type, format, volume of data

Storage, backup, access

Ethical, legal issues

Sharing
UOTTAWA: ETHICS REVIEW APPLICATION

REB submissions:

Length of data conservation
UOTTAWA: ETHICS REVIEW APPLICATION

REB submissions:

Length of data conservation

Safeguarding data
UOTTAWA: ETHICS REVIEW APPLICATION

REB submissions:

Length of data conservation

Safeguarding data

Issues of anonymity or confidentiality
TOOLS

DMP Assistant (Canada): portagenetwork.ca

DMP Online (UK): dmponline.dcc.ac.uk

DMP Tool (USA): dmptool.org
Consent Forms

- Two-tiered
  1) Taking part in study
  2) Use of data

- Examples: UK Data Service

  Google: UK Data Services consent forms
4 SIMPLE THINGS

Make a Plan

Create a System
Still setting up my computer. In 2015, I had a "Super U" computer. Now, I have a "Super U" computer. I need to tune it.

In my "Super U" computer, I have a funny photo:

Chris Buddle
@CMBuddel

#protip
NEVER name folders "stuff", "miscellaneous", "important stuff", "misc stuff"

Today, I realize this is what I do.

#ScienceInAction
ORGANIZE YOUR FOLDERS

Project Name

Surveys
Poster
Paper
FILE NAMING EXAMPLE

Project name
Standardized date format
Description of file content
Versioning Information

Project_YYYYMMDD_ContentDescription_Version.ext
NAMING FILES

Versioning

FileName_v1.doc
FileName_v2.doc
NAMING FILES

Avoid special characters: #, %
Use capitals or underscores
Example: FileName.doc
Standardized (ISO) date: YYYYMMDD
Important amongst researchers if sharing
Important amongst researchers if sharing

Describe the data for future use
Batman movie
Amazon.ca

Movies & TV

1-24 of 765 results

Search for: Batman

Video Format: Blu-ray | DVD | VHS | HD DVD | Universal Media Disc | See more

Amazon's Choice

Starring: Various
Blu-ray Batman / Batman Returns / Batman Forever / Batman & Robin - Bilingual
CDN$ 50.92 prime
Get it by Tomorrow, Oct 5

More buying choices
CDN$ 27.99 (12 used & new offers)
Eligible for FREE Shipping

5 stars
128

Batman: Gotham by Gaslight 2018
Starring: Bruce Greenwood
Blu-ray BD/ DVD/ UV/ BIL
CDN$ 19.99 prime
Get it by Tomorrow, Oct 5

More buying choices
CDN$ 17.99 (4 used & new offers)
Eligible for FREE Shipping

Star Rating: 5 stars
19

Batman vs. Two-Face 2017
Starring: Adam West
Blu-ray
CDN$ 14.99 prime
Get it by Tomorrow, Oct 5

More buying choices

Batman Ninja 2018
Starring: Fred Tatasciore
Blu-ray BIL/ UV/ DVD/ BD - Steelbook
CDN$ 39.99 prime
Only 8 left in stock - order soon.

More buying choices

Movie & TV Show Original Language
Clear

English
French
Spanish
Italian
German

Blu-ray Region
Region A/1
Region B/2
Region C/3
Title: Batman Returns

Release Date: 19 June 1992

Alternate Title: Batman 2

Runtime: 126 minutes

Director: Tim Burton

Storyline: Batman battles Penguin. Catwoman shows up to help.
Batman 1996
Starring: Michael Keaton
DVD
CDN$ 28.81
Only 1 left in stock - order soon.
More buying choices
CDN$ 1.00 (29 used & new offers)

Batman: le défi 1992
Blu-ray
CDN$ 35.66
Only 1 left in stock - order soon.
FREE Shipping

Batman Returns 1992
Blu-ray Import
CDN$ 46.81
Only 1 left in stock - order soon.
FREE Shipping
Suitability for your study: understand data in detail
Suitability for your study: understand data in detail

Other researchers can find, use, properly cite your data
CITING DATA

Data cited similar to a journal citation
4 SIMPLE THINGS

Make a Plan

Create a System

Secure Your Data
3-2-1 RULE

3 copies of your data
3-2-1 RULE

3 copies of your data

2 different locations
3-2-1 RULE

3 copies of your data
2 different locations
More than 1 type of storage hardware
BACK-UP OPTIONS

Personal computer hard drive
External hard drives
Departmental servers

uOttawa contact: Jarno van der Kolk
Senior Scientific Computing Specialist
jvanderk@uottawa.ca
ENCRIPTION

One option = programs such as:

- 7-zip (free at: 7-zip.org)
- Foxit PhantomPDF
  (request education discount $10 US per year at: support@foxitsoftware.com)
ENCRYPTION

- Computer: Disk encryption
  https://it.uottawa.ca/security/disk-encryption
- USB Drives (office supply retailers)
- Secure file transfer: uOttawa LiquidFiles
  https://it.uottawa.ca/email/liquidfiles

uOttawa contact: Jarno van der Kolk
Senior Scientific Computing Specialist
jvanderk@uottawa.ca
SAFE & SECURE STORAGE

Dropbox or Google Drive (generic) are **NOT** secure storage
SAFE & SECURE STORAGE

Dropbox or Google Drive (generic) are NOT secure storage

Option:

uOttawa Google Drive File Stream

uOttawa contact: Jarno van der Kolk / jvanderk@uottawa.ca
SAFE & SECURE STORAGE

Option:
Sync.com
- Canadian data residency
- Do not claim ownership to data
SAFE & SECURE STORAGE

Sensitive data:

- Encryption
- Do not use cloud services
4 SIMPLE THINGS

Make a Plan
Create a System
Secure Your Data
Long-Term Access
DATA MUST BE READABLE

<table>
<thead>
<tr>
<th>No Proprietary formats</th>
<th>Yes Open formats</th>
</tr>
</thead>
<tbody>
<tr>
<td>.doc (Word)</td>
<td>.txt [text]</td>
</tr>
<tr>
<td>.xlsx (Excel)</td>
<td>.csv [spreadsheets]</td>
</tr>
<tr>
<td>.mov (Quicktime)</td>
<td>.m4a (M-PEG 4) [audio]</td>
</tr>
<tr>
<td>.gif or .jpg (images)</td>
<td>.tif [images]</td>
</tr>
</tbody>
</table>
DATA MUST BE READABLE

<table>
<thead>
<tr>
<th>Proprietary formats</th>
<th>Open formats</th>
</tr>
</thead>
<tbody>
<tr>
<td>× No</td>
<td>✓ Yes</td>
</tr>
<tr>
<td>.doc (Word)</td>
<td>.txt [text]</td>
</tr>
<tr>
<td>.xlsx (Excel)</td>
<td>.csv [spreadsheets]</td>
</tr>
<tr>
<td>.mov (Quicktime)</td>
<td>.m4a (M-PEG 4) [audio]</td>
</tr>
<tr>
<td>.gif or .jpg (images)</td>
<td>.tif [images]</td>
</tr>
</tbody>
</table>
DATA MUST BE ACCESSIBLE

Data must include notes, details
DATA MUST BE ACCESSIBLE

Data must include notes, details

Others should be able to understand
### DATA DICTIONARY: QUANTITATIVE

<table>
<thead>
<tr>
<th>Variable</th>
<th>Variable Name</th>
<th>Variable Type</th>
<th>Variable Width</th>
<th>Values / Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant ID number</td>
<td>ID</td>
<td>Numeric</td>
<td>3</td>
<td>001-900</td>
</tr>
<tr>
<td>Date of birth</td>
<td>DOB</td>
<td>yyyy/mm/dd</td>
<td></td>
<td>1900-2010/1-12/1-31</td>
</tr>
</tbody>
</table>
| Status            | STAT          | Numeric       | 1              | 1 = alive
                                                                              |
|                   |               |               |                | 2 = deceased                 |
| Hemoglobin        | HB            | Numeric       | 2.1            | 4.0 – 18.0                   |
## DATA DICTIONARY: QUALITATIVE

<table>
<thead>
<tr>
<th>Code</th>
<th>Code Name</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clarity</td>
<td>CLA</td>
<td>Coherence of components</td>
</tr>
<tr>
<td>Structure</td>
<td>STR</td>
<td>Arrangement between component parts</td>
</tr>
<tr>
<td>Navigation</td>
<td>NAV</td>
<td>Accurately ascertaining position and planning for movement through information</td>
</tr>
<tr>
<td>Saliency</td>
<td>SAL</td>
<td>Quality by which item stands out in relation to its neighbours</td>
</tr>
<tr>
<td>Flow</td>
<td>FLO</td>
<td>Moving along in a logical, steady manner</td>
</tr>
</tbody>
</table>
DATA REPOSITORIES

Secure place for research data
DATA REPOSITORIES

Secure place for research data

Can impose appropriate access restriction and/or embargoes
DATA REPOSITORIES

- **Subject-specific**: re3data.org
- **Multidisciplinary** (Zenodo; Figshare; Dryad)
- **UOttawa**: Dataverse
  
  https://dataverse.scholarsportal.info/dataverse/ottawa

**Contact**: Felicity.Tayler@uottawa.ca, E-Research Librarian
Get into groups of 2 or 3
ACTIVITY

1) Read: Sample Data Management Plan
2) Fill out: Assessment Sheet
ACTIVITY

Section 1.3: ‘…how much data (volume) will be produced’

- Do not state sample size
- Do not describe the length of survey
ACTIVITY

Section 2.2: ‘…data formats created or used’

- “…data formats generally accepted…including SPSS, STATA”

- (Survey data not mentioned)
ACTIVITY

Section 4.4: ‘…communicating policies’

- Data management plan developed with guidance from UNC policies
- Not clear if these cover guidelines on reuse, redistribution, etc.
ACTIVITY

Section 5.4: ‘…identifies a timeframe for how long…archived’

- Implied given that repository (Dataverse) is at UNC
biblio.uottawa.ca/en/services/faculty/research-data-management

What is research data?

Research data are data "used as primary sources to support technical or scientific enquiry, research, scholarship, or artistic activity". They are the raw data used to generate results and conclusions in research.

Why manage research data?

Many publishers and granting agencies are asking authors and grant recipients to include a data management plan in their grant applications and publications. As a result, the responsible management of research data is becoming an increasingly important aspect of the research process.

Data management planning

Data management planning includes consideration for many factors, including the careful planning of data creation, collection, storage, and preservation. For more information, see the University of Ottawa Response to the Draft Tri-Agency Research Data Management Policy.
Research data management

What is data management?
Data management involves the organization and maintenance of data during the active stages of a project, and suitable archiving of the data at the project’s completion. It is an on-going activity throughout the data lifecycle.

What is research data?
Research data are data “used as primary sources to support technical or scientific enquiry, research, scholarship, or artistic activity.”

Why manage research data?
Many publishers and granting agencies are asking authors and applicants to document, disseminate, or share their data. You should manage your research data to make them discoverable beyond your life and career. To manage your research data, you need to:
- plan;
- create;
- store;
- share;
- discover;
- cite.

Data management planning
Data management planning includes consideration for many topics, for example:
- storage;
- access;
- preservation;
- disposal;
- standardization.

University of Ottawa Response to the Draft Tri-Agency Research Data Management Policy

Read more
QUESTIONS

Felicity Tayler,
Research Data Management Librarian
Felicity.Tayler@uottawa.ca

Laure Perrier
l.perrier@utoronto.ca