Data Management in the Classroom

Thea P Atwood, University of Massachusetts - Amherst

Available at: https://works.bepress.com/tpatwood/2/
Data Management in the Classroom (and in practice)

Thea P. Atwood

tpatwood@library.umass.edu

@librarianThea
Today’s Agenda...

1. Introduction to Data Management
2. Break
3. Your own stories & experiences
4. Data sharing in the humanities
5. Break
6. Workshop wrap-up
Today’s Agenda...

1. Introduction to Data Management
2. Break
3. Your own stories & experiences
4. Data sharing in the humanities
5. Break
6. Workshop wrap-up
**Why manage data?**

- **Saves you time**
  - Easier to find and use data
  - Tell what you last worked on, what you need to do
  - Find & use files in the future
  - Focus on your *project*, not requests

- **In compliance with funder requirements**
  - *keep an eye on the NIH & Wellcome Trust*
Why manage data?

● Increase your reputation
  ○ e.g. Piwowar, Day, & Fridsma, 2007, Sharing Detailed Research Data is Associated with Increased Citation Rate

● Facilitate Progress
  ○ Sharing & preserving your data is easier
  ○ Help others find & understand your data, so they can build on your research

● Be prepared for catastrophes!
What is data?

From Circular A-110 - Section 36.d.2.i:

(i) *Research data* is defined as the **recorded factual material commonly accepted** in the scientific community **as necessary to validate research findings**, but not any of the following: preliminary analyses, drafts of scientific papers, plans for future research, peer reviews, or communications with colleagues. This "recorded" material excludes physical objects (e.g., laboratory samples). Research data also do not include:

(A) Trade secrets, commercial information, materials necessary to be held confidential by a researcher until they are published, or similar information which is protected under law; and

(B) Personnel and medical information and similar information the disclosure of which would constitute a clearly unwarranted invasion of personal privacy, such as information that could be used to identify a particular person in a research study.
What is **data**?

Do you have a different understanding of data?

Does this definition apply to your own research and research assets?

What does this definition miss or fail to cover?
The Data Lifecycle

From DataONE: http://www.dataone.org/best-practices

From Digital Curation Centre: http://www.dcc.ac.uk/resources/curation-lifecycle-model
The Data Lifecycle

What does your research or data lifecycle entail?
Data Management Plans

Defined:
A formal document which outlines what you will do with your research assets during and after you complete your project.
Data Management Plans

Most funders provide guidance/guidelines for how to comply with their requirements, but not always clear. However -- lots of other sources that can help!

Think about the end of your project, and work towards now.
Data Management Plans

Some tips:

● Cover as much as you can about a project before you begin collecting research assets.
● Write up your plan so you can refer back to it instead of trying to re-interpret what you’ve done.
● Think about all the steps you need to archive & preserve your data.
● What does it take to make your research assets “preservation ready?”
Data Management Plans

Examples are out there!

- If your library offers data management services, may have samples available
- Your Office of Research Compliance* may have copies stored (though you’ll probably have to ask)
- DMPTool has some examples based on funding agency, including one for NEH

*actual name may vary
Nuts & Bolts

Managing data is different for everyone... but there are some basics!

- Have a **consistent** organization & filing system
  - Having something is better than nothing!
  - Use **non-proprietary** formats *as much as possible*
  - Other best practices: naming convention, filing convention
Nuts & Bolts

Example from *Teaching Students to Document Their Empirical Data*:

- Main folder: “ThesisDocumentation”
  - Document: readme.pdf
    - Sub-folder: Data
      - Sub-folder: RawData
      - Sub-folder: ImportableData
    - Sub-folder: Metadata
    - Sub-folder: DoFiles
Nuts & Bolts

Example from *Teaching Students to Document Their Empirical Data:*

Bonus: uploaded to explore: https://drive.google.com/folderview?id=0B6Sp5TsYUNuWVXQwSGhyTGQ3WUk&usp=sharing
Nuts & Bolts

Document, document, document!
i.e., Metadata, data about data, documentation,…

- Outline as much as is necessary for someone else to understand your project:
  - What you collected
  - Why you collected it
  - The original purpose of the data
  - When you collected

- Lots of options for execution: pdf, txt, xml, tei, etc…
Nuts & Bolts

Storage & backups: here, near, and far away

- Local hard drive (*here*)
- External hard drive (*near*)
- Institutional repository (*near*)
- Partner institution or discipline-specific repository (*far away... as long as it is in a different region/time zone*)
Tying in data management

Haverford TIER program: AWESOME model, and a reminder that you don’t have to do it alone!

(and they have offered training -- watch http://www.haverford.edu/TIER/ for opportunities)
Tying in data management

Smaller steps:

● When introducing research skills, introduce best practices (your own tried-and-true methods or something from your discipline)

● Encourage students to use tools to help them manage other aspects of their academic lives, e.g., Zotero

● Remind students to backup their work consistently - automated or otherwise!

● Your ideas?
More information & help

Librarians!

Online training!
+ MANTRA http://datalib.edina.ac.uk/mantra/

Workshops!
+ DH Curation Institute: http://www.dhcuration.org/institute/
More information & help

Guides!
+ DH Digital Curation: http://www.dhcuration.org/

Data Management Guidance!
+ DMPTool: https://dmp.cdlib.org/ (DMPTool2: May 29!)

And more! Keep your eyes peeled -- and be an advocate for your discipline to offer data management at your next regional or national meeting!
Today’s Agenda...

1. Introduction to Data Management
2. Break
3. Your own stories & experiences
4. Data sharing in the humanities
5. Break
6. Workshop wrap-up
Today’s Agenda...

1. Introduction to Data Management
2. Break
3. Your own stories & experiences
4. Data sharing in the humanities
5. Break
6. Workshop wrap-up
Today’s Agenda...

1. Introduction to Data Management
2. Break
3. Your own stories & experiences
4. Data sharing in the humanities
5. Break
6. Workshop wrap-up
Today’s Agenda...

1. Introduction to Data Management
2. Break
3. Your own stories & experiences
4. Data sharing in the humanities
5. Break
6. Workshop wrap-up
Today’s Agenda...

1. Introduction to Data Management
2. Break
3. Your own stories & experiences
4. Data sharing in the humanities
5. Break
6. Workshop wrap-up
Thanks!

More questions? Email!

tpatwood@library.umass.edu

Occasional posts on Data Management and other topics? Follow me on Twitter!
@librarianThea