



University
of Idaho

Organizing your Research and Data Management

Jeremy Kenyon

jkenyon@uidaho.edu

Research Librarian,
UI Library



Summary

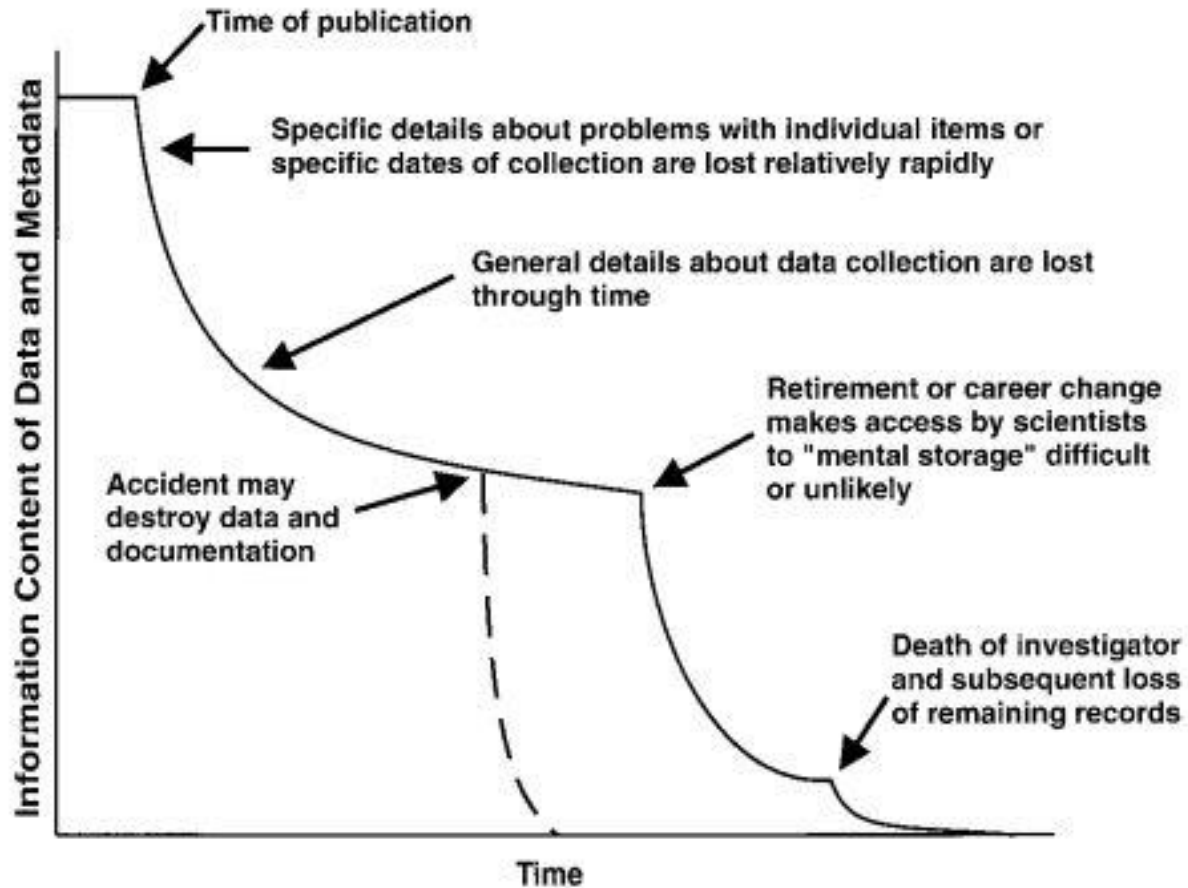
- Three reasons to develop a system of organization
- Dealing with files
- Data management planning
- Dealing with notes



Reasons to develop a system

- Problems of memory
 - Can you remember everything that happened during a project? How about after 3 years?
- Problems of time
 - Do you have time to go back and do things over again? How often do you want to re-organize things?
- Problems of scale
 - Does your approach work for everyone? Does it work for 3 people, for 5, or for 10?

Problems of Memory





Problems of Time

- 80% of analysts' time is spent discovering and preparing data for analysis (DalleMule and Davenport, 2017)
- 50-80% of time “mired in mundane labor of collecting and preparing unruly digital data” (New York Times, 2014)
- Lost data, lack of documentation, and ambiguity about responsibilities (in teams) also add to this problem.



Problems of Scale

- We tend to use idiosyncratic, “it-works-for-me” methods of organization, reliant on individual tics and traits that do not work for others or teams.
- We also tend to create “data swamps” rather than “data lakes”, due to a lack of documentation and governance of our data storage.
- We need a model for organizing files and data in a manner that enables efficiency, recall, and optimization for both individuals and teams.



A model for organizing

1. Collect

1. Attendance #'s
 - Spreadsheet/CSV
 - Metadata
2. Surveys
 - Paper forms
 - Metadata
3. Enrollment #'s
 - Spreadsheet/CSV
 - Metadata
4. Literature review
 - references (Zotero)
5. Notes from lit review
 - free text (OneNote)
6. Analysis code
 - R scripts

3. Describe & Manage

- Security & Access Rules
 - any rules, IRB, laws, etc.
- Sharing Rules
 - how public can it be, partners' access
- Metadata Rules
 - who, what, when, where, why, description

2. Store & Process

- OneDrive – Files (1, 2, 3, 5)
 - Zotero - References (4)
 - Github – Code (6)
-
- Derived files: cleaned data, analyzed data, reformatted data

4. Apply

- Papers
- Presentations
- Posters
- Data Sharing
- Future Projects



A STORY TOLD IN FILE NAMES:

Location: C:\user\research\data

Filename	Date Modified	Size	Type
data_2010.05.28_test.dat	3:37 PM 5/28/2010	420 KB	DAT file
data_2010.05.28_re-test.dat	4:29 PM 5/28/2010	421 KB	DAT file
data_2010.05.28_re-re-test.dat	5:43 PM 5/28/2010	420 KB	DAT file
data_2010.05.28_calibrate.dat	7:17 PM 5/28/2010	1,256 KB	DAT file
data_2010.05.28_huh??.dat	7:20 PM 5/28/2010	30 KB	DAT file
data_2010.05.28_WTF.dat	9:58 PM 5/28/2010	30 KB	DAT file
data_2010.05.29_aaarrgh.dat	12:37 AM 5/29/2010	30 KB	DAT file
data_2010.05.29_#\$\$@*&!!.dat	2:40 AM 5/29/2010	0 KB	DAT file
data_2010.05.29_crap.dat	3:22 AM 5/29/2010	437 KB	DAT file
data_2010.05.29_notbad.dat	4:16 AM 5/29/2010	670 KB	DAT file
data_2010.05.29_woohoo!!.dat	4:47 AM 5/29/2010	1,349 KB	DAT file
data_2010.05.29_USETHISONE.dat	5:08 AM 5/29/2010	2,894 KB	DAT file
analysis_graphs.xls	7:13 AM 5/29/2010	455 KB	XLS file
ThesisOutline!.doc	7:26 AM 5/29/2010	38 KB	DOC file
Notes_Meeting_with_ProfSmith.txt	11:38 AM 5/29/2010	1,673 KB	TXT file
JUNK...	2:45 PM 5/29/2010		Folder
data_2010.05.30_startingover.dat	8:37 AM 5/30/2010	420 KB	DAT file



File Naming

File naming convention (FNC):

a framework for naming your files in a way that describes what they contain and how they relate to other files (Brandt, 2017)

Things to consider:

- What's your primary unit of analysis: spatial, time, subject/topic?
- Add the date in the filename as YYYY-MM-DD
- Use underscores for different elements

Files might be named:

2019-08-27_MicaCreekStreamTemp_raw.csv

2019-09-05_MicaCreekStreamTemp_clean.csv



File Naming Resources

DataOne. 2019. Assign descriptive file names. *Best Practices*. <https://www.dataone.org/best-practices/assign-descriptive-file-names>

Santaguida, V. 2016. Folder and File Naming Convention – 10 Rules for Best Practice
<http://www.exadox.com/en/articles/file-naming-convention-ten-rules-best-practice>



Storage Options

The University of Idaho official resources is OneDrive

- Everyone has 1 TB of space

Others:

- Google Drive – 15 GB (incl. your gmail/photos)
- Dropbox – 2 GB
- OwnCloud -

Also consider the [Open Science Framework](#)



Data Management Planning

When starting a project – create a plan!

Use the model we worked through and consider:

1. Roles and responsibilities
2. Backup and redundancy
3. Security and regulations
4. Metadata creation
5. Filenaming conventions



Dealing with Notes

Notecards & the Zettelkasten Method:

- Both focus on the connection between reading, writing, and thinking
 - Bibliographic Reference Cards
 - Direct Quotations Cards
 - Summary Cards

Digital Tools:

- EverNote, OneNote, some specialized software tools, e.g. Bear and tools at the Zettelkasten site

Resources:

- <http://www.raulpacheco.org/2018/11/note-taking-techniques-i-the-index-card-method>
- <https://zettelkasten.de>



Annotations

Aside from the techniques above, consider annotation tools:

- Mendeley.com's application has a built-in annotator
- Hypothes.is provides a browser extension for annotating documents



Graduate Student Essentials

9/3	Research Refresher
9/10	Tips and Tricks for Word, Excel, and PowerPoint
9/17	Scholarly Presence
9/24	Citation Management
10/1	Organizing Your Research and Data Management



References

1. Michener, et al. 1997. [https://doi.org/10.1890/1051-0761\(1997\)007\[0330:NMFTE\]2.0.CO;2](https://doi.org/10.1890/1051-0761(1997)007[0330:NMFTE]2.0.CO;2)
2. DalleMule, L. & T.H. Davenport. 2017. *What's your data strategy?* *Harvard Business Review*, May-June, pp. 112-121.
3. Lohr, S. 2014. For Big-Data Scientists, 'Janitor Work' is Key Hurdle to Insights. *New York Times*, Aug. 17, Section B, Page 4, New York edition. Retrieved from: <https://www.nytimes.com/2014/08/18/technology/for-big-data-scientists-hurdle-to-insights-is-janitor-work.html>
4. Brandt, D.S. 2017. Data Management for Undergraduate Researchers: File Naming Conventions. Purdue University Libraries. Retrieved from: <https://guides.lib.purdue.edu/c.php?g=353013&p=2378293>