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

Specific details about problems with individual items or specific dates of collection are lost relatively rapidly.

General details about data collection are lost through time.

Accident may destroy data and documentation.

Retirement or career change makes access by scientists to "mental storage" difficult or unlikely.

Death of investigator and subsequent loss of remaining records.

(Michener, et al. 1997)
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 individualistics 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

2. Store & Process
   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

3. Describe & Manage
   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:

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


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:
- 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
<table>
<thead>
<tr>
<th>Date</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>9/3</td>
<td>Research Refresher</td>
</tr>
<tr>
<td>9/10</td>
<td>Tips and Tricks for Word, Excel, and PowerPoint</td>
</tr>
<tr>
<td>9/17</td>
<td>Scholarly Presence</td>
</tr>
<tr>
<td>9/24</td>
<td>Citation Management</td>
</tr>
<tr>
<td>10/1</td>
<td>Organizing Your Research and Data Management</td>
</tr>
</tbody>
</table>
References


