Hello
my name is
Data Management Workshop Series

• **Introduction to Data Management**
  - Feb 10\textsuperscript{th} 4PM – 5PM
  - Apr 6\textsuperscript{th} 1PM – 2PM

• **Documentation and Organization for Data and Processes**
  - Feb 17\textsuperscript{th} 4PM – 5PM
  - Apr 13\textsuperscript{th} 1PM – 2PM

• **Making Research Data Public: Why, What, and How**
  - Feb 24\textsuperscript{th} 4PM – 5PM
  - Apr 20\textsuperscript{th} 1PM – 2PM
Objectives

1. Consider and adopt an overall organizational structure for project and data files
2. Learn how to visualize your research processes through a workflow diagram
3. Identify what kind of data documentation is beneficial for your research
Reality

From PhD Comics (http://www.phdcomics.com/comics/archive.php?comicid=1323)
Process Documentation
Workflow Diagrams

- good starting point for mapping out elements of a project
- provide a visual of the entire process
- help you (and others) understand the steps needed to acquire and analyze your data
Example Workflow Diagrams

Doesn’t have to be pretty
Example Workflow Diagrams

Can be simple
Example Workflow Diagrams

Or very detailed

http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0080278
Don’t forget project level documentation

- Context of the project
- Funding & IRB review
- Personnel
- Instruments
- Decisions made
Basic Organization
Controlled Vocabulary

- Take the guess work out of choosing between:
  - a preferred spelling
    - behavior vs behaviour
Controlled Vocabulary Examples

• MeSH – Medical Subject Headings

• WordNet – English Lexicon Database

• ISO 3166 – Codes for countries & their regions

• ISO 639.2 – Abbreviations for language names
Consistent File and Folder Naming

For quickly finding and sorting files and folders, the names should be consistent but unique. Avoid special characters.

- project name/acronym
- experiment/instrument type
- site location information (if applicable)
- researcher initials
- date (consistently formatted, i.e. YYYYMMDD)
- version number (w/ leading zeros)
ISO Date Standard

PUBLIC SERVICE ANNOUNCEMENT:

Our different ways of writing dates as numbers can lead to online confusion. That’s why in 1988 ISO set a global standard numeric date format.

This is the correct way to write numeric dates:

2013-02-27

The following formats are therefore discouraged:

02/27/2013 02/27/13 27/02/2013 27/02/13
20130227 2013.02.27 27.02.13 27-02-13
27.2.13 2013. II. 27. 2/2-13 2013.158904109
MMXIII-II-XXVII MMXIII 2013158904109
((3+3)×(111+1)-1)×3/3-1/3³ 2013
10/11011/1101 02/27/20/13 01237
Power Naming

Bacillus anthracis methyltransferase co-expression activity experiment 01 September 04 2014.txt
Bacillus anthracis methyltransferase co-expression activity experiment 02 September 04 2014.txt
Bacillus anthracis methyltransferase co-expression activity experiment 03 September 04 2014.txt
Versioning

HJI BAM Co-Exp Run 01 20140904 RAW.txt
HJI BAM Co-Exp Run 02 20140904 RAW.txt
HJI BAM Co-Exp Run 03 20140904 RAW.txt

HJI BAM Co-Exp Run 01 20140904.txt
HJI BAM Co-Exp Run 02 20140904.txt
HJI BAM Co-Exp Run 03 20140904.txt

HJI BAM Co-Exp Run 01 20140904 V01.exls
HJI BAM Co-Exp Run 02 20140904 V01.exls
HJI BAM Co-Exp Run 03 20140904 V01.exls

HJI BAM Co-Exp 2014 09 Analysis V01.enz
HJI BAM Co-Exp 2014 09 Analysis V10.enz

Raw data (no manipulation)

Working data

Versioned working data

Analyzed Data
Storage & Back-Up Best Practices

- Security
- Robustness
- Number of copies
- Location of copies
- Scheduling
- Efficacy
Data Documentation
“Metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource. Metadata is often called data about data or information about information” (NISO, Understanding Metadata 2004;1).
Data Documentation Continuum

- Low-BARRIER
  - Fast
  - Easy

- Low-Quality
  - Irregular
  - Incomplete

- High-Quality
  - Standardized
  - Rich

- High-BARRIER
  - Slow
  - Skilled
Why take the time?

• find data from other researchers to support your research;
• use the data that you do find;
• help other professionals to find, understand, and use data from your research; and
• use your own data in the future when you may have forgotten details of the research.
• save time when depositing your research
ReadMe Files

Simple text file that accounts for all files and folders in a dataset

- basic information about the dataset (author, year, associated publication as appropriate)
- explanation of naming conventions
- relationship between directory structure and the data
- additional information for re-use (see extensive example)
Codebooks

Primary source of data documentation

- Study-level information
- descriptions and annotations of the variable and data items
Data Dictionaries

provides a detailed description for each element or variable in your dataset and data model
- descriptive name
- the data type
- allowed values, units, and text description
- additional information for re-use
Metadata Schemas/Standards

• Some disciplines already have systems
  – Ecological Metadata Language (EML)
  – Data Documentation Initiative (DDI)

• Generic standards to consider
  – Dublin Core

• Sometimes multiple/no standards
Objectives

1. Consider and adopt an organizational structure
2. Visualize research processes through a workflow diagram
3. Identify what kind of data documentation is beneficial for your research
Contact

Peg Burnette
Biomedical Librarian
phburn@illinois.edu

Ayla Stein
Metadata Librarian
astein@illinois.edu

Heidi Imker
Director, Research Data Service
imker@illinois.edu

visit:
310-312 Main Library

call:
(217) 300-3513

website:
researchdataservice@illinois.edu
Data Management Workshop Series

• Introduction to Data Management
  – Feb 10\textsuperscript{th} 4PM – 5PM
  – Apr 6\textsuperscript{th} 1PM – 2PM

• Documentation and Organization for Data and Processes
  – Feb 17\textsuperscript{th} 4PM – 5PM
  – Apr 13\textsuperscript{th} 1PM – 2PM

• Making Research Data Public: Why, What, and How
  – Feb 24\textsuperscript{th} 4PM – 5PM
  – Apr 20\textsuperscript{th} 1PM – 2PM
Scholarly Commons
http://www.library.illinois.edu/sc/index.html

Library Home Page
http://www.library.illinois.edu/

Subject Specialists

<table>
<thead>
<tr>
<th>Subject Area</th>
<th>Library</th>
<th>Subject Specialist</th>
<th>Phone (217)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accounting</td>
<td>Business Information Services</td>
<td>Cerissa Phillips</td>
<td>333-0169</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Becky Smith</td>
<td>244-0388</td>
</tr>
<tr>
<td>Advertising</td>
<td>Communications Library</td>
<td>Lisa Romero</td>
<td>333-6348</td>
</tr>
<tr>
<td>African Studies</td>
<td>International &amp; Area Studies Library</td>
<td>Atoma Batoma</td>
<td>333-2904</td>
</tr>
<tr>
<td>Afro-American Studies</td>
<td>African American Research Center in the History, Philosophy, &amp; Newspaper Library</td>
<td>Thomas Weissinger</td>
<td>333-3006</td>
</tr>
<tr>
<td>Aging/Gerontology</td>
<td>Social Sciences, Health, and Education Library</td>
<td>JJ Pionke</td>
<td>265-0002</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Peg Burnette</td>
<td>300-5942</td>
</tr>
</tbody>
</table>

http://www.library.illinois.edu/administration/collections/subjectdirectory.html
Useful?
Tell your friends and colleagues!

Not Useful?
Tell us!