Introduction to Data Management

Elizabeth Wickes
Christie Wiley
Data Management Workshop Series

• **Introduction to Data Management**
  – February 16\(^{th}\) 10AM – 11AM

• **Documentation and Organization for Data and Processes**
  – February 23\(^{rd}\) 10AM – 11AM

• **Making Research Data Public: Why, What, and How**
  – March 1\(^{st}\) 10AM – 11AM
Intro to Data Management Objectives

1. Learn the benefits of data management
2. Learn the components of data management
3. Identify gaps in your current practices
4. Know where to find help
Data Management Practices

• complete records
• find
• understand
• back-up
• migrate
Data Management Practices

• I’ve lost two patent claims because we didn’t have complete records in lab notebooks.

• I panic every time I try to find my students’ data.

• I know I won’t be able to publish anything after my grad student leave because it’s too hard to understand what he/she did.

• We don’t have the assay results with substrate X, because I didn’t back-up my computer. And we’re out of substrate X.

• The only reason I was able to do the analysis was because on a whim I decided to migrate some of my data off of floppy disks to CDs back in the 90s. I wish I had done all of it.
Benefits of Good Data Management

- Better understand your data now and in the future
- Access your data in the future
- Save time and effort
- Meet funder and publisher requirements
What are the elements of good data management?
Plan
Data Management Planning Basics

- Data Management Plans (DMPs)
  - What data produced, how will you make it accessible, how will you preserve it, what restrictions are there on access

- NSF - 2011
- NIH - 2003 (>500K)
- DOE - 2014
DMPTool – http://www.dmptool.org/
Organize
Organization

- Distinguish Raw Data
- Be Consistent in File Naming and Directory Structures
- Define and Exercise Version Control
**Date Tip**

**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. 27/2-13 2013.158904109 MMXIII-II-XXVII MMXIII 4/2/355555555555555555555555555

(3+3×(11+1)-1×3-1/3) 2013

10/1101 1101 02/27/2013 20140904 20140904 20140904 20140904

VS.

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

Run 1 B anth meth Sept 4 .txt
BAM Rxn 2 2014_09_04.txt
20140904_meth_3.txt
Describing Data

• Enables someone unfamiliar with your data to find, evaluate, understand, and reuse your data
• Helps YOU to remember details about your data after time has passed

Examples

ReadMe File:  http://goo.gl/eOzQOO
Data Dictionaries:  Page 4-1 of this User Guide  http://goo.gl/Ugqiqo
Codebook:  http://www.icpsr.umich.edu/icpsrweb/ICPSR/help/cb9721.jsp
Basic Documentation

• Generic Information:
  • title, key dates, creators, subjects, rights, files, formats, versions, project & funder

• Additional Information:
  • source of the data, how analyzed, software and hardware requirements, resulting publications
Storage & Backup
Storage & Back-Up Best Practices

- Security
- Robustness
- Number of copies
- Location of copies
- Scheduling
- Efficacy
## Considerations for Sensitive Data

<table>
<thead>
<tr>
<th>Data Type</th>
<th>Requirement</th>
</tr>
</thead>
<tbody>
<tr>
<td>protected human information (PHI)</td>
<td>contingent on compliance with FERPA/HIPAA/IRB regulations and guidelines</td>
</tr>
<tr>
<td>personally identifiable</td>
<td>contingent on ability to de-identify</td>
</tr>
<tr>
<td>proprietary data</td>
<td>contingent on campus IP regulations</td>
</tr>
<tr>
<td>classified data</td>
<td>contingent on campus export control</td>
</tr>
<tr>
<td>local sensitivity</td>
<td>contingent on lab/collaboration culture</td>
</tr>
</tbody>
</table>
Anticipate the Long Term

• What happens to your data in...
  • 2 years?
  • 5 years?
  • 20 years?
• Will the formats be valid?
• Will the hardware be defunct?
• Will anyone know where it is?
Anticipate the Long Term

- More Control
- More Responsibility
- Less Control
- Higher Potential Impact

Keep Private

Make Public

re3data.org
Intro to Data Management Objectives

1. Learn the benefits of data management
2. Learn the components of data management
3. Identify gaps in your current practices
4. Know where to find help
Contact

Elizabeth Wickes
Data Curation Specialist
wickes1@illinois.edu

Christie Wiley
Engineering Data Services Librarian
cawiley@illinois.edu

visit:
310-312 Main Library

call:
(217) 300-3513

website:
researchdataservice.illinois.edu

Twitter:
@ILresearchdata
Data Management Workshop Series

• Introduction to Data Management
  – February 16\textsuperscript{th} 10AM – 11AM

• Documentation and Organization for Data and Processes
  – February 23\textsuperscript{rd} 10AM – 11AM

• Making Research Data Public: Why, What, and How
  – March 1\textsuperscript{st} 10AM – 11AM
Scholarly Commons
http://www.library.illinois.edu/sc/index.html

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

Bonus Tips!

Paste DOI here.
No VPN needed for access to full text (just netid/password)!
Useful?
Tell your friends and colleagues!

Not Useful?
Tell us!