Introductions!
Research Data Service (RDS)

The Research Data Service provides the Illinois research community with expertise, tools, and infrastructure to manage and steward research data.

- Knowledge around data policies, resources, archiving, & preservation
- Consultation for data management planning & implementation
- Workshops on data management, documentation, and data publishing
- Data Management Plan reviews and DOI minting services
- Solutions for public access to research data
- Centralized, private storage for active (“working”) data (with NCSA)

visit: researchdataservice.illinois.edu or email: researchdata@library.illinois.edu
What do we do?

**Expertise**
- Knowledge around data policies, tools, resources, archiving, and preservation
- Consultation and workshops for data management planning and implementation

**Tools**
- Data Management Plan creation wizard (DMPTool.org)
- Tools for data citation (DOI minting)

**Infrastructure**
- Illinois Data Bank (self-deposit institutional data repository)
Workshop goals

• Understand what documentation can look like
• Choose what is relevant
• Come up with a relevant action plan
• Start an outline
Documentation Content

Detail

Project
Dataset
Data file
Datum

Workflow

Experimental Procedures
Transformations
Workflows
Analysis
Complexity

- Generally, the more complex the project, the more complex the documentation
Documentation can be for...

- Maintaining consistency of data
- Training new staff/students
- Assessing data for reuse
- Assistance in actual reuse
- Efficiency in archiving
Activity 1: Using Documentation

Step 1: Go to these dataset pages


Han, Xueying; Appelbaum, Richard; Stocking, Galen; Gebbie, Matthew. International STEM Graduate Student in the United States Survey 2015. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2015-08-10. http://doi.org/10.3886/E43668V1
Activity 1: Using Documentation

Step 2:

• Review the ICPSR dataset pages, any documentation files, etc.

• Download the data files.

• What documentation is there? How many participants did each study have? What was the gender breakdown for each?

• There are some curveballs here! If you get stuck, move on to the other one. This isn’t a a test, don’t stress too much. Take notes on what was helpful and what was confusing.
Activity 1: Using Documentation

Answers!

• Meili (2015):
  • There were interviews and coded case results. The data file from the interviews doesn’t list the demographics, but the article does.
  • From the associated article: “Thirty-five were men and sixteen were women.” (page 148)
    • https://wp0.its.vanderbilt.edu/jotl/2015/03/article-do-human-rights-treaties-help-asylum-seekers-lessons-from-the-united-kingdom/

• Han, Appelbaum, Stocking, & Gebbie (2015)
  • Q3: 1="Male"; 2="Female", 3="Other", 4="I do not wish to respond”
  • Domestic: 742 male, 769 female, 8 other, 16 refused
  • International: 482 male, 292 female, 13 refused
  • Total: 1,224 male, 1,061 female, 8 other, 29 refused
Discussion

• What was similar and different about the documentation for these datasets?

• What uncertainties, questions, or confusion points did you encounter in determining your answers?

• What did you find helpful, convenient, or crucial in determining your answers?

• How did having a numerical code versus text content change your ability to work with the data?

• What was the most minimal piece of information you needed to answer the questions?
Levels of documentation

• Each dataset has unique needs:
  • Project:
    • What was done, with what instrument, to what, etc...
  • Dataset:
    • Manifest of files in the package, groups, etc...
  • Data files:
    • Contents and file names
  • Data point:
    • Codebook of text content, units, etc...
Minimum viable documentation

• What documentation doesn’t need to be:
  • A dissertation on the project
    • Leave that to your publication and other project documentation
  • Overly detailed for the people who aren’t going to use it
    • You can presume they have similar technical/domain knowledge

• What documentation should be:
  • **Enough** information,
  • about the **project, methods, and materials**
  • such that the information is **maintainable** over time,
  • in an **accessible format**,
  • and valuable for those **who need it**.
Use the tools you have

• Many tools can capture information about your data and store it with the data (e.g. a readme tab in an Excel file)

• Built in metadata functionality:
  - Equipment: cell phones, cameras, scanners
  - Software: ArcGIS, Microsoft Word, Adobe Photoshop

• Common metadata tools:
  - Spreadsheet software: Google Sheets, Microsoft Excel, OpenOffice/ LibreOffice Calc
  - Text editors: Notepad, Notepad++, Atom, Microsoft Word

• Many generic and discipline-specific tools. What’s common in your field?
  - E.g., ArcCatalog, Dublin Core Generator, Colectica
Examples of Documentation

• Readme Files
  • Text files that provides basic information about a dataset, such as:
  • accounts for all files and folders in a dataset
  • High level info: author, year, associated publication as appropriate
  • explanation of naming conventions
  • relationship between directory structure and the data

• Data Dictionaries/Codebooks
  • “Provides a detailed description of each element or variable in your dataset”. –
  • See examples linked in handout
Activity 2: Begin sketching your documentation

• Just for clarification, we don’t expect you to finish all these activities. So just try to get started on them, and these materials are yours to take home.

• These worksheets are meant to be prompts to help you go think about these things, and not meant to be your complete documentation.
Activity 2: Begin sketching your documentation

• Step 1: Try to think of a specific dataset you are working with.
  • You may also answer these questions for the general type of data that you work with.
  • Alternatively, use one of the datasets from Activity 1.
• In the space provided, write down the name of the project or dataset you will be using for this activity.
Activity 2: Begin sketching your documentation

- Step 2: Determine the audience of your data.
- This may be just you in the short term, but could potentially include others. Think through the future of your data for the short, medium, and long term.
- Place a checkmark in the table to indicate the timespan and the audience.
- Blanks are provided and you may chance any wording as necessary.
Activity 2: Begin sketching your documentation

• Step 3: Identify the things that make up your project. Think of all the specific devices, services, physical materials, and digital files used for your project.

• Use the spaces provided to jot them down. If applicable, use arrows to connect items to indicate a workflow.

• Next, use the grid to document where everything is, what it is called, authors, etc. Change these fields as desired.
Activity 2: Begin sketching your documentation

- Step 4: Identify the relevant sections for your documentation file.
- Check all that apply.
- Use other marks to indicate uncertainty and areas where you need to ask more questions of your team.
Activity 2: Begin sketching your documentation

• Step 5: Review some example readme files and other online documentation.

• Cornell’s “Guide to writing "readme" style metadata”
  • http://data.research.cornell.edu/content/readme

• ICPSR’s Data Preparation Guide: Important Metadata Elements (Social science)
  • https://www.icpsr.umich.edu/icpsrweb/content/deposit/guide/chapter3docs.html

• Find a data repository with data in your research area. Review some of their guides or a few popular deposits.
Activity 2: Begin sketching your documentation

• Step 6 and homework: Begin writing! Start with the small sections you’re sure of, identify sections where you’ll need to get the input of others, and just start writing! Use a blank piece of paper or a computer.

• The room is booked for another 30 minutes if you want to stay to work more.