Fall 2016
Hello
my name is
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)

Infrastructure
• Illinois Data Bank (self-deposit institutional data repository)
  • http://databank.illinois.edu
Workshop goals

• Explore the data sharing landscape and issues related to data deposit
• Learn basic assessment of repositories related to your discipline
• Understand the importance of data preparation for sharing and deposit
• Plan for deposit of your research data
Data Repositories

• Disciplinary
  – GenBank (Genetic Sequencing, 1982)
  – ICPSR (Social Sciences, 1962)

• Multidisciplinary
  – Figshare (2011)
  – Zenodo (2013)

• Institutional
  – Illinois Data Bank (UIUC, 2016)

How do I find a repository?
re3data.org
Considerations for choosing a repository

• Mandates
  – Funding agency
  – Publisher

• Sensitivity/Privacy
  – Student data (FERPA)
  – Health information (HIPAA)
  – Patent (Patent Law)

• Audience or Users

• Scope

• Standard practices in your discipline

• Fees
Evaluating a repository

• How do you know that a repository is “good”? Reliable? Won’t disappear tomorrow (and if it does, what happens to your data)?
• Follows disciplinary/data curation best practices?
• Indicators of stability and disciplinary support or endorsement?
• Preservation services offered? Commitment to long-term preservation?
• Persistent Identifiers
Activity 1: Exploring data repositories

**Data Project and Repository Features Comparison Tool** worksheet

[http://www.re3data.org/](http://www.re3data.org/) online directory of data repositories

**Step 1:** Use [http://service.re3data.org/browse/by-subject/](http://service.re3data.org/browse/by-subject/) to select one data repository that might be appropriate for your current or upcoming research project.

→ Consider subject coverage and scope, and types of data/files accepted.

**Step 2:** Using the **Data Project and Repository Features Comparison Tool** worksheet, and identify at least three criteria to investigate, and complete the “Repository” column for each of your selected criteria.

→ Information may be found in the repository’s description in re3data.org repository description, and/or in the documentation and policies available on the repository’s webpage.

**Bonus:** Complete the “My Project” column for your selected criteria, and then complete the “Repository Features Match Project Needs?” checklists for these criteria.
Report & Discuss

- Limitations?
- Benefits?
- What support is there
  - for researchers?
  - for depositors?
  - for data users?
Preparing your data

• What file types, formats, and sizes
  – Check repository specifications
• Specialized software needed to read files
  – Include Proprietary and Open Source formats
• What metadata is required
• Access Restrictions
  – Length of time (embargo)
  – Users
  – Environmental controls
Ethical Considerations

Human Subjects

• Institutional Review Board (IRB)
• Privacy, confidentiality (Informed Consent)
  – Strip identifiable information
  – Decrease geographic granularity
  – Generalize categories
  – Top code
  – And ask for help!
Intellectual Property, Copyright, and Licensing

- Who owns and controls the data
- What are the PI’s restrictions on data?
- Is it yours to share?
- Are you using data from a third party?
- Are you working with an external party, particularly a commercial entity or industry partner?

creativecommons.org

Credit: Michael Porter

Attribution-NonCommercial-NoDerivs 2.0
Activity 2

• Begin completing the deposit considerations form for a dataset you have in mind.
• This worksheet is designed to help you capture information you know and help you think through topics.
• Not 100% for all domains, and not a decision tree.
Wrap Up

• **Assess** your project’s data sharing needs, requirements, restrictions.

• **Identify and evaluate** suitable data sharing/publication options.

• **Compare** individual project needs with features of data sharing/publication options, and select “best” option.
Open Lab

Feel free to stay and keep working!
We will be here to answer questions.