

# Data Management Workshop 3: Preparing for Data Sharing

Elizabeth Wickes, Data Curation Specialist ([wickes1@illinois.edu](mailto:wickes1@illinois.edu))

Beth Sheehan, Social Sciences Research Services Librarian ([edivince@illinois.edu](mailto:edivince@illinois.edu))

Karen Hogenboom, Head of Scholarly Commons ([hogenboo@illinois.edu](mailto:hogenboo@illinois.edu))

Peg Burnette, Biomedical Sciences Librarian ([phburn@illinois.edu](mailto:phburn@illinois.edu))

Qian Zhang, Post-doctoral Research Associate ([zqian1@illinois.edu](mailto:zqian1@illinois.edu))

## About this workshop

Making research data public is becoming a reality for many disciplines, but for many researchers and disciplines there is a complicated set of issues to consider before publication or sharing of data. This workshop will cover the basic steps of research data sharing & publication, from initial considerations to depositing. Participants will work through guidance to help them make decisions about when and how to publish or share data. An optional 30 minutes of extra lab time will be available for participants.

## Upcoming workshops

### Data Rescue (Data Management Workshop 1)

October 4, 12-1pm, Grainger 335

Registration link: <http://illinois.edu/calendar/detail/4068?eventId=33240142>

Have a mess of data floating around your computer or lab? Data management is an essential task for students and faculty but hard to get started. Go no further! In this workshop you will identify, group, and plan on how to reorganize your current and future data. We will discuss strategies for organization, folder structure, and create an organizational plan. This workshop is the first in a three-part series about data management.

### Creating Data Documentation (Data Management Workshop 2)

October 11, 12-1pm, Grainger 335

Registration link: <http://illinois.edu/calendar/detail/4068?eventId=33244239>

Writing project, code, and data documentation doesn't need to be the worst part of your day. This hands on workshop will give you experience using various types of documentation, discuss strategies for writing documentation, and get you started writing a template for your projects. Bring a dataset you'd like to work with but examples will be provided.

### Smart and Simple Data Management

October 7, 1-2:30pm, Funk/ACES 509

Registration link: <http://illinois.edu/calendar/detail/4068?eventId=33240146>

October 12, 1-2:30pm, Grainger 335

Registration link: <http://illinois.edu/calendar/detail/4068?eventId=33240144>

This session aims to provide you with data management best practices and tools to increase your research efficiency and impact. We'll present a basic introduction to data management using a data management plan framework, hands on activities, and discuss how to find and vet resources for making data publicly accessible.

### Savvy Researcher Workshop Series

Find more workshops on tools to improve your research at: <http://illinois.edu/calendar/list/4068>



## Activity 1: Exploring Data Repositories

1. Use re3data.org to select one data repository that might be appropriate for your current or upcoming research project. Initially, consider subject coverage and scope, and the types of data/files accepted.
  - a. Subject Directory: <http://service.re3data.org/browse/by-subject/>
  - b. Content Type Directory: <http://service.re3data.org/browse/by-content-type/>
  
2. Review the **Data Project and Repository Features Comparison Tool** worksheets, and identify at least three criteria to investigate. Try to pick the three that are your biggest concerns about the repository? Write down the three you select below, or write in a concern that you have but don't see on the worksheet.
  - 1.
  
  - 2.
  
  - 3.
  
3. Complete the "My Project" column information for the three criteria with actual or anticipated specifications for your current or upcoming research project. Compare your project's needs to the information about the repository, and complete the "Repository Features Match Project Needs?" checklist for each criterion.



	MY PROJECT	REPOSITORY	Repository Features Match Project Needs?
A.	<b>Publisher/Funder Data Sharing Requirements</b>	Features Meeting Requirements; Concerns or Unmet Requirements	<p><input type="checkbox"/> Yes: Repository meets all requirements, OR project has no publisher/ funder requirements.</p> <p><input type="checkbox"/> No: Repository does not meet all publisher/ funder requirements.</p> <p><input type="checkbox"/> Partial/Maybe: Publisher/ funder has not yet been chosen OR repository meets mandatory requirements but does not meet optional/ preferred.</p>
B.	<b>File types and Formats Number and Size of Files</b>	File types, Formats, Sizes Accepted/Supported	<p><input type="checkbox"/> Yes: Accepts project file types, and number and size of files.</p> <p><input type="checkbox"/> No: Does not accept/support project file types, number and size, OR is not accepting new deposits.</p> <p><input type="checkbox"/> Partial: Does not accept existing file types, but files could be converted without loss of utility</p>
C.	<b>Discipline(s) or area(s) of research</b>	Disciplines or areas of research covered (if multidisciplinary, note quantity of related research area data)	<p><input type="checkbox"/> Yes: Disciplinary match, OR significant amount of same discipline</p> <p><input type="checkbox"/> No: Not a disciplinary match, OR minimal data from related disciplines, OR repository scope eliminates project eligibility.</p> <p><input type="checkbox"/> Partial: Disciplinary match, but minimal data from related topics and scope.</p>
D.	<b>Funding Available for Deposit Fees</b>	Cost to Depositor	<p><input type="checkbox"/> Yes: Cost is within budget, OR there is no cost</p> <p><input type="checkbox"/> No: Cost to deposit is higher than budgeted funds.</p> <p><input type="checkbox"/> Partial: Additional funding may be available to cover funding cost, but is not guaranteed.</p>
E.	<b>Data Availability Period Requirements</b>	Data Retention or Data Availability Guarantee	<p><input type="checkbox"/> Yes: The repository clearly states that the data will be made available for the time period I need.</p> <p><input type="checkbox"/> No: The repository doesn't clearly state how long data will be retained or won't retain data for the necessary time period.</p> <p><input type="checkbox"/> Partial: Retention policies are unclear or certain conditions must be met for data to be retained long enough for the project.</p>
F.	<b>Intellectual Property (IP) and Licensing Permissions or Restrictions</b>	Licensing Requirements or Options	<p><input type="checkbox"/> Yes: Licensing options match project/IP needs</p> <p><input type="checkbox"/> No: Repository requires a more permissive license than can be granted for the project data</p> <p><input type="checkbox"/> Partial/Maybe: Licensing requirements are unclear or absent; project licensing permissions or restrictions are currently unknown; or other IP issues prevent data deposit into this repository</p>
G.	<b>Privacy Concerns</b>	Features Addressing Privacy Concerns	<p><input type="checkbox"/> Yes: Features and requirements adequately address privacy concerns, OR project has no privacy concerns</p> <p><input type="checkbox"/> No: Data cannot be deposited in this repository without violating privacy concerns</p> <p><input type="checkbox"/> Partial: Privacy protection is sufficient if sensitive aspects are not deposited</p>
H.	<b>Data Access Restrictions Required</b>	Access Restriction Options	<p><input type="checkbox"/> Yes: Repository options satisfy project access restriction needs, or no access restrictions necessary</p> <p><input type="checkbox"/> No: Repository options do not support project's access restriction requirements</p> <p><input type="checkbox"/> Partial: Access options are sufficient if sensitive aspects of project data are not deposited</p>
I.	<b>Data Use/ Reuse Statistics Desired</b>	Data Use/ Reuse Statistics Tracking and Reporting	<p><input type="checkbox"/> Yes: Desired use stats tracked, OR data use tracking is not a priority.</p> <p><input type="checkbox"/> No: Use stats are not tracked, OR data use stats are tracked internally but will not be shared with depositor or others</p> <p><input type="checkbox"/> Partial/Maybe: Some use stats are tracked but not all types desired, OR availability of use stats is unclear or inconsistent</p>



A. Publisher/Funder Data Sharing Requirements:

My Project: Do you plan to publish in a journal with a data deposit requirement? And/or is the data funded by an agency that requires data sharing? If so, what are the stipulations?

Repository: Which repository features satisfy your publisher/ funder requirements? Are there any requirements that cannot be satisfied by this repository, and are these requirements mandatory, or optionally preferred or recommended?

B. Data Files Description (File types, formats, sizes):

My Project: What are the file types and formats of your data? What is the approximate number and size of data files. What software or analysis tools are required to view and/or manipulate your data?

Repository: What file types and formats are accepted by the repository? Is there a limit to the number of files and/or the individual or cumulative file size accepted? If they do not support your file types or formats, does the repository provide support for converting files to alternative formats? Would your data still be useable if converted to their preferred types/formats?

C. Data Scope (Disciplines or Areas of Research):

My Project: What disciplines or areas of research would be most interested in your data? What aspects of your research scope might help researchers discover and locate your data if it was deposited in a repository with similar scope (for example, similar geography, species, time period, data collection tools)? This may vary by discipline.

Repository: What discipline(s) are served by the repository? If it is multidisciplinary, how much of current holdings is in the same discipline/subdiscipline as your project, or have a similar research scope or topic? Is there anything that they will not accept due to the scope, and could this disqualify your project from inclusion?

D. Cost to Deposit:

My Project: Do you have any funding available to pay for deposit of data, if a fee is required?

Repository: Does the repository charge a fee to the depositor? If so, how is the fee determined? Can you obtain an estimate in advance?

E. Data Availability Period Requirements:

My Project: Does your funder, publisher, institution, or other related entity require a specific timespan for your data to be retained or remain publicly available? You may also know or feel that your data has a value period of a certain number of years.

Repository: Does the repository clearly state their data retention policy? Do they mention the specific number of years that data will remain hosted and available from the repository? Some repositories may have stipulations that data must be deposited with documentation and other supporting information in order to be retained for the maximum number of years.

#### F. Licensing Requirements:

**My Project:** Did your data originate from a third party, and if so, what are the agreements or requirements for use and reuse of the data? Do you have permission to share the data and/or the products of data analysis and processing, and are there any requirements or restrictions on licensing or reuse of the original or processed data? Did you collaborate with another researcher who also needs to give permission before sharing the data? Do you have the authority to assign a specific level of licensing for use/ reuse for shared project data? Who “owns” the project data, and are there any limits or specifications for licensing the data due to publisher/ sponsor/ institutional requirements?

**Repository:** Does the repository address issues of intellectual property in their instructions to depositors, and provide guidelines or options for handling intellectual property issues? Does the repository require data to be deposited under a specific type(s) of license for use/ reuse and attribution (for example: Creative Commons 0, Creative Commons Attribution, etc). If so, do you have the authority to grant the required level of license for deposited data?

#### G. Privacy Concerns:

**My Project:** Would sharing all or any portion of your data violate terms approved by the IRB, laws, or disciplinary/ institutional guidelines protecting confidentiality of research subjects? What would you need to do to clean, redact, or anonymize data before it could be shared? Would the data still be useable if shared without this information?

**Repository:** Does the repository accept data that has been cleaned, redacted or anonymized? Do they provide any support to the researcher to facilitate scrubbing of data, or include features in repository policies or processes that provide privacy protection?

#### H. Data Access Restrictions:

**My Project:** Do you need to restrict access to all or part of your data? If so, which parts, for how long, and to whom should it be restricted? Consider questions of intellectual property, privacy and cultural sensitivity, impact of data availability prior to publication, licensing requirements, ownership of data, etc.

**Repository:** Are options provided to restrict access to all or part of the data? Will these options meet the access restriction needs of your project? Can these settings be altered later, and who can make these changes?

#### I. Data Use/Reuse and Impact:

**My Project:** Is the ability to track use/reuse of your project data important to you? What metrics or information about the use and reuse of your data would be helpful in measuring the impact of your research?

**Repository:** Does the repository assign a persistent identifier (ie. DOI, ARK) to the data? Does the repository track data use statistics at the project and/or file level, and what do they track (for example: number of views, number of downloads, etc)? How do they make these statistics available, and to whom? Does the repository provide features to facilitate linking of data records with associated publications or citations derived from that data?



## Activity 2: Sharing Considerations Worksheet

<p><b>1) Identify the repository for deposit</b></p> <p>Use the Activity 1 worksheet to identify and evaluate a relevant repository</p>	Name:			
	URL:			
	Is there a charge for deposit, curation, file storage, etc.?		Who will be responsible for paying?	
	Who will be making the actual deposit or leading the deposit process?			
	When does the data need to be available from the repository?			
	What type of persistent identifier will the repository issue for your data?			
	Write a brief description of the deposit process. Provide a link to information on the deposit process if there is one.			
<p><b>2) Write a title</b></p> <p>This title will appear in your dataset's citation. You may repeat parts of the associated article's title, e.g. "Dataset for..."</p>				
<p><b>3) Write a description</b></p> <p>Think of this as the abstract for your dataset. This may include portions of your associated article's abstract, but does not need to be a full copy of your documentation file.</p>				
<p><b>4) Write keywords</b></p> <p>You may want to use similar keywords to those assigned to related publications. Try to aim for five or fewer.</p>				

5) Collect author information	#	D?	CA?	Author name	Email and/or ORCID number	Other information	
<p>These are the authors who will be listed in the dataset's citation.</p> <p>Column header key:</p> <p>#: Author number</p> <p>D?: Indicate who will perform the actual deposit process. This is usually one person, but some workflows require several.</p> <p>CA?: Indicate who will be the corresponding author.</p> <p>Other information: Use to record institution names, titles, or other required information for authors.</p>							

<p><b>6) Select a license to release the data under</b></p> <p>This section has some guidelines to help you think more about selecting a license, but selecting a license can be a complex process that cannot be appropriately captured in a worksheet. <b>We strongly recommend reaching out for help with professionals who can help navigate funders, publishers, copyright, patent, etc. issues.</b></p> <p>Contact the Research Data Service (<a href="http://researchdataservice.illinois.edu/contact-us/">http://researchdataservice.illinois.edu/contact-us/</a>) for more help. Members of the RDS staff can get you connected with a network of experts to help with your data submission process.</p>	<p>Are you accessing or using your data under terms of a contract or other license? Do any of those items have specific license requirements?</p>
	<p>Does your funder, publisher, or institution have requirements about which licenses you can release data under?</p>
	<p>Are there any other entities or groups that will need to comment or approve a license selection?</p>
<p><b>7) Repository timespans and publication delays</b></p>	<p>Is your data required (for example, by a publisher) to remain unavailable until the time of article publication or sometime after? Is your data required to be made available within a certain time period after a publication?</p>
	<p>Is your data required (for example, by a funder or institution) to remain available for a certain amount of time?</p>

<p><b>8) Identify any other information or items necessary for deposit</b></p> <p>Think of items like additional descriptive elements required by the repository. For example, GIS coordinates, documentation files, a readme, etc.</p> <p>Help on writing documentation is available in the handout.</p>	Item	Required?	Who will prepare this?	Due date
<p><b>9) Identify any related materials that the dataset should be linked to</b></p> <p>These may be articles, publications, code repositories, other datasets, theses, etc.</p>	Item	URL or other location		Kind

## Homework!

1. Complete the remaining criteria in the Data Project and Repository Features Comparison Tool with information about your project and the selected repository. OR  
If you have determined that the initial repository would not be appropriate, select a new repository to examine and complete the comparison tool worksheet.
2. Visit the data repository's website, and attempt to download and use a dataset. Try to find one that is similar to your own project's data. Reflect:
  - a. Are the search, browse and discover features of that repository easy to use? How easy or difficult is it to find the data that you want?
  - b. Does the repository provide the necessary documentation in order to interpret and use the data? If not, what is missing?
3. Review all guidelines and instructions for depositors provided on the repository's website. What support or assistance, if any, is provided for depositors? What support or assistance will you need, internally and/or externally, to prepare and deposit your data?
4. What are the benefits to you and your research if you deposit in this repository? What are the limitations or risks?

## Other important considerations

- These worksheets are meant to represent the most general situations and are not meant to capture every possible scenario, need, or consideration.
- Data sharing practices across disciplines have differing best practices, repositories, licensing suggestions, etc. Look to your community and consult with domain experts to get a fuller picture.
- Many of these decisions should happen in consultation with team members, PIs, etc. Creating a standard practices for sharing for your team can save time and makes it easier for newer members to understand these choices.
- Seeking expert help, particularly when considering any required release timelines, publication delays, licensing, and other complex issues can make the decision making process easier. The University Library has experts in these areas, including domain experts, who can help you navigate the data sharing process. Contact the Research Data Service at [researchdata@library.illinois.edu](mailto:researchdata@library.illinois.edu) to get connected to these relevant experts.

## Relevant Links

- Illinois Data Bank: <https://databank.illinois.edu>
- IDEALS: <https://www.ideals.illinois.edu/>
- Research Data Service: <http://researchdataservice.illinois.edu/>
- Scholarly Commons: <http://www.library.illinois.edu/sc/>
- Office for the Protection of Research Subjects: <https://opr.s.research.illinois.edu/>

## Need more help after this workshop?

The Research Data Service (<http://researchdataservice.illinois.edu/>) team is happy to provide more information about all data management topics and offers personal data consultations. They can be reached at [researchdata@library.illinois.edu](mailto:researchdata@library.illinois.edu) or on Twitter @ILresearchdata. This workshop and many others can be run for your research group or lab by request. All Research Data Service consultations are free and confidential to all members of the Urbana-Champaign campus community.