

## Syllabus: LIS 341, Research Data Management Across the Disciplines

School of Library and Information Studies

University of Wisconsin-Madison

Summer 2012

Dorothea Salo (please call me "Dorothea")

Office address: 4261 Helen C. White Hall

Jabber/GTalk: salo@wisc.edu

salow@wisc.edu, 608-265-4733

Office Hours: by appointment

Skype: dorotheasalo



To the extent possible under law, the person who associated CC0 with this work has waived all copyright and related or neighboring rights to this work. This work is published from: United States.

### Course Objectives

- Understand why managing research data is important
- Navigate data-ownership, data-security, and sensitive-data issues
- Be aware of national law, funder policies, and institutional policies around research data
- Construct and implement a sensible project workflow for data
- Avoid common scenarios of data loss and data leakage
- Triage data and metadata formats for sustainability and usefulness
- Package and license data for publication and sharing
- Cite datasets
- Write a data-management plan for a research funder

### Course Policies

\*\*I wish to fully include persons with disabilities in this course. Please let me know within two weeks if you require special accommodation. I will try to maintain the confidentiality of this information.\*\*

Academic Honesty: I follow the academic standards for cheating and plagiarism set forth by the University of Wisconsin.

### Readings

There are no required textbooks for this course. As useful links come up, they will be posted to the class link page (<http://pinboard.in/u:dsalo/t:341>) so that participants can bookmark them for future use.

## MONDAY: Bootstrapping

*Learning objectives: Why research-data management matters. Law and policy associated with data management (HIPAA, FERPA, FISMA, OMB Circular A-110, federal-agency guidelines, data-management plan requirements, institutional policy). Data ownership (copyright, patent law, institutional policy).*

Linklists: <http://pinboard.in/u:dsalo/t:nsf>, <http://pinboard.in/u:dsalo/t:horrorstories/>, <http://pinboard.in/u:dsalo/t:dataownership>

**9:00-9:15 Introductions; how the class will work**

**9:15-9:45 What's the big deal about data management? Why does it matter?**

**9:45-10:15 What are all these data-management policies I've been hearing about? Which ones apply to me?**

**10:15-10:30 BREAK**

**10:30-11:30 Who owns the data I generate? How do I keep my data secure?**

**HOMEWORK: Data "elevator pitch." Data inventory.**

## TUESDAY: Keeping it together

*Learning objectives: Data lifecycles. Project management. Workflow change. Organizing data files while a project is in progress. File versioning. File-naming conventions. Short-term backup (physical storage, cloud storage, local storage).*

Linklists: <http://pinboard.in/u:dsalo/t:datamodels>, <http://pinboard.in/u:dsalo/t:projectmanagement>, <http://pinboard.in/u:dsalo/t:filenaming>

9:00-9:15 Pitch your data!

9:15-10:15 Data lifecycles; project management tools; changing workflows

10:15-10:30 BREAK

10:30-11:30 File management; file-naming conventions; short-term backup

**HOMEWORK: Data lifecycle. File-naming conventions. Project README file.**

## WEDNESDAY: Bits and bytes

*Learning objectives: File formats. Assessing file formats for sustainability. Research-data-specific file formats. Basic data security. Metadata. Metadata types. Finding, assessing, and using metadata standards.*

Linklists: <http://pinboard.in/u:dsalo/t:fileformats>, <http://pinboard.in/u:dsalo/t:metadata>

9:00-9:15 Data lifecycle Q&A

9:15-10:15 File formats; assessing risks

10:15-10:30 BREAK

10:30-11:30 Metadata; metadata standards

**HOMEWORK: Data-files risk assessment. Metadata for your data.**

## THURSDAY: The long haul

*Learning objectives: Why backup is not the same as preservation. What you should and shouldn't keep. Finding a long-term home for data. Data repositories. Institutional repositories. Data journals. Disciplinary differences with respect to data repositories, metadata standards, funder policies, common practice, etc.*

Linklists: <http://pinboard.in/u:dsalo/t:digitalpreservation>, <http://pinboard.in/u:dsalo/t:datapublishing>, <http://pinboard.in/u:dsalo/t:repositories>

9:00-9:15 Risk and metadata Q&A

9:15-10:15 Long-term data preservation. Data repositories (disciplinary and institutional). Data journals. Data as "supplementary materials."

10:15-10:30 BREAK

10:30-11:30 Data in your discipline; catch-up time and Q&A if needed

**HOMEWORK: Data-management plan**

## FRIDAY: Sharing, finding, reusing, and citing data

*Learning objectives: Data sharing. Open Notebook Science. Dataset identifiers (DOIs for data). Data citation. Trends, general and disciplinary.*

Linklists: <http://pinboard.in/u:dsalo/t:datasharing>, <http://pinboard.in/u:dsalo/t:datacitation>, <http://pinboard.in/u:dsalo/t:opendata>

9:00-9:15 What have you heard about data sharing?

9:15-10:15 Licensing data. Citing data. Open science; open notebook science. Data use metrics; bibliometric impact of data publishing.

10:15-10:30 BREAK

10:30-11:30 Resource roundup!

## Grading Schema

| Assignments:         | Percentage | Due Date        |
|----------------------|------------|-----------------|
| Data-management plan | 50%        | Friday, June 15 |
| Class participation  | 25%        |                 |
| Daily homework       | 25%        |                 |

Final grade scale: 100-93.5 A; 93.4-89.5 AB; 89.4-83.5 B; 83.4-79.5 BC; 79.4-73.5 C and so on... No extra credit opportunities are available in this class. No assignment grades are dropped. Perfection is not the goal of any assignment; learning is. Mistakes and difficulties are to be expected, and will not count against your grade.

All assignments for this course should relate to a real-world research project you are working on. It may be your dissertation research, a project currently in planning stages (this would be ideal!), a project in your lab, or a project for which you provide technical or research support. If you do not have a project, I will provide you with a case study.

## MONDAY HOMEWORK

### PROJECT ELEVATOR PITCH

Write a brief (one-page double-spaced *maximum*) “elevator pitch” explaining the significance and power of your data. Imagine that you are speaking to another researcher *in a field not your own* who might find your data useful if you can interest her in it.

### DATA INVENTORY

Make a list of the data that your project will generate, keeping in mind class discussions about what is and isn't data. For each type of data, write down its format (“analog” if not kept on a computer; otherwise, file format). If your data change over the course of your project (e.g. through cleanup, analysis, or visualization), list the results of such changes and *their* formats as well.

## TUESDAY HOMEWORK

### DATA LIFECYCLE

Using your data inventory, draw a data lifecycle for your project. You may use any of the lifecycle models shown in class as a model, or you may depart from them if you prefer (a plain flowchart is fine!). Take special note of any lifecycle stages you are currently uncertain about; we will discuss them briefly in class Wednesday.

### FILE-NAMING SCHEME

For at least one of the data types you listed in your inventory (and preferably all of them), devise a file-naming scheme. Document it as though you were writing training materials for a new collaborator or assistant.

### PROJECT README.txt FILE

Given your data inventory and file-naming scheme, write a README.txt file as though you were putting it in the top folder containing all the files for your project. Email the file to me (sa1o@wisc.edu) before class on Wednesday.

## WEDNESDAY HOMEWORK

### DATA FILES RISK ASSESSMENT

First, write an overall risk assessment for the digital data arising from your project, including:

- poor-practice risks (be honest!)
- organizational and business risks (including to any web-based services you plan to use)
- technological risks (not specific to any particular file or file format)
- loss, attack, or accident risks

Then, for each type of file in your data inventory, write an individual risk assessment, including:

- business risks related to software or hardware producers
- hardware/software obsolescence risks

### METADATA HUNT

Find a DESCRIPTIVE metadata standard appropriate to your project. Using that standard's documentation and/or an example document from that standard, make minimal metadata for some or all of the data arising from your project. If you honestly cannot find a descriptive metadata standard for your discipline (not all disciplines have them!), use DC.

## THURSDAY HOMEWORK

### DATA-MANAGEMENT PLAN

Boil down the writing you have done over the course of this week to a funder-appropriate data-management plan. You may write to any of the guidelines discussed in class (if you choose NSF, you *must* choose a directorate or division), or another if you provide me a link to the applicable guidelines. You *must* observe length limits provided in funder guidelines! If there are none, two pages maximum.