NTRES 6600: Research Data Management Seminar

Spring 2017
Location: Fernow G24

First Class Meeting Monday, January 30, 2017
Class Meeting time: Monday 4:40 - 6:00

Instructors:   Cliff Kraft (cekk7)
               Erica Johns (emj73)

Office Hours:  by appointment

Grading:       S/U only, 1 credit

I. Rationale:

In an era that confronts us with an overwhelming capacity to collect data, knowing how to manage data is essential. Not only does the National Science Foundation (NSF) require a data management plan with all grant applications, a growing number of funders and publishers have data sharing requirements. Data management is equally important for the individual researcher trying to document, organize, and evaluate empirical information. In this course, students will learn how to organize data for efficient statistical and graphical analysis, develop metadata describing data content, and explore best practices for sharing and archiving research data.

II. Course Aims and Objectives:

Aims

This course is designed to introduce students in the environmental sciences to best practices regarding data management and to help students manage their data effectively and efficiently while meeting funder and publication requirements. Emphasis is on concepts rather than application, but specific examples of tools and resources for hands-on practice and application will be provided at every opportunity. In-class exercises will reinforce the ideas brought up throughout the course.

Specific Learning Objectives:

By the end of this course, students will be able to:

- Describe their research data lifecycle in order to identify areas for improvement in the research and data management process;
- Follow best practices involved in creating a flat database in order to maximize the likelihood of long-term preservation and potential for reuse;
- Follow best practices in structuring relational databases in order to make analysis and retrieval easier/more efficient for long-term studies;
- Interact with SQLite for analysis and visualization;
- Evaluate disciplinary data repositories in order to determine requirements and suitability for data deposit;
- Use the documentation accompanying a data set in order to evaluate data quality;
- Create appropriate documentation in order to prepare data for effective long-term use and reuse by themselves or others;
- Evaluate Data Management Plans (DMP) and have the tools to prepare an effective DMP that meets funder requirements.
III. Format and Procedures:

This class will meet once a week for 80 minutes, with a total of six class sessions. The class will be largely discussion-based, with readings available for each class via Blackboard. To facilitate meaningful discussions, participants are expected to read or otherwise prepare for each discussion before coming to class. Communication will also take place via the class Blackboard site, and participation will include posting a reaction or reflection on each day’s discussion.

IV. Our Assumptions:

As graduate researchers, you are experts in a particular subject area who are conducting research to accomplish a variety of goals. Our aim is to facilitate the research you are already doing and bring to bear our experience in data management to enhance your research. Therefore, this course is intended to help you frame the questions and issues you need to think about in relation to managing data for your particular research project. With this in mind we will focus on “hands-on” examples and associated discussions to help you apply these examples to your data. We do not expect you to become skilled data managers during this course. Instead, you will have an opportunity to expand your awareness of the emerging best practices in the field of data management, as well as develop a framework for thinking about data management for any research project in which you are involved.

V. Course Requirements:

1. Class attendance and participation policy:

Since this is a short course, we hope that you can participate in all class sessions.

2. Course readings:

There is no required text for this course. However, we will be asking you to read or watch materials posted on the course’s Blackboard site on a regular basis, which will be announced in class and through the Blackboard announcement email system. In addition, there will be a lesson about relational databases that will need to be reviewed prior to the 2nd class, in order to fully understand the hands-on exercise (2nd class meeting) and SQLite instruction (3rd class meeting).

3. Weekly Assignments:

Most weeks, we ask students to have completed background readings before coming to class, which will help facilitate useful discussion during class. You will get the most out of this class by being an active participant; that is your motivation and your reward.

VIII. Tentative Course Schedule: (May change to accommodate guest presenters & student needs)

Class 1  Introduction to Data Management

Jan. 30  We will use the first class session for introductions and logistics. The instructors will give a brief explanation of the data lifecycle and how the framework can help us to identify data management needs. We’ll then have a group discussion of research interests, data problems encountered, and data management needs.

Readings:


Class 2 & 3

Data Organization & Analysis

Feb. 6, 13

Organizing your data at the front end of a research project will save time and increase your ability to analyze data. This session will introduce you to the principles involved in creating a flat database (with Excel), reasons for taking the next step towards linking tables in a relational database, and will provide examples to help you organize your own data in this manner. Topics will include best practices for data types, how to handle missing data, organization by data type, and data file formats. The second session will include a hands-on introduction to SQLite. Bring your laptops to class 3, February 13, 2017.

Assignment:

1. Review the slide deck in Blackboard titled “Data Organization & Analysis” prior to class 2: will quickly review content before an applied exercise.
2. Download Mozilla Firefox and follow “SQLite manager” add-on instructions from Blackboard (upload by 2/6/2017) by class 3, 2/13/2017.
3. Bring your laptop to class 3.

Readings:


Class 4  
Data Sharing  

Feb. 27  
The NSF and other funding agencies have already adopted data sharing policies. Publishers also have data sharing requirements, whether they host data themselves, or expect researchers to deposit data in a data center or to make it available upon request. As a result, key questions arise, such as where to share data, what to share, and how to share? During this class session we'll discuss disciplinary databases, Cornell's eCommons digital repository, and some other sharing strategies, and will discuss evaluation criteria upon which to base your decision about where to share your data.

Readings:


Compare:


Resource:

Research data management service group - Sharing Data. http://data.research.cornell.edu/content/sharing-data
**Class 5**

**Data Quality and Documentation**

March 6

While written documentation -- for example, in a lab notebook -- is still important, the platforms on which modern researchers are working and collecting data are increasingly complex. How do you document your digital data and the steps you take to analyze it? Are your files sufficiently organized and well-described so that others can interpret what you’ve done? Or will you be able to interpret what you did three months from now?

During this class session on data documentation, we’ll discuss the challenge of remembering details relevant to interpreting your data and offer some best practices and strategies to adopt to organize and describe your data for yourself and others.

**Readings:**


**Resources:**

Disciplinary Metadata | Digital Curation Centre - [http://www.dcc.ac.uk/resources/metadata-standards](http://www.dcc.ac.uk/resources/metadata-standards)

Kozlowski, Wendy. (2014). Guide to writing “readme” style metadata. [http://data.research.cornell.edu/content/readme](http://data.research.cornell.edu/content/readme)

**Class 6**

**Data Management Planning (DMP)**

March 13

**Assignment:**

Review an assigned data management plan (provided via Blackboard by 2/20/2017). Meet with group, work through questions below, come prepared to discuss in class.

- Use this guidance (RDMSG - Data Management Planning [http://data.research.cornell.edu/content/data-management-planning](http://data.research.cornell.edu/content/data-management-planning)) to prepare your review.
- What information is missing that should be included in a comprehensive data management plan?
- What information is included that is extraneous or wrong? What modifications do you suggest?
- How do you think a funder’s proposal review board would respond to this data management plan? Would they approve it, send back to the PI for clarification/edits, or fail to accept it?

**Readings:**