

Research Data Management

Syllabus for GRAD 521

Course Information

2 credits (CRN 38469)

Offered: winter term (annually)

Meeting location: Furman Hall 105

Meeting time: 10 – 10:50 a.m., Tuesday & Thursday

Prerequisites: None, but **students should have a good idea of what his/her M.S./Ph.D. research project will be, if not already be engaged in the work.** This course is applied, and you will need to understand how to relate the material to your research workflow.

Web site: <http://guides.library.oregonstate.edu/grad521>

Instructor Information

Amanda L. Whitmire, Ph.D.

Valley Library, 4th Floor

Voice: 541.737.3133

amanda.whitmire@oregonstate.edu

Office hours: By appointment. Please send me an email or give me a call when you have questions or something you want to discuss. We can also meet before or after class.

Course Description

Careful examination of all aspects of research data management best practices. Designed to prepare students to exceed funder mandates for performance in data planning, documentation, preservation and sharing in an increasingly complex digital research environment. Open to students of all disciplines.

Course Content

This is a lecture-discussion course in which the instructor presents topics, discussion assignments are explained, and students complete assigned projects both during class periods and outside of class. This course is designed to incorporate substantial active learning approaches; lecture will be punctuated by individual and group activities. Significant student participation is expected. We assume that students generally have not had prior data management training, but students with prior experience can still expect to benefit significantly from this course. *The purpose of the course is to enable students to properly manage research data as a means to foster efficiency, ensure proper stewardship of data, and exceed funder requirements for planning, preservation and sharing.* This is *not* a course in data science or data analysis, and is not designed to prepare students to be database administrators.

Major course content includes the following seven topics. (See detailed course outline below under “Weekly Schedule”)

1. Overview of research data management, definitions and best practices
2. Types, formats and stages of research data
3. Data storage, backup and security
4. Metadata (data documentation)

5. Legal and ethical considerations of research data
6. Data sharing and reuse
7. Archiving and preservation

Measurable Student Learning Outcomes

By participating fully in this curriculum, the student should be able to:

1. Explain the need for managing/sharing research data, relevant public policies, and the lifecycle continuum for managing and preserving research data;
2. Identify potential re-users, communicate the value of their research data for re-use, and formulate a dissemination strategy;
3. Formulate an abbreviated data management plan or data curation profile to manage their research project data and define roles/responsibilities of research staff;
4. Explain the range of research data types, stages, formats, and relevant software that may need to be managed and preserved in future research efforts;
5. Propose what descriptive data needs to be documented in a standard way via metadata to allow research data sets to be managed and preserved;
6. Plan how to handle issues involved in securely storing research data in central databases, archives and/or repositories, backing it up, and managing access to data;
7. Explain legal (ownership) and ethical considerations related to data-sharing;
8. Plan for issues related to long-term preservation, discovery, and re-use.

Evaluation of Student Performance & Grading Plan

Students will be assigned a letter grade (A-F) based on the following:

Homework assignments:	20% / 50 pts.
Class participation:	10% / 25 pts.
Mid-term (a Data Curation Profile):	30% / 75 pts.
Final exam (a data management plan):	<u>40% / 100 pts.</u>
	100% / 250 pts.

Homework Assignments

Homework assignments will be assigned periodically throughout the quarter, and may include reading and reflective writing, reviewing case studies, etc. Homework is intended to reinforce or expand upon material covered in lecture, and prompts each student to think critically about the material and apply it to their own discipline and/or individual experience. This is how we connect discipline-agnostic material to your discipline-specific workflows.

Class Participation

Student engagement in group activities is important and productive. As such, participation will be a component of the final grade. Participation can be earned in several ways. Some of these capitalize on opportunities to create a discourse community in the class setting: asking questions during class, offering opinions and observations, and responding to fellow students' questions and ideas. Some affirm the knowledge-building of peer-to-peer learning: small-group discussions, small-group presentations, comparing notes with a partner, etc. Finally, engagement with the course may also be demonstrated via

writing: short in-class writing assignments, discussion board posts, and reflection exercises. Although students may earn participation via several means, two constraints determine what constitutes appropriate use of these methods. First, students are obliged to spread their efforts across methods over the entire term. In other words, one will not earn an 'A' in participation simply by contributing to class discussion often for one or two weeks, nor simply by submitting a few discussion board posts but never speaking. Second, the quality of the contribution impacts the grade. "Because I just think so", "No, really?" or similar vague or loosely formulated statements will not count toward the participation grade, nor will discussion that does not inform the topic for the day or the overall course theme.

Midterm Exam (take home)

As a means to understand the culture of practice within your research discipline, you will be tasked with interviewing someone within your field (your advisor, a favorite researcher in your department, etc.) in order to create a Data Curation Profile (DCP). A DCP is essentially an outline of the "story" of a dataset or collection, describing its origin and lifecycle within a research project. A DCP can, for a given research project, provide insight into areas of attention for data management, help assess information needs related to datasets created, help identify data services needed, and create a starting point or roadmap for curating a dataset for archiving and preservation. <http://datacurationprofiles.org/>

Final Exam (take home)

The primary scholarly output for a student taking this course is his or her individual data management plan (DMP). Ideally, this plan would address the research project that each student is involved in for his/her thesis or dissertation research, but plans describing a fictional research project from the student's discipline will also be accepted. This is a take-home 'exam,' and grading criteria will be provided. The activity of creating the plan compels each student to assimilate course content and apply it to his or her own research project. As such, the DMPs will serve as a direct measure of how well you have understood the material presented in the course, and if the proposed learning outcomes were achieved. A component of your final exam process will also be critiquing DMPs from other students in the class.

Learning Resources

There are no required texts or supplies for this course. We will supply web links or PDF copies of selected reading assignments to the students as needed. See the list of required readings in the Appendix.

Weekly Schedule

Week 1

Tuesday, 6 January: Overview of Research Data Management

- Introduction to course, instructor and fellow students
- Explain what research data is/are
- Describe the value and relative importance of data management to the success of a research project.

Thursday, 8 January: Introduction to data management plans & the research lifecycle

- Explain the lifecycle continuum to manage and preserve research data
- Understands US federal agency context for data management and sharing
- Identify data management plan (DMP) requirements used to characterize and plan for the lifecycle of research data

Week 2

Tuesday, 13 January: *Types, Formats, and Stages of Data*

- Explain what a research data set is, and the range of data types and formats
- Understands which data types are appropriate for answering different types of research questions
- Understands the lifecycle of data; Can identify stages of research data, and which data management actions are associated with lifecycle stages

Thursday, 15 January: *Data organization & documentation*

- Develops appropriate file- and folder-naming conventions
- Organizes data and documentation in a logical manner
- Creates standard operating procedures for data management and documentation, including proper use of a field or laboratory notebook

Week 3

Tuesday, 20 January: *Intro to your midterm project*

- Introduction to Data Curation Profiles

Thursday, 22 January: *-OPEN LAB-*

Week 4

Tuesday, 27 January: *Data Storage, Backup and Security*

- Understand why data storage, backup and security of research data are important
- Understand best practices methods for research data storage, backup, access control, migration to newer storage media and security of research data
- Formulate an approach to creating a data storage, backup and security plan for your project

Thursday, 29 January: *Metadata: Contextual Details Needed to Make Data Meaningful to Others*

- Understands the concept of, and rationale for, metadata
- Identify and list the types of information typically included in metadata records
- Identify applicable standards for documenting and capturing metadata in your discipline

Week 5

Tuesday, 3 February: *(in Autzen Classroom, Valley Library)*

- Familiar with tools for creating metadata appropriate for your discipline and data type
- Formulate an approach to creating metadata for a project
- Develops an ability to read and interpret metadata from external disciplinary sources

Thursday, 5 February: *Data entry and manipulation*

- Recognize inconsistencies that can make a dataset difficult to understand and/or manipulate
- Identify data entry tools
- Identify validation measures that can be performed as data is entered

Week 6

Tuesday, 10 February: *Intellectual Property, Copyright & Data Licensing*

- Explain ownership considerations related to data sharing
- Explain and evaluate potential legal issues connected to your data; intellectual property, copyright claims, licenses needed for use, monetary charges for data

- Understand publisher and licensing restrictions on re-use of data, analysis software and instrumentation (difference between CC0, Public Domain and OpenData Licenses)
- Explain ethical considerations related to data sharing

Thursday, 12 February: *Legal & Ethical Considerations for Research Data*

GUESTS: Jillian and Amy from the Institutional Research Board

- Understand privacy levels for research data as required by potential funding agencies
- Recognize the importance of privacy with some forms of research data (HIPAA)
- Understand the importance of removing key personal identifiers to facilitate confidentiality

Week 7

Tuesday, 17 February: *Data Sharing*

GUEST: Sue Kunda, Digital Scholarship Librarian w/OSU Libraries

- Discuss issues/obstacles related to reuse and sharing
- Describe steps to make data sharing easier

Thursday, 19 February: *Data Sharing and Reuse*

- Address re-use/sharing requirements from granting agencies or sponsors
- Understand options for maximizing data reuse
- Understand data identifiers and their use
- Understand benefits of using a unique researcher ID in metadata (e.g. ORCID or ISNI)

Week 8

Tuesday, 24 February: *Plan for Archiving and Preservation of Data*

- Explain options for a long-term sustainable preservation strategy/policy for your data (e.g., discipline specific, institutional, departmental, individual).
- Address the need for conversion to standard formats needed for re-use
- Perform basic archival processes: checksum, auditing, format migration, etc.
- Understand costs & timelines for data storage, management tools and services

Thursday, 26 February (in Autzen Classroom, Valley Library)

- Identify types of available repositories/archives (discipline-based, institutional, etc.)
- Choose appropriate option for long-term storage of data
- Understand process issues for depositing data in repository

Week 9

Tuesday, 3 March: OPEN LAB (in Autzen Classroom, Valley Library)

- Open time to work on metadata, data organization, or any other topic for which you would like assistance. Bring your data and get some help!
- Make up a computer lab exercise that you missed

Thursday, 5 March: *Data Visualization*

Week 10 – Dead week

Tuesday, 10 March

- Use class period to work on your DMP, get help from instructor, etc.

- Draft DMP should be completed by end of class period

Thursday, 12 March

- Exchange a copy of your DMP with at least 2 other students; critique and be critiqued
- Use feedback to prepare for submission of your final DMP by the scheduled time period for our final exam

Week 11 – Finals week

16 – 20 March

Students turn in a detailed data management plan for their thesis/dissertation research project.

Statement Regarding Students with Disabilities

"Accommodations are collaborative efforts between students, faculty and Disability Access Services (DAS). Students with accommodations approved through DAS are responsible for contacting the faculty member in charge of the course prior to or during the first week of the term to discuss accommodations. Students who believe they are eligible for accommodations but who have not yet obtained approval through DAS should contact DAS immediately at 737-4098."

Expectations for Student Conduct, i.e., cheating policies

<http://oregonstate.edu/studentconduct/regulations/index.php>

Appendix: Weekly Readings¹

—Subject to revision—Check Canvas for latest reading assignments

Week 1: Overview of Research Data Management

1. Committee on Ensuring the Utility and Integrity of Research Data in a Digital Age; National Academy of Sciences. 2013. (pages 95-99) Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age. Washington, D.C.: The National Academies Press.
http://books.nap.edu/openbook.php?record_id=12615&page=95
2. Interagency Working Group on Digital Data to the Committee on Science of the National Science and Technology Council. 2009. "Harnessing the Power of Digital Data for Science and Society." Introduction: A Revolution in Science: p. 3–5 (p. 11-13 of PDF).
<http://www.nitrd.gov/Publications/PublicationDetail.aspx?pubid=25>
3. Whitmire, Amanda. "LibGuides. Research Data Services. Funder Mandates."
<http://guides.library.oregonstate.edu/data-management-funder-mandates>
4. Whitmire, Amanda. 2013. "Data Management Throughout the Research Lifecycle" (August 16).
<http://dx.doi.org/10.6084/m9.figshare.774628>
5. Choose an example DMP to read from: https://dmptool.org/public_dmps

Week 2 – Types, Formats, and Stages of Data

1. Create and Manage Data: Formatting your Data: File Formats & Software (UK Data Archive):
<http://data-archive.ac.uk/create-manage/format/formats>
2. Data Types and File Formats: (Oregon State University Libraries):
<http://guides.library.oregonstate.edu/data-management-types-formats>
3. File Naming & Versioning (Oregon State University Libraries):
<http://guides.library.oregonstate.edu/data-management-file-organization>

Week 3 – Data Management Planning, For YOUR Research

1. Carlson, Jake (2010). Data Curation Profiles Toolkit User Guide, Purdue University Libraries / Distributed Data Curation Center. Retrieved 10 October 2012 from:
<http://datacurationprofiles.org/download>
2. Carlson, Jake (2010). Data Curation Profiles Toolkit: Interviewer's Manual, Purdue University Libraries / Distributed Data Curation Center. Retrieved 10 October 2012 from:
<http://datacurationprofiles.org/download>
3. Carlson, Jake (2010). Data Curation Profiles Toolkit: Interview Worksheet, Purdue University Libraries / Distributed Data Curation Center. Retrieved 10 October 2012 from:
<http://datacurationprofiles.org/download>
4. Carlson, Jake (2010). Data Curation Profiles Toolkit: The Profile Template, Purdue University Libraries / Distributed Data Curation Center. Retrieved 10 October 2012 from:
<http://datacurationprofiles.org/download>
5. Free choice: any of the public Data Curation Profiles listed in the DCP Directory:
<http://docs.lib.purdue.edu/dcp/> Be sure to browse around by year – there are many DCPs to choose from.

Week 4 – Data Storage, Backup and Security

¹ Adapted from "New England Collaborative Data Management Curriculum" (Lamar Soutter Library, University of Massachusetts Medical School, 2012), <http://library.umassmed.edu/necdmc/index>.

1. Backing Up Data (UK Data Archive): <http://www.data-archive.ac.uk/create-manage/storage/back-up>
2. Data security, protection & confidentiality (University of Edinburgh): <http://www.ed.ac.uk/schools-departments/information-services/services/research-support/data-library/research-data-mgmt/data-security>
3. Information Security Primer (Security overview Section 3, Section 6.1.2: Identification and Authentication of Users, Section 6.1.3: Cryptography): http://www.seattle.gov/informationSecurity/pdf/EPRI_securityPrimer.pdf
4. "NASA networks open to cyber attacks" (Help Net Security, March 2011): <http://www.net-security.org/secworld.php?id=10824>
5. Version Control and Authenticity (UK Data Archive): <http://www.data-archive.ac.uk/create-manage/format/versions>

Week 5 – Contextual Details Needed to Make Data Meaningful to Others

1. What is Metadata? (David Woodbury: 5 minutes): <http://vimeo.com/3161893>
2. Introduction to Metadata: Setting the Stage (Getty Research Institute): http://www.getty.edu/research/publications/electronic_publications/intrometadata/setting.html
3. Whitmire, Amanda. "LibGuides. Research Data Services. Metadata/Documentation." <http://guides.library.oregonstate.edu/data-management-metadata>
4. Seeing Standards: A Visualization of the Metadata Universe (Jenn Riley, Indiana University Libraries): <http://www.dlib.indiana.edu/~jenrile/metadatamap/>

Week 6 – Legal and Ethical Considerations for Research Data

1. Guidelines for Responsible Data Management in Scientific Research, p. 6-8 (US DHHS): <http://ori.hhs.gov/education/products/clinicaltools/data.pdf>
2. "Who Owns Research Data?" (US DHHS): http://ori.dhhs.gov/education/products/columbia_wbt/rcr_data/case/index.html#2
3. Westra, Brian. 2013. "Sharing or Publishing Your Research Data: Constructing Access Provisions." Accessed November 14. <http://libweb.uoregon.edu/datamanagement/sharingdata.html#three>
4. "Prison for HIPAA Privacy Violator" (Health Data Management Magazine, 06/01/2010): http://www.healthdatamanagement.com/issues/18_6/hipaa-prison-for-hipaa-privacy-violator-40382-1.html
5. Baylor University (no date), Patent Law Basics for University Researchers, <http://www.baylor.edu/research/vpr/files/patentlawbasics.pdf>

Week 7 – Data Sharing & Reuse Policies

1. White EP, Baldrige E, Brym ZT, Locey KJ, McGlinn DJ et al. (2013) Nine simple ways to make it easier to (re)use your data. PeerJ PrePrints 1:e7v2 <http://dx.doi.org/10.7287/peerj.preprints.7v2>
2. Whitmire, Amanda. "LibGuides. Research Data Services. IP & Licensing Data." 2013. <http://guides.library.oregonstate.edu/data-licensing-ip>.

Week 8 – Plan for Archiving and Preservation of Data

1. Whyte, A. & Wilson, A. (2010). "How to Appraise and Select Research Data for Curation". DCC How-to Guides. Edinburgh: Digital Curation Centre. <http://www.dcc.ac.uk/resources/how-guides>