1 Course Description

Current trends in research funding policy and big data will continue to drive demand for data management skills at public and private institutions, as well as academia. Key concepts, tools, and practices for each stage of the data management life cycle are the foundations for managing data in personal, research, and institutional contexts. The goal of this course is developing the knowledge base for effective data management in the age of big data, and familiarity with related ethical, legal, and social considerations.

The course will focus primarily on management practices in research data contexts from the humanities to the sciences, with attention to trends in the private sector, government, and library data services. Several data management technologies will be presented to demonstrate the features and functionality of current tools to prepare students to identify suitable tools for a specific context of practice.

2 Learning Outcomes

Upon completion of the course, students are expected to be able to:

1. Explain data life cycle concepts to non-technical audiences;
2. Demonstrate familiarity with commonly-used data management software, standards, and practices;
3. Evaluate the resource, process, and cultural requirements for data management;
4. Recommend short- and long-term preservation strategies for data sets;
5. Identify relevant ethical, legal, and social implications of data management in context; and
6. Develop a data management plan.
3 Learning Agreement

3.1 Prerequisites

This course has no prerequisites.

3.2 Course Materials

Text

The required course text (referred to as “RDM” in the weekly schedule) is approximately $25 on Amazon.


Additional articles will be provided via Canvas.

Systems

Canvas (http://elms.umd.edu/) will be the authoritative source for the course syllabus, schedule, presentation materials, announcements, and assignment details. Course assignments will be submitted on Canvas.

We will also use a virtual machine to access software for demo purposes. Details on accessing the virtual machine will be shared by email and via Canvas.

Due to extensive in-class activities, bringing a laptop (or checking one out from IT Services) for class every week is strongly recommended. For some activities and demos, a tablet may suffice, but many demos will require you to follow along using software from the virtual machine.

3.3 Syllabus Change Policy

This syllabus is accurate as of the date on the first page; however, it is subject to change with advance notice. In most cases, syllabus changes will be made by agreement between the instructor and students.

3.4 Academic Integrity

The University of Maryland has a nationally recognized Code of Academic Integrity administered by the Student Honor Council (http://www.studenthonorcouncil.umd.edu/code.html). This Code sets standards for academic integrity at Maryland for all undergraduate and graduate students. As a student, you are responsible for upholding these standards for this course. It is very important for you to be aware of the consequences of cheating, fabrication, facilitation, and plagiarism. As defined by the University of Maryland, academic dishonesty includes the following activities:

1. “CHEATING: intentionally using or attempting to use unauthorized materials, information, or study aids in any academic exercise.

2. FABRICATION: intentional and unauthorized falsification or invention of any information or citation in an academic exercise.

3. FACILITATING ACADEMIC DISHONESTY: intentionally or knowingly helping or attempting to help another to violate any provision of this Code.

4. PLAGIARISM: intentionally or knowingly representing the words or ideas of another as one’s own in any academic exercise.”
Academic dishonesty also includes buying assignments, submitting the same paper more than once, forging signatures, bribery, and other acts that deceive others about your academic work or record. You may also find this definition from the Office of Student Conduct helpful: http://osc.umd.edu/OSC/AcademicDishonesty.aspx.

If you have any questions about this policy or how to properly cite materials, please use all available resources, including the library, websites, and the instructor.

3.5 Attendance

Regular class attendance is obligatory. Since in-class participation is part of the course evaluation, missing class will negatively affect your course grade. If you must miss class, notify me in advance by email and check with your classmates afterward so that you can catch up.

3.6 Classroom Environment

In this class, everyone is treated with mutual respect. As a graduate student, I expect you are fully capable of behaving professionally in the classroom, being a team player in group activities, and asking questions when you need assistance. Due to our small class size and hybrid seminar-lab format, it is very important that you arrive to class prepared to discuss the readings and participate in in-class activities.

3.6.1 Excused Absences

In compliance with University policy, you may excuse yourself from one class session for medical reasons, making a reasonable effort to inform me in advance. More than two absences for medical reasons requires documentation from a health care provider in order to avoid penalties on participation grades.

In addition, it is the student’s responsibility to inform me of any intended absences for religious observances within the first two weeks of class to avoid penalties on participation grades.

Students may also be excused for participation in University activities at the request of university authorities; written documentation of such an event is required to avoid penalties on participation grades.

3.6.2 Inclement Weather

Official closures and delays are announced on the campus website at umd.edu and snow phone line (301-405-SNOW), as well as on local radio and TV stations. Unless there is an official closure or delay, you should assume that class will meet.

3.6.3 Emergency Preparedness

If a public emergency arises, please see the University’s Emergency Preparedness Website at http://www.umd.edu/emergencypreparedness/ for information about the current status of the campus. If a class session needs to be rescheduled, I will email you as soon as possible.

3.7 Communications

Communication outside of class will use Canvas or your umd.edu email account. Course announcements will be posted on Canvas and individual correspondence will be conducted via email. I will make every effort to send announcements with adequate advance notice; failure to receive email announcements will not be considered a suitable excuse for not being informed. I will typically
reply in two business days, usually less, but do not expect a reply over the weekend. Telephone is not an effective way to contact me.

3.8 CourseEvalUM

Course evaluations are a part of the process by which the University of Maryland seeks to improve teaching and learning. CourseEvalUM provides a standard, online, University-wide course evaluation instrument, and is linked in Canvas. Each course evaluation contains a set of universal questions, and some are supplemented by questions from specific colleges. All information submitted to CourseEvalUM is confidential. Instructors and academic administrators can only view summarized evaluation results after final grades have been submitted.

3.9 Academic Assistance

If you experience difficulties keeping up with the academic demands of this course, consider contacting the Learning Assistance Service, 2202 Shoemaker Building, 301-314-7693. Their educational counselors can help with time management, reading, math learning skills, note-taking, and exam preparation skills. All services are free to UMD students.

3.10 Disability Accommodations

The University is legally obligated to provide appropriate accommodations for students with disabilities. The campus’ Disability Support Services Office (DSS) works with students and faculty to address a variety of issues ranging from test anxiety to physical and psychological disabilities (http://www.counseling.umd.edu/DSS). If a student or instructor believes that the student may have a disability, they should consult with DSS (4-7682, email Dissup@umd.edu). Note that to receive accommodations, students must have disabilities documented by DSS before receiving an Accommodation Letter to present to course instructors regarding needed accommodations.

3.11 Intellectual Property

The University of Maryland’s official policy is that copyright for all course materials is held by the professor. Because I hold the intellectual property rights under this policy, my materials are freely available via a Creative Commons 3.0 BY-NC-SA license, http://creativecommons.org/licenses/by-nc-sa/3.0/.

With respect to ownership of student work, I may request written permission to use exceptional work as examples for future classes, but you hold all copyright to your own work and may decide whether or not to permit such use. Please see the special note on IP for the course project.

3.12 Open Collaboration Data Factories

The OCDF is a group of researchers and practitioners who are working on the development of data infrastructure and services to support research about open online communities and teaching data skills. The OCDF resources serve as a shared context to develop effective training experiences in basic data management principles through hands-on experience, while leveraging student work to evolve a useful, usable research resource for studying open online communities. As such, many course activities and deliverables will pertain to OCDF data and products, and will contribute to their development via class activities and feedback.
4 Course Schedule

Note the course schedule is subject to change. Your assignments are due on the dates we have class; due dates are also posted in Canvas. Written assignments will be submitted on Canvas and deliverables to share with classmates will be posted as links in Canvas Discussion threads.

Readings are listed for the dates for which they should be completed; for example, you should read the selections listed under Week 2 by class time on February 1, etc.

Week 1, January 25: All About Data (CANCELED DUE TO SNOWZILLA)

- **Activity** Online discussion: data management & you

Week 2, February 1: All About Data & Data Management Contexts

- **Topics** Course intro & administration; data & data management ecosystems; context of data management today
- **Guest Lecture** Susan Borda, Montana State University
- **Activity** Define your data space

Week 3, February 8: Data Management Planning & Data Collection

- **Topics** Data management life cycles & planning; data entry and file naming standards; data structures & containers
- **Readings** Laney et al. (2015); Vertesi & Dourish (2011); RDM Chapter 3: The Use of Life Cycle Models in Developing and Supporting Data Services
- **Activities** Intro to OCDF (watch video before class); data set hunting or lifecycle mapping
- **Demo** DMPTool & DMVitals

Week 4, February 15: Assessment & Accession

- **Topics** Data sources & streams; attributes of data to acquire & retain; appraisal & curation decision-making
- **Readings** NOAA (2008), RDM Chapter 8: Data Citation
- **Demo** SQL & Quarry
- **Guest Lecture** Aaron Halfaker, Wikimedia Foundation
- **Activities** Quarry querying; review OCDF records for deaccession
Week 5, February 22: Data Quality

- **Topics** Definitions of quality; QA vs QC; types of error & error handling strategies; data rescue
- **Readings** Marchionini et al. (2012), pp. 5–18 and one position paper from each category (4 in total)
- **Guest Lecture** Denise Hills, Program Director, Energy Investigations, Geological Survey of Alabama
- **Activity** QC on OCDF records
- **Assignment** DMP Questionnaire due

Week 6, February 29: Digital Curation

- **Topics** Digital curation tools & technologies
- **Activity** Class meets at Digital Curation Innovation Center

Week 7, March 7: Metadata, Schemas, & Data Description

- **Topics** Metadata, controlled vocabularies, taxonomies, tags, Linked Open Data
- **Readings** RDM Chapter 7: Metadata Services; Gilliland (2008)
- **Activity** Create new OCDX files in JSON
- **Guest Lecture** Kristen Schuster, PhD Candidate, University of Missouri

March 15—Spring Break—no class

Week 8, March 21: Sharing & Discovery

- **Topics** Sharing data; data search tools
- **Readings** Tenopir et al. (2011), Ubaldi (2013)
- **Demo** Dash, DataONESearch
- **Activities** Mid-semester feedback; data search feature comparison & recommendations
- **Guest Lecture** Lisa Federer, National Institutes of Health

Week 9, March 28: Integrating Data

- **Topics** Remixing data; updating/migrating legacy databases; entity resolution & record linking; challenges of syntax, schema, & semantics
- **Readings** Halem et al. (1999), Brizan & Tansel (2006)
- **Guest Lecture** Dr. Jen Hammock, Encyclopedia of Life & Smithsonian Institution, National Museum of Natural History
- **Demo** Google Fusion Tables, R Studio
- **Activity** TBA
Week 10, April 4: Analyzing Data

- **Topics** Scripts & workflow tools; managing processes like data; replicability; analysis types
- **Readings** Jain et al. (2015), Hettne et al. (2014)
- **Demo** GitHub
- **Assignments** Draft Timeline
- **Guest Lectures** Dr. Alyson Young, UMBC & Dr. Arfon Smith, GitHub

Week 11, April 11: Preserving Usefulness

- **Topics** Storage; short-term preservation (backups); long-term preservation (repositories); system-level preservation
- **Guest Lecture** Dr. Adam Kriesberg, University of Maryland iSchool
- **Demo** Figshare, DataHub.io, DRUM
- **Activities** Mid-semester feedback; registry/repository analysis

Week 12, April 18: Ethics, Legal, & Social Implications

- **Topics** Ethical considerations in data management; disclosure & compliance; chain of custody
- **Readings** Sagoyo et al. (2014), Zimmer (2010), Madden (2008)
- **Guest Lecture** Dr. Nicholas Proferes, University of Maryland iSchool
- **Assignments** Draft Budget Advisory
- **Demo** EU Data Privacy Compliance Check
- **Activity** ELSI audits for OCDF records

Week 13, April 25: Data Policies & Governance

- **Topics** Terms of use & access control; intellectual property (copyright, license, patents, liability); sensitive & protected data; risk management; evaluation
- **Readings** RDM Chapter 2: Data Governance; RDM Chapter 6: Copyright, Open Data, and the Availability-Usability Gap; Bowser et al. (2013); Terms of Use at [http://www.usa-npn.org](http://www.usa-npn.org)
- **Guest Lecture** Alyssa Rosemartin, USA-National Phenology Network
- **Demo** Online policy generators
- **Activity** Policy audits for OCDF records
Week 14, May 2: Future Challenges & Careers

- **Topics** Careers in data management; future directions of practice and research
- **Readings** RDM Chapter 19: The Next Generation of Challenges in the Curation of Scholarly Data; Bakhshi et al. (2014)
- **Guest Lecture** Karl Nilsen, Research Data Librarian, McKeldin Library
- **Activity** Job posting analysis

Week 15, May 9: Final reports

- **Activity** Presentations of data management plans
- **Assignments** Data Management Plan; completed Data Management Portfolio

5 Assessment

This course provides an overview and introduction to key topics in data management. To practice valuable professional skills, class members will engage in discussions, readings, and collaborative and individual assignments. Discussions will help you develop your ability to reflect about practical issues and discuss these with colleagues. Readings will provide an introduction to topics and exposure to current issues, debates, issues, and solutions. Written and group assignments serve as skill building exercises.

<table>
<thead>
<tr>
<th>Type</th>
<th>Assignment</th>
<th>Due</th>
<th>Points</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual</td>
<td>Draft Timeline</td>
<td>Week 10</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Individual</td>
<td>Draft Budget Advisory</td>
<td>Week 12</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Individual</td>
<td>Data Management Plan &amp; Presentation</td>
<td>Week 15</td>
<td>30</td>
<td>2,3,4,5,6</td>
</tr>
<tr>
<td>Individual</td>
<td>Data Management Portfolio</td>
<td>Week 15</td>
<td>30</td>
<td>all</td>
</tr>
<tr>
<td>Individual</td>
<td>In-class Participation</td>
<td>throughout</td>
<td>15</td>
<td>all</td>
</tr>
<tr>
<td>Team</td>
<td>Tool Demo</td>
<td>as assigned</td>
<td>15</td>
<td>2,3,6</td>
</tr>
</tbody>
</table>

**Total** 100

5.1 Guidelines for preparing assignments

*Follow all instructions carefully*, and ask questions as soon as they arise if you are uncertain about requirements. Failure to meet document requirements will be penalized as specified in the rubrics for assignments; content that does not match formatting requirements will be subject to additional scrutiny for potential plagiarism.

In addition to punctuality, the grammar, presentation and your ability to follow instructions are very important, as in the real world. It is essential that you spell check and proofread your documents; proofreading a *printed* copy is especially effective for finding errors that you might overlook on the screen.

- **Use 11pt Times New Roman** for body text.
• All text must be in black. You may use other text styles for title text and section or table headers, so long as all text is black. Most Word styles fail this requirement.

• Documents must use 1.5 line spacing with 1” margins on all sides in 8.5” x 11” (US letter) format.

• Paragraph formatting must use standard block format with a full blank line between paragraphs or leading indents for all paragraphs following the first paragraph in a section.

• On every page, document headers must include your name and UMD email address or team name on the left and page numbering on the right.

• Use APA format for citations and references. Web-only resources must always include the URL and date accessed.

• Every graded team assignment must include a contributorship statement as an appendix. This statement provides a simple, brief summary of which team members undertook which roles in the preparation of the assignment.

• See assignment descriptions on Canvas for accepted file types, grading rubrics, and to submit assignments.

When you prepare assignments or post on the discussion boards be sure to provide proper, complete bibliographical information for any sources referenced, for direct quotations, and for the sources of key concepts or ideas. Check the UMD citation guide for more details: http://www.lib.umd.edu/ues/guides/citation-tools. I will not strictly enforce formatting guidelines, but if your references are not retrievable with the information provided, you will lose points. Librarians from our university libraries are always available to help with academic search and reference tools, so if you have questions about what must be cited or how to cite, please ask.

5.2 Grades and Grading

Assignments are due as defined in the syllabus unless otherwise specified. The penalty for late assignments will be 10% within the first 24 hours, and an additional 25% for each week thereafter. An exception is possible in an extreme circumstance in which there is no reasonable way to anticipate or control the situation. Computers crashing, viruses, lost files, etc. are specifically not grounds for an extension.

Detailed grading rubrics for most assignments will be provided in advance on Canvas; please take advantage of them as you prepare your assignments to check whether your work meets grading criteria.

If you wish to discuss a grade, submit a written explanation of your argument (email) and arrange for a private conversation. Except for unusual circumstances, no appeals will be considered more than two weeks after the graded paper is returned. For final course grades, no appeal will be considered more than two months after the final day of classes.

Unless announced otherwise, assignments submitted by the due date will be graded within 1–2 weeks. Assignments submitted late will receive lower priority and so will take longer to grade. Final grades will be computed based on the scale below and partial points/percentages will be rounded for final grades.

100%+: A+ (4.0)
96 - 99%: A (4.0)
92 - 95%: A- (3.7)
88 - 91%: B+ (3.3)
84 - 87%: B (3.0)
80 - 83%: B- (2.7)
75 - 79%: C+ (2.3)
70 - 74%: C (2.0)
66 - 69%: D (1.0)
0 - 65%: F (0.0)

5.3 Participation 15%
Class discussions are an important way to learn and demonstrate learning; everyone is expected to partake in discussion of readings, presentations, and in-class activities. Non-attendance will be reflected in a decrease in this grade (and likely other grades as well). You can earn 1 point (percent) each week.

5.4 Tool Demo 15%
In pairs or small groups, students will present live demos of data management tools and provide a tip sheet for new users. The demos will be presented in class, and should be structured to walk class members through key activities for using the software, e.g., loading in a data file and viewing distributions in R, or setting up a GitHub repository and syncing files with the desktop client. Presentations will start in the fourth week of class; total presentation time should be no more than 15 minutes.

This assignment has three parts:

- Introduce the tool with a brief presentation (up to 5 minutes) highlighting the key features of the tool, and identifying competitor products and reasons for/against choosing this tool.
- Prepare a demo (up to 10 minutes) on the assigned tool. This means you may need to prepare and share sample files to use during the demo so that classmates can follow along with you.
- Create a two-page tip sheet that includes: an annotated screenshot of 1–2 main interfaces, and 5–7 annotated links to web-based resources for new users, such as how-to blog posts, reference sites, and tutorials.

Links to files shared as part of the tool demo (including the tip sheet) must be posted in the proper Canvas Discussion thread by 1:45 PM on the day of the presentation.

5.5 Data Management Portfolio 30%
Each week, we will do hands-on exercises and activities in class; these activities will involve portfolio entries that will accumulate into a final portfolio of your data management knowledge and skills. Guidelines will be provided for each portfolio entry, which may include demonstrating evidence of completed data management tasks and brief reflective essays. The portfolio will demonstrate your understanding and ability to apply general principles of data management to practical tasks and questions.

At the end of the semester, the accumulated entries will be compiled into a portfolio to be submitted as a stand-alone document in Canvas. While you have the option of leaving this task to the last minute, you are strongly encouraged to complete each portfolio entry as we progress.
through the semester, so that you only need to do light editing at the end to compile a professional final document.

5.6 Data Management Plan

You will develop a data management plan for either a research/project proposal or a client organization, to be provided to the client for implementation. Several researchers volunteered to serve as clients and descriptions are posted on Canvas, but you may otherwise solicit your own client (e.g., a current employer, your own project, a unit within UMD, etc.)

5.6.1 Assignment Options

If you choose the research/project data management plan option: You will interview a researcher (PI, principal investigator) about their research plans, specifically probing to discover their data management needs. You will use what you learn from the PI to develop a funder-compliant data management plan. If the funder does not currently require a data management plan, use the NSF templates and guidelines as a starting point. Ensure that the plans you develop adhere to the best practices in the PI’s field of practice.

If you choose the organizational data management plan option: You will interview at least one individual in the organization about how they generate and use data, specifically probing to discover their data management needs. You will use what you learn from the client to develop a set of recommendations that includes the key elements of a data management plan. You may use NSF templates and guidelines as a starting point for developing a document. Ensure that the plans you develop adhere to the best practices in the client’s domain.

5.6.2 Draft Timeline

A draft timeline for your plan implementation is due by Week 10. The timeline of implementation tasks must illustrate when key tasks must be completed to fully support the data management plan as specified. This deliverable should include a graphical representation of the timeline of key data management activities, with a brief textual overview of the timeline.

5.6.3 Draft Budget Advisory

A draft budget advisory document for your place is due by Week 12. The budget advisory document must include an itemized list of all expected expenses for supporting the data management plan, including (but not limited to) the staff time, tools and infrastructure, licenses, and any other related fees. Note that you will not need to include dollar figures (although estimates are fine and your clients will welcome them), but will need to specify every item that must be budgeted in order to fully support the data management plan as specified.

5.6.4 Final Data Management Plan

Your final data management plan is due on Week 15. It must include the description of the plan, with components and formatting as required by the funder or client needs, as well as the implementation timeline and budget advisory.

5.6.5 Presentation

You will share the highlights of the data management plan in a 15-minute presentation on the last day of class, on Week 15. More details on the presentation requirements will be provided in class.
Special note on intellectual property and privacy: This assignment requires you to contribute your work toward a client deliverable. The data management plan you create can be considered a form of compensation for the client’s participation in your project (therefore, everyone wins, but no one gets paid). You may include the data management plan in your professional portfolio in an anonymized version, i.e., without the names of specific individuals or organizations. If you have any concerns about your IP or privacy rights as relate to this (or any other) assignment, please talk to me by Week 2.

6 Course Readings


Zimmer, M. (2010). “But the data is already public”: On the ethics of research in Facebook. Ethics and Information Technology, 12, 313-325.