Data Management and Practice

Christine L. Borgman, University of California, Los Angeles

Available at: https://works.bepress.com/borgman/400/
Data Management and Practice

Winter, 2017, UCLA Information Studies 262A
Thurs, 9am-12:20pm, IS Room 245, January 12 through March 23 (exam week)
Christine L. Borgman, Distinguished Professor & Presidential Chair
235 GSE&IS bldg, 310-825-6164; Christine.Borgman@ucla.edu

Contents

Course Description: 262A, 262B ................................................................. 2
Course Objectives .......................................................... 3
Course Materials ........................................................... 3
Office Hours ................................................................. 3
Grading ................................................................. 3
Summary of Assignment Due Dates ........................................... 4
Topics, Readings, and Guest Speakers ........................................ 5
Week 1: Overview of Data Management and Practices, January 12 .......... 5
Week 2: What are data? January 19 .................................................. 5
Week 3: Public policies for research data, requirements for researchers, January 26 .... 6
Week 4: Data management plans and processes, February 2 .................... 6
Week 5: Data sharing and reuse: Practice and policy, February 9 ................ 7
Week 6: Data practices in the scientific domains, February 16 ..................... 7
Week 7: Data practices in the social scientific and humanities domains, February 23 .... 8
Week 8: Data citation, credit, and discovery, March 2 .......................... 9
Week 9: Data Management by research teams, libraries, and archives, March 9 .......... 9
Week 10: Student presentations, March 16 ........................................ 10
Syllabus References .......................................................... 11
Course Description: 262A, 262B

Information is the foundation of scholarship. Data is a particular class of information. Once considered primarily a part of the scholarly process, data are now viewed as products to be shared, mined, combined, managed, and sustained for reuse. Data scientists are information professionals who manage data, whether in science, social sciences, humanities, arts, medicine, law, government, or private institutions. As the practical and political roles of research data advance, so does scholarship on data practices, policies, and technologies.

These two courses prepare graduate students for professional positions in data management in all fields and for research on data practices. The job market is expanding rapidly for data science professionals at both the master’s and PhD research level, providing many employment opportunities. The Harvard Business Review named data scientist as “the sexiest job of the 21st century.” Course topics survey the landscape of data management, practices, services, and policy across fields and sectors, focusing primarily on scholarly applications. Themes include data management practices (e.g., metadata, provenance, technical standards); national and international data policy (e.g., intellectual property, release policies, open access, economics); management of data by research teams, data centers, libraries, and archives; and data curation, preservation, and stewardship.

Managing data is a complex process, involving expertise in knowledge organization, information policy, technology, and in the specific research domain. The courses are intended for graduate students in information studies and any other domain that requires the management of research data. By bringing together students from across campus, these seminar courses will engage students in practical, professional, and theoretical challenges in the use and reuse of research data. Assignments include hands-on analyses of data archives, data management plans, curating data for a research team, and domain-specific activities. Students will work in teams on real-world problems with UCLA researchers and will make class presentations.

Data management and practice (262A in winter 2017) provides a basic foundation for the data sciences. We focus on practical concerns, engaging with faculty research teams to address their data management requirements. At least two guest speakers will join us to discuss current issues in their domains. Data curation and policy (262B in spring 2017) builds upon this foundation to examine longer time issues of curation, stewardship, and knowledge infrastructure. We combine practical, policy, and research concerns with an advanced project to broker partnerships between faculty research teams and data repositories. Several guest speakers – national and international – will represent stakeholders in areas such as government data policy, publishing technologies, and ethics in data.

Data management and practice (262A) is a pre-requisite for 262B; students may choose to take only 262A or both courses. Thus, 262B in spring 2017 is open to students who completed 262A in 2016 or 2017.

NB: These courses are not scheduled to be taught in 2017-2018, so this is the only offering of these courses for the current cohort of Information Studies MLIS and PhD students.
Course Objectives

1. Students will gain an understanding of professional roles in managing research data.
2. Students will gain understanding of core research questions in data practice and policy.
3. Students will become familiar with the literature of data practices in information studies, social studies of science, and computer supported cooperative work.
4. Students will gain awareness of current and emerging data issues in national and international information policy.
5. Students will learn professional criteria for managing, selecting, and appraising data.
6. Students will learn to use and assess data collections, repositories, and services.
7. Students will gain technical skills in managing data in specific research settings.
8. Students will gain a basic knowledge of practices to curate digital data.

Course Materials

Two books are required for purchase: (Borgman, 2015a; Ray, 2014). Paperback editions of both books will be available for purchase at the LuValle Commons bookstore. *Big Data, Little Data, No Data: Scholarship in the Networked World* (Borgman, 2015) is also available online in hardcover and digital editions.

All other course materials will be posted on or linked from the CCLE. Enrolled students have access to the course site at [http://www.ccle.ucla.edu](http://www.ccle.ucla.edu).

Office Hours

Tuesdays, 3pm to 5pm. Please sign up in advance by Doodle ([http://doodle.com/poll/m4mf47veuzr8cxub6](http://doodle.com/poll/m4mf47veuzr8cxub6)) While you are welcome to stop by during office hours, most slots fill up in advance. If you are not able to keep an appointment, please cancel it on the Doodle as early as possible so that someone else may have the slot.

Grading

Assignment 1 (individual work): 25%
Term project (team work): 50%
Class participation and analysis of readings: 25%

Details of the assignments are provided on separate documents.

Students are expected to complete all assigned readings prior to each week’s class sessions and come prepared to discuss them. Your preparation and contributions to the discussion are the

---

1 Winner, 2016 American Publishers’ PROSE award for Best Book in the Computing and Information Sciences.
basis for 20% of your grade. Written assignments are to be submitted electronically to the CCLE site and on paper at the beginning of the class session, as noted. Assignments will be marked down 2 points for each day late. No assignments will be accepted after 5pm on Thursday, March 23.

**Summary of Assignment Due Dates**

<table>
<thead>
<tr>
<th>Assignment Given</th>
<th>Due Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assignments given:</td>
<td>January 12 (Week 1)</td>
</tr>
<tr>
<td>Project teams formed:</td>
<td>January 19 (Week 2)</td>
</tr>
<tr>
<td>Asst 1 topic description due:</td>
<td>January 19 (Week 2)</td>
</tr>
<tr>
<td>Bring sample of “data”:</td>
<td>January 19 (Week 2)</td>
</tr>
<tr>
<td>Term project proposal due:</td>
<td>January 26 (Week 3)</td>
</tr>
<tr>
<td>Asst 1 report due:</td>
<td>February 8 (Week 5); to be discussed in class on Feb 9</td>
</tr>
<tr>
<td>Teams meet with instructor:</td>
<td>Weeks 6-8</td>
</tr>
<tr>
<td>Project report outline due:</td>
<td>February 23 (Week 7)</td>
</tr>
<tr>
<td>Class presentation:</td>
<td>March 16 (Week 10)</td>
</tr>
<tr>
<td>Final project due to CCLE:</td>
<td>March 21 (exam week)</td>
</tr>
</tbody>
</table>
Topics, Readings, and Guest Speakers

Readings are to be completed in advance of each class session. Please come to class prepared to discuss the material and its relationship to larger issues in the course. Prepare some talking points as part of your reading and studying. We will spend minimal time lecturing and most of the time discussing the readings and issues that arise from them.

Week 1: Overview of Data Management and Practices, January 12

We will devote the first week of class to an overview of the concepts of data, management, and applications in scholarship, media, and policy. These readings span the scholarly and popular press, indicating the difficulties of acquiring and interpreting data effectively. We will begin to form project groups for the term.

Assignments: Assignment #1 and the term project will be distributed and explained.

Readings
(Ray, 2014), Introduction, pages 1-21
(Keller & Neufeld, 2014) Understanding our role in the world of big data (cartoon book)
(Lohr & Singer, 2016) How data failed us in calling an election
(Lazer, Kennedy, King, & Vespignani, 2014) Google Flu trends
(Davenport & Patil, 2012) Data scientist: Sexiest Job
(Rosenberg, 2003) History of information overload
Recommended:
(Ayres, 2007), Introduction, p 1-18

CERN Video (4:49); will be shown and discussed at the beginning of class: http://www.youtube.com/watch?v=j50ZssEojtM

More videos here: https://www.youtube.com/user/CERNTV

Week 2: What are data? January 19

“Data” is a far more ambiguous concept than is immediately apparent. Decisions about what data are to be managed, shared, and curated depend heavily on how the concept is defined. We will explore definitions and facets of “data” as a basis for discussion throughout the term. Project teams and assignments will be made today.

Assignments:
• Bring in a sample today of something that you consider to be data. We will discuss them in class.
• Due today: Assignment 1 topic description
Readings
(Borgman, 2015a), Chapters 1 to 4, Provocations; What Are Data?; Data scholarship; Data diversity
(Edwards et al., 2013); see also Knowledge Infrastructures site: http://knowledgeinfrastructures.org
(Rosenberg, 2013) History of “data”
(Laney, 2001) Volume, variety, velocity

Week 3: Public policies for research data, requirements for researchers, January 26

Researchers’ rights and responsibilities for data management are codified in public policies, funding agency contracts, and publishing requirements. Adoption and implementation vary widely by research domain, university, country, and many other factors. Similarly, the knowledge infrastructures to support data management vary in scope, scale, and specifics by similar factors. Policy, law, and economic consequences can be local, national, and international.

Assignments:
• Term project proposals due today.
• Class activity: each team will present a brief oral report on the policies that may apply to the research domain of their term project.

Readings
(Ray, 2014) part 1: Understanding the policy context (2 chapters)
(Boulton et al., 2015) International accord on open data
(Organization for Economic Cooperation and Development, 2007) International policy on data sharing
(National Science Foundation, 2011a, 2011b) NSF policies on data sharing

Recommended:
(Fox & Harris, 2013) International policy analysis for science
(European Commission High Level Expert Group on Scientific Data, 2010) Riding the wave report

Week 4: Data management plans and processes, February 2

“Data management” encompasses activities performed throughout a research project and well beyond, and can refer to specific plans that are mandated by funding agencies. Some activities are local and ad hoc; some are distributed and standardized; most fall somewhere in between. Researchers, data scientists, data librarians, repository staff, publishers, and many other stakeholders may be involved in data management. This week we address basic principles and components of the planning process.

Further readings on data management plans are listed in assignment 1.
Readings
(Ray, 2014), Part 2, Planning for data management, chapters 3-5
(“UC3: University of California Curation Center,” 2016) Explore their services and tools
(“Digital Curation Centre | because good research needs good data,” 2016) Explore their services and tools
(“European Landscape Study of Research Data Management,” 2013) European study of data management needs of researchers
(Steinhart, Chen, Arguillas, Dietrich, & Kramer, 2012) Researchers’ responses to data management plan requirements
(Kimpton & Morris, 2013) Local repository and cloud-based practices (Ray, Ch 11)

Week 5: Data sharing and reuse: Practice and policy, February 9

Despite the proliferation of data sharing policies, many factors augur against data sharing and reuse. Incentives to release data often run counter to the reward systems of scholarship; skills and resources are lacking; and suitable repositories may not exist. This week we will discuss practice, policy, and perspectives of the many stakeholders in data sharing and reuse.

Assignments:
• Class activity: each team will present a brief oral report on the reuse potential for data from the research domain of their term project.
• Assignment #1 is due Feb 8 to CCLE; to be discussed in class today.

Readings
(Borgman, 2015a), Chapter 8: releasing, sharing, and reusing data
(Borgman, 2016) Social Science Research Council discussion paper on data sharing
(Pasquetto, Sands, Darch, & Borgman, 2016) Open data principles
(Ray, 2014), Chapter 19, Clifford Lynch
(Wessels et al., 2014) Issues in the development of open access to research data; European perspective

Recommended:
(Borgman, 2015b) Short essay on issues in data sharing for EU community

Week 6: Data practices in the scientific domains, February 16

Notions of data vary greatly by context, discipline, time, and place. We will spend weeks 6 and 7 exploring case studies in multiple fields. Much research policy and data management practice is based on scientific data, thus we start with the sciences.

Dr. Francoise Genova of the Strasbourg Astronomical Data Center http://cdsweb.u-strasbg.fr/ will be our guest today.
Readings
(Borgman, 2015a), Chapter 5, Science cases
(Borgman, Darch, Sands, & Golshan, 2016) Astronomy knowledge infrastructure
(Landais et al., 2015; Perret et al., 2015) Astronomy work of Dr. Genova
(National Science Board, 2005) Major policy report on data, repositories, and practices
(Ribes & Jackson, 2013) Data in science collaborations
(Edwards, Mayernik, Batcheller, Bowker, & Borgman, 2011) Data as glue and friction

Websites to explore:
(ADS, 2016; CDS, 2016; “NASA/IPAC Extragalactic Database (NED),” 2016)

Recommended:
(Borgman, Wallis, & Mayernik, 2012) Science and computer science practices

Week 7: Data practices in the social scientific and humanities domains, February 23

Data management practices in the social sciences and humanities tend to be much different from those in the sciences. UCLA has deep expertise in these areas; hence leaders in these areas are invited for a class discussion this week.

Assignments: Project report outline due.

Panel Session (invited):
Dr. Lisa Snyder, Institute for Digital Research and Education, UCLA
https://idre.ucla.edu/people/profiles/lisa-snyder
Dr. Miriam Posner, Program Coordinator, Digital Humanities Program, UCLA
http://miriamposner.com/about.html

Readings
(Borgman, 2015a) Chapter 5: Social sciences cases; Chapter 6, Humanities cases
(Ray, 2014), Chapter 10, Social Science Data
(Dombrowski, 2014) Outcomes of Project Bamboo, well known digital humanities project
(Vardigan & Whiteman, 2007) ICPSR and OAIS
(King, 2011) Stewardship of social sciences data

Recommended:
(“Archaeology Data Service,” 2013) Best practices guide
(arts & humanities research council, 2012) Technical planning guide
(“Berkeley Initiative for Transparency in the Social Sciences,” 2014) UCB project
(Inter-university Consortium for Political and Social Research, 2012) Best practices guide for social sciences data
Week 8: Data citation, credit, and discovery, March 2

Citing publications is one of the most important ways of giving credit to scholars for their work. Citations to data are often proposed as a parallel means to giving credit for sharing data. However, data are much more difficult to cite, due to a lack of mechanisms, incentives, and practices. The ability to discover data depends heavily on the availability of metadata, such as data citations. Citation metrics, whether to publications, data, or other scholarly products, are highly contentious, as they are easily gamed and misused. We will discuss basic issues of data citation, credit, and metrics this week.

Assignments: Class activity: Each team will provide an oral report about the data citation and discovery practices of the domain of their research project.

Readings
(Borgman, 2015a), Chapter 9: Credit, attribution, and discovery
(Hicks, Wouters, Waltman, De Rijcke, & Rafols, 2015) Leiden manifesto for research metrics
(Priem, Taraborelli, Groth, & Neylon, 2010) Original altmetrics manifesto
(Brand, Allen, Altman, Hlava, & Scott, 2015) Publishers’ perspectives on scholarly metrics
(Kratz & Strasser, 2014) Overview of “data publication” issues

Week 9: Data Management by research teams, libraries, and archives, March 9

We will conclude this term and lay the foundation for Part II of this course with a discussion of the workforces and institutional activities associated with managing data.

Rebekah Cummings (University of Utah) and Julian Gautier (Harvard University), UCLA alumni who are now data librarians (and who have taken this course) will participate by video to lead a discussion on the role of librarians in data management and library practice. Bring your questions for them.

Readings
(Borgman, 2015a) Chapter 10, What to keep and why
(Ray, 2014), Chapters 7-10, Managing project data; digital repositories
(Arlitsch, 2014) Libraries, data, and interoperability
(“European Landscape Study of Research Data Management,” 2013) SURF study
(Hanson, Surkis, & Yacobucci, 2012) Cartoon video (4:40) about data management by teams
(Akers, Sferdean, Nicholls, & Green, 2014) Building Support for Research Data Management: Biographies of Eight Research Universities.

Recommended:
(Fearon, Gunia, Lake, Pralle, & Sallans, 2013) ARL survey of data management planning; read Executive Summary
Week 10: Student presentations, March 16

See project assignment for details. We will devote the last class session to a public presentation of student projects and a general discussion of project findings. Faculty partners are encouraged to attend and participate in the discussion.

To do: invite your faculty partners to come to your presentation.

Final projects due Tuesday, March 21, to CCLE
Syllabus References


