Title
Syllabus for Data, Data Practices, and Data Curation, Part I, Winter 2014, UCLA Information Studies

Author
Borgman, Christine L.

Publication Date
2014

Data, Data Practices, and Data Curation

Part I, Winter, 2014, UCLA Information Studies
Weds 1:30-4:50, IS Room 245, January 8 through March 19 (exam week)
Christine L. Borgman, Professor & Presidential Chair
215 GSE&IS bldg, 310-825-6164; borgman@gseis.ucla.edu

Course Description, Parts I and II

In today's technology-intensive research environments, petabytes of data may be produced in a matter of hours, days, or weeks. Those data may be lost in a similar amount of time if they are not captured, curated, and marked up in ways that allow for discovery and reuse by others. Datasets large and small can be very useful not only to researchers, but also to students, to the general public, and to policy makers. Among the classes of data of broad general interest are scientific records of the climate, the skies and galaxies, plant and animal species, social and economic observations, and cultural and historical records. Research policy by governments and funding agencies encourages – and increasingly requires – that investigators make plans for data management, curation, and dissemination.

The National Science Foundation announced a new requirement in 2010 for all grant proposals: they now must include data management plans. In 2012, the White House initiated open access requirements on publications and data for all large federal funding agencies. Similar requirements are in place in the U.K., European Union, and elsewhere. These policies are causing a mad scramble for compliance by researchers, universities, librarians, and archivists. The library and archive communities have issued several reports on their role in data management, adding to a plethora of policy reports in this area. More courses and curricula for data management professionals are being developed. Data management is clearly a growth area for information studies graduates.

These two courses (winter and spring) survey the rich landscape of data practices and services, including data as evidence and their role in research; data-intensive research methods; social studies of data practices; national and international data policy (e.g., intellectual property, release policies, open access, economics); comparisons between disciplines; management of data by research teams, data centers, libraries, and archives; technical standards for data and metadata; and data curation. Part I (winter) lays the foundation for data practices and services across the disciplines. Part II (spring) builds upon this background to provide practical experience in data curation. One large project will be undertaken across the two terms plus several smaller assignments. The courses will be graded separately. Part I is a pre-requisite for Part II. However, by taking Part I, you are not obligated to take Part II.

These courses will be a mix of readings, discussion, practicum, field trips, and guest lectures. Invited speakers include local experts and distinguished guests, some via videoconference.

Librarians, archivists, and other information professionals bring essential skills to the realm of research data. Information activities related to data include developing metadata, standards, and systems of classification, establishing archival plans for data selection, migrating data to new platforms and standards, creating finding aids for multiple user communities, and developing databases and technologies to support data creation, preservation, discovery, and reuse. Data management is a growth area in academic and special libraries, and will be an increasingly important set of skills for librarians and archivists in all sectors.

This is an introductory graduate course, suitable for masters and doctoral students in information studies and in data-intensive research fields. The course is open to practicing librarians and archivists through concurrent enrollment, with instructor's permission. Upper division undergraduates may enroll with instructor permission.

The two-part sequence of courses in Data, Data Practices, and Data Curation was developed, starting in 2009, with contributions by UCLA doctoral students Jillian Wallis and Laura Wynholds. Student participation also contributes to the development and adaption of these courses. Thanks also are due to instructors of similar courses at other universities who shared their syllabi and course materials, especially Margaret Hedstrom and Ann Zimmerman of the University of Michigan, Carole Palmer and Melissa Cragin at the University of Illinois, and Carolyn Hank at the University of North Carolina.

Course Objectives

1. Students will learn to distinguish among many forms of data, factors by which they vary, how data change in meaning between contexts and over time.
2. Students will learn professional criteria for selecting and appraising data.
3. Students will learn to distinguish among different types of data collections, repositories, and services.
4. Students will learn roles that data play in research collaborations.
5. Students will gain a basic knowledge of practices to curate digital data.
6. Students will learn basic principles of public policies for data.

Course Materials

One book is required for purchase: (Ray, 2014)

All other course materials will be posted on or linked from the Moodle site for this course. Enrolled students have access to the site at http://www.ccle.ucla.edu.

Office Hours

Mondays and Wednesdays, 5-6pm (link posted on CCLE), other times by appointment, and by email.

Grading

Short paper assignment 25%
Term project 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 basis for 25% of your grade. Written assignments are due at the beginning of the class session, on paper, and are to be submitted electronically to the CCLE / Moodle site. Assignments will be marked down 2 points for each day late. No assignments will be accepted after midnight on Wednesday, March 19. No exceptions, as the instructor is leaving the country on March 20.

Overview:

Topics, Readings, and Guest speakers

Week 1: Overview of Data, Data Practices, and Data Curation, January 8

We will devote the first week of class to an overview of the concept of data and its manifestations across scholarly disciplines and in public parlance. We will begin to form project groups for the term. Assignment #1 and the term project will be distributed and explained.

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 and the curriculum. Prepare some talking points as part of your reading and studying.

Readings
(Borgman, forthcoming), Chapters 1 and 2, Big and Little Scholarship; Digital Scholarship
(Ray, 2014), Introduction, pages 1-21
(Ayres, 2007), Introduction, p 1-18

Video (4:49): http://www.youtube.com/watch?v=j50ZssEojtM

Field trip: We'll break at 4pm to walk to the CNSI auditorium (far end of science quad; bring your walking shoes) for the Jon Kleinberg distinguished lecture on **bursts, cascades, and hot spots: A glimpse of some on-line social phenomena at global scales**

More info here:
*www.**ipam**.ucla.edu/programs/sl2014/lecture.pdf*
http://www.ipam.ucla.edu/abstract.aspx?tid=11674
https://www.ipam.ucla.edu/schedule.aspx?pc=sl2014

Week 2: What are data? January 15

"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 devote today to exploring some of the many definitions and facets of "data." We will form project teams and assign them to faculty members today.

Assignment: Bring in a sample today of something that you consider to be data. We will discuss them in class.

Readings
(Borgman, forthcoming), Chapter 3, What are data?

(Edwards et al., 2013); see also Knowledge Infrastructures site: http://knowledgeinfrastructures.org

Week 3: Public policies for research data, January 22

The ability to deposit, discover, share, retrieve, reuse, and curate data all depend upon public policies about rights and responsibilities. These policies have legal and economic aspects and vary widely around the world, although many international agreements are in place. Term project proposal is due Friday of this week. Please make office hour appointments to discuss your project in the weeks ahead.

Readings
(Ray, 2014) part 1: Understanding the policy context (2 chapters)
(Organization for Economic Cooperation and Development, 2007)
(Wood et al., 2010)

Week 4: Planning for data management, January 29

Data management requires considerable planning on the part of the research team and on the part of repositories. We'll start with some basic principles and components of the planning process. Assignment #4 is due at the start of class today.

Readings
(Ray, 2014), Part 2, Planning for data management, chapters 3-5
(Abrams, Cruse, & Kunze, 2009)

Week 5: Data in the sciences, February 5

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

Speaker (by video): Prof. Borgman, from Melbourne, Australia
Local host and speaker: Dr. Peter Darch, UCLA Information Studies

Readings
(Borgman, forthcoming), Chapter 5, Science cases
(National Science Board, 2005)
(Fortson et al., 2011)

Week 6: Data in the social sciences, February 12

The social sciences have a very long history of data management and archiving. UCLA's own Libbie Stephenson is an international leader in data archives for the social sciences.

Speaker: Elizabeth Stephenson, Director, Institute for Social Research Data Archive, UCLA

Readings
(Borgman, forthcoming) Chapter 5: Social sciences cases
(Ray, 2014), Chapter 10, Social Science Data
(Vardigan & Whiteman, 2007)
(*Guide to Social Science Data Preparation and Archiving: Introduction*, 2012)
(King, 2011)

Week 7: Data in the humanities, February 19

Data is a much different notion in the humanities than in other disciplines. UCLA's digital humanities program and the IDRE-HASIS have addressed campus data management issues for several years already. Team Project Report outline due today.

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

Readings
(Borgman, forthcoming), Chapter 6, Humanities cases
("Archaeology Data Service," 2013)
("AHRC Technical Plan," 2012)
(Kouw, Van den Heuvel, & Scharnhorst, 2013)
(Borgman, 2009)

Week 8: Team workshops, February 26; IDCC in San Francisco

Prof. Borgman and other members of the faculty and IS student body will be in San Francisco this week participating in the International Digital Curation Conference. Please use this time to work together on your team projects. The classroom will be available to you for group work.

Week 9: Data sharing and reuse, March 5

The ability of libraries and archives to manage data is predicated upon their release by researchers. Relatively few data are released and even fewer are reused. This week will discuss the reasons why that is the case and the role that librarians and archivists may plan in the process.

Readings
(Borgman, forthcoming), Chapter 7 and 8, Supply and Demand for Data
(Ray, 2014), Chapter 19, Clifford Lynch

Week 10: Data Management by libraries and archives, March 12

We will conclude this term and lay the foundation for Part II of this course with a deeper discussion of the activities of librarians and archivists in managing data. The remainder of the Joyce Ray book will be covered in Part II.

Readings
(Ray, 2014), Chapters 7-10, Managing project data; digital repositories

Week 11 (Exam Week): Student presentations, March 19

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.

Wednesday, March 19: Final projects due, 5pm, to instructor's mailbox and by PDF to CCLE. NB: No extensions as Prof. Borgman is leaving the U.S. on March 20.

Syllabus References:

Abrams, S., Cruse, P., & Kunze, J. (2009). Preservation is not a place. *International Journal of Digital Curation*, *4*. Retrieved from http://www.ijdc.net/index.php/ijdc/article/viewFile/98/73

Archaeology Data Service / Digital Antiquity: Guides to Good Practices. (2013). Retrieved June 11, 2013, from http://guides.archaeologydataservice.ac.uk/

Arts & Humanities Research Council - Technical Plan. (2012). Retrieved July 2, 2013, from http://www.ahrc.ac.uk/Funding-Opportunities/Research-funding/RFG/Application-guidance/Pages/Technical-Plan.aspx

Ayres, I. (2007). *Super Crunchers: Why thinking-by-numbers is the new way to be smart*. New York: Bantam.

Borgman, C. L. (forthcoming). *Big Data, Little Data, No Data: Scholarship in the Networked World*. Cambridge MA: MIT Press.

Borgman, C. L. (2009). The digital future is now: A call to action for the humanities. Digital Humanities Quarterly, 3. Retrieved from http://digitalhumanities.org/dhq/vol/3/4/000077/000077.html

Edwards, P. N., Jackson, S. J., Chalmers, M. K., Bowker, G. C., Borgman, C. L., Ribes, D., … Calvert, S. (2013). *Knowledge Infrastructures: Intellectual Frameworks and Research Challenges* (p. 40). Ann Arbor, MI: University of Michigan. Retrieved from http://deepblue.lib.umich.edu/handle/2027.42/97552

Fortson, L., Masters, K., Nichol, R., Borne, K., Edmondson, E., Lintott, C., … Wallin, J. (2011). *Galaxy Zoo: Morphological Classification and Citizen Science* (arXiv e-print No. 1104.5513). Retrieved from http://arxiv.org/abs/1104.5513

Guide to Social Science Data Preparation and Archiving: Introduction (No. Fifth edition). (2012). Retrieved from https://www.icpsr.umich.edu/icpsrweb/content/deposit/guide/index.html

King, G. (2011). Ensuring the Data-Rich Future of the Social Sciences. *Science*, *331*, 719–721. doi:10.1126/science.1197872

Kouw, M., Van den Heuvel, C., & Scharnhorst, A. (2013). Exploring uncertainty in knowledge representations: Classifications, simulations, and models of the world. In P. Wouters, A. Beaulieu, A. Scharnhorst, & S. Wyatt (Eds.), *Virtual knowledge: experimenting in the humanities and the social sciences* (pp. 127–149). Cambridge, Mass.: MIT Press.

National Science Board. (2005). *Long-Lived Digital Data Collections*. Retrieved from http://www.nsf.gov/pubs/2005/nsb0540/

Organization for Economic Cooperation and Development. (2007). *OECD Principles and Guidelines for Access to Research Data from Public Funding* (pp. 1–24). Retrieved from www.oecd.org/dataoecd/9/61/38500813.pdf

Ray, J. M. (2014). *Research data management: practical strategies for information professionals*. West Lafayette, Ind.: Purdue University Press.

Vardigan, M., & Whiteman, C. (2007). ICPSR Meets OAIS: Applying the OAIS Reference Model to the Social Science Archive Context. Retrieved from http://deepblue.lib.umich.edu/handle/2027.42/60440

Wood, J., Andersson, T., Bachem, A., Best, C., Genova, F., Lopez, D. R., … Hudson, R. L. (2010). *Riding the wave: How Europe can gain from the rising tide of scientific data*. Final report of the High Level Expert Group on Scientific Data. Retrieved from http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf