Title
Syllabus for Data Management and Practice, Part I, Winter 2016, UCLA Information Studies

Author
Borgman, Christine L.

Publication Date
2016

Data Management and Practice

Winter, 2016, UCLA Information Studies 262A
Thurs, 9am-12:20pm, IS Room 121, January 7 through March 15 (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

Data are both process and products of the research enterprise. Funding agencies are requiring data management plans as part of grant proposals, journals are requiring the release of data to reviewers and readers alike, and libraries and archives are adding data to their collections. Researchers are acquiring skills in data handling as part of graduate programs. Managing data is a complex process, involving expertise in technology, knowledge organization, information policy, and in the research domain. These two Information Studies courses (262A and 262B) survey the landscape of data management, practices, services, and policy, including the uses of data in the sciences, social sciences, and humanities; 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. The courses are intended for graduate students in information studies and in any 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 analyses of data archives, data management plans, curating data for a research team, and domain-specific activities. Students will work in teams with UCLA researchers and will make class presentations. The course provides basic foundation on data issues for masters and PhD students concentrating in libraries, archives, informatics, information policy, scholarly communication, and to students in other disciplines.

Students taking 262A in 2016 can take 262B in 2017 (not required, but 262A is pre-requisite for 262B). Both courses will be taught in 2016-2017.

Course Objectives

1. Students will learn to distinguish among many forms of data and factors that influence their interpretation in different contexts and over time.
2. Students will learn professional criteria for managing, selecting, and appraising data.
3. Students will learn to use and assess data collections, repositories, and services.
4. Students will gain technical skills in managing data in specific research settings.
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

Two books are required for purchase: (Borgman, 2015a; Ray, 2014)

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.

Office Hours

Tuesdays, 3pm to 5pm. Please sign up in advance by Doodle (http://doodle.com/poll/sfv7cq59uwe6tbmg). 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): 30%
Term project (team work): 50%
Class participation and analysis of readings: 20%

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 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 17.

Summary of Assignment Due Dates

Assignments given:	January 7 (Week 1)
Project teams formed:	January 14 (Week 2)
Asst 1 topic description due:	January 14 (Week 2)
Bring sample of "data":	January 14 (Week 2)
Term project proposal due:	January 21 (Week 3)
Asst 1 report due:	February 3 (Week 5); to be discussed in class on Feb 4
Teams meet with instructor:	Weeks 6-8
Project report outline due:	February 18 (Week 7)
Class presentation:	March 15 (Exam week)
Final project due to CCLE:	March 17

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.

Week 1: Overview of Data Management and Practices, January 7

We will devote the first week of class to an overview of the concepts of data, management, and applications across scholarly disciplines. We will begin to form project groups for the term. Assignment #1 and the term project will be distributed and explained.

Readings
(Ray, 2014), Introduction, pages 1-21
(Ayres, 2007), Introduction, p 1-18
(Cukier & Mayer-Schoenberger, 2013) Overview of their book (Mayer-Schonberger & Cukier, 2013)
(Lazer, Kennedy, King, & Vespignani, 2014) Google Flu
(Rosenberg, 2003) History of information overload

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

Week 2: What are data? January 14

"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.

Assignment: 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 21

Researchers' rights and responsibilities for data management are codified in public policies. These policies have legal and economic aspects that vary widely; many international agreements also are in place.

Readings
(Ray, 2014) part 1: Understanding the policy context (2 chapters)
(Steinhart, Chen, Arguillas, Dietrich, & Kramer, 2012) Researchers' responses to data management plan requirements
(Organization for Economic Cooperation and Development, 2007) International policy on data sharing
(Wood et al., 2010) EU Riding the Wave report
(National Science Foundation, 2011a, 2011b) NSF policies on data sharing
(Australian National Data Service, 2014; National Health and Medical Research Council, 2007) Australian policies on data management
(Fox & Harris, 2013) International policy analysis for science

Week 4: Data management plans and processes, January 28

"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.

Readings
(Ray, 2014), Part 2, Planning for data management, chapters 3-5
("UC3: Data Management Guidelines," 2014) California Digital Library
("European Landscape Study of Research Data Management," 2013) European study of data management needs of researchers
(Fox & Harris, 2013) Summary of international planning
(Kimpton & Morris, 2013) Local repository and cloud-based practices (Ray, Ch 11)

Week 5: Data practices in the scientific domains, February 4

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.

Assignment #1 is due Feb 3; to be discussed in class today.

Readings
(Borgman, 2015a), Chapter 5, Science cases
(National Science Board, 2005) Major policy report on data, repositories, and practices
(Borgman, Wallis, & Mayernik, 2012) Science and computer science practices
(Ribes & Jackson, 2013) Data in science collaborations

(Edwards, Mayernik, Batcheller, Bowker, & Borgman, 2011) Data as glue and friction

Week 6: Data sharing and reuse: Practice and policy, February 11

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.

Readings

(Meyer and Schroeder, 2015) Chapter 5: Distributed Data. From Knowledge Machines: Digital Transformations of the Sciences and Humanities
(Borgman, 2015a), Chapter 8: releasing, sharing, and reusing data
(Ray, 2014), Chapter 19, Clifford Lynch
(Wessels et al., 2014) Issues in the development of open access to research data
(Borgman, 2015b) Short essay on issues in data sharing for EU community
("Reproducible Research Tools | Department of Statistics," 2014) List of tools, readings

Week 7: Data practices in the social scientific and humanities domains, February 18

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.

Panel Session:

Elizabeth Stephenson, Director, Institute for Social Research Data Archive, UCLA
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
(Inter-university Consortium for Political and Social Research, 2012) Best practices guide for social sciences data
(King, 2011) Stewardship of social sciences data
("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

Week 8: Data citation, credit, and discovery, February 25

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.

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: No class meeting, March 3

Prof. Borgman will be traveling this week. Please work with your teams and faculty partners to finish your term project and presentations.

Week 10: Data Management by research teams, libraries, and archives, March 10

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.

Panel of library staff to be invited.

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) Short clever story (video, 4:40) about data management by teams
(Fearon, Gunia, Lake, Pralle, & Sallans, 2013) ARL survey of data management planning; read Executive Summary

Week 11 (Exam Week): Student presentations, March 15

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.

Final projects due noon, March 17, to CCLE

Syllabus References

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

Arlitsch, K. (2014). Being Irrelevant: How Library Data Interchange Standards Have Kept Us Off the Internet. *Journal of Library Administration*, *54*(7), 609–619. http://doi.org/10.1080/01930826.2014.964031

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

Australian National Data Service. (2014). [Home page]. Retrieved January 24, 2014, from http://www.ands.org.au/

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

Berkeley Initiative for Transparency in the Social Sciences. (2014). Retrieved December 16, 2014, from http://bitss.org/

Borgman, C. L. (2015a). *Big Data, Little Data, No Data: Scholarship in the Networked World*. Cambridge MA: MIT Press.

Borgman, C. L. (2015b, January). If data sharing is the answer, what is the question? *ERCIM News*, (100). Retrieved from http://ercim-news.ercim.eu/

Borgman, C. L., Wallis, J. C., & Mayernik, M. S. (2012). Who's Got the Data? Interdependencies in Science and Technology Collaborations. *Computer Supported Cooperative Work*, *21*(6), 485–523. http://doi.org/10.1007/s10606-012-9169-z

Brand, A., Allen, L., Altman, M., Hlava, M., & Scott, J. (2015). Beyond authorship: attribution, contribution, collaboration, and credit. *Learned Publishing*, *28*(2), 151–155. http://doi.org/10.1087/20150211

Cukier, K. N., & Mayer-Schoenberger, V. (2013, May 1). The Rise of Big Data. *Foreign Affairs*, (May/June 2013). Retrieved from http://www.foreignaffairs.com/articles/139104/kenneth-neil-cukier-and-viktor-mayer-schoenberger/the-rise-of-big-data?page=show#

Dombrowski, Q. (2014). What Ever Happened to Project Bamboo? *Literary and Linguistic Computing*, *29*(3), 326–339. http://doi.org/10.1093/llc/fqu026

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*. Ann Arbor: University of Michigan. Retrieved from http://deepblue.lib.umich.edu/handle/2027.42/97552

Edwards, P. N., Mayernik, M. S., Batcheller, A. L., Bowker, G. C., & Borgman, C. L. (2011). Science Friction: Data, Metadata, and Collaboration. *Social Studies of Science*, *41*, 667–690. http://doi.org/10.1177/0306312711413314

European Landscape Study of Research Data Management. (2013). Retrieved September 27, 2013, from http://www.surf.nl/en/actueel/Pages/EuropeanLandscapeStudyofResearchDataManagement.aspx

Fearon, D. J., Gunia, B., Lake, S., Pralle, B. E., & Sallans, A. L. (2013). Research Data Management Services, SPEC Kit 334 (July 2013). Retrieved from http://publications.arl.org/Research-Data-Management-Services-SPEC-Kit-334/

Fox, P., & Harris, R. (2013). ICSU and the Challenges of Data and Information Management for International Science. *Data Science Journal*, *12*, WDS1–WDS12. Retrieved from https://www.jstage.jst.go.jp/article/dsj/12/0/12_WDS-001/_article

Hanson, K., Surkis, A., & Yacobucci, K. (2012). *Data Sharing and Management Snafu in 3 Short Acts* [Film and Animation]. Retrieved from http://www.youtube.com/watch?v=N2zK3sAtr-4&sns=em

Hicks, D., Wouters, P., Waltman, L., De Rijcke, S., & Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature*, *520*(7548). Retrieved from http://www.nature.com/news/bibliometrics-the-leiden-manifesto-for-research-metrics-1.17351

Inter-university Consortium for Political and Social Research. (2012). *Guide to Social Science Data Preparation and Archiving: Best Practice Throughout the Data Life Cycle* (No. 5th edition). Ann Arbor, MI: ICPSR. Retrieved from https://www.icpsr.umich.edu/icpsrweb/content/deposit/guide/index.html

Kimpton, M., & Morris, C. M. (2013). Managing and Archiving Research Data: Local Repository and Cloud-Based Practices. In J. M. Ray (Ed.), *Research Data Management: Practical Strategies for Information Professionals* (pp. 223–238). West Lafayette, IL: Purdue University Press.

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

Kratz, J., & Strasser, C. (2014). Data publication consensus and controversies. *F1000Research*. http://doi.org/10.12688/f1000research.3979.3

Laney, D. (2001). 3D Data Management: Controlling Data Volume, Velocity, and Variety. Retrieved July 10, 2013, from http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The Parable of Google Flu: Traps in Big Data Analysis. *Science*, *343*(6176), 1203–1205. http://doi.org/10.1126/science.1248506

Mayer-Schonberger, V., & Cukier, K. (2013). *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Boston: Houghton Mifflin Harcourt.

National Health and Medical Research Council. (2007). Australian Code for the Responsible Conduct of Research. Retrieved January 25, 2014, from http://www.nhmrc.gov.au/guidelines/publications/r39

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

National Science Foundation. (2011a). *NSF Data Management Plans*. Washington, DC: NSF. Retrieved from http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/gpg_2.jsp#dmp

National Science Foundation. (2011b). *NSF Data Sharing Policy*. Washington, DC: NSF. Retrieved from http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/aag_6.jsp#VID4

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

Priem, J., Taraborelli, D., Groth, P., & Neylon, C. (2010). altmetrics: a manifesto – altmetrics.org. Retrieved May 25, 2013, from http://altmetrics.org/manifesto/

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

Reproducible Research Tools | Department of Statistics. (2014). Retrieved December 16, 2014, from http://www.stat.wisc.edu/reproducible#workflow

Ribes, D., & Jackson, S. J. (2013). Data bite man: The work of sustaining a long-term study. In L. Gitelman (Ed.), *"Raw Data" is an Oxymoron* (pp. 147–166). Cambridge, MA: MIT Press.

Rosenberg, D. (2003). Early Modern Information Overload. *Journal of the History of Ideas*, *64*(1), 1–9. http://doi.org/10.1353/jhi.2003.0017

Rosenberg, D. (2013). Data before the fact. In L. Gitelman (Ed.), *"Raw Data" is an Oxymoron* (pp. 15–40). Cambridge, MA: MIT Press.

Steinhart, G., Chen, E., Arguillas, F., Dietrich, D., & Kramer, S. (2012). Prepared to Plan? A Snapshot of Researcher Readiness to Address Data Management Planning Requirements. *Journal of eScience Librarianship*, *1*(2). http://doi.org/10.7191/jeslib.2012.1008

UC3: Data Management Guidelines. (2014). Retrieved May 14, 2014, from http://www.cdlib.org/uc3/datamanagement/documenting.html

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

Wessels, B., Finn, R. L., Linde, P., Mazzetti, P., Nativi, S., Riley, S., … Wyatt, S. (2014). Issues in the development of open access to research data. *Prometheus*, *32*(1), 49–66. http://doi.org/10.1080/08109028.2014.956505

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