Title

Syllabus for Data, Data Practices, and Data Curation, Part II, Spring 2014, UCLA Information Studies

Author

Borgman, Christine L.

Publication Date

2014

Data, Data Practices, and Data Curation

Part II, SPRING 2014, UCLA Information Studies
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 days, weeks, or months. 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. This requirement is causing a mad scramble for compliance by researchers, universities, librarians, and archivists. The Association for Research Libraries already has issued several reports on the library's role in data management, adding to a plethora of policy reports in this area. The Institute for Museum and Library Services is funding curriculum development for data librarians. Data management is clearly a growth area for information studies graduates.

These two courses (winter and spring) will 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 for this term include local experts and distinguished guests from the National Academy of Sciences and the Library of Congress by 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. Funding agencies and faculty are looking to libraries for leadership for the management, curation, hosting, and dissemination of research data. Data librarianship 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.

The two-part sequence of courses in Data, Data Practices, and Data Curation was developed with the substantial contributions of UCLA doctoral students Jillian Wallis and Laura Wynholds and guidance from students enrolled in prior offerings of the course in 2010 and 2011. 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 between the many forms of data, how data vary by scholarly discipline, and how they are used throughout the scholarly life cycle.
2. Students will learn some 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 the roles that data play in research collaborations.
5. Students will gain a basic knowledge of data curation practices in the library and archive fields.
6. Students will learn basic principles of public policies for data.

Course Materials

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

Office Hours

Wednesdays, 3:00-4:30pm (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 the 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 site. Assignments will be marked down 2 points for each day late. No assignments will be accepted after midnight on Weds, JUNE 7.

Overview:

Topics, Readings, and Guest speakers

Week 1, April 2: Selection and appraisal across domains

Week 2: April 9: Data archives and repositories

Week 3: April 16: Economics of data: preservation, access, and sustainability

Week 4: April 23: Managing the Data Deluge: NISO seminar
http://www.niso.org/news/events/2014/virtual/data_deluge/

Week 5: April 30: Provenance in data: WWW and archival approaches, Paul Groth, University of Amsterdam, by video

Week 6: May 7: U.S. and international data policy: Paul Uhlir, National Academies of Science, by video

Week 7: May 14: Intellectual property in data: Maureen Whalen, Getty, in person and Peter Hirtle, Berkman Center, Harvard, by video

Week 8: May 19: Research Libraries Respond to Data Policy: MacKenzie Smith, University Librarian, UC-Davis (Special Session)

Week 9, May 28: Technologies of persistence and identification

Week 10: June 7: Project presentations

Week 11 (Exam Week): June 11: Projects due

Week 1, April 2: Selection and appraisal

How do we determine what data are worth keeping? What are the criteria? To what extent are these scholarly, disciplinary, policy, or professional questions? Who decides? What role do scholars play in determining what is kept and what are professional roles?

Readings

(Borgman, 2015), Chapter 10, What to Keep and Why
(Faniel & Jacobsen, 2010)
(Gutmann et al., 2009)
(Harvey, 2008)
("Strategic Directions: Appraisal Policy," 2007)
(Goodman et al., 2014)
("Australian Code for the Responsible Conduct of Research," 2007)
(Pfeiffenberger & Carlson, 2011)

References

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

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

Faniel, I. M., & Jacobsen, T. E. (2010). Reusing Scientific Data: How Earthquake Engineering Researchers Assess the Reusability of Colleagues' Data. *Journal of Computer-Supported Cooperative Work*, *19*, 355–375. doi:10.1007/s10606-010-9117-8

Goodman, A. A., Pepe, A., Blocker, A. W., Borgman, C. L., Cranmer, K., Crosas, M., … Slavkovic, A. (2014). 10 Simple Rules for the Care and Feeding of Scientific Data. *PLoS Computational Biology*. Retrieved from https://www.authorea.com/users/3/articles/3410/_show_article

Gutmann, M., Abrahamson, M., Adams, M., Altman, M., Arms, C. R., & King, G. (2009). From Preserving the Past to Preserving the Future: The Data-PASS Project and the Challenges of Preserving Digital Social Science Data. *Library Trends*, *57*, 315–337.

Harvey, R. (2008). Appraisal and Selection. Digital Curation Centre. Retrieved from http://www.dcc.ac.uk/resources/briefing-papers/introduction-curation/appraisal-and-selection

Pfeiffenberger, H., & Carlson, D. (2011). "Earth System Science Data" (ESSD) - A Peer Reviewed Journal for Publication of Data. *D-Lib Magazine*, *17*. doi:10.1045/january2011-pfeiffenberger

Strategic Directions: Appraisal Policy. (2007). *National Archives and Records Administration*.

Week 2: April 9: Data archives and repositories

Data collections and repositories take many forms and reside in many places. Libbie Stephenson will provide an introduction to the technical aspects of data ingest and professional aspects of repository management.

Speaker: Libbie Stephenson, Director, UCLA Social Science Data Archive

Readings (and sites to visit)

Only a couple of these are narrative readings (Greenberg, White, Carrier, & Scherle, 2009; Kimpton & Morris, 2013; Peer, Green, & Stephenson, 2014). The COAR report ("COAR: The Current State of Open Access Repository Interoperability," 2012) should be skimmed, and we'll return to it later when we discuss technologies for access to data. The rest are sites that you should visit and browse to get a sense of the array of data repositories and issues such as standards and certifications.

Open Data
Open Data Foundation http://www.opendatafoundation.org/?lvl1=resources&lvl2=papers
OpenDOAR - Open Access Countries and Organisations - Europe. (2013). Retrieved June 18, 2013, from http://www.opendoar.org/countrylist.php?cContinent=Europe#United%20Kingdom
Dealing with Data. (2013). *OpenAIRE*. Retrieved June 16, 2013, from http://www.openaire.eu/en/component/content/article/9/462
Directory of Open Access Repositories. (2007). Retrieved from www.opendoar.org

Standards
Trustworthy Repositories. (2014). *Digital Curation Centre*. Retrieved January 28, 2014, from http://www.dcc.ac.uk/resources/repository-audit-and-assessment/trustworthy-repositories
Data Seal of Approval. (2014). Retrieved March 31, 2014, from http://www.datasealofapproval.org/en/

Metadata
Greenberg, J., White, H. C., Carrier, S., & Scherle, R. (2009). A Metadata Best Practice for a Scientific Data Repository. *Journal of Library Metadata*, *9*, 194–212.
DataCite Schemas repository. (2014). Retrieved February 12, 2014, from http://schema.datacite.org/meta/kernel-3/index.html
COAR » The Current State of Open Access Repository Interoperability (2012). (n.d.). Retrieved November 2, 2012, from http://www.coar-repositories.org/working-groups/repository-interoperability/coar-interoperability-project/the-current-state-of-open-access-repository-interoperability-2012/

Repositories – Management issues
(Kimpton & Morris, 2013; Peer et al., 2014)

Repositories - offering services
Australian National Data Service. (2014). Retrieved January 24, 2014, from http://www.ands.org.au/
UK Data Service http://ukdataservice.ac.uk/

Inter-university Consortium for Political and Social Research (ICPSR) http://www.icpsr.umich.edu
DataBib http://databib.org/
Registry of Research Data Repositories http://www.re3data.org/
OpenContext http://opencontext.org/ and The Digital Archaeological record (tDAR) https://www.tdar.org/

Repositories - self-deposit
Dryad http://datadryad.org/
figshare http://figshare.com/
DataShare/U.S. http://datashare.ucsf.edu/xtf/search
DataShare/U.K. http://datashare.is.ed.ac.uk/
openICPSR http://openicpsr.org/

Repository Systems
Islandora http://islandora.ca/
Duraspace http://www.duraspace.org/
Hydra Project http://projecthydra.org/
California Digital Library Merritt http://www.cdlib.org/uc3/merritt/

Recommended Readings

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

Data Seal of Approval. (2014). Retrieved March 31, 2014, from http://www.datasealofapproval.org/en/

Dealing with Data. (2013). *OpenAIRE*. Retrieved June 16, 2013, from http://www.openaire.eu/en/component/content/article/9/462

Directory of Open Access Repositories. (2007). Retrieved from www.opendoar.org

OpenDOAR - Open Access Countries and Organisations - Europe. (2013). Retrieved June 18, 2013, from http://www.opendoar.org/countrylist.php?cContinent=Europe#United%20Kingdom

Trustworthy Repositories. (2014). *Digital Curation Centre*. Retrieved January 28, 2014, from http://www.dcc.ac.uk/resources/repository-audit-and-assessment/trustworthy-repositories

Week 3: April 16: Economics of data: preservation, access, and sustainability

In managing and curating research data, the economic issues are unavoidable. What are the economic models for sustaining access to data? How can projects be costed and budgeted? How do economic considerations influence choices of what to keep, for how long, and why? We will consider several approaches to the economics of research data.

Readings

(Beagrie & Houghton, 2014)
(Berman et al., 2010)
(David, 2009)

(Hess & Ostrom, 2007)
(Kelty, 2012)
(Rosenthal, 2010)

References

Beagrie, N., & Houghton, J. (2014). The Value and Impact of Data Sharing and Curation. JISC. Retrieved from repository.jisc.ac.uk/5568/1/iDF308_ _Digital_Infrastructure_Directions_Report%2C_Jan14_v1-04.pdf

Berman, F., Lavoie, B., Ayris, P., Choudhury, G. S., Cohen, E., Courant, P., … Van Camp, A. (2010). *Sustainable Economics for a Digital Planet: Ensuring Long-Term Access to Digital Information*. San Diego: National Science Foundation, Andrew W. Mellon Foundation, Library of Congress, Joint Information Systems Committee of the UK, Council on Library and Information Resources, National Archives and Records Administration. Retrieved from http://brtf.sdsc.edu/publications.html

David, P. (2009). *Mitigating "Anticommons" Harms to Research In Science and Technology* (Discussion Paper No. 10-009). Stanford Institute for Economic Policy Research. Retrieved from http://ideas.repec.org/p/sip/dpaper/10-009.html

Hess, C., & Ostrom, E. (2007). Introduction: An overview of the knowledge commons. In C. Hess & E. Ostrom (Eds.), *Understanding knowledge as a commons: from theory to practice* (pp. 3–26). Cambridge, Mass.: MIT Press.

Kelty, C. M. (2012). This is not an article: Model organism newsletters and the question of "open science." *BioSocieties*, *7*(2), 140–168. doi:10.1057/biosoc.2012.8

Rosenthal, D. (2010). Stepping Twice Into The Same River. Presented at the Joint Conference on Digital Libraries, 12 July 2010: ACM. Retrieved from http://blog.dshr.org/2010/06/jcdl-2010-keynote.html

Recommended Readings

Berman, F., & Cerf, V. G. (2013). Who Will Pay for Public Access to Research Data? *Science*, *341*(6146), 616–617. doi:10.1126/science.1241625

Boyle, J., & Jenkins, J. (2003). The genius of intellectual property and the need for the public domain. In *The Role of Scientific and Technical Data and Information in the Public Domain* (pp. 10–14). Washington, DC: The National Academies Press. Retrieved from http://newton.nap.edu/catalog/10785.html#toc

Data Archiving Policy, Directorate for Social, Behavioral & Economic Sciences, National Science Foundation. (2012). Retrieved November 14, 2012, from http://www.nsf.gov/sbe/ses/common/archive.jsp

Esanu, J. M., & Uhlir, P. F. (2003). *The Role of Scientific and Technical Data and Information in the Public Domain*. Washington, DC: The National Academies Press. Retrieved from http://books.nap.edu/catalog/10785.html

Esanu, J. M., & Uhlir, P. F. (2004). *Open Access and the Public Domain in Digital Data and Information for Science: Proceedings of an International Symposium*. Washington, DC: The National Academies Press. Retrieved from http://www.nap.edu/catalog.php?record_id=11030

Week 4: April 23: Managing the Data Deluge: NISO seminar

In class attendance of the **NISO Virtual Conference: Dealing with the Data Deluge: Successful Techniques for Scientific Data Management.**
http://www.niso.org/news/events/2014/virtual/data_deluge/

The class today will be held in YRL 11348, the presentation room, from 8am to 1pm. Students are encouraged to come early to hear the keynote speaker (8-9am) and stay for the closing session if they can. Because Prof. Borgman is among the presenters (11:15-11:45am), the fees are waived for participants.

Readings for this week are by the presenters or on their topics, so that you will have a full framing of the issues being addressed by these leaders in the field.

About the Virtual Conference (NISO text)

With the expansion of digital data collection and the increased expectations of data sharing, researchers are turning to their libraries or institutional repositories as a place to store and preserve that data. Many institutions have created such data management services and see the data curation role as a growing and important element of their service portfolio. While some of the experience in managing other types of digital resources is transferrable, the management of large-scale scientific data has many special requirements and challenges. From metadata collection and cataloging data sources, to identification, discovery, and preservation, best practices and standards are still in their infancy.

This Virtual Conference will explore in greater depth than traditional webinars some of the practical lessons from those who have implemented data management and developed best practices, as well as provide some insight into the evolving issues the community faces. It will include discussions related to certification of trusted repositories, provenance and identification issues around data, data citation, preservation, and the work of several repository networks to advance distribution of scientific information.

Readings

(CODATA-ICSTI Task Group on Data Citation Standards and Practices, 2013)
(Crosas, 2011)
(Datacitation Synthesis Group, 2014)
(DataCite, 2013)
("The Dataverse Network: Presentations," 2012)
(*The Purdue University Research Repository (PURR)*, 2013)
(Holdren, 2013a, 2013b)

References

CODATA-ICSTI Task Group on Data Citation Standards and Practices. (2013). Out of Cite, Out of Mind: The Current State of Practice, Policy, and Technology for the Citation of Data. *Data Science Journal*, *12*, 1–75. Retrieved from https://www.jstage.jst.go.jp/article/dsj/12/0/12_OSOM13-043/_article

Crosas, M. (2011). The Dataverse Network®: An Open-Source Application for Sharing, Discovering and Preserving Data. *D-Lib Magazine*, *17*. doi:10.1045/january2011-crosas

Datacitation Synthesis Group. (2014). Joint Declarion on Data Citation Principles - Final. *Force11: The Future of Research Communications and Scholarship*. Retrieved February 12, 2014, from http://www.force11.org/datacitation

DataCite. (2013). Retrieved September 10, 2013, from http://www.datacite.org/

Holdren, J. P. (2013a, February 22). Increasing Access to the Results of Federally Funded Scientific Research. Executive Office of the President, Office of Science and Technology Policy. Retrieved from http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf

Holdren, J. P. (2013b, February 22). Memorandum for the Heads of Executive Departments and Agencies. Executive Office of the President, Office of Science and Technology Policy. Retrieved from http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf

The Dataverse Network: Presentations. (2012). *Projects at Harvard*. Retrieved November 13, 2012, from http://thedata.org/presentations

The Purdue University Research Repository (PURR). (2013). Retrieved from http://www.youtube.com/watch?v=Yw0IJj7FqA8&feature=youtube_gdata_player

Week 5: April 30: Provenance in data: WWW and archival approaches, Paul Groth, University of Amsterdam, by video

Archival and Internet approaches, Paul Groth, University of Amsterdam, discussant, 9am-10:30am by video. Groth will discuss which of the many competing definitions of provenance is most useful for data, what aspects of provenance are most applicable to data management, to curation, to discovery, and to curation, what are the points of intersection between archival, historical, and technical concepts of provenance, especially with respect to data, and what are the practical considerations in documenting provenance for data stewardship.

Readings

(Barbier, Feng, Gundecha, & Liu, 2013)
(Bunn, 2014)
(Cheny, Gil, Groth, & Miles, 2011)
(Garber, 2012)
(Goodman et al., 2014)
(Groth & Moreau, 2013)
(Meng, 2011)
(Niu, 2013)

References

Barbier, G., Feng, Z., Gundecha, P., & Liu, H. (2013). Provenance Data in Social Media. *Synthesis Lectures on Data Mining and Knowledge Discovery*, *4*(1), 1–84. doi:10.2200/S00496ED1V01Y201304DMK007

Bunn, J. (2014). Questioning autonomy: an alternative perspective on the principles which govern archival description. *Archival Science*, *14*(1), 3–15. doi:10.1007/s10502-013-9200-2

Cheny, J., Gil, Y., Groth, P., & Miles, S. (2011). *Requirements for Provenance on the Web.* Retrieved from http://www.w3.org/2005/Incubator/prov/wiki/User_Requirements

Garber, M. (2012, March 15). Scholars: Yes, We Need Better Attribution Systems (but No, We Don't Know How to Make Them, Either). *The Atlantic*. Retrieved March 31, 2014, from http://www.theatlantic.com/technology/archive/2012/03/scholars-yes-we-need-better-attribution-systems-but-no-we-dont-know-how-to-make-them-either/254527/

Groth, P., & Moreau, L. (2013). PROV-Overview. Retrieved April 14, 2014, from http://www.w3.org/TR/2013/NOTE-prov-overview-20130430/

Meng, X.-L. (2011). Multi-party inference and uncongeniality. In M. Lovric (Ed.), *International Encyclopedia of Statistical Science* (pp. 884–888). Berlin: Springer-Verlag.

Niu, J. (2013). Provenance: crossing boundaries. *Archives and Manuscripts*, *41*(2), 105–115. doi:10.1080/01576895.2013.811426

Recommended Readings

Archer, D. W., Delcambre, L. M. L., & Maier, D. (2013). User Trust and Judgments in a Curated Database with Explicit Provenance. In V. Tannen, L. Wong, L. Libkin, W. Fan, W.-C. Tan, & M. Fourman (Eds.), In Search of Elegance in the Theory and Practice of Computation (pp. 89–111). Springer Berlin Heidelberg. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-41660-6_5

Foster, I., & Moreau, L. (2006). *Provenance and Annotation of Data*. Springer: Heidelberg. Retrieved from http://www.w3.org/2011/prov/wiki/Connection_Task_Force_Informal_Report

Hastings, J., Chepelev, L., Willighagen, E., Adams, N., Steinbeck, C., & Dumontier, M. (2011). The Chemical Information Ontology: Provenance and Disambiguation for Chemical Data on the Biological Semantic Web. *PLoS ONE*, *6*(10), e25513. doi:10.1371/journal.pone.0025513

Moreau, L., & Missier, P. (Eds.). (2013, April 30). PROV-DM: The PROV Data Model. Retrieved from http://www.w3.org/TR/prov-dm/

Moreau, L. (2010). The Foundations for Provenance on the Web. *Foundations and Trends® in Web Science*, *2*(2-3), 99–241. doi:10.1561/1800000010

Pearson, S. H. (2012). Three legal mechanisms for sharing data. In P. F. Uhlir (Ed.), *For Attribution -- Developing Data Attribution and Citation Practices and Standards: Summary of an International Workshop* (pp. 71–76). National Academies Press. Retrieved from http://www.nap.edu/openbook.php?record_id=13564&page=71

Zhao, J., Miles, A., Klyne, G., & Shotton, D. (2009). Linked data and provenance in biological data webs. *Briefings in Bioinformatics*, *10*(2), 139–152. doi:10.1093/bib/bbn04

Week 6: May 7: U.S. and international data policy: Paul Uhlir, National Academies of Science, invited

Paul Uhlir will discuss the most pressing policy questions for research data in the U.S., how U.S. research policies for data differ from those of other countries and regions, such as UK, EU, and Australia, and his recent book on governance policy for microbial data.

Readings

(Anderson, 2013)
(Boulton, 2012)
(Burwell, VanRoekel, Park, & Mancini, 2013)
("Data Archiving Policy," 2012)
("Joint Data Archiving Policy," 2012)
(Duke & Porter, 2013)
(Holdren, 2013a, 2013b)
(Lide & Wood, 2012)
(Uhlir, 2007)

References

Anderson, K. (2013, June 4). Joining a CHORUS, Publishers Offer the OSTP a Proactive, Modern, and Cost-Saving Public Access Solution. *The Scholarly Kitchen*. Retrieved from http://scholarlykitchen.sspnet.org/2013/06/04/joining-a-chorus-publishers-offer-the-ostp-a-proactive-modern-and-cost-saving-public-access-solution/

Boulton, G. (2012). Open your minds and share your results. *Nature*, *486*(7404), 441–441. doi:10.1038/486441a

Burwell, S. M., VanRoekel, S., Park, T., & Mancini, D. J. (2013, May 9). Open Data Policy-Managing Information as an Asset. Executive Office of the President, Office of Management and Budget. Retrieved from www.whitehouse.gov/.../omb/memoranda/2013/m-13-13.p...

Data Archiving Policy, Directorate for Social, Behavioral & Economic Sciences, National Science Foundation. (2012). Retrieved November 14, 2012, from http://www.nsf.gov/sbe/ses/common/archive.jsp

Duke, C. S., & Porter, J. H. (2013). The Ethics of Data Sharing and Reuse in Biology. *BioScience*, *63*(6), 483–489. doi:10.1525/bio.2013.63.6.10

Holdren, J. P. (2013a, February 22). Increasing Access to the Results of Federally Funded Scientific Research. Executive Office of the President, Office of Science and Technology Policy. Retrieved from http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf

Holdren, J. P. (2013b, February 22). Memorandum for the Heads of Executive Departments and Agencies. Executive Office of the President, Office of Science and Technology Policy. Retrieved from http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf

Joint Data Archiving Policy. (2012). *Dryad*. Retrieved from http://datadryad.org/jdap

Lide, D. R., & Wood, G. H. (2012). *CODATA @ 45 Years: 1966 to 2010*. Paris: CODATA. Retrieved from http://www.codata.org/about/CODATA@45years.pdf

Uhlir, P. F. (2007). Open Data for Global Science: A Review of Recent Developments in National and International Scientific Data Policies and Related Proposals. *Data Science Journal*, *6*, 1–3. Retrieved from http://www.codata.org/dsj/special-open-data.html

Recommended Readings

AURA/NOAO Data Rights Policy. (2013). Retrieved March 14, 2013, from http://www.noao.edu/noaoprop/help/datarights.html

Boulton, G. (2012b). Science as an open enterprise. *The Royal Society*. Retrieved May 24, 2013, from http://royalsociety.org/policy/projects/science-public-enterprise/report/

CODATA-ICSTI Task Group on Data Citation Standards and Practices. (2013). Out of Cite, Out of Mind: The Current State of Practice, Policy, and Technology for the Citation of Data. *Data Science Journal*, *12*, 1–75. Retrieved from https://www.jstage.jst.go.jp/article/dsj/12/0/12_OSOM13-043/_article

Curtis, J., Harding, A., & Reynolds, F. (2011). *History for the taking? Perspectives on material heritage*. London: British Academy. Retrieved from http://www.britac.ac.uk/policy/History-for-the-taking.cfm

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

Expanding Public Access to the Results of Federally Funded Research. (2013). *The White House*. Retrieved May 20, 2013, from http://www.whitehouse.gov/blog/2013/02/22/expanding-public-access-results-federally-funded-research

Haeussler, C. (2011). Information-sharing in academia and the industry: A comparative study. *Research Policy*, *40*, 105–122. doi:10.1016/j.respol.2010.08.007

Howard, J. (2013a, February 22). White House Delivers New Open-Access Policy That Has Activists Cheering. *The Chronicle of Higher Education*. Retrieved from http://chronicle.com/article/White-House-Delivers-New/137549/?cid=at&utm_source=at&utm_medium=en

Howard, J. (2013b, June 7). Universities and Libraries Envision a "Federated System" for Public Access to Research. *Wired Campus*. The Chronicle of Higher Education. Retrieved from http://chronicle.com/blogs/wiredcampus/universities-and-libraries-envision-a-federated-system-for-public-access-to-research/44147

Journal Research Data Policy Bank. (2013). Retrieved May 24, 2013, from http://jordproject.wordpress.com/category/project-information/

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

Mathae, K. B., & Uhlir, P. F. (Eds.). (2012). *The Case for International Sharing of Scientific Data: A Focus on Developing Countries: Proceedings of a Symposium*. Washington, D.C.: The National Academies Press.

Nash, V. (2013, September 19). Responsible research agendas for public policy in the era of big data. *Policy and Internet*. Retrieved from http://blogs.oii.ox.ac.uk/policy/responsible-research-agendas-for-public-policy-in-the-era-of-big-data/

National Information Standards Organization. (2013). Recommended Practice for Online Supplemental Journal Article Materials. Retrieved September 10, 2013, from http://www.niso.org/publications/rp/rp-15-2013

NIH Public Access Policy. (2013). Retrieved from http://publicaccess.nih.gov/submit_process.htm

Pienta, A. M., Alter, G. C., & Lyle, J. A. (2010). The Enduring Value of Social Science Research: The Use and Reuse of Primary Research Data. Presented at the The Organisation, Economics and Policy of Scientific Research, Torino, Italy. Retrieved from http://deepblue.lib.umich.edu/bitstream/handle/2027.42/78307/pienta_alter_lyle_100331.pdf?sequence=1

RECODE Project. (2013). Policy RECommendations for Open Access to Research Data in Europe. *Policy RECommendations for Open Access to Research Data in Europe*. Retrieved September 11, 2013, from http://recodeproject.eu/resources-project-publications-other-projects-links-etc/

Research Councils UK. (2011). RCUK Common Principles on Data Policy. Retrieved February 9, 2013, from http://www.rcuk.ac.uk/research/Pages/DataPolicy.aspx

Research Councils UK announces new Open Access policy. (2012, July). Retrieved November 16, 2012, from http://www.rcuk.ac.uk/media/news/2012news/Pages/120716.aspx

Research Data Alliance. (2013). Retrieved September 30, 2013, from https://rd-alliance.org/node

Schofield, P., Eppig, J., Huala, E., Hrabe de Angelis, M., Harvey, M., Davidson, D., … Hancock, J. (2010). Sustaining the data and bioresource commons. *Science*, *330*, 592–593.

The Royal Society. (2010). *The Scientific Century: Securing Our Future Prosperity*. Royal Society: London. Retrieved from http://royalsociety.org/policy/publications/2010/scientific-century/

Uhlir, P. F. (2012). *The Future of Scientific Knowledge Discovery in Open Networked Environments: Summary of a Workshop*. Washington, D.C.: The National Academies Press. Retrieved from http://www.nap.edu/catalog.php?record_id=18258

Uhlir, P. F., & Levey, C. W. (2010). *The Changing Role of Libraries in Support of Research Data Activities: A Public Symposium. Meeting Recap*. National Academies of Science, Board on Research Data and Information. Retrieved from http://sites.nationalacademies.org/PGA/brdi/PGA_047286

Uhlir, P. F., & Schröder, P. (2007). Open Data for Global Science. *Data Science Journal*, *6*, 36–53. Retrieved from http://www.codata.org/dsj/special-open-data.html

Week 7: May 14: Intellectual property in data: Maureen Whalen, Getty, and Peter Hirtle, Berkman Center, Harvard, by video

Policies and practices for open data tend to say little about the problems of intellectual property rights. The OECD report (read earlier in the term) is an exception. Yet rights in data are unclear and often contested. Investigators may not know what rights they have in data, and they may be obtaining data from other sources for which they have little control over rights, formats, reuse, deposit, etc. Complicating matters further are the license stacking problems, wherein each dataset has particular rights associated with it, making data difficult to combine.

Readings

(Ball, 2012)
("Defining 'non-consumptive' use," 2014)
("LIBER response to Elsevier's text and data mining policy | LIBER," 2014)
(Hirtle, 2011)
(Scotchmer, 2003)
(Shillum, 2014)

References

Ball, A. (2012). How to License Research Data. *Digital Curation Centre*. Retrieved March 3, 2013, from http://www.dcc.ac.uk/resources/how-guides/license-research-data

Defining “non-consumptive” use. (2014). *Australian Law Reform Commission*. Retrieved January 15, 2014, from http://www.alrc.gov.au/publications/8-non-consumptive-use/defining-%E2%80%98non-consumptive%E2%80%99-use

Hirtle, P. B. (2011). Introduction to Intellectual Property Rights in Data Management. Retrieved March 3, 2013, from https://confluence.cornell.edu/display/rdmsgweb/introduction-intellectual-property-rights-data-management

LIBER response to Elsevier’s text and data mining policy | LIBER. (2014, March 28). Retrieved May 12, 2014, from http://www.libereurope.eu/news/liber-response-to-elsevier%E2%80%99s-text-and-data-mining-policy

Scotchmer, S. (2003). Intellectual property -- When is it the best incentive mechanism for S&T data and information? In *The Role of Scientific and Technical Data and Information in the Public Domain* (pp. 15–18). Washington, DC: The National Academies Press. Retrieved from http://books.nap.edu/catalog/10785.html

Shillum, C. (2014, January 31). Elsevier updates text-mining policy to improve access for researchers. *Elsevier Connect*. Retrieved May 12, 2014, from http://www.elsevier.com/connect/elsevier-updates-text-mining-policy-to-improve-access-for-researchers

Recommended Readings

Boyle, J., & Jenkins, J. (2003). The genius of intellectual property and the need for the public domain. In *The Role of Scientific and Technical Data and Information in the Public Domain* (pp. 10–14). Washington, DC: The National Academies Press. Retrieved from http://newton.nap.edu/catalog/10785.html#toc

Committee on Ensuring the Utility and Integrity of Research Data in a Digital Age; National Academy of Sciences. (2009). *Ensuring the Integrity, Accessibility and Stewardship of Research Data in the Digital Age*. National Academy of Science: Washington. Retrieved from http://www.nap.edu/catalog.php?record_id=12615

David, P. A. (2004). Can “Open Science” be protected from the evolving regime of intellectual property protections. *Journal of Institutional and Theoretical Economics*, *160*. Retrieved from http://www-econ.stanford.edu/faculty/workp/swp03011.pdf

Elliott, R. (2005). Who owns scientific data? The impact of intellectual property rights on the scientific publication claim. *Learned Publishing*, *18*, 91–94.

Gadd, E., Oppenheim, C., & Probets, S. (2003). The intellectual property rights issues facing self-archiving: Key findings of the RoMEO project. *D-Lib Magazine*, *9*. Retrieved from http://www.dlib.org/dlib/september03/gadd/09gadd.html

Guilbault, L. (2013). Licensing research data under open access conditions. In D. Beldiman (Ed.), *Information and Knowledge: 21st Century Challenges in Intellectual Property and Knowledge Governance*. Cheltenham: Edward Elgar. Retrieved from http://www.ivir.nl/staff/guibault.html

Hirtle, P. B., Hudson, E., & Kenyon, A. T. (2009). *Copyright and Cultural Institutions: Guidelines for U.S. Libraries, Archives, and Museums*. Retrieved from http://hdl.handle.net/1813/14142

Kelty, C. M. (2012). This is not an article: Model organism newsletters and the question of “open science.” *BioSocieties*, *7*(2), 140–168. doi:10.1057/biosoc.2012.8

Loring, J. F., & Campbell, C. (2006). Intellectual property and human embryonic stem cell research. *Science*, *311*, 1716–1717.

SHERPA/RoMEO: Publisher copyright policies & self-archiving. (2006). Retrieved from http://www.sherpa.ac.uk/romeo.php

The Hargreaves Review of Intellectual Property: Where Next? (2012). London: The Stationery Office.

The Royal Society. (2003). *Keeping science open: the effects of intellectual property policy on the conduct of science*. Royal Society: London. Retrieved from http://royalsociety.org/uploadedFiles/Royal_Society_Content/policy/publications/2003/9845.pdf

UC and the Google Book Settlement: Frequently Asked Questions. (2009). Retrieved from http://osc.universityofcalifornia.edu/google/faq.html

Week 8: May 19: Research Libraries Respond to Data Policy: MacKenzie Smith, University Librarian, UC-Davis

Special session, Monday, May 19, 12:30-3:30pm, in GSEIS room 121 with respondents Dr. Vessela Ensberg, Data Curation Analyst, UCLA Libraries, and Edson Smith, data curation strategist and technical lead on data curation, UCLA Libraries. Will discuss how universities are adapting to an environment of open access and open data mandates, how library associations such as ARL are addressing data management, what approaches UC campuses are taking, what role libraries play in research policy compliance, and similar topics.

Readings

(Crosas, 2013)
(IAU Working Group Libraries, 2013)
(Kenney, 2014)
(Smith, 2010)
(Soehner, Steeves, & Ward, 2010)

References

Crosas, M. (2013). A Data Sharing Story. *Journal of eScience Librarianship*, *1*(3). doi:http://dx.doi.org/10.7191/jeslib.2012.1020

IAU Working Group Libraries. (2013). Best Practices for Creating a Telescope Bibliography. *IAU-Commission5 - WG Libraries*. Retrieved May 7, 2013, from http://iau-commission5.wikispaces.com/WG+Libraries

Kenney, A. R. (2014). Leveraging the Liaison Model. Retrieved March 31, 2014, from http://www.sr.ithaka.org/blog-individual/leveraging-liaison-model

Smith, M. (2010, December 9). *IDCC10 Keynote by MacKenzie Smith*. Retrieved from http://vimeo.com/17662208

Soehner, C., Steeves, C., & Ward, J. (2010). *E-Science and Data Support Services: A Study of ARL Member Institutions* (pp. 1–74). Washington, DC: Association of Research Libraries. Retrieved from http://www.arl.org/bm~doc/escience_report2010.pdf

Recommended Readings

Anderson, R., Frost, H., Hoebelheinrich, N., & Johnson, K. (2005). The AIHT at Stanford University: Automated Preservation Assessment of Heterogeneous Digital Collections. *D-Lib Magazine*, *11*(12). doi:10.1045/december2005-johnson

Association of Research Libraries. (2009). *The Research Library's Role in Digital Repository Services*. Washington, DC. Retrieved from http://www.arl.org/bm~doc/repository-services-report.pdf

Howard, J. (2013a, June 7). Universities and Libraries Envision a "Federated System" for Public Access to Research. *Wired Campus*. The Chronicle of Higher Education. Retrieved from http://chronicle.com/blogs/wiredcampus/universities-and-libraries-envision-a-federated-system-for-public-access-to-research/44147

Howard, J. (2013b, December 9). Digital Public Library of America: Young but Well Connected. *The Chronicle of Higher Education*. Retrieved from http://chronicle.com/article/Digital-Library-of-America/143489/?cid=wc&utm_source=wc&utm_medium=en

Lynch, C. A. (2013). Defining a National Library in a Digital World: Dame Lynne Brindley and the British Library. *Alexandria: The Journal of National and International Library and Information Issues*.

Nielsen, H. J., & Hjørland, B. (2014). Curating research data: the potential roles of libraries and information professionals. *Journal of Documentation*, *70*(2), 221–240. doi:10.1108/JD-03-2013-0034

Rimkus, K., Padilla, T., Popp, T., & Martin, G. (2014). Digital Preservation File Format Policies of ARL Member Libraries: An Analysis. *D-Lib Magazine*, *20*(3/4). doi:10.1045/march2014-rimkus

Romanowsky, K. (2013, September 4). Introducing the Open Access Policy Implementation (OAPI) project: California Digital Library. California Digital Library. Retrieved from http://www.cdlib.org/cdlinfo/2013/09/04/introducing-the-open-access-policy-implementation-oapi-project/?utm_source=twitterfeed&utm_medium=twitter

Rowe, C. (2013). The new library of Babel? Borges, digitisation and the myth of a universal library. *First Monday*, *18*(2). doi:10.5210/fm.v18i2.3237

Special Issue on Institutional Repositories. (2008). *Library Trends*, *57*. Retrieved from http://muse.jhu.edu/journals/library_trends/toc/lib.57.2.html

Willett, P. (2011). Data Management Guidelines. Retrieved from http://www.cdlib.org/services/uc3/datamanagement/index.html

Week 9: May 28: Technologies of persistence and identification

The technologies associated with data curation and management vary widely by domain, method, and many other factors. We will focus on generic tools and principles for technology as an entry point for developing a broader background. These readings and the video address the following topics:

Persistent identification
Unique identification of digital objects
Digital Object Identifiers
CrossRef
Open Archives Initiative protocols OAI-PMH and OAI-ORE
Research Objects and provenance
Linked open data for data
Namespaces, URLs, and versions of record

Readings

(Bechhofer et al., 2013)
(Bechhofer, De Roure, Gamble, Goble, & Buchan, 2010)

(CrossRef, 2009)
(*CNI*, 2013)
("OpenURL and CrossRef," 2006)
(Goodman et al., 2014)
(Pepe, Mayernik, Borgman, & Van de Sompel, 2010)
(Sanderson & Van de Sompel, 2012)
(Simons, 2012)
(Weigel, Lautenschlager, Toussaint, & Kindermann, 2013)

References

Bechhofer, S., Buchan, I., De Roure, D., Missier, P., Ainsworth, J., Bhagat, J., … Goble, C. (2013). Why linked data is not enough for scientists. *Future Generation Computer Systems*, *29*(2), 599–611. doi:10.1016/j.future.2011.08.004

Bechhofer, S., De Roure, D., Gamble, M., Goble, C., & Buchan, I. (2010). Research Objects: Towards Exchange and Reuse of Digital Knowledge. *Nature Precedings*. doi:10.1038/npre.2010.4626.1

CrossRef. (2009). *The Formation of CrossRef: A Short History*. Retrieved from http://www.crossref.org/01company/02history.html

From the Version of Record to a Version of the Record. (2013). Retrieved from http://www.youtube.com/watch?v=fhrGS-QbNVA&feature=youtube_gdata_player

Goodman, A., Pepe, A., Blocker, A. W., Borgman, C. L., Cranmer, K., Crosas, M., … Slavkovic, A. (2014). Ten Simple Rules for the Care and Feeding of Scientific Data. *PLoS Comput Biol*, *10*(4), e1003542. doi:10.1371/journal.pcbi.1003542

OpenURL and CrossRef. (2006). Retrieved from http://www.crossref.org/02publishers/16openurl.html

Pepe, A., Mayernik, M. S., Borgman, C. L., & Van de Sompel, H. (2010). From Artifacts to Aggregations: Modeling Scientific Life Cycles on the Semantic Web. *Journal of the American Society for Information Science and Technology*, *61*, 567–582. doi:10.1002/asi.21263

Sanderson, R., & Van de Sompel, H. (2012). Cool URIs and Dynamic Data. *IEEE Internet Computing*, *16*(4), 76–79. doi:10.1109/MIC.2012.78

Simons, N. (2012). Implementing DOIs for Research Data. *D-Lib Magazine*, *18*(5/6). doi:10.1045/may2012-simons

Weigel, T., Lautenschlager, M., Toussaint, F., & Kindermann, S. (2013). A Framework for Extended Persistent Identification of Scientific Assets. *Data Science Journal*, *12*, 10–22. doi:10.2481/dsj.12-036

Recommended Readings

Bekaert, J., & Van de Sompel, H. (2006). *Augmenting Interoperability across Scholarly Repositories*. New York, NY: A meeting sponsored and supported by Microsoft, the Andrew W. Mellon Foundation, the Coalition for Networked Information, the Digital Library Federation, and the Joint Information Systems Committee. Retrieved from http://msc.mellon.org/Meetings/Interop/FinalReport

Sompel, H. V. de. (2012). The Web-Based Scholarly Record: Identification, Persistence, Actionability. *Libraries in the Digital Age (LIDA) Proceedings*, *12*(0). Retrieved from http://ozk.unizd.hr/proceedings/index.php/lida2012/article/view/60

Sompel, H. van de, & Lagoze, C. (2009). All Aboard: Toward a Machine-Friendly Scholarly Communication System. In *The Fourth Paradigm: Data-Intensive Scientific Discovery* (pp. 1–8). Retrieved from http://research.microsoft.com/en-us/collaboration/fourthparadigm/4th_paradigm_book_part4_complete.pdf

Van de Sompel, H., & Beit-Arie, O. (2001). Open linking in the scholarly information environment: Using the OpenURL framework. *D-Lib Magazine*, *10*. Retrieved from http://www.dlib.org/dlib/march01/vandesompel/03vandesompel.html

Van de Sompel, H., Hammond, T., Neylon, E., & Weibel, S. L. (2006). RFC 4452: The "info" URI Scheme for Information Assets with Identifiers in Public Namespaces. *Requests for Comments*. Retrieved from http://www.rfc-archive.org/getrfc.php?rfc=4452

Van de Sompel, H., Lagoze, C., Nelson, M. L., Warner, S., Sanderson, R., & Johnston, P. (2009). Adding eScience Assets to the Data Web. Presented at the Linked Data on the Web, CEUR-WS.

Van de Sompel, H., Lagoze, C., Nelson, M., Warner, S., Sanderson, R., & Johnston, P. (n.d.). Adding eScience Assets to the Data Web. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=pubmed&cmd=Retrieve&dopt=AbstractPlus&list_uids=related:zKJkxCNLPY4J:scholar.google.com/

Van de Sompel, H., Sanderson, R., Klein, M., Nelson, M. L., Haslhofer, B., Warner, S., & Lagoze, C. (2012). A Perspective on Resource Synchronization. *D-Lib*

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

Week 10: June 7: Project presentations

See project assignment for details. We will devote the last class session to a public presentation of student projects and to a general discussion of the data curation issues identified in each project.

Week 11 (Exam Week): June 11: Projects due

Final projects are due by 5 p.m. on Wednesday, June 11.