

LIS 668, Digital Curation

School of Library and Information Studies
University of Wisconsin-Madison
Spring 2014

Dorothea Salo (please call me “Dorothea”)
Office address: 4261 Helen C. White Hall
Course link page: <http://pinboard.in/u:dsalo/t:668>

salo@wisc.edu, 608-265-4733
Office Hours: by appointment
Skype: dorotheasalo

Course Objectives

- Assess, plan for, manage, and execute a small-scale data-management or digital-archiving project.
- Assess digital data for preservability; make yes-or-no accessioning decisions.
- Understand (and where relevant, apply) technological, economic, and social models of digital preservation and sustainability.
- Understand forms, formats, and lifecycles of digital data across a wide breadth of contexts.
- Evaluate software and hardware tools relevant across the data lifecycle.
- Construct a current-awareness strategy; assimilate substantial amounts of relevant writing.
- Self-sufficiently acquire technical knowledge.

This course is designed to assess student progress in the following SLIS program-level outcomes: 1b, 2a, 2b, 3a, 3b, 3d, 4a, and 4b.

Course Policies

I wish to fully include persons with disabilities in this course. Please let me know within two weeks if you require accommodation. I will try to maintain the confidentiality of this information.

Academic Honesty: I follow the academic standards for cheating and plagiarism set forth by the University of Wisconsin.

An explicit goal of this course is self-sufficiency in acquiring knowledge about novel technology. To that end, I will NOT handhold you through every technology we look at. You are expected to exhaust normal information channels before you approach classmates or (especially) me with nuts-and-bolts technology questions.

Readings

There are no required textbooks for this course. Please use Learn@UW for links to all readings; students whose reading is not evident *through Learn@UW* will lose readings-and-participation points from their final grade.

Contacting me

For any difficulty with the course that is not private or confidential, including group-project issues, please use the Learn@UW help forum; *I will not answer such questions by email*. Please also do your best to assist your classmates on the forum. I am not available weekends; otherwise, I do my level best to answer forum questions and email within two business days. If you need to speak with me, please make an appointment with me *directly on WiscCal*, which will email me the appointment information and help ensure I'm not double-booked.

Should you see dead links (it does happen, usually with no notice), weird due dates, or other syllabus problems, please post them to the “Syllabus problems” forum on Learn@UW.

Course week

Our course week, for convenience, runs from Monday to Monday beginning January 21 (so the first week is a day short). All assignments will be due on Mondays by 5 pm CT unless otherwise specified; late assignments will be penalized one final-grade percentage point per day or fraction thereof late. I will allow revision and resubmission at my sole discretion and on my schedule only; any student resistance will remove the opportunity.

Most weeks' readings come with one or more "linklists." These are for your further enrichment and edification; you are NOT expected to read everything on them, because they represent several years of my accumulated reading, far too much to get through in a single semester.

Specialty areas

Your first duty in this course is to choose a "specialty area" which will govern some of your reading/lecture and assignment work through the semester. You are obviously free to learn about other specialty areas if you wish to, but you are not obliged to do so. Please declare your specialty area in the designated Learn@UW forum by Monday of the second week of class (or as soon as possible if you add the class late). If after reading Week 1's assignments you still have questions about which specialty areas best suit your career goals, please ask on the Learn@UW help forum.

Available specialty areas:

- Research-data management
- E-records management/digital asset management (Combining these is an experiment on my part; what I perceive they have in common is the bounded lifespan of managed digital objects.)
- Website preservation
- Personal digital archiving

For purely logistical reasons I cannot allow specialties not listed above; I apologize. If you have a different interest, please do let me know so that I can take it into account in future semesters. I may also be able to throw you some relevant reading!

Optional readings: For a few weeks, I have added readings relevant to specific SLIS interest/affinity groups (e.g. TLAM). All such readings are clearly labeled and *optional*.

Unit 1: Bootstrapping

Week 1: Course overview

Learning objectives: What is digital curation? Digital preservation? What is each specialty area about?

Linklists: <http://pinboard.in/u:dsalo/t:datacuration>, <http://pinboard.in/u:dsalo/t:webarchiving>, <http://pinboard.in/u:dsalo/t:personalarchiving>, <http://pinboard.in/u:dsalo/t:datacuration>

Walters and Skinner, "Digital Curation for Preservation." http://www.arl.org/storage/documents/publications/nrnt_digital_curation17mar11.pdf (pp 5-30 required; rest recommended)

Posner. "Embarrassments of riches: managing research assets." <http://miriamposner.com/blog/?p=982>

Summers, "The web as a preservation medium." <http://inkdroid.org/journal/2013/11/26/the-web-as-a-preservation-medium/>

Prom, "Making digital curation a systematic institutional function." *International Journal of Digital Curation* 6:1 (2011). <http://ijdc.net/index.php/ijdc/article/view/169>

Hunt, "The role of librarians in DAM and your organization." <http://www.slideshare.net/debhunt/the-role-of-librarians-in-dam-and-in-your-organization-createsphere-2012>

Ashenfelder, "Public library activism: Jon Eriksen's personal digital archiving campaign." <http://blogs.loc.gov/digitalpreservation/2013/09/public-library-activism-jon-eriksens-personal-digital-archiving-campaign/>

Week 2: Project management. Working with stakeholders. Keeping current.

Learning objectives: Classical project-management techniques. Agile project-management techniques. Project planning. Dealing with stakeholders. Critical path analysis. Budgeting and cost estimates. Monitoring progress. Running meetings. Common pitfalls. Keeping current.

Linklist: <http://pinboard.in/u:dsalo/t:projectmanagement>, <http://pinboard.in/u:dsalo/t:agile>

Wamsley. "Controlling project chaos: project management for library staff." *PNLA Quarterly* 73:2 (2009). http://www.pnla.org/assets/documents/Quarterly/pnla_winter09.pdf (pp. 5-6, 27)

Leon. "Project management for humanists." *#alt-academy* <http://mediacommons.futureofthebook.org/alt-ac/pieces/project-management-humanists>

Vinopal. "Project portfolio management for academic libraries." <http://crl.acrl.org/content/early/2011/08/26/crl-277.short>

Csaba. "SCRUM: The story of an agile team." <http://net.tutsplus.com/articles/editorials/scrum-the-story-of-an-agile-team/>

Research-data management specialists:

Witt and Carlson. "Conducting a data interview." http://docs.lib.purdue.edu/lib_research/81/
Data Curation Profiles Toolkit. <http://www4.lib.purdue.edu/dcp/> (Please register for the site, download and read all the materials linked from <http://www4.lib.purdue.edu/dcp/download>, and read at least two "Completed Profiles.")

Web archiving and personal digital archiving specialists:

Phillips et al. "The NDSA levels of digital preservation." http://www.digitalpreservation.gov/ndsa/working_groups/documents/NDSA_Levels_Archiving_2013.pdf
"Sustainability health check tool for digital content projects." http://sca.jiscinvolve.org/wp/files/2013/01/sustainability_healthcheck_tool.pdf

E-records/DAM specialists:

Redwine et al. "Born digital: guidance for donors, dealers, and archival repositories." <http://www.clir.org/pubs/reports/pub159/pub159.pdf> (Introduction, sections 1, 3, 4)
Phillips et al. "The NDSA levels of digital preservation." http://www.digitalpreservation.gov/ndsa/working_groups/documents/NDSA_Levels_Archiving_2013.pdf

Unit 2: The data environment

Week 3: Types of data. Data lifecycle models. Repository audit standards.

Learning objectives: OAIS model. DCC data-lifecycle model. Significant properties. Risk assessment, analysis, and mitigation. TRAC, CRL, "trusted digital repository." DRAMBORA. SPOT.

Linklists: <http://pinboard.in/u:dsalo/t:standards> (skim for relevant tags), <http://pinboard.in/u:dsalo/t:ois>
Ockerbloom, John Mark. "What repositories do: the OAIS model." <http://everybodyslibraries.com/2008/10/13/what-repositories-do-the-oais-model/>
Lavoie, Brian. "The Open Archival Information System Reference Model: Introductory Guide." http://www.dpconline.org/docs/lavoie_OAIS.pdf
DRAMBORA Interactive. "DRAMBORA: About." <http://www.repositoryaudit.eu/about/> (Please register for the site and download the entire toolkit to skim it.)
Vermaaten, Lavoie, and Caplan. "Identifying threats to successful digital preservation: the SPOT model for risk assessment." D-Lib Magazine 18:9/10. <http://www.dlib.org/dlib/september12/vermaaten/09vermaaten.html>
CRL. "Trustworthy Repositories Audit & Certification: Criteria and Checklist (TRAC)." http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf (Skim this.)
CRL. "Report on Portico audit findings." <http://www.crl.edu/sites/default/files/attachments/pages/CRL%20Report%20on%20Portico%20Audit%202010.pdf> (Skim this, so you know what an audit report looks like.)
Webb, Pearson, and Koerbin. "Oh, you wanted us to preserve that?!" Statements of preservation intent for the National Library of Australia's digital collections." D-Lib Magazine 19:1/2 (2013) <http://www.dlib.org/dlib/january13/webb/01webb.html>

Research-data management specialists:

DCC Curation Lifecycle Model. <http://www.dcc.ac.uk/sites/default/files/documents/publications/DCCLifecycle.pdf>
Lifecycle Model FAQs <http://www.dcc.ac.uk/resources/curation-lifecycle-model/lifecycle-model-faqs>
NSF. "Digital data collections by categories." http://nsf.gov/pubs/2005/nsb0540/nsb0540_11.pdf

Web archiving specialists:

Bragg and Hanna. "The Web Archiving Life Cycle Model." https://archive-it.org/static/files/archiveit_life_cycle_model.pdf

E-records/DAM specialists:

Lappin, "Why a link between MoReq2010 and the OAIS model would benefit both records managers and archivists." <http://thinkingrecords.co.uk/2012/07/13/why-a-link-between-moreq2010-and-the-oais-model-would-benefit-both-records-managers-and-archivists/>

Personal digital archiving specialists:

Ashenfelder, "When I go away: getting your digital affairs in order." <http://blogs.loc.gov/digitalpreservation/2011/07/when-i-go-away-getting-your-digital-affairs-in-order/>

Week 4: Sustainability and economic models

Learning objectives: Macro-economics of digital preservation. Perils of grant funding. Perils of governmental funding. Perils of institutional funding.

Linklist: <http://pinboard.in/u:dsalo/t:sustainability>

“Sustainable Economics for a Digital Planet.” http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf

Ithaka. “Sustaining our digital future: institutional strategies for digital content.” <http://www.sr.ithaka.org/research-publications/sustaining-our-digital-future>

Bretz, Brown, and McGregor. “Lasting change.” <http://www.cwrc.ca/wp-content/uploads/2010/12/Lasting-Change-Knowledge-Synthesis.pdf>

DPC Online. “Digital preservation business case toolkit.” <http://wiki.dpconline.org/index.php?>

title=Digital_Preservation_Business_Case_Toolkit (click through the whole thing; read at least one case study)

Research-data management specialists:

Timmer, John. “How science funding is putting scientific data at risk.” <http://arstechnica.com/science/news/2010/10/how-science-funding-is-putting-scientific-data-at-risk.ars>

Goldstein, Serge J., and M. Ratliff. “DataSpace: a funding and operational model.” <http://arks.princeton.edu/ark:/88435/dsp01w6634361k>

Wilson et al. “Developing infrastructure for research data management at the University of Oxford.” *Ariadne* 65 (2010). <http://www.ariadne.ac.uk/issue65/wilson-et-al/>

E-records/DAM specialists:

Rein, “Costs soaring for Archives’ digital library, auditors say.” <http://www.washingtonpost.com/wp-dyn/content/article/2011/02/04/AR2011020402286.html>

Week 5: The legal and regulatory environment. Data ethics.

Learning objectives: Open movements (open source, open access, open data, open government data, open notebook science). Copyright and digital preservation. Reuse rights. The dangers of “non-commercial” and “share-alike” licenses. Human-subjects research and data confidentiality. Dangers to privacy in digital-data retention, reidentification, and other data ethics issues.

Linklists: <http://pinboard.in/u:dsalo/t:copyright>, <http://pinboard.in/u:dsalo/t:dataethics>, <http://pinboard.in/u:dsalo/t:datareuse>, <http://pinboard.in/u:dsalo/t:creativecommons>

Salo, “Battle of the opens.” <http://scientopia.org/blogs/bookoftrogool/2010/03/15/battle-of-the-opens/>

“International study on the impact of copyright law on digital preservation.” http://www.digitalpreservation.gov/documents/digital_preservation_final_report2008.pdf (Part 1 and Part 5; skim the rest)

Madison, “Knowledge curation.” http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1848086

Fleishman, “The sound of silence in the National Library.” <http://boingboing.net/2013/01/09/soundofsilence.html>

Richards and King, “Three paradoxes of big data.” http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2325537

Research-data management specialists:

NIH. “Frequently asked questions about the NIH Public Access Policy.” <http://publicaccess.nih.gov/FAQ.htm>

NSF. “Dissemination and sharing of research results.” http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/aag_6.jsp#VID4

NSF. “Dissemination and sharing of research results.” <http://www.nsf.gov/bfa/dias/policy/dmp.jsp> (Please skim all directorates’ guidance. Pay special attention to guidance in any area where you have disciplinary expertise.)

Panton Principles. <http://pantonprinciples.org/> and <http://pantonprinciples.org/faq/>

Web archiving specialists:

Grotke, “Legal issues in web archiving.” <http://blogs.loc.gov/digitalpreservation/2012/05/legal-issues-in-web-archiving/>

Anderson, “Copy some webpages, owe more than the national debt.” <http://boingboing.net/2013/01/09/soundofsilence.html>

Personal digital archiving specialists:

Talbot, “Data discrimination means the poor may experience a different Internet.” <http://www.technologyreview.com/news/520131/data-discrimination-means-the-poor-may-experience-a-different-internet/>

Leber, “A new underclass.” <http://www.fastcoexist.com/3017102/a-new-underclass-the-people-who-big-data-leaves-behind?partner=rss&asid=b886252e>

E-records/DAM specialists:

Collins, “Uncharted territory.” <http://www.wired.com/wiredscience/2013/08/power-of-amateur-cartographers/>

Burn-Murdoch, “The problem with algorithms.” <http://www.theguardian.com/news/datablog/2013/aug/14/problem-with-algorithms-magnifying-misbehaviour>

Robertson, “Your medical records are for sale.” <http://www.businessweek.com/articles/2013-08-08/your-medical-records-are-for-sale>

OPTIONAL for those interested in TLAM:

WIPO, "Intellectual property and the safeguarding of traditional cultures." http://www.wipo.int/export/sites/www/freepublications/en/tk/1023/wipo_pub_1023.pdf

Christen and Merrill, "Open to the public: indigenous collections and the ethics of openness." <https://archive.org/details/OpenToThePublicIndigenousCollectionsAndTheEthicsOfOpenness>

Week 6: Data practices and needs

Learning objectives: Attitudes toward digital preservation. Attitudes toward data sharing, and how they differ across disciplines. Personal digital archiving, and how practices bleed into other environments. Attitudes toward librarians and archivists. Data security. Crowdsourcing data transcription and analysis.

Linklists: <http://pinboard.in/u:dsalo/t:rschbehavior>, <http://pinboard.in/u:dsalo/t:datapactices>, <http://pinboard.in/u:dsalo/t:horrorstories>, <http://pinboard.in/u:dsalo/t:datapublishing>, <http://pinboard.in/u:dsalo/t:datacitation>

Raloff. "Galaxy Zoo's blue mystery." ScienceNews. http://www.sciencenews.org/view/generic/id/33403/title/Science_%2B_the_Public_Galaxy_Zoos_blue_mystery_%28part_1%29 and http://www.sciencenews.org/view/generic/id/33436/title/Galaxy_Zoos_blue_mystery_%28part_2%29

Thomas, Jones, and Witmer. "History Harvests: what happens when students collect and digitize the people's history?" http://www.historians.org/Perspectives/issues/2013/1301/History-Harvests_Students-Digitize-Peoples-History.cfm

Marshall, Catherine C. "Rethinking personal digital archiving." <http://www.dlib.org/dlib/march08/marshall/03marshall-pt1.html> and <http://www.dlib.org/dlib/march08/marshall/03marshall-pt2.html>

Asher and Jahnke, "Curating the ethnographic moment." <http://www.archivejournal.net/issue/3/archives-remixed/curating-the-ethnographic-moment/> (Compare with Marshall!)

Web archiving specialists:

Yoon, "Defining what matters when preserving web-based personal digital collections." <http://www.ijdc.net/index.php/ijdc/article/view/8.1.173>

Personal digital archiving specialists:

Yoon, "Defining what matters when preserving web-based personal digital collections." <http://www.ijdc.net/index.php/ijdc/article/view/8.1.173>

"Perspectives on personal digital archiving." http://www.digitalpreservation.gov/documents/ebookpdf_march18.pdf?loclr=blogsig pp. 32-57

Research data management specialists:

Timmer, John "Jaz drives, spiral notebooks, and SCSI: how we lose scientific data." <http://arstechnica.com/science/news/2010/11/preserving-science-how-data-gets-lost.ars>

Feijen, Martin. "What researchers want." http://www.surffoundation.nl/nl/publicaties/Documents/What_researchers_want.pdf

Brown, C. Titus. "My data management plan -- a satire." <http://ivory.idyll.org/blog/may-10/data-management.html>

"ViDaaS researcher requirements report." <http://vidaas.oucs.ox.ac.uk/docs/VIDaaS%20Researcher%20Requirements%20Report.pdf>

E-records/DAM specialists:

Gallagher, "For some government agencies, it's only official if it's on floppies." <http://arstechnica.com/information-technology/2013/12/for-some-government-agencies-its-only-official-if-its-on-floppies/>

Unit 3: The service environment

Week 7: Library and archive preparedness.

Learning objectives: Staffing models in libraries and archives. Job opportunities in digital curation. Embedded librarianship. Starting a brand-new digital-curation service. Digital preservation needs and strategies in public libraries. Infrastructure. Funding (grant earmarks, budget and position shifting). Service outreach and marketing.

Kennedy and McGovern, "The five organizational stages of digital preservation." <http://quod.lib.umich.edu/cgi/t/text/text-idx?c=spobooks;idno=bbv9812.0001.001;rgn=div1;view=text;cc=spobooks;node=bbv9812.0001.001%3A11>

"Digital preservation: a role for public libraries." <http://agogified.com/1075>

NDSA. "Staffing for effective digital preservation." <http://digitalpreservation.gov/ndsas/documents/NDSA-Staffing-Survey-Report-Final122013.pdf?loclr=blogsig>

Meyer, "Safeguarding collections." <http://www.arl.org/bm~doc/safeguarding-collections.pdf>

Research data management specialists:

Newton, Mark P., C. C. Miller, and Marianne Stowell Bracke. "Librarian roles in institutional repository data set collecting." *Collection Management* 36:1 (2011). <http://dx.doi.org/10.1080/01462679.2011.530546>

Salo, "Retooling libraries for the data challenge." *Ariadne* 64 (2010). <http://www.ariadne.ac.uk/issue64/salo/>

Westra, "Data services for the sciences: a needs assessment." *Ariadne* 64 (2010). <http://www.ariadne.ac.uk/issue64/westra/>

E-records/DAM specialists:

"The role of librarians in DAM and in your organization." http://www.slideshare.net/debhunt/the-role-of-librarians-in-dam-and-in-your-organization-createsphere-2012?from_search=57

Week 8: Data literacy and data/digital-preservation training.

Learning objectives: What is data literacy? Doing data-management training. Curricula. Learners' incoming skill. Data literacy and information literacy. Personal digital archiving training.

Watch two or three of: Team Digital Preservation videos, <http://www.youtube.com/user/wepreserve>

Davies, "Data literacy: educational opportunities." <http://www.opendataimpacts.net/2011/03/data-literacy-educational-opportunities/>

Salo, Dorothea. "What I learned from running data-management bootcamps." (See Learn@UW.)

Research data management specialists:

Watch: reBIND video, http://rebind.bgbm.org/rebind_movie

Molloy and Snow. "The data management skills support initiative: Synthesising postgraduate training in research data management." <http://www.ijdc.net/index.php/ijdc/article/view/233>

Personal digital archiving specialists:

Digital Doghouse, "Someone help." <http://thedoghousesdiaries.com/5170>

"Perspectives on personal digital archiving." http://www.digitalpreservation.gov/documents/ebookpdf_march18.pdf?loclr=blogsig pp. 58-77

Ashenfelder, "Public library activism: Jon Eriksen's personal digital archiving campaign." <http://blogs.loc.gov/digitalpreservation/2013/09/public-library-activism-jon-eriksens-personal-digital-archiving-campaign/>

Web archiving specialists:

Ashenfelder, "Personal archiving in the cloud." <http://blogs.loc.gov/digitalpreservation/2011/06/personal-archiving-in-the-cloud/>

Fallows, "Hacked!" http://m.theatlantic.com/magazine/archive/2011/11/hacked/8673/?single_page=true

SPRING BREAK: enjoy!

Week 9: Assessing, collecting, publishing, licensing, and citing digital materials

Learning objectives: Data assessment. Gauging importance and preservability. Collection-development policies. Publishing digital collections. Reuse and citation of digital collections in publications. Licensing for (and against) reuse.

Linklists: <http://pinboard.in/u:dsalo/t:datacitation>, <http://pinboard.in/u:dsalo/t:datapublishing>

Skinner and Schultz, "Preserving Our Collections, Preserving Our Missions." http://www.metaarchive.org/sites/default/files/GDDP_Educopia.pdf (pp. 1-9)

Conway and Proffitt, "Taking stock and making hay: archival collections assessment." <http://www.oclc.org/content/dam/research/publications/library/2011/2011-07.pdf> (While written for analog collections, much of this applies to born-digital and neglected digitized collections as well.)

Dempsey. "How to cite GIS materials." <http://www.gislounge.com/how-to-cite-gis-materials/>

Hilton and Thompson. "Further experiences in collecting born digital archives at the Wellcome Library." *Ariadne* 53 (2007). <http://www.ariadne.ac.uk/issue53/hilton-thompson>

Hagedorn et al. "Creative Commons licenses and the non-commercial condition." http://www.pensoft.net/J_FILES/1/articles/2189/2189-G-1-layout.htm (Read for the reuse analysis of various CC licenses, which is true for far more than biodiversity information.)

Research data management specialists:

Timmer. "Preserving science: what data do we keep?" <http://arstechnica.com/science/news/2010/11/preserving-science-choosing-what-data-to-discard.ars>

Strasser. "The skinny on data publication." <http://datapub.cdlib.org/?p=301>

Penev. “Data paper’: the data publishing project of Pensoft.” <http://www.openaire.eu/en/component/content/article/76/399>

DataCite. “What do we do?” and “Why cite data?” <http://datacite.org/whatwedo> and <http://www.datacite.org/whycitedata>

Callaghan et al. “Data citation and publication by NERC’s Environmental Data Centres.” *Ariadne* 68. <http://www.ariadne.ac.uk/issue68/callaghan-et-al>

Web archiving specialists:

“Linkrot across the disciplines and a new library-driven solution.” <http://librarypublishing.wordpress.com/2013/10/21/linkrot-across-the-disciplines-and-a-new-library-driven-solution/> (Read this for how they developed this project’s scope.)

Weissman, “Scoping a web archiving project around *The Hobbit*.” <http://blogs.loc.gov/digitalpreservation/2013/01/there-and-back-again-scoping-a-web-archiving-project-around-the-hobbit/>

Lazorchak, “Web archive preservation planning.” <http://blogs.loc.gov/digitalpreservation/2011/08/web-archive-preservation-planning/>

Personal digital archiving specialists:

“Personal archiving.” <http://www.digitalpreservation.gov/personalarchiving/email.html?loclr=blogsig>

E-records/DAM specialists:

Pinsent, “The AIDA toolkit.” http://www.slideshare.net/SteveHitchcock/the-aida-toolkit-assessing-institutional-digital-assets-by-ed-pinsent?from_search=35

“AIDA self-assessment toolkit Mark II.” http://aida.jiscinvolve.org/wp/files/2009/05/self_assessment_toolkit_ii.doc and scorecard at http://aida.jiscinvolve.org/wp/files/2009/05/weighted_scorecard_example.xls

OPTIONAL for those interested in TLAM:

“Local Contexts.” <http://www.localcontexts.org/>

Week 10: Discovery of digital collections

Learning objectives: Finding digital collections, open and subscription. Licensing issues with data archives. Digital collections, the ILS, and discovery layers. Google and digital collections; search-engine optimization. Digital divides and digital collections.

Linklists: <http://pinboard.in/u:dsalo/t:datause>, <http://pinboard.in/u:dsalo/t:datasources>, <http://pinboard.in/u:dsalo/t:datasharing>

Partlo, “The pedagogical data reference interview.” <http://iassistdata.org/iq/pedagogical-data-reference-interview>

Carleton College. “Data, Datasets, and Statistical Resources.” <http://gouldguides.carleton.edu/content.php?pid=65030&sid=480389> (please look through all the tabs)

Dempsey, “Discovery vs. discoverability.” <http://orweblog.oclc.org/archives/002206.html>

Dempsey, “Thirteen ways of looking at libraries, discovery, and the catalog.” <http://www.educause.edu/ero/article/thirteen-ways-looking-libraries-discovery-and-catalog-scale-workflow-attention>

Chapman, “Evaluating the effectiveness of manual metadata enhancements for digital images.” <https://staff.lib.ncsu.edu/confluence/display/MNC/Evaluating+the+effectiveness+of+manual+metadata+enhancements+for+digital+images>

Lichtenstein. “Why open data alone is not enough.” http://m.wired.com/magazine/2011/06/st_essay_datafireworks/

Research data management specialists:

Berman, “We need a research data census.” *Communications of the ACM* 53:12 (2010). <http://dx.doi.org/10.1145/1859204.1859220>

Strasser. “Thanks in advance for sharing your data.” <http://datapub.cdlib.org/?p=1297>

Borgman. “The conundrum of sharing research data.” <http://works.bepress.com/borgman/244/>

Hogenboom, Teper, and Wiley. “Collecting small data.” <http://publications.arl.org/1h7vog.pdf>

“About Databib.” <http://databib.org/about.php>

Unit 3: Tools of the trade

Week 11: File formats and their sustainability.

Learning objectives: Evaluating file formats for preservation. Lossy vs. lossless formats. Open vs. proprietary formats.

“Preservation copy,” “digital surrogate.” Image formats (JPEG, TIFF, JPEG 2000, PNG, GIF, RAW). Audio and video formats

(codecs, sampling rate/bitrate, WAV, AIFF, mp3, MPEG4). Specialized file formats. File-format auditing tools. Migration and emulation.

Linklists: <http://pinboard.in/u:dsalo/t:fileformats>, <http://pinboard.in/u:dsalo/t:webarchiving>, <http://pinboard.in/u:dsalo/t:audio>, <http://pinboard.in/u:dsalo/t:video>

Timmer, John. "Changing software, hardware a nightmare for tracking scientific data." <http://arstechnica.com/science/news/2010/11/changing-software-hardware-a-nightmare-for-tracking-scientific-data.ars>

ICPSR, "Digital Preservation Tutorial," section 3 "Obsolescence": "File Formats and Software" and "Hardware and media" http://www.icpsr.umich.edu/dpm/dpm-eng/eng_index.html

Cornell, "Digital Imaging Tutorial." <http://www.library.cornell.edu/preservation/tutorial/contents.html> (Skim.)

Rutgers, Video Object Standards Analysis, http://rucore.libraries.rutgers.edu/collab/ref/dos_avwg_video_obj_standard.pdf

Rutgers, Audio Object Standards Analysis, http://rucore.libraries.rutgers.edu/collab/ref/dos_avwg_audio_obj_standard.pdf

Pilgrim, Mark. "Video on the web." <http://diveintohtml5.ep.io/video.html>

"Converting audio cassette tapes to CD, MP3, and other digital formats." <http://www.andybrain.com/archive/convert-cassette-to-cd-digital.htm>

File Information Tool Set. <http://code.google.com/p/fits/wiki/tools>

"MUPPET: MUlti Pass file Properties Extraction Tool." <http://www.openplanetsfoundation.org/blogs/2011-10-28-muppet-multi-pass-file-properties-extraction-tool>

Research-data management specialists:

"Why HDF?" http://www.hdfgroup.org/why_hdf/

Web archiving specialists:

Farrell, Susan ed. "A guide to web preservation." <http://jiscpowr.jiscinvolve.org/wp/files/2010/06/Guide-2010-final.pdf>

Personal digital archiving specialists:

"Guidelines for creators of personal archives." <http://www.paradigm.ac.uk/workbook/appendices/guidelines-tips.html> (Section 3 especially.)

E-records/DAM specialists:

Lazorchak, "The how of email archiving." <http://blogs.loc.gov/digitalpreservation/2013/07/the-how-of-email-archiving-more-launching-points-for-applied-research/> (Click links!)

Week 12: Metadata

Learning objectives: Descriptive, technical, administrative, and structural metadata. Crosswalking and other metadata tools (Google Refine, DataUp). Ontologies. Explaining metadata to non-librarians. Getting metadata from non-librarians. Coping with spreadsheets. Identifiers and their importance.

Qin, Ball, and Greenberg. "Functional and architectural requirements for metadata: supporting discovery and management of scientific data." <http://www.slideshare.net/jqin/functional-requirements-14181727> (ponder slide 24 with special attention!)

Wilson, Andrew. "How much is enough: metadata for preserving digital data." *Journal of Library Metadata* 10:2 (2010). <http://dx.doi.org/10.1080/19386389.2010.506395>

Riley, "Seeing Standards." <http://www.dlib.indiana.edu/~jenlrile/metadatamap/> (Download the poster and read the legend and definitions carefully.)

Kennedy, "Nine questions to guide you in choosing a metadata schema." <https://journals.tdl.org/jodi/article/viewArticle/226/205>

Strasser. "Ontologies and data." <http://datapub.cdlib.org/?p=341>

Nguyen. "Using Google Refine to clean messy data." <http://www.propublica.org/nerds/item/using-google-refine-for-data-cleaning>

Hickey, "VIAF and other IDs." <http://outgoing.typepad.com/outgoing/2011/07/viaf-and-other-ids.html>

Week 13: Digital forensics

Learning objectives: Digital forensics and related tools and techniques. Remanence. Recovering data from old physical media. Recovering data from hard drives. Recovering "deleted" data. Ethics of digital forensics.

Linklists: <http://pinboard.in/u:dsalo/t:digitalforensics>, <http://pinboard.in/u:dsalo/t:recordsmgmt>

Ross, Seamus. "Preservation pressure points." <http://www.repositoryaudit.eu/images/PreservationPressurePoints.pdf>

Ross, Seamus, and Ann Gow. "Digital archaeology: rescuing neglected and damaged data resources." <http://eprints.erpanet.org/47/>

Kirschenbaum, Matthew G., Richard Ovenden, and Gabriela Redwine. "Digital forensics and born-digital content in cultural heritage collections." <http://www.clir.org/pubs/reports/pub149/pub149.pdf>

Johnston, "Digital forensics and digital preservation." <http://blogs.loc.gov/digitalpreservation/2011/06/digital-forensics-and-digital-preservation/>

Rashid, "SSDs harder to securely purge of data than HDDs." <http://www.eweek.com/c/a/Data-Storage/SSDs-Harder-to-Securely-Purge-of-Data-than-HDDs-200129/>

"BitCurator." <http://wiki.bitcurator.net/index.php?title=Description>

E-records/DAM specialists:

"Digital records preservation: where to start guide." <http://isotc.iso.org/livelink/livelink?func=ll&objId=10083866&objAction=Open&nexturl=%2F%2Flivelink%2F%2Flivelink%3Ffunc%3D%2F%2F%3D%2F%2F%26objAction%3Dbrowse%26sort%3Dname>

Briston, Heather, and Karen Estlund. "From passive to active preservation of electronic records." *Ariadne* 65 (2010). <http://www.ariadne.ac.uk/issue65/briston-estlund/>

Week 14: Storage, backup, and replication

Learning objectives: Types of storage and backup hardware and filesystems; their relative strengths and weaknesses. MBTF, why it matters, and how to work around it. Cost models for storage. Storage and backup architectures. Cloud storage for digital curation; cloud storage risks. Methods to manage geographic replication.

Buchanan. "Everything you need to know about hard drives." <http://gizmodo.com/5106129/giz-explains-everything-you-need-to-know-about-hard-drives>

Salter, "Bitrot and atomic COWs: inside next-gen filesystems." <http://arstechnica.com/information-technology/2014/01/bitrot-and-atomic-cows-inside-next-gen-filesystems/>

Gallagher. "The great disk drive in the sky: how web giants store big -- and we mean big -- data." <http://arstechnica.com/business/2012/01/the-big-disk-drive-in-the-sky-how-the-giants-of-the-web-store-big-data/>

Storer. "Secure, energy-efficient, evolvable, long-term archival storage." <http://www.ssrc.ucsc.edu/pub/storer09-tr0901.html>

Murray, Peter. "Options in storage for digital preservation." <http://dltj.org/article/preservation-storage-options/> (following links strongly encouraged)

Aitken et al., "Digital curation and the cloud." <http://eprints.gla.ac.uk/60659/>

Rosenthal, "Estimating storage costs." <http://blog.dshr.org/2013/11/estimating-storage-costs.html#more>

Rodrigues, "Rackspace, OpenStack, and evolving cloud standards." <http://www.techrepublic.com/blog/the-enterprise-cloud/rackspace-openstack-and-evolving-cloud-standards/5531/?tag=nl.e550>

Butler, "Report: some cloud providers have 'dirty disks.'" <http://www.networkworld.com/news/2012/042612-cloud-dirty-disks-258731.html>

E-records/DAM specialists:

NARA. "NARA Bulletin 2010-05." <http://www.archives.gov/records-mgmt/bulletins/2010/2010-05.html> (especially section 6 and onward)

Week 15: Hardware and software platforms for digital archival and preservation

Learning objectives: Hardware and its durability. Institutional repository platforms (DSpace, EPrints, Fedora, BePress Digital Commons, CONTENTdm). Digital-library platforms (Greenstone, ContentDM, Omeka). Born-digital assessment and archiving tools (BitCurator, Archivematica). Other relevant platforms. Curation microservices. Organizing files; choosing filenames. Versioning. Geographic dispersal techniques (LOCKSS, cloud storage, DuraCloud).

Linklist: <http://pinboard.in/u:dsalo/t:software/t:668>

Kaplan, "Choosing a digital asset management system that's right for you." http://suberic.net/~deborah.kaplan/joao.7_1.kaplan.pdf

"About LOCKSS." http://www.lockss.org/lockss/About_LOCKSS

"Top reasons to use DSpace." <http://www.dspace.org/why-use> (read skeptically!)

EPrints. <http://www.eprints.org/software/>

"Getting started with Fedora." <https://wiki.duraspace.org/display/FCR30/Getting+Started+with+Fedora>

"About Islandora." <http://islandora.ca/about> and http://islandora.ca/solution_packs

"Advantages of Digital Commons." <http://www.bepress.com/ir/advantages.html>

"CONTENTdm overview." <http://www.oclc.org/contentdm/overview/default.htm>

"About Greenstone." <http://www.greenstone.org/>

“Omeka.” <http://omeka.org/> (click around a bit)

“iRODS Overview.” https://www.irods.org/pubs/iRODS_Overview_0903.pdf

“Tranche Project.” <https://trancheproject.org/>

“HubZero.” <http://hubzero.org/>

DuraCloud. “Introduction.” <https://wiki.duraspace.org/display/duracloud/DuraCloud> (please read Features and Services also)

“Curation micro-services.” <http://www.cdlib.org/services/uc3/curation/> (follow links, please)

“Merritt: An emergent micro-services approach to digital curation infrastructure.” <https://confluence.ucop.edu/download/attachments/13860983/Merritt-latest.pdf>

“Archivematica.” <http://archivematica.org/> and <http://archivematica.org/wiki/index.php?title=Overview>

Assignments

Assignments	Percentage	Due Date
Specialty area chosen	n/a	27 January
Individual assignments:		
Horror-story analysis	5%	27 January
DMP/preservation-plan evaluation	10%	14 April
Compare/contrast data curation guides/curricula	10%	17 March
Tool/service review	10%	12 May
Final project		
Project and PM chosen	n/a	27 January
Problem statement/Project plan/Work agreement	15%	17 February
Mid-semester check-in	10%	17 March
Project poster or video	10%	5 May
Final project report	10%	12 May
360 evaluation	10%	16 May
Readings and participation	10%	Throughout

Final grade scale: 100-93.5 A; 93.4-89.5 AB; 89.4-83.5 B; 83.4-79.5 BC; 79.4-73.5 C, 73.4-69.5 D, 69.4 or below F.

No extra credit opportunities are available in this class. No assignment grades are dropped. Any student failing entirely to turn in an individual assignment listed above will automatically fail the course. Particularly with the group project, total perfection is not the goal; learning is. Mistakes and retrenchments are to be expected, and usually will not count against your grade.

FINAL PROJECT

For your final project, you will work in a group to help solve a digital-curation problem. You will determine the nature and extent of the problem, make a plan to solve it, agree with your client and me about how much of the problem your group can solve over the course of the semester, and work to the resulting schedule.

Your group should immediately select a Project Manager. The PM is responsible for all communications about the project to me *and to the client*. (The PM may use discretion about including other group members. One exception: all group members should participate in the data interview, remotely if necessary!) The PM is also responsible for keeping the group “on time and under budget.” S/he may come to me at any time with concerns about group progress or group dynamics. Other group members with concerns should approach the PM first for resolution. PM and group are responsible for ensuring that the PM is not overloaded. (The PM doing the entire group project is a failure, not a success!)

At the end of the semester, everyone will post to a locked Learn@UW forum a short “360 evaluation” of the other members of their group: a brief description of the contributions of all other group members, including the PM. I will use this information to assign individual project-participation grades as I see fit; only I will see the posts. I will also check with clients about your group’s professionalism, competence, and accomplishments before I assign grades.

Project components:

- Data-curation profile and project plan: This should be a plan through which you and the client will solve the *entire problem* presented to you. Don’t worry; you will not necessarily be expected to complete the entire plan in one short semester! To construct this plan, you should approach the client to perform a data interview. A data-curation profile

(or similar assessment) should form part of this plan as well. You may revise this plan over the course of the semester, as you work through it and learn more! (This is normal in the real world.)

- Work agreement: Present the client with a semester-long work schedule that also indicates the deliverables your group pledges to complete by semester's-end. Turn this in to the appropriate Learn@UW dropbox along with the project plan; the PM should also send it to the client.
- Midterm check-in/agreement revision: Mid-semester, the PM should gauge where the group is and assess the likelihood of completing the promised work agreement. Renegotiating the agreement with the client is acceptable, as is rewriting the plan! Turn in a progress report and the revised agreement and/or plan (if any) to the appropriate Learn@UW dropbox.
- Project marketing: You have two options: see below for details.
- Project wrapup: A *brief* (six pages is too many; two might be enough) statement of the problem presented, the nature of the solutions suggested and deployed, the progress made over the semester, and any larger issues brought to light during the process. If you have revised your client's project plan, please provide a copy of the revised plan as well (this does not count against your pagecount).

Project marketing options:

- Produce a conference-quality poster about your project; your imaginary target conference might be ACRL, WAAL (which has a student poster prize!), or a discipline-appropriate conference (consult with your client; perhaps you'll arrange to present at a real conference) and post it to the appropriate Learn@UW dropbox as a PDF.
- Produce a short (less than five minutes, please!) but high-quality (music, editing, pleasant narration, interesting images) video explaining the project to a lay audience. By all means consult with your client about how they might use the video (e.g. fundraising, crowdsourcing, outreach), and do your best to make it useful for them. Post it to the online video streaming service of your choice, and put a link in the Learn@UW dropbox (memo field) along with the video file (.mp4, .mov, or .wmv preferred).

Grading rubric:

- Client problem profile: understanding and using the chosen interview instrument well
- Project plan: understanding of the problem, appropriateness of suggested solutions, clear expression
- Check-in: evidence of good project management, work getting done, clear expression
- Poster/video: attractive and professional design, clear expression, public-worthy
- Wrapup: accomplishment, overcoming obstacles, professional relationship with client, clear expression
- Individual contribution: per 360 evaluation

On group projects: The idea that group projects are uniquely designed to torture library-school students is a snare and a delusion. Librarianship generally and data curation specifically include immense amounts of collaborative work, from local committees and task forces to involvement in national professional organizations and everything in between. None of the obstacles to working in groups – scheduling, free riders, personality conflicts, geography – disappears when you receive your degree. If you are not good at working in a team, now is the time to learn!

OTHER ASSIGNMENTS

Horror-story analysis

Choose a horror story *relevant to your chosen specialty* to report on from my linklist at <http://pinboard.in/u:dsalo/t:horrorstories>. (Lost/stolen computers are off limits as too simplistic.) Please try to address a story that no one else has posted about yet. Write a substantive post in the designated Learn@UW forum addressing the following:

- What are/were the digital materials at issue?
- What went wrong (or could go wrong) with the data management?
- What are the consequences of the poor data-handling to the people responsible for the data?
- What are the consequences of the poor data-handling for users or potential users?
- Could this problem have been avoided? How?

You are expected to read all your colleagues' posts (and my feedback to them) within one week of the due date for posting, and you are encouraged to react substantively to your colleagues' stories as well. I will check your reading via Learn@UW reports, and you will lose two full percentage points on this assignment if you do not read.

Digital-curation plan response

Real-world examples (anonymized!) of digital-curation plans will be posted to Learn@UW. You are the librarian contacted for an evaluation of the plan. Read and evaluate the plan according to available guidance and your own sense of what is necessary, come up with no more than three recommendations for improvement (yes, you must prioritize! huge infodumps will be penalized one point) and at least one apropos and specific suggestion for further reading. Post your evaluation to the designated Learn@UW forum, writing as though you were emailing it to the sender. (This forum will be kept dark until the due date has passed; you won't be able to see your own post. Don't panic—nobody in any of my classes has lost a forum post yet—but keeping a backup copy might be wise.) *Research-data specialists*: your plan will be an NSF data-management plan. You are responsible for guessing the applicable NSF division or directorate and employing its specific guidance.

Grading rubric: did you make appropriate, well-prioritized recommendations? was your expression cordial, professional, and above all *brief*?

Compare/contrast curation guides

I will provide a list of digital-curation guides and curricula on Learn@UW. Choose two (of the same type, please) *relevant to your chosen specialty*, and compare them. Criteria may include (but are not limited to):

- intended audience
- suitability for that audience given what you know or can find out about existing curation knowledge therein
- coverage, gaps, appropriateness of topic choice
- usefulness and appropriateness of provided readings and learning objects, if any
- ease of use (how much work would an instructor have to do to teach from this? is it good for ready reference?)
- quality and usefulness of suggested assignments/tasks (where applicable)

Conclude by indicating which you would prefer to teach from or recommend to a reference-desk patron. If this assignment takes you more than four double-spaced pages, you are spending too much time on it. Grading rubric: have you decided what the audience needs to know? how and how well these curricula address that? did you read these well enough to teach or do reference from one?

Tool/service review

Choose a tool or service *appropriate to your chosen specialty* from Dorothea's linklist at <http://pinboard.in/u:dsa1o/t:tools> or another available online list (some such lists will be suggested on Learn@UW). Using the tool reviews at Chris Prom's Practical E-Records weblog (<http://e-records.chrisprom.com/tag/software-evaluation>) as examples, write a review of the tool or service and post it to the designated forum in Learn@UW (as plain text/HTML, not a Word or PDF attachment). Address the following parameters (as relevant and appropriate):

- Tool/service's intended purpose, and fitness for that purpose or others
- Tool/service's intended userbase, and fitness for that userbase or others
- Features (what the tool/service does, what problem(s) it solves) and limitations (what it does not do), considered from the perspective of the full lifecycle of digital materials in your specialty
- State of development (research project? alpha/beta? mature and usable? no longer under development?) and future prospects (would you entrust a years- or decades-long project to this?)
- Cost (money, staff, time)
- Software/hardware prerequisites (what do you need before you can use this tool or service?)
- Ease of installation (you may rely on third-party reports for this)
- Ease of use (you may rely on third-party reports for this)

Grading rubric: did you address all relevant parameters? would your review be useful to a potential tool adopter?

SLIS Goals	668 Objectives	668 Measurable Outcomes (see assignment rubrics)
1b. Students apply key concepts with respect to theories and practices of literacies, reading, and information use.	Understand (and where relevant, apply) technological, economic, and social models of digital preservation. Understand forms, formats, and lifecycles of digital data across a wide breadth of contexts.	Horror-story analysis (including systemic failures). Compare/contrast digital-curation guides.
2a. Students evaluate and debate information policy and ethics applicable in local, national, or global contexts.	Understand (and where relevant, apply) technological, economic, and social models of digital preservation.	Horror-story analysis (including data fraud, human-subjects ethics violations, security problems).
2b. Students apply core ethical principles to professional practice.	Assess, plan for, manage, and execute a small-scale data-management or digital-archiving project.	Final project: assessing rights and privacy issues, considering crowdsourcing, deciding about open access/open data.
3a. Students organize and describe print and digital information resources.	Assess, plan for, manage, and execute a small-scale data-management or digital-archiving project.	Final project invariably includes metadata consulting, creation, and management.
3b. Students search, select, and evaluate print and digital information resources.	Understand forms, formats, and lifecycles of digital data across a wide breadth of contexts.	Curation-guide assignment
3d. Students understand and use appropriate information technologies.	Assess, plan for, manage, and execute a small-scale data-management or digital-archiving project.	Final project: helping clients choose appropriate preservation and access modalities.
4a. Students participate effectively as team members to solve problems.	Assess, plan for, manage, and execute a small-scale data-management or digital-archiving project.	Final project: 360 evaluation.
4b. Students demonstrate good oral and written communication skills.	Assess, plan for, manage, and execute a small-scale data-management or digital-archiving project.	Final project poster/video.