

# LIS 855, Digital Curation

School of Library and Information Studies  
University of Wisconsin-Madison  
Spring 2013: Online

Dorothea Salo (please call me “Dorothea”)  
Office address: 4261 Helen C. White Hall  
Course link page: <http://pinboard.in/u:dsalo/t:855>

[salo@wisc.edu](mailto:salo@wisc.edu), 608-265-4733  
Office Hours: by appointment  
Skype: dorotheasalo



To the extent possible under law, the person who associated CCo with this work has waived all copyright and related or neighboring rights to this work. This work is published from: United States.

## 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.
- Understand research-data forms and formats and research-data lifecycles across scholarly disciplines.
- Evaluate software 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: 1a, 1b, 2a, 2b, 3b, 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 Fridays or 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 22 (so the first week is a day short). All assignments will be due on Mondays by 5 pm CT; 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 come with one or more “linklists.” These are for your further enrichment and edification; you are NOT expected to read everything on them, because that would be insane!

## Unit 1: Bootstrapping

### Week 1: Course overview

*Learning objectives: What is research data, and how does it fit into so-called “cyberinfrastructure” or “e-research”? Why is data suddenly getting attention? What is data curation? How is it like and unlike informatics? Data science?*

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

NYU. “Data sharing and management snafu in three short acts.” <http://www.youtube.com/watch?v=N2zK3sAtr-4>

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

Swan, Alma. “Skills, role & career structure of data scientists & curators.” <http://www.jisc.ac.uk/publications/reports/2008/dataskillscareersfinalreport.aspx>

Gold. “Cyberinfrastructure, data, and libraries.” D-Lib Magazine 13:9/10 (2007). <http://www.dlib.org/dlib/september07/gold/09gold-pt1.html> and <http://www.dlib.org/dlib/september07/gold/09gold-pt2.html>

Walters and Skinner, “Digital Curation for Preservation.” [http://www.arl.org/bm-doc/nrnt\\_digital\\_curation17mar11.pdf](http://www.arl.org/bm-doc/nrnt_digital_curation17mar11.pdf) (pp 5-30 required; rest recommended)

### Week 2: Project management. Data interviews. 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. Data interviews. 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.” PNLQ Quarterly 73:2 (2009). [http://www.pnla.org/assets/documents/Quarterly/pnla\\_winter09.pdf](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/>

Witt and Carlson. “Conducting a data interview.” [http://docs.lib.purdue.edu/lib\\_research/81/](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.”)

## Unit 2: The data environment

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

*Learning objectives: OAIS model. DCC data-lifecycle model. Significant properties. Types and sources of scientific data. Examples of quantitative and qualitative social-science data. Examples and uses of humanities data. 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:oais>  
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](http://www.dpconline.org/docs/lavoie_OAIS.pdf)

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>

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](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>

NSF. "Digital data collections by categories." [http://nsf.gov/pubs/2005/nsb0540/nsb0540\\_11.pdf](http://nsf.gov/pubs/2005/nsb0540/nsb0540_11.pdf)

#### **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>

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>

"Sustainable Economics for a Digital Planet." [http://brtf.sdsc.edu/biblio/BRTF\\_Final\\_Report.pdf](http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf)

Ithaka. "Funding sustainable digital resources." <http://www.ithaka.org/ithaka-s-r/research/funding-sustainable-digital-resources>

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

Ray, Kate. "Five open questions for data.gov." <http://techcrunch.com/2011/04/13/fiv-questions-data-gov-savethedata/>

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/>

#### **Week 5: The legal and regulatory environment around data and digital preservation**

*Learning objectives: Open movements (open source, open access, open data, open government data, open notebook science). Funder mandates (NIH Public Access Policy, NSF DMPs). Journal open-data mandates. Copyright and data. Patents and data. Panton Principles, CCo. The dangers of "non-commercial" and "share-alike" licenses. Human-subjects research and data confidentiality. Data licensing.*

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

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](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/>

#### **Week 6: Researcher data practices and needs**

*Learning objectives: Scholarly attitudes toward data sharing, and how they differ across disciplines. Personal digital archiving, and how practices bleed into the scholarly environment. Researcher 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:datapracitices>, <http://pinboard.in/u:dsalo/t:horrorstories>, <http://pinboard.in/u:dsalo/t:datapublishing>, <http://pinboard.in/u:dsalo/t:datacitation>

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>

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>

Borgman, Christine L. "Research data: who will share what, with whom, when, and why?" <http://works.bepress.com/borgman/238/>

Feijen, Martin. "What researchers want." [http://www.surffoundation.nl/nl/publicaties/Documents/What\\_researchers\\_want.pdf](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>

OCLC. "Registering researchers in authority files." <http://www.oclc.org/research/activities/registering-researchers.html>

Raloff, Janet. "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](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](http://www.sciencenews.org/view/generic/id/33436/title/Galaxy_Zoos_blue_mystery_%28part_2%29)

Mueller, Martin. "Getting undergraduates and amateurs into the business of re-editing our cultural heritage." <http://literaryinformatics.net/2011/01/07/getting-undergraduates-and-amateurs-into-the-business-of-re-editing-our-cultural-heritage-for-a-digital-world/>

## Unit 3: The service environment

### Week 7: Library and archive preparedness.

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

Rusbridge, Chris. "Tomorrow, and tomorrow, and tomorrow: poor players on the digital curation stage." <http://www.era.lib.ed.ac.uk/handle/1842/2150/>

Pryor, Graham. "Librarians doing data -- a paradox?" [http://www.dcc.ac.uk/webfm\\_send/319](http://www.dcc.ac.uk/webfm_send/319)

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

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

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

Watch: reBIND video, [http://rebind.bgbm.org/rebind\\_movie](http://rebind.bgbm.org/rebind_movie)

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

### Week 8: Data literacy and data training.

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

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>

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

University of Massachusetts. "Data-management curriculum for the sciences." [http://library.umassmed.edu/data\\_management\\_frameworks.pdf](http://library.umassmed.edu/data_management_frameworks.pdf)

### Week 9: Assessing, collecting, publishing, and citing data

*Learning objectives: Data assessment. Gauging importance and preservability. Collection-development policies. Data publishing. Data citation and credit (Datacite, ORCID). Journals and data.*

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

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

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

Whyte. "Appraise & select research data for curation." <http://www.dcc.ac.uk/resources/how-guides/appraise-select-research-data>

Faundeen and Oleson. "Scientific data appraisals: the value driver for preservation." [http://www.pv2007.dlr.de/Papers/Faundeen\\_AppraisalsValue\\_for\\_Preservation.pdf](http://www.pv2007.dlr.de/Papers/Faundeen_AppraisalsValue_for_Preservation.pdf)

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>

## SPRING BREAK: enjoy!

### Week 10: Data sharing. Existing data archives. Data discovery and reuse.

*Learning objectives: Finding disciplinary data archives, open and subscription. Licensing issues with data archives. Data, the ILS, and discovery layers. Google and data. Data sharing. Data reuse. Digital divides and data. Hathi Trust. Digital libraries as humanities-data archives.*

Linklists: <http://pinboard.in/u:dsalo/t:datause>, <http://pinboard.in/u:dsalo/t:datasources>, <http://pinboard.in/u:dsalo/t:datasharing>  
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/>  
Piowar. "Who shares? Who doesn't? Factors associated with openly archiving raw research data." *PLoS ONE*. <http://dx.doi.org/10.1371/journal.pone.0018657>  
Hogenboom, Teper, and Wiley. "Collecting small data." <http://publications.arl.org/1h7vog.pdf>  
"About Databib." <http://databib.org/about.php>  
Partlo. "The pedagogical data reference interview." <http://iassistdata.org/iq/pedagogical-data-reference-interview>  
ICPSR. "Data-driven learning guides." <http://www.icpsr.umich.edu/icpsrweb/ICPSR/OLC/guides> (Read through one or two on topics of interest.)  
Xia and Liu. "Usage patterns of open genomic data." <http://crl.acrl.org/content/early/2012/01/09/crl-324.short>  
Carleton College. "Data, Datasets, and Statistical Resources." <http://gouldguides.carleton.edu/content.php?pid=65030&sid=480389> (please look through all the tabs)  
Lichtenstein. "Why open data alone is not enough." [http://m.wired.com/magazine/2011/06/st\\_essay\\_datafireworks/](http://m.wired.com/magazine/2011/06/st_essay_datafireworks/)

### Week 11: Digital humanities, GIS

*Learning objectives: What GIS is and how it works (raster vs. vector images, topology, projections, attributes, georectification). How various disciplines use GIS. GIS tools and file types. GIS data and metadata standards. GIS preservation challenges. The "digital humanities" and their history. How libraries are supporting the digital humanities. Digital-humanities data sources. Digital-humanities preservation challenges. Documenting traditional cultural expressions.*

Linklists: <http://pinboard.in/u:dsalo/t:gis>, <http://pinboard.in/u:dsalo/t:digiHum>. Also skim <http://digitalscholarship.wordpress.com/2011/10/14/getting-started-in-the-digital-humanities/>  
Sutton, Dassau, and Sutton. "A Gentle Introduction to GIS." <http://linfiniti.com/dla/AGentleIntroductionToGIS.pdf>  
GIS DataDepot. "GIS data formats." <http://data.geocomm.com/helpdesk/formats.html>  
Play with: the NYPL Map Rectifier. <http://maps.nypl.org/warper/>  
Jackson, K. "Introducing digital humanities." <http://prezi.com/r7rmqbxifpq9/introducing-digital-humanities-full/>  
"Interview with Brett Bobley." <http://blogs.loc.gov/digitalpreservation/2011/10/interview-with-brett-bobley/>  
"Library spaces for the scholarship process." <http://www.educause.edu/EDUCAUSE+Review/EDUCAUSEReviewMagazineVolume46/LibrarySpacesfortheScholarship/231829>  
World Intellectual Property Organization. "Intellectual property and the safeguarding of traditional cultures." [http://www.wipo.int/freepublications/en/tk/1023/wipo\\_pub\\_1023.pdf](http://www.wipo.int/freepublications/en/tk/1023/wipo_pub_1023.pdf)

## Unit 3: Technology of digital preservation

### Week 12: File formats and their sustainability.

*Learning objectives: Evaluating file formats for preservation. Lossy vs. lossless formats. Open vs. proprietary formats. File formats in instrument science. Quantitative-science file formats and tools (SPSS, Stata, R, Matlab). Image formats (JPEG, TIFF, JPEG 2000, PNG, GIF, RAW). Audio and video formats (codecs, sampling rate/bitrate, WAV, AIFF, mp3, MPEG4). HDF. GIS formats. "Preservation copy," "digital surrogate." Compound objects; archiving websites; BagIt. File-format auditing tools.*



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](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](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](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>

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

"Why HDF?" [http://www.hdfgroup.org/why\\_hdf/](http://www.hdfgroup.org/why_hdf/)

BagIt specification. <https://confluence.ucop.edu/display/Curation/BagIt> (please download and read the spec)

JHOVE. <http://jhove.sourceforge.net/>

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>

### Week 13: Metadata

*Learning objectives: Descriptive, technical, administrative, and structural metadata. Discipline-specific metadata standards. Codebooks and data dictionaries. Crosswalking and other metadata tools (Google Refine, DataUp). Ontologies. Explaining metadata to non-librarians. Getting metadata from non-librarians. Coping with spreadsheets. Dataset identifiers.*

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>

EZID. "Understanding identifiers." <http://n2t.net/ezid/home/understanding>

DataUp. "DataUp features." [http://dataup.cdlib.org/dataup\\_features.html](http://dataup.cdlib.org/dataup_features.html)

DDI FAQ. <http://www.ddialliance.org/resources/faq.html>

Getting started with DDI. <http://www.ddialliance.org/resources/getting-started>

### Week 14: Digital forensics and its technology. E-records management. Archiving born-digital materials.

*Learning objectives: Digital forensics and related tools. E-records management. 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/>

“Digital records preservation: where to start guide.” <http://isotc.iso.org/livelink/livelink?func=ll&objId=10083866&objAction=Open&nexturl=%2Flivelink%2Flivelink%3Ffunc%3D11%26objId%3D8800147%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/>

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>

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

## 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, Archivemata). 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:855>

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

“About LOCKSS.” [http://www.lockss.org/lockss/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](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](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>

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

## Assignments

### Grading Schema and Due Dates

Assignments:	Percentage	Due Date
Individual assignments:		
Horror-story assessment	5%	28 January
NSF DMP critique	10%	4 March
Compare/contrast data-management curricula	10%	18 March
Add two data repositories to Databib	10%	8 April
Blog post	10%	13 May (earlier if you wish)
Final project		
DCP/Project plan/Work agreement	15%	18 February
Mid-semester check-in	5%	18 March
Project poster or video	10%	6 May
Final project report	10%	13 May
Individual contribution	10%	
Readings and participation	5%	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 and so on...

No extra credit opportunities are available in this class. No assignment grades are dropped. Any student failing entirely to turn in an 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 segments:

- Data-curation profile and project plan: This should be a plan for solving the *entire problem* as presented to you by the client. 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 should form part of this plan as well. You may decide to revise this plan over the course of the semester!
- 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.
- 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! Turn in a progress report and the revised agreement (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, 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. If you can, please print the poster and get it to me by the due date, but if you have no group members on campus, I will print (I’ll ask you to chip in for the printing). This poster will be exhibited as part of the Holz Series session on May 7. If possible, please have one or more group members (it need not be the PM) present to talk to brownbag attendees about the project.
- 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 if you must).

Grading rubric:

- Data-curation profile: understanding and using the DCP 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 – 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 to report on from my linklist at <http://pinboard.in/u:dsalo/t:horrorstories>. (Off-limits: “lost or stolen computer.”) First-come, first-served; please 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 data at issue?
- What went wrong (or could go wrong) with the data-handling?
- What are the consequences of the poor data-handling to the research?
- What are the consequences of the poor data-handling for the researcher(s)?
- Could this problem have been avoided? How?

You are expected to read all your colleagues’ posts within one week of the due date for posting; I will check this via Learn@UW reports, and your grade on this assignment will be penalized if you do not.

### NSF data-management plan

A real-world example (anonymized!) of an NSF data-management plan draft will be posted to Learn@UW. You are the librarian contacted by the NSF applicant for an evaluation of the plan. Read it, find the correct NSF directorate’s guidance for it, evaluate the plan according to that guidance and your own sense of what is necessary, come up with no more than three recommendations for improvement (yes, you must prioritize! infodumps will be heavily penalized) 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 NSF grant applicant. (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.)

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

### Compare/contrast data-management curricula

I will provide a list of data-management curricula on Learn@UW. Choose two, 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 data-management 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?)
- quality and usefulness of suggested assignments

Conclude by indicating which you would prefer to teach from. 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 walk into a classroom and teach from one?

### Add two data repositories to Databib

Sign up for an account at [databib.org](http://databib.org). Using the techniques discussed in class (or any others you think of yourself), find two data repositories not yet available there, and add them. Post the database names and URLs to the designated forum on Learn@UW, with a short description of how you found them; this is necessary because they won’t show up on Databib right away owing to the review step.

## Blog post

Write a blog post suitable for Research Data Services's weblog! Use the list of blogs on Learn@UW as inspiration, and as a style/tone/length guide. Suggested topics:

- Tools and services specific to UW-Madison
- Tools and services on the open web
- A quick how-to for a common data-management problem or failing
- A lesson from a horror story (n.b.: don't use the horror story you already posted about; that's double-dipping)
- A well-worded explanation of difficult terminology or techniques
- Open-data advocacy (make it appealing! zealotry isn't appealing)
- Positive case studies on how good data-management practices (or data sharing) helped research

RDS will decide whether to add your post to the blog; if we do, you will be appropriately credited.

SLIS Goals	855 Objectives	855 Measurable Outcomes
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 research-data forms and formats and research-data lifecycles across scholarly disciplines.	Horror-story analysis (including systemic failures). Compare/contrast data-management curricula.
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 research-data forms and formats and research-data lifecycles across scholarly disciplines.	Databib 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.	Blog post. Final project poster/video.