Syllabus: LIS 855, Digital curation
School of Library and Information Studies
University of Wisconsin-Madison
Spring 2012: Online

Dorothea Salo (please call me "Dorothea")
Office address: 4261 Helen C. White Hall
Jabber/GTalk: salo@wisc.edu
Course link page: http://pinboard.in/u:dsalo/t:855
salo@wisc.edu, 608-265-4733
Office Hours: M 11:30-1:30 and 4-6
Skype: dorotheasalo

Course Objectives

- Assess, plan for, manage, and execute a small-scale digital-preservation 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 digital-preservation software tools.
- Read and write XML metadata, particularly METS, MODS, and PREMIS.
- 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 special 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. I have placed Ross Harvey's *Digital Curation* (Neal-Schuman 2010) on reserve, and recommend but do not require its purchase. 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
Please use the Learn@UW help forum *before* emailing me; please also do your best to assist your classmates there. I am not available Fridays or weekends; otherwise, I do my level best to answer email within two business days. If you need to speak with me but cannot make my office hours, please make an appointment with me *directly on WiscCal*, which will email me the appointment information and help ensure I'm not double-booked.

If 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 23. All assignments (save one; see the grading schema) 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.

Unit 1: Bootstrapping

Week 1: Course overview and XML introduction

Learning objectives: XML well-formedness rules. DTDs, schemas, XML validity. XML editors. Declaring and using XML namespaces.

Linklists: http://pinboard.in/u:dsalo/t:xml, http://pinboard.in/u:dsalo/t:datacuration

Recommended: Harvey, chs. 1 and 2.

Gold, Anna. "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 (pp 5-9 required; rest recommended)

w3schools.com. "XML tutorial." http://www.w3schools.com/xml/default.asp (Introduction, How to Use, Syntax, Elements, Attributes, then below to XML Namespaces and XML Summary)

"Comparison of XML editors." http://en.wikipedia.org/wiki/List_of_XML_editors

Week 2: Project management and service development. 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. Service development in libraries. Outreach and marketing in libraries. Data interviews. Keeping current.

Linklist: http://pinboard.in/u:dsalo/t:projectmanagement

Wamsley, Lori H. "Controlling project chaos: project management for library staff." PNLA Quarterly 73:2 (2009). http://www.pnla.org/quarterly/Winter2009/PNLA_Winter09.pdf (pp. 5-6, 27)

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

Denning, Steve. "The death and reinvention of management." http://stevedenning.typepad.com/steve_denning/2010/11/the-deathand-reinventionof-management-a-draft-synthesis.html (only the section entitled "Shift #3: New coordination: From bureaucracy to dynamic linkage.")

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

Witt, Michael, and Jake R. 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.")

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

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

Unit 2: The sociology of digital data management and preservation

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

Learning objectives: OAIS model. DCC data-lifecycle model. Types and sources of scientific data. Reference, research, and resource data collections. Examples of quantitative and qualitative social-science data. Examples and uses of humanities data. Crowdsourcing data transcription and analysis.

Linklists: http://pinboard.in/u:dsalo/t:standards (skim for relevant tags)

Recommended: Harvey chs. 3 and 4.

Ockerbloom, John Mark. "What repositories do: the OAIS model." http://everybodyslibraries.com/2008/10/13/what-repositories-do-the-oais-model/

OAIS Reference Model. http://public.ccsds.org/publications/archive/650x0b1.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

ICPSR. "Guide to social science data preparation and archiving." http://www.icpsr.umich.edu/files/ICPSR/access/dataprep.pdf

Cragin, Melissa, and Kalpana Shankar. "Scientific data collections and distributed collective practice." Computer Supported Cooperative Work 15:2/3 (2006). http://dx.doi.org/10.1007/s10606-006-9018-z

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_I%29 and 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/

Week 4: 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 program. Digital preservation needs and strategies in public libraries. Infrastructure. Funding (grant earmarks, budget and position shifting). Liaison librarians and research-data curation. Research data as "the new special collections."

Recommended: Harvey ch. 5.

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

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

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/

Week 5: Researcher practices and needs

Learning objectives: Scholarly attitudes toward data sharing, and how they differ across disciplines. Personal digital preservation, and how practices bleed into the scholarly environment. Researcher attitudes toward librarians and archivists. Data security. Data citation and credit (Datacite, ORCID). Journals and data. Peer review and data.

Linklists: http://pinboard.in/u:dsalo/t:rschbehavior, http://pinboard.in/u:dsalo/t:datapractices, 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

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

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

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

Karger, David. "Why all your data should live in one application." http://groups.csail.mit.edu/haystack/blog/2010/10/20/why-all-your-data-should-live-in-one-application/

"What is DataCite?" http://datacite.org/whatisdc.html

Fenner, Martin. "ORCID or how to build a unique identifier for scientists in 10 easy steps." http://blogs.nature.com/mfenner/2010/01/03/orcid-or-how-to-build-a-unique-identifier-for-scientists-in-10-easy-steps

Week 6: 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. "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/

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

Week 7: The legal and regulatory environment

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, CC0. The dangers of "non-commercial" and "share-alike" licenses. Human-subjects research and data confidentiality.

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

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.)

Piwowar, Heather. "Research Works Act attacks data dissemination too." http://researchremix.wordpress.com/2012/01/07/rwa-data/

"Share the data: making large-scale proteomics data widely available." Bio-IT World. http://www.bio-itworld.com/2010/08/25/open-proteomics-comment.html

Nguyen, Thinh. "Remembering Babel: open data sharing & integration." http://sciencecommons.org/weblog/archives/2009/11/19/remembering-babel-open-data-sharing-integration/

Nguyen, Thinh. "Freedom to research: keeping scientific data open, accessible, and interoperable." http://sciencecommons.org/wp-content/uploads/freedom-to-research.pdf

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

Week 8: Assessing data. Collection-development and digital-preservation policies.

Learning objectives: Data assessment. Gauging importance and preservability. Collection-development policies.

Recommended: Harvey ch. 7, ch. 11.

Timmer, John. "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 (pp. 1-9)

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

Faundeen, John L., and Lyndon R. Oleson. "Scientific data appraisals: the value driver for preservation." http://www.pv2007.dlr.de/Papers/Faundeen_AppraisalsValue_for_Preservation.pdf

Beagrie et al. "Digital preservation policies study." http://www.jisc.ac.uk/media/documents/programmes/preservation/jiscpolicy_p1finalreport.pdf

Week 9: Existing data archives. Data discovery.

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

Linklist: http://pinboard.in/u:dsalo/t:datause

Berman, Francine. "We need a research data census." Communications of the ACM 53:12 (2010). http://dx.doi.org/10.1145/1859204.1859220

Hogenboom, Teper, and Wiley. "Collecting small data." http://publications.arl.org/1h7vog.pdf

Partlo, K. "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, J. "Why open data alone is not enough." http://m.wired.com/magazine/2011/06/st_essay_datafireworks/

Week 10: Special topics: 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

Bivens-Tatum, Wayne. "Librarians and traditional cultural expressions." http://dx.doi.org/10.1080/10477845.2010.508693

World Intellectual Property Organization. "Intellectual property and the safeguarding of traditional cultures." http://wipo.int/export/sites/www/tk/en/publications/1023.pdf

SPRING BREAK: enjoy!

Unit 3: The technology of digital preservation

Week 11: Digitization, file formats, and digital 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.

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

Recommended: Harvey ch. 4, ch. 9

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

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/

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

Week 12: Metadata

Learning objectives: Descriptive, technical, administrative, and structural metadata. MODS, METS, PREMIS, Dublin Core. Discipline-specific metadata standards. Codebooks and data dictionaries. Crosswalking and crosswalking tools (Google Refine). Explaining metadata to non-librarians. Getting metadata from non-librarians. Coping with spreadsheets. DDI.

Recommended: Harvey ch. 6

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

Guenther, McCallum, "New metadata standards for digital resources: MODS and METS." http://findarticles.com/p/articles/mi_qa3991/is_200212/ai_n9150534

Cundiff and Trail, "Using METS and MODS..." http://www.loc.gov/standards/mods/presentations/mets-mods-morgan-ala07/

"Guidelines for using PREMIS with METS for exchange." http://www.loc.gov/standards/premis/guidelines-premismets.pdf

DDI FAQ. http://www.ddialliance.org/resources/faq.html
Getting started with DDI. http://www.ddialliance.org/resources/getting-started

Week 13: Threat models. Auditing. Organization. Versioning.

Learning objectives: Business-model and organizational threats to digital data. Risk assessment, analysis, and mitigation. Migration and emulation, including tools. "Trusted digital repository." Dark archive. Two-tier archive. TRAC, DRAMBORA. File-format auditing tools (JHOVE, FITS). Bit-auditing and checksums.

Recommended: Harvey ch. 13
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/
DRAMBORA Interactive. "DRAMBORA: About." http://www.repositoryaudit.eu/about/ (Please register for the site and download the entire toolkit.)
CRL. "Trustworthy Repositories Audit & Certification: Criteria and Checklist (TRAC)." http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf
CRL. "Report on Portico audit findings." http://www.crl.edu/sites/default/files/attachments/pages/CRL%20Report%20on%20Portico%20Audit%202010.pdf
JHOVE. http://hul.harvard.edu/jhove/
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 14: 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). Other relevant platforms. Curation microservices. Relational databases, XML databases, RDF triplestores. Organizing files; choosing filenames. Versioning. Geographic dispersal techniques (LOCKSS, cloud storage, DuraCloud).

Linklist: http://pinboard.in/u:dsalo/t:software
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
"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

Week 15: E-records and records management

Learning objectives: Growth in e-records. Differences between e-records-management and other kinds of data management/archiving. Authenticity. Digital forensics. Strategies and tools for assessing, deduplicating, and accessioning e-records. Duke Data Accessioner. Archivematica. Institutional repositories for records archival.

"Digital records preservation: where to start guide." http://isotc.iso.org/livelink/livelink?func=ll&objId=10083866&objAction=Open&nexturl=%2Flivelink%2Flivelink%3Ffunc%3Dll%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/

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

"Data accessioner." http://library.duke.edu/uarchives/about/tools/data-accessioner.html

"Archivematica." http://archivematica.org/ and http://archivematica.org/wiki/index.php?title=Overview

Assignments

Grading Schema and Due Dates

Assignments:	Percentage	Due Date
Individual assignments:	40%	
XML well-formedness exercise		30 January
Wiki link: one relevant blog		6 February
Blog post summary		6 February
Wiki link: one 2011/2012 data-management-related conference		6 February
Wiki page: one disciplinary data/metadata standard		13 February
Wiki link: two research-data-management websites		20 February
Data-management website compare/contrast		20 February
Two job postings to Learn@UW		20 February
Wiki page: one discipline-specific data-management resource		27 February
Wiki link: one university research-data policy		19 March
Wiki page: one data-management software tool		16 April
Producing a BagIt bag from a METS file		30 April
Attending one Second Monday Data Brownbag		(a second Monday)
NSF DMP critique	5%	12 March
Final project	50%	
DCP/Project plan		13 February
Project checkins		5 March, 9 April
Project poster		4 May (*note Friday deadline*)
Final project report		14 May
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 technical assignments, perfection is not the goal; learning is. Mistakes and retrenchments are to be expected, and often 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 plan.

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 email me 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 raise or lower individual project-participation grades as I see fit; only I will see the email. I will also check with clients about your group's professionalism, competence, and accomplishments before I assign grades.

Project segments:

- 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!
- Project checkins: On the project-discussion forum in Learn@UW, please let me know what you're doing and how it's going.
- Project poster: 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). If you can, please print the poster and get it to me by the due date. If you have no group members on campus, email me the poster file (PDF, please) by the due date (I'll ask you to chip in for the printing). This poster will be exhibited as part of the Second Monday Data Brownbag 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.
- 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 project plan, please provide a copy of the revised plan as well (this does not count against your pagecount).

Grading rubric:

- Project plan: 40% (understanding of the problem, appropriateness of suggested solutions, clear expression)
- Checkins: 10% (evidence of good project management, clear expression)
- Poster: 30% (attractive and professional design, clear expression, conference-worthy)
- Wrapup: 20% (accomplishment, overcoming obstacles, professional relationship with client, clear expression)

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!

WIKI ASSIGNMENTS

We will be seeding a data-curation wiki built on MediaWiki at `http://digcurwiki.dsalo.info/`. For each assignment, I will make a sample of what I expect you to produce. Do not omit category assignment; it's a key aspect of wiki gardening! I will not teach you MediaWiki syntax; I expect advanced library-school students to be able to use help files to learn it on their own. It will help my grading, particularly on links pages, if you leave your initials after each link or page you post.

Blog: Find one blog that you believe worth following for research-data managers and/or digital preservationists. Post a link to the wiki with a brief description. Then post a link-and-summary of one particularly trenchant post from that blog, with any reactions or questions you have about it, to the appropriate Learn@UW forum.

Conference: Find one 2011 or 2012 conference that focuses on digital preservation or research-data management, *or* has a full track relevant to these issues. Post a link (in date order by conference start date, please) to the wiki.

Website: Find TWO university or university-library sites offering research-data management advice to researchers (sites specific to NSF data-management plans are fine). Post links to the wiki. Then, compare and contrast the better (in your opinion) of the two sites you found with Research Data Service's website (`http://researchdata.wisc.edu/`), making at least three site-improvement suggestions to RDS. If this takes more than two pages, you're spending too much time on it.

Discipline-specific data or metadata standard: Find one standard specific to a particular discipline or subdiscipline (science, social-science, humanities all acceptable) and add a page on it to the wiki. On that page, please describe the standard briefly (you may cut-and-paste from the standard's documentation if you attribute), explain what kind of standard it is (data-file format? metadata? both?), and assign it to one or more disciplines via MediaWiki's category mechanism.

Discipline-specific resource: Find one useful website or document offering research-data management advice specific to a particular discipline or specialty. You are welcome to use a discipline you have educational or experiential experience in.

Software tool: Find one software tool that helps with a research-data management problem (a discipline-specific problem is fine). Post a page describing and linking to the tool to the wiki.

Policy: Find an official university or library policy governing retention and/or ownership of research data. (It's fine if this is part of a broader intellectual-property, records, or preservation policy; just be prepared to point/link to which sections of the policy are germane.) Link to it from the wiki, adding any explanation or extra pointers you think necessary.

OTHER ASSIGNMENTS

Second Monday Data Brownbag
SLIS and Research Data Services sponsor a brownbag series on the second Monday of each month (exception: May's is May 7, to avoid exam week) at noon in the SLIS Commons. Brownbags are also broadcast online. Attend one of these events, in person or online, and post a well-informed question or reaction to Learn@UW. (Alternately, I will give you credit if you ask a worthwhile question of the speaker during the session.)

Well-formed XML instance
You will find a non-well-formed instance of XML on Learn@UW. Make it well-formed. Those of you with existing XML experience may wish to know that this instance purported to be TEI (`http://tei-c.org/`); by all means make it a valid (if bare-bones) TEI instance.

Grading rubric: will it parse?

Job postings
Find two examples of jobs concerned with digital preservation and/or research-data management, and post them to the appropriate forum on Learn@UW, with a candid-but-brief evaluation of how well you fit them. One of the two job postings must NOT be in an academic library.

Grading rubric: suitable jobs? honest self-analysis?

NSF data-management plan
A real-world example (anonymized!) of an NSF data-management plan draft will be posted to Learn@UW. 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, and post your evaluation to Learn@UW, writing as though you were emailing it to the NSF grant applicant. *Your evaluation should not be longer than the plan*; you *will* lose points for excessive prolixity.

Grading rubric: did you make good, well-prioritized recommendations? was your wording professional, but not stiff?

METS file to BagIt
You will choose a digital object from a list of METS files available on Learn@UW. You will assemble the files referred to from the METS file into a "bag" valid according to the latest BagIt spec (`https://confluence.ucop.edu/display/Curation/BagIt`). (NOTE: So-called "holey bags" are acceptable ONLY if the files are not web-downloadable.)

Grading rubric: did you construct the bag accurately? is the bag valid? did you capture as much information as the BagIt spec will let you?