

DIGITAL PRESERVATION
NYU MIAP , Fall 2011
Monday 6-10PM, 721 Broadway, room 670

Version 2 - 09/08/2011

Instructors:

Chris Lacinak
chris@avpreserve.com
Phone: 917-548-8632
Skype: clacinak
GTalk: chris@avpreserve.com

Kara Van Malssen
kara@avpreserve.com
Phone: 917-842-9586
Skype: kara_vanmalssen
GTalk: kvanmalssen@gmail.com

ABSTRACT: This class will address the use of digital files and infrastructure as preservation media, and will investigate current theories and practices for the conservation and preservation of both digitized and born digital materials. Students will learn the details of how digital repositories work, and what infrastructure and procedures need to be a part of a repository in order to make it preservation compliant. Students will gain practical skills with identification and risk assessment for works as a whole, their component parts, and associated software and metadata. Initiatives by broadcasters, the Library of Congress and other national archives, digital libraries and others will be explored as examples of the architecture and attributes of digital repositories. Emphasis will be placed on how archivists may interact with these repositories as part of their preservation practice. Students will also develop an increased understanding of metadata and tools to help manage digital materials.

EXPECTATIONS: Each student will do extensive readings and three assignments: an individual review of a repository or other digital preservation project, a computer programming assignment, and a final project to be negotiated with one of the professors. For details, see “Assignments” section at end of this syllabus. Attendance at all classes is expected; more than one unexcused absence will affect grading. Grades will be based on a combination of class preparedness and participation (40%); and assignments (40% for the final project, 15% for assignment 1, and 5% for assignment 2).

Texts: There are no required texts for this class. Most readings can be found online. Texts that are not available online will be on reserve in the Bobst Library and the Cinema Studies Film Study Center.

Twitter: A Twitter list has been set up for the class called miap-digitalpreservation. You are required to follow this list, managed by @kvanmalssen. The list can be accessed from: <http://twitter.com/#!/kvanmalssen/miap-digitalpreservation>. Students are expected to suggest Twitter users to add to the list each week, and be prepared to discuss topics, articles, tools, projects, and other posts that users we are following post to the list.

Class 1 (Sep 12): Preservation in the Analog World and Intro to Digital

Topics

- Introductions (Group, 30 min)
- Summary & Course Requirements (KVM, 20 min)
- Exercise: Interactive review of analog and digital technical history, structures, major developments and terminology. (C L, 90 min)
- Introduction to basic issues (KVM, 70 min)
 - Video: *IDC 2011 Digital Universe Study: Extracting Value from Chaos*. Executive

Summary, June 2011. <http://www.emc.com/collateral/demos/microsites/emc-digital-universe-2011/index.htm>

- Video: *Why Digital Preservation is Important for Everyone*, Library of Congress, 2010 (<http://www.digitalpreservation.gov/videos/digipres/>)
- Video: *Keeping Up With Data Rot*, CBS News, CBS Sunday Morning, March 4, 2009 (<http://www.cbsnews.com/video/watch/?id=4836762n>)
- Digitization and born-digital Digital vs analog preservation
- Risk factors for digital objects
- Digital preservation basic requirements
- Preservation strategies
- Digital Asset Management (DAM), Content Management System (CMS), Collection Management System, Media Asset Management, Production Asset Management

Topics brought up by class to focus on this semester:

- * Exhibition Issues
- * Video Games
- * Tools list - via wiki - MIAP students to create and share with us
- * Resources on digital storage cost calculators
- * Small organizations - asset management
- * Interim solutions for small organizations

Class 2 (Sep 19): The Transition (CL Out)

Topics

- Presentation on terminology & principles - (KVM, 1 hr)
- Case Studies (KVM 2 hrs).
 - Guest Lecture: Yvonne Ng
 - Presentation of a case study and open 1 up for discussion.
 - Custom vs. off-the-shelf
 - Open source vs. commercial
 - Standards vs. custom
 - Policies and technologies
- Student choices for assignment 1

Required Readings

- Cornell University Library Research Department. "Setting the Stage: Introduction" in *Digital Preservation Management: Implementing Short-term Strategies for Long-term Problems*. Accessed 9/3/09 at <http://www.icpsr.umich.edu/dpm/dpm-eng/timeline/index.html>
- JISC. *Digital Preservation Briefing Paper* (http://www.jisc.ac.uk/publications/briefingpapers/2006/pub_digipreservationbp.aspx)
- Cornell University Library Research Department. Sections Titled: "Obsolescence and Physical Threats" & "Terms & Concepts" -> "Strategies" in *Digital Preservation Management: Implementing Short-term Strategies for Long-term Problems*. Accessed 9/3/09 at <http://www.icpsr.umich.edu/dpm/dpm-eng/oldmedia/index.html> <http://www.icpsr.umich.edu/dpm/dpm-eng/terminology/strategies.html>
- Arms, Caroline R. and Carl Fleischhauer. "Sustainability Factors" in *Sustainability of Digital Formats: Planning for Library for Congress Collections*. Accessed 9/3/09 at <http://www.digitalpreservation.gov/formats/index.shtml>
- Definition of digital preservation Accessed 9/3/09 at <http://www.ala.org/ala/mgrps/divs/alcts/resources/preserv/defdigpres0408.pdf>
- Besser, Howard. "Digital Longevity" in **Handbook for Digital Projects: A Management Tool for Preservation and Access**, <http://www.nedcc.org/resources/digitalhandbook/ix.htm>
- **Sustainable Preservation Practices and the Rhizome ArtBase**, Accessed 8/25/11 <http://media.rhizome.org/blog/8332/rhizome-digital-preservation-practices.pdf>

- Lavoie, Brian. “13 Ways of Looking at Digital Preservation”. 2004. *D-Lib Magazine*. <http://www.dlib.org/dlib/july04/lavoie/07lavoie.html>
- IDC Digital Universe Studies
 - 2011: <http://www.emc.com/collateral/demos/microsites/emc-digital-universe-2011/index.htm>
 - 2008: <http://www.emc.com/collateral/analyst-reports/diverse-exploding-digital-universe.pdf>
- Start regularly reading the Library of Congress digital preservation blog *The Signal*: <http://blogs.loc.gov/digitalpreservation/>

Recommended Readings

- “Preserving Digital Information, Report of the Task Force on Archiving of Digital Information” (1996) Donald Waters and John Garrett, Eds. The Commission on Preservation & Access and the Research Libraries Group. <http://www.oclc.org/programs/ourwork/past/digpresstudy/final-report.pdf>
- Basic introduction to PLANETS/PLATO in the video *Preserving the British Library’s C19 Newspaper Collection with Planets: a short film* (<http://www.youtube.com/watch?v=K6NnFcSpAh8>) (importance of context)
- “The Evidence in Hand: Report of the Task Force on the Artifact in Library Collections” (2001). Stephen G. Nichols & Abby Smith, Eds. Council on Library & Information Resources (<http://www.clir.org/pubs/reports/pub103/pub103.pdf>)
- Technical Advisory Services for Images. “File Formats and Compression”. Last update 2005. <http://www.tasi.ac.uk/advice/creating/fformat.html>
- Rice, David & Lacinak, Chris. “Digital Tape Preservation Strategy: Preserving Data or Video?” 2009. <http://www.avpreserve.com/dvanalyzer/dv-preservation-data-or-video/>
- Kurt D. Bollacker. “Avoiding a Digital Dark Age,” *American Scientist*, May-June 2010. <http://www.americanscientist.org/issues/pub/2010/3/avoiding-a-digital-dark-age>

Class 3 (Sep 26): Metadata

Topics

- Metadata & Packages Overview (KVM 75 min)
 - Review metadata types
 - Metadata Object Description Schema—MODS <http://www.loc.gov/standards/mods/>
 - PREMIS <http://www.loc.gov/standards/premis/>
 - Metadata Encoding & Transmission Standard -- METS <http://www.loc.gov/standards/mets/>
 - Walk through sample XML documents
- Activity: Metadata generation and XML (120 min)
 - Intro lecture: Technical preservation metadata derivation and capture (CL, 30 min)
 - PREMIS group exercise (90 min)

*Let people know about bringing laptops

Required Readings

- Gilliland, Anne. *Setting the Stage*, in Gill et. al. Introduction to Metadata Version 3.0 (http://www.getty.edu/research/publications/electronic_publications/intrometadata/pdf.html)
- Caplan, Priscilla. *Understanding PREMIS*, Washington DC: Library of Congress Network Development and MARC Standards Office, Feb 1, 2009 (<http://www.loc.gov/standards/premis/>)
- W3C *XML Essentials* (<http://www.w3.org/standards/xml/core>)
- Look over Jenn Riley’s *Glossary of Metadata Standards*, 2010 (<http://www.dlib.indiana.edu/~jenrile/metadatamap/>)
- Dappart, A. & Enders, M. (2010). Digital Preservation Metadata Standards. *Information Standards Quarterly*, Spring. http://www.loc.gov/standards/premis/FE_Dappert_Enders_MetadataStds_isqv22no2.pdf
- “Toward a Library of Virtual Machines” *The Signal*, Sept 21, 2011 <http://blogs.loc.gov/digitalpreservation/2011/09/toward-a-library-of-virtual-machines-insights-interview-with-vasanth-bala-and-mahadev-satyanarayanan/>

Recommended Readings

- Howard D. Wactlar and Michael G. Christel, Digital Video Archives: Managing through Metadata http://www.digitalpreservation.gov/library/docs/es_archives.pdf
- Follow up from Class 2:
 - Functional Requirements: http://en.wikipedia.org/wiki/Functional_requirements
 - Use case: http://en.wikipedia.org/wiki/Use_case
 - User story: http://en.wikipedia.org/wiki/User_story
 - API (Application Programming Interface): <http://en.wikipedia.org/wiki/Api>

Class 4 (Oct 3): Practical Work with Digital Files and Embedded Metadata

Topics/Activities

- Working with files, performing routine tasks of identification, creating checksums, documentation and parsing of metadata for capture, exporting metadata, creating preservation object deliverables. (CL & KVM 4 hrs)

Required Readings

- Meta 101. PhotoMetadata.org
<http://www.photometadata.org/META-101>
<http://www.photometadata.org/META-101-metadata-Q-and-A>
<http://www.photometadata.org/META-101-metadata-classes>
<http://www.photometadata.org/META-101-metadata-types>
<http://www.photometadata.org/META-101-metadata-structure>
- ARSC Technical Committee Study on Embedded Metadata in Wave Files (To be provided)
- Lacinak, Chris. "Embedded metadata in WAVE files: a look inside issues and tools" (To be provided)
- Zetter, Kim. "Metadata in State Records is Public Record, Court Rules". Wired Magazine, Threat Level Blog. Accessed 6/22/2010
<http://www.wired.com/threatlevel/2009/10/metadata/>
- Lacinak, Chris. "A Primer on Codecs for Moving Image and Sound Archives". 2010.
http://www.avpreserve.com/wp-content/uploads/2010/04/AVPS_Codec_Primer.pdf

Review

- Wikipedia Checksum Entry: <http://en.wikipedia.org/wiki/Checksum>
- MediaInfo: <http://mediainfo.sourceforge.net/en>
- BagIt—simple protocol for transferring files and metadata
 - Video: Bagit: Transferring Content for Digital Preservation, Library of Congress, 2009 (<http://www.digitalpreservation.gov/videos/bagit0609.html?rec=4629>)
 - Summary (http://www.digitalpreservation.gov/news/2008/20080602news_article_bagit.html)
 - Specifications (<http://www.digitalpreservation.gov/library/resources/tools/docs/bagitspec.pdf>)
- ExifTool: <http://www.sno.phy.queensu.ca/~phil/exiftool/>
- DV Analyzer: <http://www.avpreserve.com/dvanalyzer/>
- BWF MetaEdit: <http://sourceforge.net/projects/bwfmetaedit/>
- The public forum for the Visual Resources Association Embedded Metadata Working Group <http://metadatadeluxe.pbworks.com/>

Class 5 (Oct 17): Repository Architecture and Systems

Topics

- Guest lecture: Dirk Van Dall, Broadway Video Digital Media
 - Storage media and terms
 - Storage architectures: HSM, distributed, grid, etc.
 - Software and hardware

- Virtualization, partitioning, etc.
- Real world examples from large and small organizations
- Repository and Standards Lecture (KVM 120 min)
 - Essential Repository Functions: What does a Digital Repository do?
 - Staffing Roles, Responsibilities and Stakeholders
 - OAIS Model, SIP, DIP, AIP
 - Survey of tools – FEDORA, DSpace, Archivematica, CDL microservices, Rosetta,
 - IRODS and SRB

Required Readings

- Rosenthal, David, et al, “Requirements for Digital Preservation Systems” D-Lib Magazine, November 2005. <http://www.dlib.org/dlib/november05/rosenthal/11rosenthal.html>
- Besser, Howard. Moving from Digital Collections to Interoperable Digital Libraries http://www.firstmonday.org/issues/issue7_6/besser/

Review

- “Reference Model for an Open Archival Information System.” CCSDS 650.0-B-1. Blue Book. Issue 1. <http://public.ccsds.org/publications/archive/650x0b1.pdf>
- Related architectures
 - Dspace <http://www.dspace.org/>
 - Fedora <http://www.fedora-commons.org/>
 - LOCKSS <http://lockss.stanford.edu>
 - Universal Virtual Computer see Raymond Lorrie’s “A Project on Preservation of Digital Data” RLG DigiNews 5:3 <http://www.rlg.org/preserv/diginews/diginews5-3.html>
- Digital Curation Center -- <http://www.dcc.ac.uk/>

Recommended Readings

- “The Cloud and Archival Storage” Technology at ICSP Blog, November 23, 2010. <http://techaticpsr.blogspot.com/2010/11/cloud-and-archival-storage.html>
- NARA:
 - <http://www.archives.gov/preservation/products/>
 - <http://www.archives.gov/era/era-approach.html>
 - http://www.archives.gov/electronic_records_archives/
 - <http://www.archives.gov/era/about/status.pdf>

Class 6 (Oct 24): Production Workflows and Systems; Intro to Computing Environments

Topics

- Workflows and Systems Lecture (CL 90 min)
 - Preservation oriented production
 - Types of elements produced (ie. elements vs. final works, versions)
 - Ingest Process
 - File Format identification and validation
 - Obsolescence Monitoring vs. Normalization
 - Prioritization/Selection
 - Migration and Refereshing
- Introduction to computing environments
 - Demystify programming and become conversant with developers
 - Command-line tools for digital preservation
 - Navigating the command line in UNIX/LINUX
 - Review of Unix Tutorial (to be completed before class). <http://www.ee.surrey.ac.uk/Teaching/Unix/index.html>

Required Readings

- UNIX Tutorial for Beginners. <http://www.ee.surrey.ac.uk/Teaching/Unix/index.html>. Lessons: Intro, 1, 2, 3, 4 and 7. Use server metadatansandbox.com. Login info has been emailed to you.

- Bradley, K (editor). (2009). "Guidelines on the Production and Preservation of Digital Audio Objects". IASA TC04, second edition. International Association of Sound and Audiovisual Archives. www.iasa-web.org/tc04/audio-preservation
- Mary Ide, Dave MacCarn, Thom Shepard, and Leah Weisse, Understanding the Preservation Challenge of Digital Television <http://www.clir.org/pubs/reports/pub106/television.html>

Class 7 (Oct 31): Multi-institutional projects & Intellectual Property Issues and Digital Rights Management

** Final project proposals due*

Topics

- Intellectual Property Issues and Digital Rights Management
 - Jane Park of Creative Commons Guest Lecture (60 min)
 - Howard Besser Guest Lecture (60 min)
- Student Presentations on Digital Preservation and Access Projects (100 min)
- General Discussion (30 min)

Required Readings

- **Multi-institutional Projects (this is the first of 2 sections with readings for this week)** (spend at least 10 minutes looking at each of these websites)
 - NDIIPP: National Digital Information Infrastructure and Preservation Program – Library of Congress <http://www.digitalpreservation.gov>
 - Digital Curation Center <http://www.dcc.ac.uk/>
 - International Preservation Consortium <http://www.netpreserve.org/>
- **Intellectual Property Issues, and Digital Rights Management**
 - Howard Besser's "Copyright and Digital Preservation" notes, 10/27/2011. <http://www.nyu.edu/tisch/preservation/program/11fall/digital-copyright.html>
 - "VIDEO AT RISK" DRAFT MEMORANDUM ON REFORMATTING CIRCULATING VIDEO COLLECTIONS (via email)
 - OCLC's Well-Intentioned Practice for Putting Digitized Collections of Unpublished Materials Online" (<http://www.oclc.org/research/activities/rights/practice.pdf>)
 - Reproduction, Copyright and Ownership in the context of digital preservation (<http://www.nyu.edu/tisch/preservation/program/05spring/digital-copyright.html>)
 - American Library Association, "Digital Rights Management and Libraries" on the web site of ALA. <http://www.ala.org/ala/issuesadvocacy/copyright/digitalrights/index.cfm>
 - Besser, Howard. The Challenge of Copyright & Difficult Materials -- besser.tsoa.nyu.edu/howard/Talks/09soima-copyright.pdf
 - Techdirt.com website, *Can a Contract Remove Fair Use Rights* <http://www.techdirt.com/articles/20110209/03531413021/can-contract-remove-fair-use-rights.shtml>

Watchaaaa

- Grace Agnew - Restrictions of Digital Rights Management (<http://www.youtube.com/watch?v=aRfX2gPwXMo>)
- The Pirate's Dilemma – Matt Mason (http://cdnlthree.kaltura.com/p/22646/sp/2264600/serveFlavor/flavorId/1_ph4u1p1j/name/1_ph4u1p1j.ogv?noVar=0?h=088cb9a191f6d46e92f30)
- 2009 Open Video Conference: Fair Use Battles: Discussion (http://www.archive.org/details/ovc_fair_use_battles_19june2009)

Recommended Readings

- "Digital Archives: Navigating the Legal Shoals", Columbia Law School, April 16, 2010 (site includes video of presentations) <http://www.law.columbia.edu/kernochan/symposia/digital-archives>

- “Undue Diligence: Seeking Low-risk strategies for Making Collections of Unpublished Materials More Accessible”, OCLC San Mateo, March 11, 2010 (site includes audio of presentation) <http://www.oclc.org/research/events/2010-03-11.htm>
- Preserving Digital Public Television. “[Intellectual Property and Copyright Issues Relating to the Preservation and Future Accessibility of Digital Public Television Programs](#)”, April 2010, pdf downloadable from <http://www.thirteen.org/ptvdigitalarchive/uncategorized/pdptv-intellectual-property-and-copyright-issues-report-released/122/>
- [International Study on the Impact of Copyright Law on Digital Preservation](#) (Joint report of JISC, NDIIPP, the OAK Law Project and the SURFfoundation, July 2008) <http://www.ithaka.org/ithaka-strategy/ithaka-case-studies-in-sustainability>
- Andrew Charlesworth. Digital Lives: Legal and Technical Issues (discussion paper) Oct 2009 (<http://britishlibrary.typepad.co.uk/files/digital-lives-legal-ethical.pdf>)
- Duncan, Charles et al. "Digital Rights Management: Final Report". JISC. 2004. http://www.intrallect.com/index.php/intrallect/knowledge_base/general_articles/jisc_drm_study_2004_1

Class 8 (Nov 7): Web and Social Media Archiving

* Assignment #2 due

Topics

- Video: **Web Archiving, Library of Congress, 2009** (<http://www.digitalpreservation.gov/videos/webarch09/index.html>)
- Guest Lecture by Patrick Davison and follow up discussion: (90 min)
 - Issues with archiving of websites and social media
 - Theories and methodologies for collecting and preserving web sites and social media
- Student Presentations (60 min)
- Group Discussion (60 min)

Required Readings

- Watch entire Webinar video on using Archive-IT <http://archive-it.org/public/webinar.html>
- Internet Archive’s Glossary of Web Archiving Terms (<https://webarchive.jira.com/wiki/display/ARIH/Glossary+of+Web+Archiving+Terms>)
- Look over what the International Internet Preservation Consortium does, review recent activities, etc. (<http://www.netpreserve.org/about/index.php>)
IIPC Project to Evaluate Emulation and Migration as Long-Term Preservation Solutions for Web Archives
- Listen to “Michael Nelson on Digital Preservation” on Surprisingly Free (stream or podcast) <http://surprisinglyfree.com/2011/09/06/michael-nelson/>
- Look over the Memento project website: <http://www.mementoweb.org/>
- Trapani, Gina. Fast Company, **How to Backup Your Social Media Life**. October 2010. <http://www.fastcompany.com/1695794/how-to-backup-your-social-media-life>
- ACT-IAC Collaboration & Transformation (C&T) Shared Interest Group (SIG), **Best Practices Study of Social Media Records Policies** <http://www.actgov.org/knowledgebank/whitepapers/Documents/Shared%20Interest%20Groups/Collaboration%20and%20Transformation%20SIG/Best%20Practices%20of%20Social%20Media%20Records%20Policies%20-%20CT%20SIG%20-%2003-31-11%20%283%29.pdf>
- Archive-It Page:
 - <https://webarchive.jira.com/wiki/display/ARIH/5+Challenges+of+Web+Archiving>
 - <https://webarchive.jira.com/wiki/display/ARIH/Archiving+Social+Networking+Sites+with+Archive-It>

Recommended Readings

- Cornell University Library. “Web Tool Evaluation Categories” in web site Virtual Remote Control. Ithaca, NY: Cornell University Library. n.d. Accessed 3/3/05 at <http://>

- irisresearch.library.cornell.edu/VRC/categories.html
- Cornell University Library. “Web Tool Resources” in web site Virtual Remote Control. Ithaca, NY: Cornell University Library. 2004 Accessed 3/3/05 at <http://irisresearch.library.cornell.edu/VRC/resources.html>
- Look over the web site: WebArchivist.org. Seattle: University of Washington and Utica, NY: State University of New York Institute of Technology. Accessed 3/5/05 at <http://www.webarchivist.org/about.htm>
- Andrew Stawowczyk Long, (Digital Preservation Standards Strategist, National Library of Australia) Long-Term Preservation of Web Archives – Experimenting with Emulation and Migration Methodologies (<http://www.netpreserve.org/publications/reports.php>)

Class 9 (Nov 14): Personal Archiving

Topics

- Personal Archiving Videos and Discussion (90 min)
 - Video: *Why Digital Preservation is Important for You*, Library of Congress, 2010 (http://www.digitalpreservation.gov/videos/personal_archiving/)
 - Video: *Greenest Grass and Bluest Skies*, Daniel Reetz at Personal Digital Archiving Conference, 2011 (<http://www.archive.org/details/PDA2011-danielreetz>)
 - Video: *The Smallest Day: A nerd and his Dad set out to digitize the family archives*, Stan James at Personal Digital Archiving Conference, 2011 (<http://www.archive.org/details/PDA2011-stanames>)
- Guest lecture from Jeff Ubois via Skype (60 min)
- Guest lecture from (60 min)

Read/Watch:

- Jonathan Good, founder of 1000memories.com at TED Talks: <http://youtu.be/sGE4HMvDe-Q>
- Library of Congress website “Personal Archiving: Preserving Your Digital Memories”, (<http://www.digitalpreservation.gov/you/>)
- Digital Lives Research Project <http://www.bl.uk/digital-lives/>
 - Read the report at <http://britishlibrary.typepad.co.uk/files/digital-lives-synthesis02-1.pdf>
- Saving Our Present for the Future: Personal Archiving 2010, The Internet Archive, San Francisco, February 16, 2010 (<http://www.personalarchiving.com/conference/>)
- Video Lecture: *Archiving Digital Photos*, Library of Congress, 2010 (http://www.digitalpreservation.gov/videos/personal_archiving/photos.html)
- Gordon Bell’s MyLifeBits
 - <http://research.microsoft.com/en-us/projects/mylifebits/>
 - Bell’s video lecture at Library of Congress (http://www.loc.gov/today/cyberlc/feature_wdesc.php?rec=4908)

Class 10 (Nov 28): Sustainability Issues & TRAC

Topics

- Sustainability Issues from micro to macro (CL, 120 min)
- Trustworthy Repositories Audit and Certification: Criteria and Checklist (KVM, 90 min)

Required Readings

- Smith, Abby, et al. “Sustainable economics for a digital planet: Ensuring long term access to digital information” (often referred to as Blue Ribbon Task Force on Digital Preservation) Accessed 6/22/2010 at http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf
- Read at least the executive summaries for:
 - Ithaka Case Studies in Sustainability 2009: Sustaining Digital Resources: An On-the-Ground View of Projects Today (July 2009) Ithaka Case Studies in Sustainability Project
 - Ithaka Case Studies in Sustainability 2011: Revenue, Recession, Reliance: Revisiting the SCAI

Ithaka S+R Case Studies in Sustainability (October 2011)

- Beagrie, Neil. Keeping Research Data Safe Factsheet: Cost Issues in Digital Preservation of Research Data, Sept 2010 (http://www.beagrie.com/KRDS_Factsheet_0910.pdf)
- Trustworthy Repositories Audit and Certification: Criteria and Checklist http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf

Recommended Readings

- Strategies for Sustainable Preservation of Born Digital Public Television -- <http://www.thirteen.org/ptvdigitalarchive/>
- Rusbridge, Chris (2006). "Excuse me... Some digital preservation fallacies?" *Ariadne* 46. (<http://www.ariadne.ac.uk/issue46/rusbridge/?ref=cia-team.com>)
- Lavoie, Brian. "The Fifth Blackbird: Some Thoughts on Economically Sustainable Digital Preservation." 2008. *D-Lib Magazine*. www.dlib.org/dlib/march08/lavoie/03lavoie.html.
- Bradley, Kevin. "Defining Digital Sustainability". LIBRARY TRENDS, Vol. 56, No. 1, Summer 2007 ("Preserving Cultural Heritage," edited by Michèle V. Cloonan and Ross Harvey), pp. 148–163. 2007. The Board of Trustees, University of Illinois. To be provided.
- Academy of Motion Picture Arts & Sciences' *The Digital Dilemma* (<http://www.oscars.org/science-technology/council/projects/digitaldilemma/>)
- MOIM <http://wiki.digitalrepositoryauditandcertification.org/bin/view/Main/WebHome>

Class 11 (Nov 29): Preservation Planning -- Lab class

Topics

- DRAMBORA Toolkit
- IT-Innovation Simulation Tool

Required Readings (read all summaries and FAQ, then spend at least 15 minutes looking over the linked details for each of these)

- DRAMBORA: <http://www.repositoryaudit.eu/about/>
- IT-Innovation PrestoPRIME **Simulation** Tool: <http://prestoprime.it-innovation.soton.ac.uk/>

Class 12 (Dec 5): Asset Management and Time-Based Metadata

Topics

- Lecture (CL 90 min)
 - Assessing the Asset Management landscape
 - Types of systems – Content Management System, Collection Management System, Repository, Digital Asset Management System, Media Asset Management System, etc.
 - Common misconceptions
 - Common issues
 - Asset management system selection process
 - Time-Based objects: what makes moving image and sound different from textual, still image and graphic based objects (20 minutes)
 - Workflow Products and Versioning (30 minutes)
 - language, audio configuration, subtitles, closed caption, edits
 - elements vs edited masters vs. Field recordings
 - Implications to preservation
 - Guest Panel
 - Panel of users who have selected and implemented systems (Panel Speakers TBD 120 min)

Required Readings

- Real Story Group, 2011 Content Technology Vendor Map, <http://www.realstorygroup.com/images/>

- [RSG-Subway-2011-Large.jpg](#)
- Real Story Group, *Intro to Digital Asset Management: Just what is a DAM?*
<http://www.realstorygroup.com/Feature/124-DAM-vs.-DM>
- Berry, Matt. Online Video Insider. *Dive Deep With Video Metadata For Major Monetization Opportunities* http://www.mediapost.com/publications/?fa=Articles.showArticle&art_aid=124244&nid=112236
- Windsor, Ralph. *What Does Timeline Metadata Mean For Video Asset Management?* <http://www.daydream.co.uk/video-asset-management-timeline-metadata.asp>

Class 13 (Dec 12): Collaborative approaches to digital preservation and access

Topics

- Review Case Studies of arrangements and agreements between organizations and repositories and group discussion (KVM and CL, 120 min)
- Hannah Frost Guest Lecture via Skype (60 min)
- Discussion about presentations and wrapping up the semester (30 min)

Required Readings

- TBD

Class 14 (Wed Dec 14): Presentations of Final Projects

DIGITAL PRESERVATION CLASS ASSIGNMENTS

ASSIGNMENT 1: REVIEW OF DIGITAL PRESERVATION AND ACCESS RELATED PROJECT (15%)

due Oct 31

2-3 page paper (single spaced) + bibliography/webography and 5-10 minute in-class presentation

Choose from our list of digital preservation and access related projects. Describe the project, the stakeholders, and the duration. Explain the scope of materials covered, and whether they are digitized, born-digital, or a mixture of both. Where applicable, note the following:

- the preservation and access goals of the project and the associated requirements
- system architecture
- selection criteria
- file formats and wrappers
- metadata
- standards referenced
- plans for migration and/or emulation
- user interface (note: Perform sample queries to see what limitations you might face)
- query structure

- quality of files accessible (eg. resolution)
- intellectual property limitations, etc.
- sustainability planning

What were the major revelations of the project, and what are future project plans. Attach a bibliography/webography of works written about the project. Be sure to look at the following: What need or problem is this project trying to solve? What schema, standards or models does it incorporate or is it unique? What special approaches or designs are incorporated that may be interesting or unique? What collections and preservation communities are involved? At what stage is the project and how is it working?

Possible projects to review:

- **PLANETS: Preservation and Long-term Access through Networked Services** (<http://www.planets-project.eu/>)
- **Europeana** -- <http://www.europeana.eu/portal/>
- **MetaArchive Cooperative** “provides low-cost, high-impact preservation services to help ensure the long-term accessibility of the digital assets of universities, libraries, museums, and other cultural heritage institutions” (<http://www.metaarchive.org/>)
- **Rhizome ArtBase** (<http://rhizome.org/artbase/featured/>)
- **PrestoCentre** (<http://www.prestocentre.eu/>)
- **Blue Ribbon Task Force on Sustainable Digital Preservation and Access** (<http://brtf.sdsc.edu/>)
- **JISC: Preservation Eprints Services - Enabling long-term open access to materials in institutional repositories (IRs), 2005-2007** (<http://www.preserv.org.uk/>)
- **Web At Risk: Preserving Our Nation’s Digital Cultural Heritage:** a project of California Digital Library, NYU, and University of North Texas (<http://www.cdlib.org/services/uc3/partners/webatrisk.html>) (see video explanation at http://www.digitalpreservation.gov/partners/web-at-risk/web-at-risk_video.html)
- **North Carolina Geospatial Data Archiving Project:** This project collects and preserves information identifying the geographic location and characteristics of natural or man-made features in the state of North Carolina. Project staff also work with state and county agencies to encourage good data creation practices that will improve preservation and access requirements. (<http://www.lib.ncsu.edu/ncgdap/>)
- **National Geospatial Digital Archive** (<http://www.ngda.org/>)
- **UDFR** (<http://www.udfr.org/>)
- **Persistent Digital Archives and Library System (PeDALS) Project** (<http://pedalspreservation.org/>)
- **CLOCKSS (Controlled LOCKSS)** “is a not for profit joint venture between the world’s leading scholarly publishers and research libraries whose mission is to build a sustainable, geographically distributed dark archive with which to ensure the long-term survival of Web-based scholarly publications for the benefit of the greater global research community.” (<http://www.clockss.org/>)
- **Data Preservation Alliance for the Social Sciences (Data-PASS)** “is a voluntary partnership of organizations created to archive, catalog and preserve

data used for social science research. Examples of social science data include: opinion polls; voting records; surveys on family growth and income; social network data; government statistics and indices; and GIS data measuring human activity.” (<http://www.icpsr.umich.edu/icpsrweb/DATAPASS/>)

- **HathiTrust** is “a digital repository for the nation’s great research libraries, HathiTrust brings together the immense collections of partner institutions. HathiTrust was conceived as a collaboration of the thirteen universities of the Committee on Institutional Cooperation and the University of California system to establish a repository for these universities to archive and share their digitized collections.” (<http://www.hathitrust.org/>)
- **WAISDA? Project** (<http://www.geheugenvannederland.nl/?/en/homepage>)
- **Portico** “is a digital preservation service provided by **ITHAKA**, a not-for-profit organization with a mission to help the academic community use digital technologies to preserve the scholarly record and to advance research and teaching in sustainable ways. (<http://www.portico.org/>)
- **Federal Agencies Digitization Guidelines Initiative**
<http://www.digitizationguidelines.gov>
- **Chronopolis at the San Diego Super Computer Center** (Digital Preservation Demonstration Project) <http://chronopolis.sdsc.edu/>
- **International Conference on the Preservation of Digital Objects iPRES** <http://www.ifs.tuwien.ac.at/dp/ipres2010/related.html>
- **MOIMS (Mission Operations Information Management Services) - Repository Audit and Certification Working Group.** This is a voluntary working group developing an ISO standard for auditing and certification of trusted digital repositories.
- **DBPedia** (<http://dbpedia.org>)
- **Duraspace** (www.duraspace.org)

ASSIGNMENT 2: COMMAND LINE ASSIGNMENT/HOMEWORK (5%)

1. Download the samples folder from dropbox to your computer
 - Generate a MD5 checksum for one file (using md5 command if you are on a Mac, or any application of your choice if you are on a PC - Karen's Hasher is free: <http://www.karenware.com/powertools/pthasher.asp>), and save it someplace on your computer (make sure you note the file name and checksum result)
2. rsync the samples folder from your machine to the server
 - `rsync -e ssh -av [full_path_to_samples_directory/] username@metadatasandbox.com:samples`
3. Find and grep:

Output a text file with a list of all the files in the sample directory

- cd to the samples directory
- find samples -type f > output.txt

Using grep and the output file:

Count the number of files in a specific format of your choice (hint: use the -c option)

4. Use ffmpeg to transcode video and/or audio

ffmpeg is already installed at metadatasandbox.com (If you want to install ffmpeg on your machine for your own use, the steps here might be helpful: <http://howto-pages.org/ffmpeg/> or <http://stephenjungels.com/jungels.net/articles/ffmpeg-howto.html>)

Use ffmpeg to do the following, using the files in the samples folder (name your resulting files the same thing as the original, but add _something to the name, where something is your choice)

- Convert a video/audio using one option (e.g. bitrate)
- Convert a video/audio using two options (e.g. target & framerate)
- Grab a still frame from a video

HINT #1: Google is your friend! Do some research to see how other people have used ffmpeg to do the above

HINT #2: type ffmpeg then return to get the command syntax, type ffmpeg -h to get a long list of options (hint #3 - these are also available on the web)

- Download your results to your local computer to check if they work! Compare them to the original file.

5. Install BagIt and bag the samples directory

- Read about the BagIt specification: <http://www.digitalpreservation.gov/library/resources/tools/docs/bagitspec.pdf>
- Download the BagIt utility from <http://sourceforge.net/projects/loc-xferutils/>
- Put the bagit zip file in your account on the server (use rsync or FTP client).
- unzip the file
- cd to the unzipped bagit directory
- cat README.txt
- cd bin and press return
- ./bag and press return, read the operation instructions
- Since you are going to create a bag, you will use the create command
- ./bag create --help and press return, read the operation instructions
- if your samples/ directory is installed in the your home directory, your command will be
 - ./bag create ~/mybag ~/samples/
- Did it work? if no, check the logs
- If it did work, cd to ~/mybag

- list the contents
- Do the contents conform to the bagit specification?
- cat manifest-md5.txt
- Does the checksum for the file from step 1 match the one you see here?

6. Export your shell text or copy and paste to a text file. If you do this in multiple sessions, make sure you save your output for each session. Send final output to Kara and Chris.

ASSIGNMENT 3: FINAL PROJECT (40%)

one-page proposal due Oct 31

paper due Dec 12; presentation in class Dec 12

term project or paper (8-15 pages single-spaced)

In class presentation (20 minutes including Q&A)

A substantive, in-depth, individual project or paper revolving around any topic we discuss in class. Instructors will provide a list of possible topics, before the end of September, or you may suggest your own. You should try to discuss your proposed topic with one of the instructors in September or October. You must turn in a formal project/paper proposal no later than Oct 31, and one of the instructors must approve your topic and scope before you start work on it.