DIGITAL PRESERVATION
NYU MIAP , Fall 2010
Monday 6-10PM, 721 Broadway, room 674

Version 2.7 11/1/10, latest version at
http://www.nyu.edu/tisch/preservation/program/10fall/digitalpreservation.pdf
Instructors: Howard Besser <howard@nyu.edu>, Chris Lacinak <chris@avpreserve.com>

<u>ABSTRACT</u>: 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 of rights management for digital materials.

<u>EXPECTATIONS</u>: Each student will do extensive readings and three assignments: a group web archiving project, an individual review of an institutional or multi-institutional repository project, 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 (30%); and assignments (35% for the final project, 15% for the review, and 20% for the web archiving project).

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.

Class 1 (Sep 13): Preservation in the Analog World and Intro to Digital
<u>Topics</u>
- Summary & Course Requirements (HB, 20 min)
- Exercise: Interactive review of analog and digital technical history, structures, major developments and terminology. (CL, 1.5 hrs)
- Introduction to basic issues (HB, 100 min)
    - 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)
    - Preservation and Access Principles
    - OAIS Basics
    - Preservation as strategy
    - Digitization Basics, Compression Artifacts
    - Reformatted vs. Born Digital
    - Web Archiving basic Issues (& dates/times for Archive-IT Webinar)

Class 2 (Sep 20): Why is Digital Preservation different?
Topics
- Formal Talk (HB, 2 hrs)
  - Concepts of changes from analog to digital as impacts objects, systems and practices
  - History of Collection DBs in other transitional periods
  - Migration vs. Emulation
- Discussion (CL, 1 hr)
- Student choice for multi-institutional project review

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
- "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

Recommended Readings
- 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. Accessed 6/22/10 at
  http://www.tasi.ac.uk/advice/creating/fformat.html

Class 3 (Sep 27): Practical Work with Digital Files and Embedded Metadata (HB gone)
Topics/Activities
- LAB: Working with files, performing routine tasks of identification, creating checksums, documentation and parsing of metadata for capture, exporting metadata, creating preservation object deliverables. (CL, 2.5 hrs)
- GROUP ACTIVITY: Walking through an XML document (CL, Kara, Walter, 1 hr)

Required Readings
- Meta 101. PhotoMetadata.org. Accessed 6/22/2010
  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

- Lacinak, Chris & Rice, David. "Assessment Report and Initial Recommendations for Embedding Metadata in WAVE Files". Accessed 6/22/2010
http://www.digitizationguidelines.gov/audio-visual/documents/AVPS_Audio_Metadata_Overview_090612.pdf
- 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/

Recommended Readings
- The public forum for the Visual Resources Association Embedded Metadata Working Group
http://metadatadeluxe.pbworks.com/

Class 4 (Oct 4): Preservation Metadata

Topics
- Howard's reports from conferences over past 10 days
    - Report from last week's LC meeting on "Designing Storage Architectures for Preservation Collections"
    - Home Movie Mania at Packard Campus, Culpeper Star Exponent, Sept 23, 2010
    http://www2.starexponent.com/entertainment/2010/sep/23/home-movie-mania-packard-campus-ar-520502/
    - Hot off the press (August 2010): Council on Library & Information Resources The state of Recorded Sound Preservation in the United States: A National Legacy at Risk in the Digital Age (commissioned by LC's National Recording Preservation Board)
    http://www.clir.org/pubs/abstract/pub148abst.html | background info
    http://www.loc.gov/rr/record/nrpb/nrpb-clir.html
    - LC's Topics in Preservation Series -- Optical Scanning Applied to Recorded Sound Preservation and Access: Status and Prospects by Carl Haber, Oct 4, 2010
    https://sas.elluminate.com/site/external/jwsdetect/playback.jnlp?psid=2010-10-04.1141.M.55FA28315348E64C3232ADCE2BC526.vcr&sid=2007050
- Metadata In General, and how it conceptually fits into preservation repositories (HB, 2 hrs)
    - o Review metadata types
    - o Related metadata (Dublin Core, EAD)
    - o 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)
    - o Structural and Administrative Metadata—Metadata Encoding & Transmission Standard -- METS http://www.loc.gov/standards/mets/
    - o Metadata Object Description Schema—MODS http://www.loc.gov/standards/mods/
    - o PREMIS   http://www.oclc.org/research/projects/pmwg/
    http://www.loc.gov/standards/premis/
    - o FOXML-Fedora Dublin Core
    - o UPF
- Technical preservation metadata, what various metadata look like, how it happens, etc (CL, 1.5 hrs)
- Choose groups and topics for Web Archiving assignment
- **NOTE: Wednesday October 6th at 6PM, required Archive-It Webinar**

Required Readings

- Preparation for Web Arhciving assignment
  - Look over the kinds of collections that are in Archive-It, what it does, etc.
    http://www.archive-it.org/
  - Go through the (slightly out-dated) mini-tour of using Arhive-It http://www.archive-it.org/public/tour.html
  - Think of a sample website or subject area that you might want to crawl, and have 1 to 3 seed URLs ready to execute that crawl
- Caplan, Priscilla. Understanding PREMIS, Washington DC: Library of Congress Network Development and MARC Standards Office, Feb 1, 2009 (available at **http://www.loc.gov/standards/premis/**)
- Gilliland, Anne. *Setting the Stage*, in Gill et. al. Introduction to Metadata Version 3.0 (http://www.getty.edu/research/conducting_research/standards/intrometadata/setting.html)
- Look over Jenn Riley's *Glossary of Metadata Standards*, 2010 (http://www.dlib.indiana.edu/~jenlrile/metadatamap/)

Recommended Readings
- Gill, Tony. *Metadata and the Web*, in Gill et. al. Introduction to Metadata Version 3.0 (http://www.getty.edu/research/conducting_research/standards/intrometadata/ metadata.html)

Class 5 (Oct 18): Repository Architecture, Institutional repositories, Scholarly Communication, Digital Curation, Data Curation (HB & CL)
Topics
- Guest Speaker : David Millman, Head of NYU Digital Library Technology Services (http://dlib.nyu.edu/dlts/projects/)
- Staffing Roles, Responsibilities and Stakeholders
- OAIS Model, SIP, DIP, AIP
- Repository attributes
- Open Archive Initiative OAI-PM

Required Readings
- Besser, Howard. Moving from Digital Collections to Interoperable Digital Libraries http://www.firstmonday.org/issues/issue7_6/besser/
- "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
- Data curation: An ecological perspective", by Sayeed Choudury http://crln.acrl.org/content/71/4/194.full

Review
- 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
- NARA: Electronic Records Archive – US National Archives and Records Administration 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 25): Multi-institutional projects & Intellectual Property Issues and Digital Rights Management (HB; CL away)

- Topics

- Student presentations on multi-institutional projects
- Howard's discussion on Multi-institutional projects
- Intellectual Property

Multi-institutional Projects (this is the first of 2 sections with readings for this week)
- Required Readings (spend at least 10 minutes looking at each of these websites)
  - NDIIP: 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
- Required Readings
  - OCLC's Well-Intentioned Practice for Putting Digitized Collections of Unpublished Materials Online", March 28, 2010 (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)
  - Besser, Howard. The Challenge of Copyright & Difficult Materials -- **besser**.tsoa.nyu.edu/howard/Talks/09soima-**copyright**.pdf
  - Howard Besser [written testimony submitted to US Copyright Office Section 108 Study Committee], April 28, 2006 (http://www.section108.gov/docs/Besser_NYU.pdf)
  - Besser, InterPARES study –
    - Howard Besser & Mahnaz Ghaznavi (with Susan Gutman, Luke Meagher and Adele Torrance) Copyright Law regarding Preservation--National Legislation & Implementation Guidelines: Annotated Bibliography, draft version 3, 2005 (http://besser.tsoa.nyu.edu/howard/Papers/interpares-copyright.pdf)
  Recommended Readings
  - Conference Material
    - "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-s-r/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)
  - American Library Association, "Digital Rights Management and Libraries" on the web site of ALA. Accessed 6/22/10 at http://www.ala.org/ala/issuesadvocacy/copyright/digitalrights/index.cfm
  - Duncan, Charles et al. "Digital Rights Management: Final Report". JISC. 2004. Accessed 6/22/10 at http://www.intrallect.com/index.php/intrallect/knowledge_base/general_articles/jisc_drm_study_2004_1

Class 7 (Nov 1): Web Archiving (HB; CL away)
- Video: *Web Archiving*, **Library of Congress, 2009**
  (**http://www.digitalpreservation.gov/videos/webarch09/index.html**)
- Finish Talks from last week (websites often outdated)
- Issues of Versioning/Editions
- Finishing and recap from previo¨s weeks: Copyright, Metadata, OAIS
- Issues with Web Archiving
- Experiences with Web Archiving
- ARCHIVE-IT http://www.archive-it.org/
- Heritrix http://crawler.archive.org/
- Theories and methodologies for collecting and preserving web sites
- Discussion of final projects
- The practicalities of risk assessment, emulation and migration of born digital materials
- Deconstructing and analyzing web site (http://besser.tsoa.nyu.edu/impact/) with multiple operating systems and browsers; diagnosing problems with and with software; diagramming structure, documenting behaviors, capturing screen shots, migrating data between media storage devices.
- ARCHIVE-IT assignment
- <u>Required Readings</u>
  - o Watch entire Webinar video on using Archive-IT http://archive-it.org/public/webinar.html
  - o Internet Archive's Glossary of Web Archiving Terms
    (https://webarchive.jira.com/wiki/display/ARIH/Glossary+of+Web+Archiving+Terms)
  - o 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

  Recommended Readings
  - o 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
  - o 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
  - o 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
  - o Andrew Stawowczyk Long, (Digital Preservation Standards Strategist, National Library of Australia) <u>Long-Term Preservation of Web Archives – Experimenting with Emulation and Migration Methodologies</u> (http://www.netpreserve.org/publications/reports.php)

Class 8 (Nov 8): Moving Image, Sound, and other Complex works
<u>Topics</u>
- Time-Based objects: what makes moving image and sound different from textual, still image and graphic based objects (20 minutes)
- Overview of digital tape-based formats (40 minutes)
  - o History of digital tape formats
  - o Accessing the video as data (ie. DigiBeta v. DV tape v. Files) and the associated implications
- Overview of Digital formats, Codecs and wrappers (60 minutes)
  - o Reformatted Digital v. Born Digital
  - o Overview of relevant formats and codecs
  - o Uncompressed vs. Lossless Compression vs. Lossy Compression
  - o File size calculations
- Overview of relevant emerging standards and initiatives (30 minutes)
- Other Complex Objects: Dynamic vs. static (30 minutes)
  (Internet Art based upon the underlying changing internet, installations that change based upon their

environment, performance art, analogy to music, video games)
- o Emulation issues
- o Media Matters: Collaborating Towards the Care of Time-Based Media Works of Art http://www.tate.org.uk/research/tateresearch/majorprojects/mediamatters/
- o Translating works--Laserdisc & software (A la rencontre de Philippe)
- o http://yalepress.yale.edu/yupbooks/viewbook.asp?isbn=0300056508
- o http://www.brynmawr.edu/llc/help/philippe/philippe.html
- o --http://web.mit.edu/cbbs/media1c.htm
- Workflow Products and Versioning (30 minutes)
  - o language, audio configuration, subtitles, closed caption, edits
  - o elements vs edited masters vs. Field recordings
  - o Implications to preservation

Required Readings
- 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
- 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
- Rice, David & Lacinak, Chris. "Digital Tape Preservation Strategy: Preserving Data or Video?" 2009. http://www.avpreserve.com/dvanalyzer/dv-preservation-data-or-video/
- 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
- Read the first section of the Academy of Motion Picture Arts & Sciences' *The Digital Dilemma* (http://www.oscars.org/science-technology/council/projects/digitaldilemma/)
- Look over the goals of the Academy's Image Interchange Framework project (http://www.oscars.org/science-technology/council/projects/iif.html) and read the 2008 presentation on this (www.digitalpreservation.gov/partners/ampas/high/image_interchange_presentation102008.pdf)

Recommended Readings
- Abby Smith, David Randal Allen, and Karen Allen. Survey of the State of Audio Collections in Academic Libraries. Washington DC: Council on Library and Information Resources. 2004. Accessed 3/3/05 at http://www.clir.org/pubs/abstract/pub128abst.html
- Symposium for the Adoption of **JPEG 2000** in Archives and Libraries.
  - o CNI summary -- Report on the Symposium for the Adoption of **JPEG 2000** in Archives and Libraries. www.cni.org/tfms/2004b.fall/ abstracts/handouts/CNI-murray-adoption.pdf
- Dance Heritage Coalition, "Digital Video Preservation Reformatting Project," 2004 http://www.danceheritage.org/preservation/digital.html
- Howard D. Wactlar and Michael G. Christel, Digital Video Archives: Managing through Metadata http://www.digitalpreservation.gov/library/docs/es_archives.pdf
- Ubois, Jeff. "New approaches to television archiving." First Monday. Volume 10, Number 3. March 2005. Accessed 3/3/05 at http://www.firstmonday.org/issues/issue10_3/ubois/

Class 9 (Nov 15): Personal Archiving and Selection Issues (HB)

Topics
- Personal Archiving
  - o Video: *Why Digital Preservation is Important for You*, Library of Congress, 2010 (http://www.digitalpreservation.gov/videos/personal_archiving/)
  - o Discussion: How do you organize your own digital material?
  - o Loss is part of a natural process
  - o Various Talks from 2010 Personal Archiving conference (http://www.personalarchiving.com/conference2010/)
  - o Gordon Bell's MyLifeBits

- PR (http://www.youtube.com/watch?v=lkNcZUT53VA)
- Bell discussing MyLifeBits (http://vimeo.com/14282561)
- Slides explaining the project (http://www.authorstream.com/Presentation/Goldye-9515-Bell-Life-Bits-Talk-030502-MyLifeBits-Attempting-realize-Memex-Vision-bell-mylifebits-talk-ppt-powerpoint/)

- Selection Issues

Personal Archiving (this is the first of 2 sections with readings for this week)
> Readings (spend at least 10 minutes on each of these bullet points)
>> o Library of Congress website "Personal Archiving: Preserving Your Digital Memories", (http://www.digitalpreservation.gov/you/)
>> o 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
>> o Saving Our Present for the Future: Personal Archiving 2010, The Internet Archive, San Francisco, February 16, 2010 (http://www.personalarchiving.com/conference/)
>> o Video Lecture: *Archiving Digital Photos*, Library of Congress, 2010 (http://www.digitalpreservation.gov/videos/personal_archiving/photos.html)
>> o 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)

Selection Issues
> Readings (spend at least 15 minutes on each of these bullet points)
>> - "Online Australia Publications: Selection Guidelines for Archiving and Preservation by the National Library of Australia" on PANDORA web site. Canberra: National Library of Australia. 2005. Accessed 3/3/05 at http://pandora.nla.gov.au/selectionguidelines.html
>> - "Recommended Appraisal Guidelines for Selecting Born-Digital Master Programs for Preservation and Deposit with the Library of Congress" (PDPTV Selection Paper) http://www.thirteen.org/ptvdigitalarchive/our-work/selection-and-appraisal/

Class 10 (Nov 22): Risk Management and Negotiation with Repositories
Topics
- Review Case Studies of arrangements and agreements between organizations and repositories (CL, 2 hrs)
- Risk Management (HB, 1.5 hrs)
- SIP Negotiation
- Student presentations of Web Archiving project
- Guest speakers on digital preservation projects involving partnering
  - o Hannah Frost
  - o Yvonne Ng

Required Readings
- Anne R. Kenney, et. al. "Preservation Risk Management for Web Resources". D-Lib Magazine. Volume 8 Number 1. January 2002. Accessed 3/3/05 at http://www.dlib.org/dlib/january02/kenney/01kenney.html
- Spend at least 15 minutes looking over "Trustworthy Repositories Audit & Certification: Criteria and Checklist" on Center for Research Libraries Website. CRL and OCLC, 2007. Accessed 6/18/10 http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf

Recommended Readings

- Anne R. Kenney, et. al. "Virtual remote Control: Building a Preservation Risk Management Toolbox for Web Resources". D-Lib Magazine. Volume 10 Number 4. April 2002. Accessed 3/3/05 at http://www.dlib.org/dlib/april04/mcgovern/04mcgovern.html

Class 11 (Nov 29): Sustainability Issues

<u>Topics</u>
- o Sustainability Issues from micro to macro, including infrastructure, content (elastic frameworks) (CL, 2 hrs)
- o Governance and Policy Issues (HB, 1.5 hrs)
- o Presentations on Web Archiving projects

<u>Required Readings</u>
- o 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
- o Strategies for Sustainable Preservation of Born Digital Public Television -- http://www.thirteen.org/ptvdigitalarchive/
- o Rusbridge, Chris (2006). "Excuse me... Some digital preservation fallacies?" *Ariadne* 46. (http://www.ariadne.ac.uk/issue46/rusbridge/?ref=cia-team.com)
- o 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)

<u>Recommended Readings</u>
- o Sustaining Digital Resources: An On-the-Ground View of Projects Today (July 2009) Ithaka Case Studies in Sustainability Project http://www.ithaka.org/ithaka-s-r/strategy/ithaka-case-studies-in-sustainability/report/SCA_Ithaka_SustainingDigitalResources_Report.pdf
- o Lavoie, Brian. "13 Ways of Looking at Digital Preservation". 2004. *D-Lib Magazine*. Accessed 6/22/10 at http://www.dlib.org/dlib/july04/lavoie/07lavoie.html
- o Lavoie, Brian. "The Fifth Blackbird: Some Thoughts on Economically Sustainable Digital Preservation." 2008. *D-Lib Magazine*. Accessed 6/22/10 at www.dlib.org/dlib/march08/lavoie/03lavoie.html.
- o 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. Find on Blackboard site for class.
- o Look over the Participant Presentations given at the Library of Congress "Designing Storage Architectures" meeting held September 27-28 2010 (http://www.digitalpreservation.gov/news/events/other_meetings/storage10/index.html)

Class 12 (Dec 6): Making a Digital Repository operationally work; Auditing & Certification

<u>Topics</u>
- Defining 'Repository' (10 minutes)
- Goals and Objectives of a Repository (20 minutes)
- Staffing: Players, Roles and Responsibilities in and around Repositories (30 minutes)
- Technology: Main components of a system (20 minutes)
- Operation: Establishing and implementing standards, workflow, start-up costs, ongoing costs. (30 minutes)
- Repository Salient Activities (20 minutes)
- Review of TRAC & Drambora (60 minutes)
- Interim/Transitional solutions for smaller institutions, Partnering (30 minutes)

- Why is planning & documentation of steps taken important? Video: ***Team Digital Preservation and the Arctic Mountain Adventure*** (http://www.youtube.com/watch?v=PGFOZLecjTc)

- What is involved in planning and running a digital repository? (establishing and implementing standards, workflow, start-up costs, ongoing costs, etc.?) How can smaller archives relate to large institutional projects?
- Automatic extraction of metadata
- Format migration
- destination formats, metadata and redundancy of data
- costing out equipment, storage, and other technical needs
- choice of destination formats; minimal descriptive standards and redundancy of data
- staffing needs
- What does an archivist need to know? What are various roles for archivists in this environment. What are the actual tasks and job functions?
- Sustainability Issues
- Trustworthiness
  - o   Auditing
  - o   Checksum validation

Required Readings (read both of these summaries, then spend at least 15 minutes looking over the linked details for each of these)
- TRAC http://www.oclc.org/research/news/2007-03-12.htm
- DRAMBORA http://www.repositoryaudit.eu/about/
- Look over *A Guide to Distributed Digital Preservation*. K. Skinner and M. Schultz, Eds. (Atlanta, GA: Educopia Institute, 2010) and read the intro http://www.metaarchive.org/GDDP

Class 13 (Dec 13): Workflows and Systems (CL; HB away)
Topics
- Preservation oriented production
- Types of elements produced (ie. elements vs. final works, versions)
- File naming and Identifiers
- File management
- Submission guidelines
- Quality Control
- Ingest Process
  - o   File Format identification and validation
- Obsolescence Monitoring vs. Normalization
- Prioritization/Selection
- Migration and Refereshing

Readings TBA

Class 14 (Wed Dec 15): Presentations of Final Projects

DIGITAL PRESERVATION
CLASS ASSIGNMENTS

REVIEW OF MULTI-INSTITUTIONAL REPOSITORY or RELATED PROJECT (15%)
due Oct 25
2-3 page paper + bibliography/webography and 5 minute in-class presentation

Choose from our list of multi-institutional repository and related projects. Describe the project, the partners, and the duration. Explain the scope of materials covered, and whether the materials ingested are digitized, born-digital, or a mixture of both. Note who decides what goes into the repository and what criteria are used (selection); what file formats, metadata, and wrappers need to be assembled for ingest, what is the basic architecture of the repository, and what plans exist for migration and/or emulation. If it is primarily an access repository, note the user interface, query structure quality of files accessible (eg. resolution), intellectual property limitations, etc., and perform sample queries to see what limitations you might face. 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 Repositories 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/)
- **PANDORA**: Australia's web archive – National Library of Australia http://pandora.nla.gov.au/index.html
- **NEDLIB - Networked European Deposit Library** http://www.kb.nl/coop/nedlib/
- **MINERVA: Mapping the Internet Electronic Resources Virtual Archive** – Library of Congress http://www.loc.gov/minerva/
- **US National Archives Electronic Records Archives (ERA)** (http://www.archives.gov/era/) See, for example the following article: Digital Library Federation, An Electronic Records Archives (ERA) Update http://www.diglib.org/preserve/ERA2004.htm

- JISC's (UK) **PRESERV Project**: 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/)
- *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/)
- **Memory of the Netherlands (Geheugen van Nederland)** "is a gigantic digital treasury, full of information about the Dutch past. Visitors to the Memory website have access to hundreds of thousands of superb images, recordings, film footage and texts that have been classified under some fifty digital collections. Together, they offer a unique and varied picture of the history and culture of the Netherlands." (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/)
- ****CASPAR: Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval** (http://www.casparpreserves.eu/)

- ****Federal Agencies Digitization Guidelines Initiative**
  http://www.digitizationguidelines.gov
- ****Digital Preservation Europe (DPE)** –
  http://www.digitalpreservationeurope.eu/
- **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

- Other initiatives
  - CURA—Community-University Research Alliance
  - International Network for the Conservation of Contemporary Art (INCCA)
    http://www.incca.org/
  - Variable Media Initiative http://variablemedia.net/
  - Interpares II   http://www.interpares.org/ip2/ip2_index.cfm
  - CEDARS: curl exemplars in digital archives
    http://www.leeds.ac.uk/cedars/ and
  - CAMiLEON: Creative Archiving at Michigan and Leeds Emulating the Old on the New
    http://www.si.umich.edu/CAMILEON/

WEB ARCHIVING group project (20%)
training deadlines in Sept and Oct (see below)
due Nov 29
extensive project with 3-5 page group paper documenting and evaluating what you did

Working in a small group (3-5 students) you will meet with a library, archive, or museum staff member and develop and execute a plan to create an ongoing archive of topical interest to them. This might be an archive on a particular topic of interest, or it might be a comprehensive archive of a portion of their own organization's website. (For examples, see http://www.archive-it.org/public/topic.html) You will meet with relevant organization staff and guide them through the process of selecting the scope of what will be archived (in terms of location, content type, media type, etc.), the frequency of crawls, and any intellectual property and privacy concerns. After you write this up as part of their policy on archiving their website (or on archiving a topic on their behalf) and have it approved by them, you will tackle the technological implementation. You will set up ARCHIVE-IT (http://www.archive-it.org/) to perform a web crawl of the site(s), to archive it, and to provide alerts and statistics. You will choose the seed pages and do several iterative crawls, examining sample archived materials as well as statistics to make sure that the scope of what you get back is neither too large nor too small. Once you have zeroed in on the appropriate seeds and checked this with your "client" organization, you will set frequency of crawl for each seed, choose the scope of the crawl with Scope-IT, set up metadata for each

seed, and set it up for access. You will then have staff at the organization you're working with access it and suggest further modifications. Your group will then summarize and evaluate the experience and present it to the class. In order to complete the assignment before the end of Nov, you will need to attend an Archive-IT Information Webinar in September, and an Arhive-IT Full Production Training session in early Oct. The Webinar and Training Sessions are only given twice per month, so missing these sessions may mean that you cannot complete your assignments.

FINAL PROJECT (35%)
one-page proposal due Nov 1
paper due Dec 15; presentation in class Dec 15
term project or paper (8-15 pages single-spaced)

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 Sept, or you may suggest your own. You should try to discuss your proposed topic with one of the instructors in Sept or Oct. You must turn in a formal project/paper proposal by Nov 1, and one of the instructors must approve your topic and scope before you start work on it.