

NYU MOVING IMAGE ARCHIVING AND PRESERVATION

DIGITAL PRESERVATION

Fall 2014

Tuesdays 6-10PM
721 Broadway, room 652

INSTRUCTOR

Kara Van Malssen
kara@avpreserve.com
Phone: 917-475-9630 x 2
Office Hours: Tuesdays 5:15 – 6pm or by appointment

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 the functions of digital preservation environments and repositories, and what infrastructure, policies, and procedures need to be a part of a repository in order to make it preservation compliant. Students will gain practical skills with identification, analysis, handling, and risk assessment for works as a whole, their component parts, and associated software and metadata. Initiatives and R&D efforts by the national, international, regional, and cooperative organizations will be explored. Digital literacy will be emphasized, and through a combination of lectures, discussion, and hands-on activities and lab exercises, students will develop an increased understanding of digital technologies, ecosystems, and requirements.

OBJECTIVES

Upon completion of this course, students will:

- Be able to understand and articulate the requirements, responsibilities, and functions of digital preservation environments;
- Understand strategies for digital preservation management
- Understand the implications of decisions made today and how these impact future accessibility of content;
- Be able to demonstrate an understanding of digital infrastructures, standards, protocols, and technologies including operating systems (UNIX/Linux), networks (HTTP, STFP, etc), file systems, storage systems, and more;
- Have improved their ability to communicate with technologists

EXPECTATIONS

The material to be covered in this semester is extensive, and cannot all be addressed during in class meeting times. Therefore, students are expected to do substantial readings each week in order to be prepared for class exercises and discussions. Readings assigned for the week should be read in ADVANCE of the class session. Students will also regularly

be assigned tutorials or homework exercises to help practice and develop skills.

There will be three assignments: an individual review of a repository or other digital preservation project, a web archiving group project, and an individual final project (of your choice) to be negotiated with the professor. Students are strongly encouraged to submit a draft of their final project in advance of the deadline. Assignments should be delivered electronically via email to the instructor. Feedback will be provided within 3 weeks after the (on-time) assignment is submitted.

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 (35%); and assignments (35% for the final project, 15% for assignment 1, and 15% for assignment 2). Class sessions will include a mix of lecture, discussion, and group or individual exercises. Your participation in these in class activities is a required part of your grade. Many in-class exercises and homework assignments will require a computer. Student laptops will be required most weeks for in-class exercises, so please plan on bringing your laptop to class. If you can't bring a laptop, you can work with others, or possibly borrow one from the department. Please inform the instructor if you regularly cannot bring a laptop so that arrangements can be made.

Part of your participation grade may include volunteering to research and present briefly on specific topics addressed throughout the semester.

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, or provided by the instructor.

The course will require the use of free or pre-licensed software, which students may be required to download in order to complete in-class and homework assignments. Please come prepared with software downloaded and installed in advance of the class when instructed.

SOCIAL MEDIA

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. Students may be asked to follow conferences throughout the semester. All students are encouraged to sign up for Twitter and begin learning how to use it in advance of the class.

Blogs: Students are expected to start regularly reading the Library of Congress digital preservation blog The Signal (<http://blogs.loc.gov/digitalpreservation/>), and the Open Planets Foundation blog (<http://www.openplanetsfoundation.org/>). Students should do independent research and suggest other blogs for the class to read.

COURSE SCHEDULE

NOTE: Readings and topics subject to change

Week 1 (Sep 2): Intro to Digital Technology & Digital Preservation

Topics

- Introductions (30 min)
- Syllabus summary & course requirements (30 min)
 - Intro to course technologies
 - Getting set up on Google Docs
- Introduction to digital technology and digital preservation (70 min)
- Group Exercise: Interactive review of analog and digital technical history, structures, major developments and terminology. (90 min)

Week 2 (Sep 9): Introduction to OAIS | Introduction to Computing Environments

***** BRING YOUR LAPTOP TO CLASS! *****

Due

- Top three choices for assignment 1
- Software installation:
 - **EVERYONE** (Mac and PC users) must *download* and *install* [VirtualBox](#) and the class Virtual Machine before this class meeting.
 - Virtual Box downloads are available from: <https://www.virtualbox.org/wiki/Downloads>. Choose the one appropriate for your computer's operating system.
 - Read the first few sections of the [user manual](#) to get an idea of what it is.
 - Download the Virtual Machine in the Google Drive class folder. The file is called DigiPresF2013_VM.ova. Please note this is a large file and may take some time to download (i.e. don't start it 5 min before class over wifi...).
 - Password for the machine if it falls asleep: miap2014
 - Install the virtual machine in VirtualBox. Open VirtualBox and then select File > Import Appliance. Choose the class file you downloaded from Google Drive and follow the instructions.
 - **WINDOWS/PC** users please also download and install [Cygwin](#) before class.

Topics

- The OAIS Reference Model (120 min)
 - Essential repository functions
 - Roles, Responsibilities and Stakeholders
 - OAIS functional and information model
- Individual guided exercise: Navigating the command line in UNIX/Linux (100 min)
 - Introduction to the UNIX command line
<http://www.ee.surrey.ac.uk/Teaching/Unix/index.html>

Required Readings

- International Data Corporation, "The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things"

- Interactive version: <http://www.emc.com/leadership/digital-universe/2014iview/index.htm>
- PDF version: <http://idcdocserv.com/1678>
- Full website: <http://www.emc.com/leadership/digital-universe/index.htm>
- Task Force on Archiving of Digital Information, "Preserving Digital Information, Report of the Task Force on Archiving of Digital Information." Donald Waters and John Garrett, Eds. The Commission on Preservation & Access and the Research Libraries Group, May 1996. <http://www.clir.org/pubs/reports/pub63>
- Cornell University Library Research Department. "Setting the Stage: Introduction" and "Obsolescence and Physical Threats" & "Terms & Concepts" -> "Strategies" in Digital Preservation Management: Implementing Short-term Strategies for Long-term Problems. <http://www.dpworkshop.org/dpm-eng/introduction.html>
- Lavoie, Brian "The Open Archival Information System: An Introductory Guide" OCLC, January 2004. www.dpconline.org/docs/lavoie_OAIS.pdf
- Rosenthal, David, et al, "Requirements for Digital Preservation Systems" D-Lib Magazine, November 2005. <http://www.dlib.org/dlib/november05/rosenthal/11rosenthal.html>

Recommended Readings

- Besser, Howard. Moving from Digital Collections to Interoperable Digital Libraries <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/958/879>

Week 3 (Sep 16): Identification & Characterization

*** BRING YOUR LAPTOP TO CLASS! ***

Due

- UNIX tutorial output
- Please *download* DROID in your Virtual Machines before class. This means you should open the VM, open a web browser in it, and download DROID: <http://nationalarchives.gov.uk/documents/information-management/droid-binary-6.1.3-bin.zip>. Then install it. Please let me know if you have any questions.

Topics

- OAIS review
- Characteristics of digital files (60 min)
 - File formats in general
 - File format identification
 - Characteristics of digital files
 - Risk factors
 - Sustainability criteria
- Individual guided exercise: Characterizing digital files (140 min)

Required Readings

- Read sections 1, 2, 3 of "Reference Model for an Open Archival Information System." CCSDS 650.0-M-2. Magenta Book, June 2012. http://www.dpconline.org/docs/lavoie_OAIS.pdf
<http://public.ccsds.org/publications/archive/650x0m2.pdf>
- Jackson, Andrew, "Formats Over Time: Exploring UK Web History" 5 Oct 2012 <http://arxiv.org/pdf/1210.1714v1.pdf>

- Then read: Rosenthal, David, “Formats through time” DSHR’s blog, Oct 9, 2012 <http://blog.dshr.org/2012/10/formats-through-time.html> (read the comments too)
- Then: Rosenthal, David, “Cleaning up the ‘Formats through time’ mess” DSHR’s blog, Oct 13, 2012 <http://blog.dshr.org/2012/10/cleaning-up-formats-through-time-mess.html> (read the comments too)
- 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
- 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>
- Review tools:
 - ExifTool: <http://www.sno.phy.queensu.ca/~phil/exiftool/>
 - MediaInfo: <http://mediainfo.net/en/MediaInfo>
 - ffmpeg: <http://www.ffmpeg.org/>
 - Imagemagick: <http://www.imagemagick.org/script/index.php>
 - PRONOM: <http://www.nationalarchives.gov.uk/PRONOM/Default.aspx>
 - DROID: <http://www.nationalarchives.gov.uk/information-management/manage-information/preserving-digital-records/droid/>

Recommended Readings

- Van Malssen, Kara “Planning Beyond Digitization: Digital Preservation for Audiovisual Collections” Zorgen voor onzichtbare assets, over het behoud van digitale AV-collecties. Annemieke de Jong, editor. Netherlands Institute for Sound and Vision, 2011. <http://www.avpreserve.com/papers-and-presentations/planning-beyond-digitization-digital-preservation-for-audiovisual-collections/>

Week 4 (Sept 23): Intro to Web Archiving | Presentation of Multi-institutional Projects (Assignment 1)

***** BRING YOUR LAPTOP TO CLASS! *****

Due

- Assignment 1 presentation and paper

Topics

- Archive-It Training – Jefferson Bailey, Partner Specialist, Internet Archive (90 min)
- Assignment 1 presentations

Required Readings

- Read about Archive-It <https://www.archive-it.org/>
 - 5 Challenges of Web Archiving: <https://webarchive.jira.com/wiki/display/ARIH/5+Challenges+of+Web+Archiving>
 - Archiving Social Media Sites w/Archive-It: <https://webarchive.jira.com/wiki/pages/viewpage.action?pageId=3113092>
 - Glossary of Web Archiving Terms: <https://webarchive.jira.com/wiki/display/ARIH/Glossary+of+Web+Archiving+Terms>
- Maureen Pennock. March 2013. “Web-Archiving: DPC Technology Watch Report” 13-01. <http://dx.doi.org/10.7207/twr13-01>
- Fino-Radin, Ben. “Sustainable Preservation Practices and the Rhizome ArtBase”

- <http://media.rhizome.org/blog/8332/rhizome-digital-preservation-practices.pdf>
- Ed Summers, “The Web as a Preservation Medium,” *Inkdroid*, November 26, 2013 <http://blogs.loc.gov/digitalpreservation/2013/08/analysis-of-current-digital-preservation-policies-archives-libraries-and-museums/>
- Melena Ryzik, “When Artworks Crash, Restorers Face Digital Test,” *The New York Times*, June 9, 2013 http://www.nytimes.com/2013/06/10/arts/design/whitney-saves-douglas-daviss-first-collaborative-sentence.html?pagewanted=all&_r=0
-

Recommended Readings

- Hany M. SalahEldeen, Michael L. Nelson, “Resurrecting My Revolution: Using Social Link Neighborhood in Bringing Context to the Disappearing Web” 10 Sept 2013 <http://arxiv.org/abs/1309.2648>

Week 5 (Sept 30): Data packaging, Fixity, and Transfer

*** BRING YOUR LAPTOP TO CLASS! ***

Due

- Assignment 2 groups
- In your virtual machines, download the following to your Downloads folder (do not open with Archive Manager when prompted upon download):
 - **Bagger:** <http://sourceforge.net/projects/loc-xferutils/files/loc-bagger/2.1.3/bagger-2.1.3.zip/download>
 - **BagIt libraries:** <http://sourceforge.net/projects/loc-xferutils/files/latest/download?source=files>

Topics

- Guest instructor: Kathryn Gronsbell, AVPreserve
- Lab class: Creating/validating checksums, packaging/structuring files, creating preservation object deliverables. (4 hrs)

Required Readings

- Jefferson Bailey, “Protect Your Data: File Fixity and Integrity,” *The Signal Blog*, April 7, 2014 <http://blogs.loc.gov/digitalpreservation/2014/04/protect-your-data-file-fixity-and-data-integrity/>
- Review the WikiPedia Checksum Entry: <http://en.wikipedia.org/wiki/Checksum>
- Review 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
 - Specification: <http://tools.ietf.org/html/draft-kunze-bagit-10>
- Fixity User Guide v0.4 <http://www.avpreserve.com/wp-content/uploads/2014/07/FixityUserGuide.pdf>
- AVPreserve Fixity study
- NDSA 2014 National Agenda for Digital Stewardship <http://www.digitalpreservation.gov/nds/nationalagenda/index.html>
- Van Malssen, Kara “Tools for Smaller Budgets” Using Open Source and Free Tools

for AV Digital Preservation Workflows.” AV Insider, Issue 2, September 2012. p 13-20. PDF available from: <http://www.prestocentre.org/library/resources/av-insider-2-preservation-times-precarity>

Required Readings

- Liptak, Adam, “In Supreme Court Opinions, Web Links to Nowhere,” *The New York Times*, September 23, 2013. http://www.nytimes.com/2013/09/24/us/politics/in-supreme-court-opinions-clicks-that-lead-nowhere.html?_r=0

Week 6 (Oct 7): NO CLASS in NYC -- MIAP students in Savannah for AMIA conference

Week 7 (Oct 14): NO CLASS -- FALL RECESS

Week 8 (Oct 21): Storage, Networking, Distribution | Intro to Preservation Metadata

Due

- Final project proposals

Topics

- Guest lecture: Dirk Van Dall, MLB Advanced Media
 - Storage media and terms: storage, file system, etc
 - Storage architectures: HSM, distributed, grid, etc.
 - Software and hardware
 - Virtualization, partitioning, etc.
 - Servers, CDNs, delivery
 - Real world examples from large and small organizations
- Intro to preservation Metadata
 - Preservation metadata
 - Databases
 - Data modeling

Required Readings: Storage

- AVPreserve storage vendor profiles: <http://www.avpreserve.com/papers-and-presentations/cloud-storage-vendor-profiles>
- “In Search of the Long-Term Archiving Solution — Tape Delivers Significant TCO Advantage over Disk,” The Clipper Group, December 23, 2010. <http://www.clipper.com/research/TCG2010054.pdf>
- Addis, Matthew, “Wash Day for your Data,” PrestoCentre Blog, February 28, 2012 <http://www.prestocentre.org/blog/wash-day-your-data>
- Rosenthal, David “Storage Roundup July 2012” DSHR’s blog, July 17, 2012 <http://blog.dshr.org/2012/07/storage-roundup-july-2012.html>
- Rosenthal, David, “Cloud vs. Local Storage Costs” DSHR blog, June 27, 2012 <http://blog.dshr.org/2012/06/cloud-vs-local-storage-costs.html>
- Addis, M. Jacyno, M. (2010) “Tools for modelling and simulating migration based preservation.” Deliverable D2.1.2 of the PrestoPRIME project. https://prestoprimews.ina.fr/public/deliverables/PP_WP2_D2.1.2_PreservationModelingTools_R0_v1.00.pdf

Required Readings: Preservation Metadata

- Review the PREMIS Data Dictionary, v. 2.2 (July 2012) <http://www.loc.gov/standards/premis/v2/premis-2-2.pdf>
- Lavoie, Brian and Gartner, Richard, "Preservation Metadata (2nd Edition)" DPC Technology Watch Report, May 2013. <http://dx.doi.org/10.7207/twr13-03>
- Caplan, Priscilla. "Understanding PREMIS," Washington DC: Library of Congress Network Development and MARC Standards Office, Feb 1, 2009 (<http://www.loc.gov/standards/premis/>)

Week 9 (Oct 28): Preservation Metadata Continued | Assignment 2 presentations

Due

- Assignment #2 paper and presentations

Topics

- Assignment 2 presentations (100 min)
- Data modeling exercise

Required Tutorial

- W3Schools XML Tutorial: <http://www.w3schools.com/xml/default.asp>

Required Readings

- W3C XML Essentials: <http://www.w3.org/standards/xml/core>
- "Embedding Metadata in Digital Audio Files" Federal Agencies Audio-Visual Working Group. Version 1.0, approved version, September 15, 2009. http://www.digitizationguidelines.gov/audio-visual/documents/Embed_Guideline_090915r.pdf
- Lacinak, Chris. "Embedded metadata in WAVE files: a look inside issues and tools" http://www.avpreserve.com/wp-content/uploads/2011/02/AVPS_Lacinak_Embedded_Metadata_MLA_2011.pdf

Week 10 (Nov 4): Digital Preservation Application Environments

***** BRING YOUR LAPTOP TO CLASS! *****

Topics

- Digital preservation systems
- Guest Lecture: Ben Fino-Radin, Museum of Modern Art
- Group activity: Using Archivematica (60 min)

Required Readings

- Read about
 - Archivematica: https://www.archivematica.org/wiki/Main_Page
 - Preservica: <http://preservica.com/>
 - Rosetta: <http://www.exlibrisgroup.com/category/RosettaOverview>
 - Fedora: <http://www.fedora-commons.org/>
- Review tool registries:
 - Open Planets Foundation Digital Preservation Tool Registry <http://wiki.opf-labs.org/display/SPR/Digital+Preservation+Tools>
 - NDIIPP Partners Tool Registry <http://www.digitalpreservation.gov/tools/>
 - CDL Micro-services <https://wiki.ucop.edu/display/Curation/Microservices>
 - Coptr: http://coptr.digipres.org/Main_Page

- California Digital Library “UC3 Curation Foundations” Rev 0.13 March 25, 2010
<https://confluence.ucop.edu/download/attachments/13860983/UC3-Foundations-latest.pdf>

Recommended Readings

- Richard N Anderson, “Digital Object Storage and Versioning in Stanford Digital Repository” (January 18, 2012)
<http://lib.stanford.edu/files/Digital%20Object%20Storage%20and%20Versioning.pdf>

These readings are primarily for next week, but I am leaving them here because I want you to be SURE to have read these by next week. It's a lot, so I'm giving you extra time.

- ***Smith, Abby, et al. “Sustainable economics for a digital planet: Ensuring long term access to digital information” Blue Ribbon Task Force on Sustainable Digital Preservation and Access: http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf
- ***Trustworthy Repositories Audit and Certification: Criteria and Checklist
http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf
 - Read up to page 8 in detail
 - Review the criteria
 - Skim everything from p 53+
- TRAC certification of the CLOCKSS Archive (July 2014)
 - CRL certification report (read post + PDF): <http://www.crl.edu/archiving-preservation/digital-archives/certification-and-assessment-digital-repositories/clockss-report>
 - LOCKSS documentation:
http://documents.clockss.org/index.php/Main_Page
 - DSHR’s Blog, “TRAC Audit Lessons Learned” August 2014:
<http://blog.dshr.org/2014/08/trac-audit-lessons.html>
- 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 SCA/Ithaka S+R Case Studies in Sustainability](#) (October 2011)

Recommended Readings

- Lavoie, Brian. "The Fifth Blackbird: Some Thoughts on Economically Sustainable Digital Preservation." 2008. D-Lib Magazine.
www.dlib.org/dlib/march08/lavoie/03lavoie.html

Week 11 (Nov 11): Sustainability, Policy, and Trust

Topics

- Guest instructor: Seth Anderson, AVPreserve (Kara away at SOIMA in Mexico)
- Sustainability and policy for digital preservation environments

Required Readings

- ***Smith, Abby, et al. “Sustainable economics for a digital planet: Ensuring long term access to digital information” Blue Ribbon Task Force on Sustainable Digital Preservation and Access: http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf

- ***Trustworthy Repositories Audit and Certification: Criteria and Checklist
http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf
 - Read up to page 8 in detail
 - Review the criteria
 - Skim everything from p 53+
- TRAC certification of the CLOCKSS Archive (July 2014)
 - CRL certification report (read post + PDF): <http://www.crl.edu/archiving-preservation/digital-archives/certification-and-assessment-digital-repositories/clockss-report>
 - LOCKSS documentation:
http://documents.clockss.org/index.php/Main_Page
 - DSHR's Blog, "TRAC Audit Lessons Learned" August 2014:
<http://blog.dshr.org/2014/08/trac-audit-lessons.html>
- 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 SCA/Ithaka S+R Case Studies in Sustainability](#) (October 2011)

Recommended Readings

- Ng, Rubin, Van Malssen, "Strategies for Sustainable Preservation of Born-Digital Public Television." NDIIPP Preserving Digital Public Television Project, February 2010.
http://www.thirteen.org/ptvdigitalarchive/files/2009/10/PDPTV_SustainabilityStrategies.pdf

Week 12 (Nov 18): Risks, Analysis, and Management

Topics

- Long-term preservation strategies
- Costs of curation
- Business analysis lecture
 - Functional Requirements
 - Use Cases
 - Agile methodologies
- Group Exercises: Technology communication (30 min)

Required Readings

- Sally Vermaaten, Statistics New Zealand; Brian Lavoie, OCLC; Priscilla Caplan, Florida Virtual Campus (FLVC), "Identifying Threats to Successful Digital Preservation: the SPOT Model for Risk Assessment," D-Lib Magazine, September/October 2012, Volume 18, No 9/10. Available from:
<http://www.dlib.org/dlib/september12/vermaaten/09vermaaten.html>
- David Rosenthal, "The Half-Empty Archive," DSHR's Blog, March 2014:
<http://blog.dshr.org/2014/03/the-half-empty-archive.html>
- Jeff Rothenberg "How Far Have we Come and What's Next" April 12, 2012, Future Perfect Conference, Wellington, NZ. YouTube video (45 min):
<http://www.youtube.com/watch?v=2ldbur1qR8I> Slides:
http://www.slideshare.net/FuturePerfect/_jeff-rothenberg-digital-preservation-

- [perspective](#)
- Lavoie, Brian. "The Fifth Blackbird: Some Thoughts on Economically Sustainable Digital Preservation." 2008. D-Lib Magazine. www.dlib.org/dlib/march08/lavoie/03lavoie.html
- Workshop Report: Preservation at Scale. International Conference on Preservation of Digital Objects. Lisbon, 2013. http://ipres2013.ist.utl.pt/workshop5_presentations/PreservationAtScaleWorkshopReport.pdf
- David Rosenthal, "The Half-Life of Digital Formats" DSHR's Blog (November 24, 2010) <http://blog.dshr.org/2010/11/half-life-of-digital-formats.html>
- Costs of Curation

Recommended Readings

- Rusbridge, Chris (2006). "Excuse me... Some digital preservation fallacies?" Ariadne 46. <http://www.ariadne.ac.uk/issue46/rusbridge/?ref=cia-team.com>
- 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
- Academy of Motion Picture Arts & Sciences' *The Digital Dilemma* <http://www.oscars.org/science-technology/council/projects/digitaldilemma/>

Week 13 (Nov 25): NO CLASS: THANKSGIVING HOLIDAY

Week 14 (Dec 2): Business Case for Digital Preservation

Topics

- Business cases
- Institutional readiness
- Group Exercise: Making a business case for digital preservation

Required Readings

- Edward Pinsent, "Digital Preservation: new assessment tools," ULCC DA Blog, 20 November 2013. Available from <http://dablog.ulcc.ac.uk/2013/11/20/spruce-project/>
 - Northumberland estates case study: http://wiki.dpconline.org/index.php?title=Northumberland_estates_case_study
- Anne R. Kenney and Nancy Y. McGovern, "The Five Organizational Stages of Digital Preservation," in *Digital Libraries: A Vision for the Twenty-first Century*, a festschrift to honor Wendy Lougee, 2003. <http://quod.lib.umich.edu/cgi/t/text/text-idx?c=spobooks;idno=bbv9812.0001.001;rgn=div1;view=text;cc=spobooks;node=bbv9812.0001.001%3A11>
- Read through the DPC Digital Preservation Business Case Toolkit: http://wiki.dpconline.org/index.php?title=Digital_Preservation_Business_Case_Toolkit

Week 15 (Dec 9): Last Class!

Topics

- Presentation of final projects

ASSIGNMENTS

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

Due: September 23

Deliverables:

- 2-3 page paper (single spaced) + bibliography/webography
 - At least ONE resource must be something other than the project website itself. It can be a paper that is hosted on the website, an article about the project, etc.
- 5-10 minute in-class presentation

Description:

Choose one project from the list of digital preservation and access related projects below. 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. What were the major revelations of the project, and what are future project plans? Attach a bibliography/webography of works written about the project. Your papers and presentations should address these questions:

- What need or problem is this project trying to solve?
- What collections and communities are involved?
- At what stage is the project and how is it working?
- What methodologies, standards and/or models does it incorporate or is it unique?
- What special approaches or designs are incorporated that may be interesting or unique?

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
- preservation plans
- user interface
- query/search structure
- quality of files accessible (e.g. resolution)
- intellectual property limitations, etc.
- sustainability planning

Please choose from the projects below (projects not on this list must be approved by the professor):

- **Open Planets Foundation** “a community hub for digital preservation” (<http://www.openplanetsfoundation.org/>). If you select this project you must also review the wiki (<http://wiki.opf-labs.org/display/REQ/Home>)
- **DPN:** A federated approach to preservation of the scholarly record. <http://www.dpn.org/>. Be sure to also read:

- <http://www.educause.edu/ero/article/case-building-digital-preservation-network>
- **Digital Public Library of America (DPLA):** <http://dp.la/>
- **Europeana** “explore Europe’s cultural collections” (<http://www.europeana.eu/portal/>). If you select this project, you must include a discussion of research and development topics found in the Professional knowledge-sharing platform (<http://pro.europeana.eu/web/guest/re-use-data>)
- **EU Screen** “discover Europe’s television heritage” (<http://www.euscreen.eu/>). If you select this project, you must include a discussion of research and development topics found on the blog (<http://blog.euscreen.eu/>)
- **Common Crawl:** “a non-profit foundation dedicated to providing an open repository of web crawl data that can be accessed and analyzed by everyone.” <http://commoncrawl.org/>
- **Digital POWRR Project:** “Preserving Digital Objects with Restricted Resources” <http://digitalpowrr.niu.edu/>
- **Rhizome ArtBase** “an online archive of new media art containing some 2159 art works, and growing” (<http://rhizome.org/artbase/featured/>). Be sure to read this report (<http://media.rhizome.org/artbase/documents/Digital-Preservation-Practices-and-the-Rhizome-ArtBase.pdf>)
- **CLOCKSS** “a not-for-profit joint venture between the world’s leading academic 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/clockss/Home>)
- **4C: Collaboration to Clarify the Costs of Curation** “4C will help organisations across Europe to invest more effectively in digital curation and preservation.” <http://www.4cproject.eu/>. Be sure to also look at the Curation Costs Exchange: <http://www.curationexchange.org/>
- **Archive Team:** “a loose collection of rogue archivists, programmers, writers and loudmouths dedicated to saving our digital heritage.” (http://www.archive-team.org/index.php?title=Main_Page)
- **ALFRED:** Archival Federal Reserve Economic Data. (<http://alfred.stlouisfed.org/>)
- **PRONOM:** an online registry for technical information about file formats (<http://www.nationalarchives.gov.uk/PRONOM/Default.aspx>)
- **Digital Preservation Coalition:** an advocate and catalyst for digital preservation in the UK. (<http://www.dpconline.org/>)
- **HathiTrust:** “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/>)
- **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/>)
- **International Internet Preservation Consortium:** A global network of experts archiving the web for future generations. <http://netpreserve.org/>
- **Chronopolis at the San Diego Super Computer Center** “Spanning academic institutions and disciplines, the Chronopolis digital preservation network provides services for the long-term preservation and curation of America's digital holdings. <http://chronopolis.sdsc.edu/>

- **DBPedia** “a community effort to extract structured information from Wikipedia and to make this information available on the Web” (<http://dbpedia.org>) and Freebase “an entity graph of people, places and things, built by a community that loves open data” (<http://www.freebase.com/>)

ASSIGNMENT 2: WEB ARCHIVING GROUP PROJECT (15%)

Due: October 28

Deliverables:

- A web-based report, roughly equivalent to a 3-5 page paper (single-spaced), using any web medium you'd like: create a website, put your paper in a github repository, etc...
- 10-15 min in-class presentation

Description:

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, 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 test 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.

A mandatory Archive-It training session is scheduled for Sept 23 during class time.

ASSIGNMENT 3: FINAL PROJECT / PAPER (35%)

Due:

- One-page proposal due Oct 21
- Presentation on last day of class - Dec 10
- Final paper due Friday Dec 13 (by 6pm)

Deliverables:

- One-page proposal
- Optional: Draft paper/outline due Nov 21
- 8-12 page paper (single spaced) + bibliography/webography
- 20 minute in-class presentation (including Q&A)

Description:

A substantive, in-depth, individual project or paper on digital preservation or access. You should try to discuss your proposed topic with the instructor in September or October. You must turn in a formal project/paper proposal no later than Oct 21, and the instructor must approve your topic and scope before you start work on it. You are strongly encouraged to turn in a draft or outline by November 21.