

# NYU MOVING IMAGE ARCHIVING AND PRESERVATION DIGITAL PRESERVATION — Spring 2015

*Syllabus version 0.993. The current version only vaguely resembles the class as it will be taught. What happens on a weekly basis is likely to morph/move/change. But the assignments will continue to remain the same, and the basic content covered will mostly remain.*

<http://www.nyu.edu/tisch/preservation/program/15spring/15dig-preserv-syllabus.docx>

CINE-GT 1807

Thursdays 6-10PM  
721 Broadway, room 652

## INSTRUCTOR

Howard Besser  
howard@nyu.edu  
Phone: 212-992-9399  
Office Hours: Thursdays 5 – 6 pm 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.

## SYLLABUS

The most current version of the syllabus will always be on Google Docs. It will change frequently. Even if you want to download it to have a local copy, you are encouraged to check the online version each week for the following week's deliverables and readings.

## 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.

**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.

## ADDITIONAL COMMITMENT

*The Personal Digital Archiving Conference is at NYU April 23-26. You **MUST attend at least one of the Workshops on Sunday April 26.** And you should attend the entire conference if you can.*

## COURSE SCHEDULE

### **NOTE: Readings and topics subject to change**

Week 1 (Jan 29): Intro to Digital Technology & Digital Preservation

#### *Topics*

- Introductions
- Syllabus summary & course requirements
  - Intro to course technologies
  - Getting set up on Google Docs or NYU Classes
- Introduction to digital technology and digital preservation
- Components/Tools/Processes
  - Repository (OAIS, SIP, DIP, trustworthiness/audit, software, hardware)
  - Transfer (METS, Bag-It)
  - File analysis tools (MedialInfo, ffmpeg, ExifTool)
  - File validation (DROID, PRENOM)
  - Tracking preservation actions to a file (PREMIS)
  - Checking file integrity (check-sums)
  - Data Modeling
  - Stakeholders, roles, responsibilities (in different configurations)
  - Sustainability (economic, bureaucratic, ...)
- History of major Digital Preservation Efforts
  - RLG/CPA 1996 (<http://www.clir.org/pubs/reports/pub63>)
  - Time & Bits (<http://longnow.org/events/01998/feb/08/time-and-bits/>)
  - NDIIPP (<http://www.digitalpreservation.gov/>) [incl PDPTV (<http://www.thirteen.org/ptvdigitalarchive/files/2009/10/>)]
  - Open Planets (<http://openpreservation.org/>)
  - Long Now Foundation (<http://longnow.org/>)
  - BRTF ([http://brtf.sdsc.edu/biblio/BRTF\\_Final\\_Report.pdf](http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf))
- The Storage Problem “red herring”
- Group Exercise: Interactive review of analog and digital technical history, structures, major developments and terminology.

Week 2 (Feb 5): Introduction to OAIS | Introduction to Computing Environments

#### *Due*

- Top three choices for assignment 1 BEFORE next class

#### *Topics*

- [Summary Slides for last week and today](http://besser.tsoa.nyu.edu/howard/Talks/) (<http://besser.tsoa.nyu.edu/howard/Talks/>)
- Howard's Report from ALA, [Digital Content Working Group \(Media sub-group\)](#)
- The OAIS Reference Model
  - Essential repository functions
  - Roles, Responsibilities and Stakeholders
  - OAIS functional and information model
- Individual guided exercise: Navigating the command line in UNIX/Linux
  - Introduction to the UNIX command line (0, 1, 2 in class)
    - <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><http://www.dlib.org/dlib/november05/rosenthal/11rosenthal.html>
- Lavoie, Brian "The Open Archival Information System: An Introductory Guide" OCLC, January 2004. [www.dpconline.org/docs/lavoie\\_OAIS.pdf](http://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>
- 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 (Feb 12): Identification & Characterization

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

#### Due

- Read Bertram Lyon's [An Introduction to Using the Command Line Interface to Work with Files and Directories](http://www.avpreserve.com/papers-and-presentations/an-introduction-to-using-the-command-line-interface-to-work-with-files-and-directories/) (<http://www.avpreserve.com/papers-and-presentations/an-introduction-to-using-the-command-line-interface-to-work-with-files-and-directories/>)
- Software installation:
  - **EVERYONE** (Mac and Windows 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 NYU Classes folder. The file is called MIAP\_DigiPres\_Spring2015.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.
- 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>. Save it.
- 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://www.dpconline.org/docs/lavoie_OAIS.pdf)  
<http://public.ccsds.org/publications/archive/650x0m2.pdf>
- Phillips, Bailey, and Owens, "The NDSA Levels of Preservation: An Explanation and Uses" (2013): [http://www.digitalpreservation.gov/ndsaworking\\_groups/documents/NDSA\\_Levels\\_Archiving\\_2013.pdf](http://www.digitalpreservation.gov/ndsaworking_groups/documents/NDSA_Levels_Archiving_2013.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](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>

#### *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/>  
<https://wiki.ucop.edu/display/Curation/Microservices>

#### *Topics*

- In preparation for assignment due next week: What is a repository? What are criteria for evaluation?
- Study Groups for technology?
- Pixar scare story [https://www.youtube.com/watch?v=EL\\_g0tyaleE](https://www.youtube.com/watch?v=EL_g0tyaleE)
- UNIX tutorial output - sections 1 and 2
  - <http://www.ee.surrey.ac.uk/Teaching/Unix/index.html>

- OAIS Review (Kara's slides)
- Characteristics of digital files
  - File formats in general
  - File format identification
  - Characteristics of digital files
  - Risk factors
  - Sustainability criteria
  - Exercise: characterizing digital files
- Archive-It background
  - Web archiving in general
  - IA background and WayBack Machine
  - Other IA cooperative ventures
    - Reed Tech & Internet Archive Announce Agreement (<http://www.reedarchives.com/2013/reed-tech-internet-archive-announce-agreement/>)
  - WAS and Archive-It
    - University of California Libraries to partner with Archive-It (<https://archive-it.org/blog/post/university-of-california-libraries-to-partner-with-archive-it>)
- Review tools:
  - ExifTool: <http://www.sno.phy.queensu.ca/~phil/exiftool/>
  - MediaInfo: <http://mediaarea.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/>

Week 4 (Feb 19): Intro to Web Archiving | Presentation of Multi-institutional Projects (Assignment 1)

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

#### *Due*

- Assignment 1 presentation and paper
- UNIX tutorial output - sections 3 and 4
  - <http://www.ee.surrey.ac.uk/Teaching/Unix/index.html> (do these exercises)

#### *Topics*

- Archive-It Training – Jefferson Bailey, Partner Specialist, Internet Archive (6:30. 90 min)
- Assignment 1 presentations
- News articles/resources
  - [In Town Where Pitcher's 266 Wins Were Part of Fabric, One Big Loss; Bob Feller Museum in Van Meter, Iowa, Will Become City Hall](#), NY Times, Feb 16, 2015
  - [A New Copyright Complaint Against Richard Prince](#), NY Times, Feb 16, 2015
  - [Toy Museum Announces Hall of Fame for Video Games](#), NY Times, Feb 17, 2015

### 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](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>  
<http://www.digitalpreservation.gov/formats/index.shtml>

### Feb 26 No Class (MIAP Students in Culpeper)

Week 5 (Mar 5): 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>
  - Fixity (v.0.4): [www.avpreserve.com/tools/fixity/](http://www.avpreserve.com/tools/fixity/) for your OS (and place this on your computer (not VM))
    - Download and install md5deep and hashdeep (v.4.4) <http://md5deep.sourceforge.net/>
      - First, open up the terminal on your virtual machine and install autoconf and g++ as super user (sudo) using the advanced packaging tool (apt). That's what these two commands do. When it asks for a password, type miap2014, and for each [y/n] prompt type y and hit enter:
        - **sudo apt-get install autoconf**
        - **sudo apt-get install g++**
      - With that done, download the latest md5deep/hashdeep source code
      - directly using this:

- **wget**
  - **<https://github.com/jessek/hashdeep/archive/release-4.4.tar.gz>**
  - Then extract the files using this:
    - **tar -xvf release-4.4.tar.gz**
  - Then change to the new directory:
    - **cd hashdeep-release-4.4**
  - and then run these four commands in order:
    - **sh bootstrap.sh**
    - **./configure**
    - **make**
    - **sudo make install**

### Topics

- Guest instructor: Kathryn Gronsbell, AVPreserve
- Lab class: Creating/validating checksums, fixity monitoring, 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](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 (when published)
- NDSA 2014 National Agenda for Digital Stewardship <http://www.digitalpreservation.gov/ndsas/nationalagenda/index.html><http://www.wired.com/threatlevel/2009/10/metadata/>
- 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>
- 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](http://www.nytimes.com/2013/09/24/us/politics/in-supreme-court-opinions-clicks-that-lead-nowhere.html?_r=0)
- Read through the NYARC Web Archiving Documentation <http://wiki.nyarc.org/web-archiving>

## Week 6 (Mar 12): Storage, Networking, Distribution | Intro to Preservation Metadata

*Due*



- An example report for your Fixity homework
  - Download Fixity for your operating system
  - Run it according to the directions that Kathryn provided
  - Forward 1 report to Howard before class

### Topics

- **Might not happen: 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>
- Altman, et al., NDSA Storage Report: Reflections on National Digital Stewardship Alliance Member Approaches to Preservation Storage Technologies, D-Lib Magazine, May/June 2013 <http://www.dlib.org/dlib/may13/altman/05altman.html><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\\_PreservationModellingTools\\_R0\\_v1.00.pdf](https://prestoprimews.ina.fr/public/deliverables/PP_WP2_D2.1.2_PreservationModellingTools_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>[http://www.getty.edu/research/publications/electronic\\_publications/intrometadata/pdf.html](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/>

### Mar 19: No Class – Spring Break

Week 7 (Mar 26): Preservation Metadata Continued

Final project proposals due

### Topics

- Finishing presentation of Assignment #1
- Status Reports from each Web Archiving project
- Report from study groups
- PDA Conference update (including Sunday Archivematica sessions, helping)
- Updates from recent news:
  - [JHOVE format validation stewardship moves to Open Preservation Foundation, and releases their stabilization plan](#) (today)
  - [Magna Carta edition found in Sandwich archive scrapbook](#), BBC, Feb 8, 2015
  - [It's a Man's \(Celluloid\) World: On-Screen Representations of Female Characters in the Top 100 Films of 2014](#), Martha M. Lauzen, SDSU
  - [Science Museums Urged to Cut Ties With Kochs](#), March 24, 2015; [Why Is Climate Denier David Koch on the Board of Top Science Museums? Letter Urges Cutting of Ties](#), Democracy Now!, March 26, 2015
  - [Howard Guttenplan, 80, Art-Film Innovator, Dies](#), NY Times, March 26, 2015
  - [Jytte Jensen obituary](#), NY Times, March 26, 2015
  - [Museum in Little Italy Seeks to Evict a Living Link to the Past](#), NY Times, March 25, 2015
  - [Tunisia's Bardo Museum in Symbolic Reopening and March After Attacks](#), NY Times, March 24, 2015
  - [Tunisians Assure Star Wars Sets Safe From Islamic State](#), March 26, 2015
  - Photos from Howard's trip
- Data modeling exercise
  - Reference: <http://www.agiledata.org/essays/dataModeling101.html>

### 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](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](http://www.avpreserve.com/wp-content/uploads/2011/02/AVPS_Lacinak_Embedded_Metadata_MLA_2011.pdf)

## Week 8 (Apr 2): Digital Preservation Application Environments, Business Case for Digital Preservation

### Topics

- Digital preservation systems
- Institutional Readiness
- Business analysis/vocabulary/lingo
  - business case
  - deliverables
  - Functional Requirements
  - Use Cases

- Agile methodologies
- Group Exercises: Technology communication
- Group Exercise: Making a business case for digital preservation
- Digital Preservation Outreach Education (DPOE)  
<http://www.digitalpreservation.gov/education/curriculum.html>
- News
  - [Jay Z Enters Streaming Music With Artist-Owned Service](#), NY Times Bits, Mar 31, 2015
  - [China Appears to Attack GitHub by Diverting Web Traffic](#), NY Times, Mar 31, 2015

### Required Readings

- Read about
  - Archivemata: [https://www.archivemata.org/wiki/Main\\_Page](https://www.archivemata.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>
  - NDIPP 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](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 three weeks from now, but I am leaving them here because I want you to be SURE to have read these by 3 weeks from now. 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](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](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](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>[http://brtf.sdsc.edu/biblio/BRTF\\_Final\\_Report.pdf](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 SCA/Ithaka S+R Case Studies in Sustainability](#) (October 2011)

*Recommended Readings* <http://archive.org/details/personaldigitalarchiving2012pt2>

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

Week 9 (Apr 9): | Assignment 2 presentations—NY State Newspaper archiving, NY State College Newspaper archiving (*maybe*)

*Due*

- Assignment #2 paper and presentations
- Business Case presentations

*Topics*

- Assignment 2 presentations
- Business Case presentations
- Archive-It updates (and transition to Bobst administration)
- Archive-It use (250G, 2,500 documents/group)
- [PDA update](#) ([preview of Bushey](#))
- Heavy readings for 2 weeks ahead
- [FIAF Congress](#)
- 

Week 10 (Apr 16): Work Session (no Howard), but the written versions of Assignment #2 must be turned in by end of classtime

Week 11 (Apr 23): Sustainability, Policy, and Trust

**Note:** *Personal Digital Archiving Conference is at NYU April 23-26. You MUST attend at least one of the Workshops on Sunday April 26. And you should attend the entire conference if you can.*

*Topics*

- Guest instructor: Seth Anderson, AVPreserve
- 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](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](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](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>[http://brtf.sdsc.edu/biblio/BRTF\\_Final\\_Report.pdf](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 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](http://www.thirteen.org/ptvdigitalarchive/files/2009/10/PDPTV_SustainabilityStrategies.pdf)

#### Week 12 (Apr 30): Risks, Analysis, and Management

##### *Topics*

- Long-term preservation strategies
- Costs of curation

##### *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 18:9/10, Sept/Oct 2012 <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](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](http://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](http://ipres2013.ist.utl.pt/workshop5_presentations/PreservationAtScaleWorkshopReport.pdf)[http://lib.stanford.edu/files/Digital Object Storage and Versioning.pdf](http://lib.stanford.edu/files/Digital_Object_Storage_and_Versioning.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](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/>

#### Additional Readings on 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](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](http://wiki.dpconline.org/index.php?title=Digital_Preservation_Business_Case_Toolkit)  
<http://omeka.org/>

<http://omeka.org/>

#### News

- Abel Ferrara Fights to Stop Re-Edited Strauss-Kahn Film, NY Times, March 26, 2015 [www.nytimes.com/2015/03/27/movies/abel-ferrara-fights-to-stop-re-edited-strauss-kahn-film.html](http://www.nytimes.com/2015/03/27/movies/abel-ferrara-fights-to-stop-re-edited-strauss-kahn-film.html)
- On Screen, Not All Artwork Is as Famous as It Appears, NY Times, March 26, 2015 <http://www.nytimes.com/2015/03/29/movies/on-screen-not-all-artwork-is-as-famous-as-it-appears.html>
- Preserving the Ghastly Inventory of Auschwitz, NY Times, April 15, 2015 <http://www.nytimes.com/2015/04/16/arts/international/at-auschwitz-birkenau-preserving-a-site-and-a-ghastly-inventory.html>
- Seller in Canceled Internment-Camp Auction Comes Forward, NY Times, April 17, 2015 <http://www.nytimes.com/2015/04/18/arts/design/seller-in-canceled-internment-camp-auction-comes-forward.html>
- America's 'national library' is lacking in leadership, yet another report finds, Washington Post, March 31, 2015 <http://www.washingtonpost.com/entertainment/museums/americas-national-library-is-behind-the-digital-curve-a-new-report-finds/2015/03/31/fad54c3a-d3fd-11e4-a62f->

- [ee745911a4ff\\_story.html](http://www.nytimes.com/2015/04/05/opinion/sunday/digital-neglect-at-the-library-of-congress.html)
- Digital Neglect at the Library of Congress, NY Times Editorial, Sunday April 4, 2015  
<http://www.nytimes.com/2015/04/05/opinion/sunday/digital-neglect-at-the-library-of-congress.html>
- Archiving: Maintaining Relevance for the Future: What the Library of Congress can teach us about preserving digital media, TV Technology, Jan 15, 2015  
<http://www.tvtechnology.com/digital-journal/0116/archiving-maintaining-relevance-for-the-future/273914>
- Comparing Formats for Video Digitization, Kate Murray, The Signal, Dec 3, 2014  
<http://blogs.loc.gov/digitalpreservation/2014/12/comparing-formats-for-video-digitization/>

Other Resources to watch:

- PDA conference (<http://personaldigitalarchiving.com/>)
  - Howard & Rick Talk (<http://www.slideshare.net/footage/personal-digital-archiving-2015-keynote-talk-by-howard-besser-rick-prelinger>)
- FADGI's Audiovisual Working Group <http://www.digitizationguidelines.gov/audio-visual/> documents currently under review:
  - **Digital File Formats for Videotape Reformatting:** The project to compare video formats for reformatting is being led by the National Archives and Records Administration with significant input from the Library of Congress. The resulting matrixes offer comparisons of the *wrappers* AVI, MOV (QuickTime), Matroska, MXF, and MPEG-2 (ad hoc file wrapper), and the following *encodings*: uncompressed (various types), lossless JPEG 2000, ffv1, and MPEG-2 (encoding). This work compliments the FADGI Audio-Visual Working Group's active contributions to the finalization of the [MXF AS-07 application specification](#) for preservation and archiving, a "pre-version" of which has been adopted by the Library of Congress as their video preservation format. Nevertheless, as the finalization of AS-07 continues, members of the Working Group and others have expressed interest in evaluating alternate digital file formats. These alternate formats may be appropriate to certain classes of content, for use as interim measures, or may be preferred by organizations where they make a better fit for existing systems and architectures. Home page: <http://www.digitizationguidelines.gov>
  - **Creating and Archiving Born Digital Video:** The born digital video exploration is being led by the Library of Congress. The four documents that comprise *Creating and Archiving Born Digital Video* provide practical technical information for both file creators and file archivists to help them make informed decisions when creating or archiving born digital video files and to understand the long term consequences of those decisions. Eight case histories document aspects of the current state of practice in six federal agencies working with born digital video. These case histories not only document deliverables and specifications but also tell the story of each project, and provide background information about the institution and the collection, as well as lessons learned. As the case histories developed, a set of high level recommended practices emerged from the collective project experiences. These recommended practices are intended to support informed decision-making and guide file creators and archivists as they seek out processes, file characteristics, and other practices that will yield files with

the greatest preservation potential. Home page:

[http://www.digitizationguidelines.gov/guidelines/video\\_bornDigital.html](http://www.digitizationguidelines.gov/guidelines/video_bornDigital.html)

- Digital Preservation Management Tools and Techniques  
<http://dpworkshop.org/workshops/management-tools>

Week 13 (May 7): Last Class!

#### *Topics*

- Presentation of final projects

## ASSIGNMENTS

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

Due: Feb 19

#### **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/http://commoncrawl.org/>
- **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.archiveteam.org/index.php?title=Main\\_Page](http://www.archiveteam.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/>)

#### Assignments:

- zz: EUScreen
- zz: Rhizome
- zz: Chronopolis
- zz: DPLA
- zz: Archive Team
- zz: Europeana
- zz: HathiTrust
- zz: DBPedia
- zz: 4C
- zz: IIPC

#### ASSIGNMENT 2: WEB ARCHIVING GROUP PROJECT (15%)

**Due:** April 9

#### **Deliverables:**

- A report, roughly equivalent to a 3-5 page paper (single-spaced)
- 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 (including by adding descriptive metadata for the collection). 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 Feb 19 during class time.

### **ASSIGNMENT 3: FINAL PROJECT / PAPER (35%)**

#### **Due:**

- One-page proposal due Mar 26
- Presentation on last day of class – May 7
- Final paper due end of day Sunday May 10

#### **Deliverables:**

- One-page proposal
- 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 Feb or March. You must turn in a formal project/paper proposal no later than Mar 26, and the instructor must approve your topic and scope before you start work on it.