DIGITAL PRESERVATION
NYU MIAP Fall 2012
Tuesdays 6-10PM, 721 Broadway, room 652

INSTRUCTOR

Kara Van Malssen
kara@avpreserve.com
Phone: 917-475-9630
Skype: kara_vanmalssen
Jabber/GTalk: kvanmalssen@gmail.com
Office Hours: Tuesdays 5:30 – 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 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 the Library of Congress and other national archives, digital libraries and more will be explored as examples of the architecture and attributes of digital repositories. Digital literacy will be emphasized, and students will develop an increased understanding of digital technologies, ecosystems, and tools. The class will include lectures, discussion, and hands-on lab exercises.

EXPECTATIONS

Each student will do extensive readings each week. Readings for the week should be read in ADVANCE of the class session. Students will 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 digital preservation workflow assignment, and a final project (of your choice) to be negotiated with the professor.

Attendance at all classes is expected; more than one unexcused absence will affect grading. Grades will be based on a combination of class preparedness and participation (40%); and assignments (40% for the final project, 15% for assignment 1, and 5% for assignment 2). 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, however, please inform the instructor if you regularly cannot bring a laptop.

Part of your participation grade will including volunteering to research and present briefly on specific topics addressed throughout the semester. Volunteering and completion of homework is mandatory in order to receive an A participation grade.

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.

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.

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.

Stack Exchange: A number of digital preservation related discussions have come up on the Stack Exchange Libraries and Information Science site. We will be following this closely, and may discuss questions that have come up here, or post answers ourselves: http://libraries.stackexchange.com/

Support your Digital Preservation community: We will be working supporting and improving documentation and development around digital preservation throughout the semester. Students should bookmark this wiki and follow these developments regularly. There will be opportunities to contribute to these efforts as in-class exercises and homework assigments. More information is available from: http://wiki.opf-labs.org/display/SPR/Support+your+digital+preservation+community

COURSE SCHEDULE

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

Class 1 (Sep 4): Intro to Digital Technology & Digital Preservation
Topics
- Introductions (30 min)
- Summary & Course Requirements (30 min)
- Introduction to Digital Technology and Digital Preservation Today (70 min)
    - Video: *IDC 2011 Digital Universe Study: Extracting Value from Chaos*. Executive Summary, June 2011. http://www.emc.com/collateral/demos/microsites/emc-digital-universe-2011/index.htm
    - Video: *Why Digital Preservation is Important for Everyone*, Library of Congress, 2010 (http://www.youtube.com/watch?v=qEmmeFFafUs)
    - Digitization and born-digital Digital vs analog preservation
    - Risk factors for digital objects

- ○ Digital preservation basic requirements
- ○ General preservation strategies
● Group Exercise: Interactive review of analog and digital technical history, structures, major developments and terminology. (90 min)
● Next 4 weeks: Each group will present in class a visual or media-based example or explanation of one of the following:
  - ○ Group 1 - HTML5
  - ○ Group 2 - Linked Data / Semantic Web
  - ○ Group 3 - Big Data
  - ○ Group 4 - Cloud

Class 2 (Sep 11): Personal Digital Archives

<u>Due</u>
Top three choices for assignment 1

<u>Topics</u>
- ● Group 1 technology examples (homework)
- ● Guest lecture – Rufus de Rham (NYU MIAP) (45 min)
- ● Exercise: Personal Digital Assessment (150 min)
- ● Student choices for assignment 1

<u>Required Readings/Videos</u>

- ● How the Internet Works in 5 minutes

General Digital Preservation:
- ● 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. 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
- ● Besser, Howard. "Digital Longevity" in Handbook for Digital Projects: A Management Tool for Preservation and Access, http://www.nedcc.org/resources/ digitalhandbook/ix.htm
- ● Lavoie, Brian. "13 Ways of Looking at Digital Preservation". 2004. *D-Lib Magazine*. http://www.dlib.org/dlib/july04/lavoie/07lavoie.html
- ● Read the IDC Digital Universe Studies
  - ○ 2011: http://www.emc.com/collateral/demos/microsites/emc-digital-universe-2011/index.htm
    2008: http://www.emc.com/collateral/analyst-reports/diverse-exploding-digital-universe.pdf

Personal Digital Archiving:

- Marshall, Cathy, "Rethinking Personal Digital Archiving" Part 1 http://www.dlib.org/dlib/march08/marshall/03marshall-pt1.html and Part 2 http://www.dlib.org/dlib/march08/marshall/03marshall-pt2.html, both *D-Lib Magazine,* March/April 2008.
- Watch: Stan James "How my family archives affected others" http://archive.org/details/personaldigitalarchiving2012pt1
- Library of Congress website "Personal Archiving: Preserving Your Digital Memories", (http://www.digitalpreservation.gov/you/)
- Trapani, Gina. *Fast Company*, "How to Backup Your Social Media Life." October 2010. http://www.fastcompany.com/1695794/how-to-backup-your-social-media-life
- Video: *Why Digital Preservation is Important for You*, Library of Congress, 2010 (http://www.youtube.com/watch?v=VWkPufGDA6o

Recommended Readings
- "Preserving Digital Information, Report of the Task Force on Archiving of Digital Information" (1996) Donald Waters and John Garrett, Eds. The Commission on Preservation & Access and the Research Libraries Group. http://www.oclc.org/programs/ourwork/past/digpresstudy/final-report.pdf
- Rice, David & Lacinak, Chris. "Digital Tape Preservation Strategy: Preserving Data or Video?" 2009. http://www.avpreserve.com/dvanalyzer/dv-preservation-data-or-video/
- Kurt D. Bollacker. "Avoiding a Digital Dark Age," *American Scientist*, May-June 2010. http://www.americanscientist.org/issues/pub/2010/3/avoiding-a-digital-dark-age
- Report and analysis on the training needs survey from DigCurV http://www.digcur-education.org/eng/Resources/Report-and-analysis-on-the-training-needs-survey

Class 3 (Sep 18): Introduction to OAIS | Introduction to Computing Environments
BRING YOUR LAPTOP TO CLASS!! BE SURE YOU HAVE VIRTUALBOX AND THE VM INSTALLED!!

Topics
- Group 2 technology presentation
- Repositories and Standards (120 min)
  - Essential Repository Functions: What does a Digital Repository do?
  - Staffing Roles, Responsibilities and Stakeholders
  - OAIS Model, SIP, DIP, AIP
- Navigating the command line in UNIX/Linux
  - Group exercise: Introduction to the UNIX command line http://www.ee.surrey.ac.uk/Teaching/Unix/index.html

Required Software Installation
- **EVERYONE** must *download* and *install* VirtualBox before this class meeting. Downloads are available from: https://www.virtualbox.org/wiki/Downloads. I will explain VirtualBox in class, but I also suggest you read the first few sections o tf 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 DigiPresF2012_VM.ova.

<u>Required Readings</u>
- Rosenthal, David, et al, "Requirements for Digital Preservation Systems" D-Lib Magazine, November 2005. 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
- 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
- "OAIS 2012 Update" Digital Preservation Seeds Blog, July 30, 2012 http://digitalpreservation.nl/seeds/standards/oais-2012-update/

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

Class 4 (Sep 25): Characteristics of Digital Files | Technical & Preservation Metadata
BRING YOUR LAPTOP TO CLASS

<u>Topics</u>
- Group 3 technology presentation
- Digital files lecture (40 min)
  - File formats in general
  - Characteristics of digital files
  - Risk factors
  - Sustainability criteria
- Metadata & Packages Overview (70 min)
  - Review SIP, AIP, DIP concepts
  - Characteriziation, significant properties and representation information
  - Review metadata types: descriptive, administrative, technical, rights, etc
  - PREMIS http://www.loc.gov/standards/premis/
  - Metadata Encoding & Transmission Standard -- METS http://www.loc.gov/standards/mets/
- Group Exercise: PREMIS (90 min)

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

<u>Required Readings</u>
- Gilliland, Anne. *Setting the Stage,* in Gill et. al. Introduction to Metadata Version 3.0 (http://www.getty.edu/research/publications/electronic_publications/intrometadata/pdf.html)

- Caplan, Priscilla. "Understanding PREMIS," Washington DC: Library of Congress Network Development and MARC Standards Office, Feb 1, 2009 (http://www.loc.gov/standards/premis/)
- W3C XML Essentials (http://www.w3.org/standards/xml/core)
- "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
- Dappart, A. & Enders, M. (2010). Digital Preservation Metadata Standards. *Information Standards Quarterly*, Spring. http://www.loc.gov/standards/premis/FE_Dappert_Enders_MetadataStds_isqv22no2.pdf
- 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

Class 5 (Oct 2): Practical Work with Digital Files and Embedded Metadata
BRING YOUR LAPTOP TO CLASS

Topics/Activities
- Guest instructors: Chris Lacinak, Seth Anderson (Kara away at iPRES)
- Lab class: Working with files, performing routine tasks of identification, creating checksums, documentation and parsing of metadata for capture, exporting metadata, creating preservation object deliverables. (4 hrs)

Required Readings
- 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
- Meta 101. PhotoMetadata.org
  http://www.photometadata.org/META-101
  http://www.photometadata.org/META-101-metadata-Q-and-A
  http://www.photometadata.org/META-101-metadata-classes
  http://www.photometadata.org/META-101-metadata-types
  http://www.photometadata.org/META-101-metadata-structure
- Embedded Metadata Manifesto website: http://www.embeddedmetadata.org/
  - Read the Manifesto, Why Metadata Matters, and Use Cases
- 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
- 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/

Review

- WikiPedia Checksum Entry: http://en.wikipedia.org/wiki/Checksum
- MediaInfo: http://mediainfo.sourceforge.net/en
- BagIt—simple protocol for transferring files and metadata
  - Video: Bagit: Transferring Content for Digital Preservation, Library of Congress, 2009 (http://www.digitalpreservation.gov/videos/bagit0609.html?rec=4629)

- ○ Summary (http://www.digitalpreservation.gov/news/2008/20080602news_article_bagit.html)
        - ○ Specifications (http://www.digitalpreservation.gov/library/resources/tools/docs/bagitspec.pdf)
- ExifTool: http://www.sno.phy.queensu.ca/~phil/exiftool/
- DV Analyzer: http://www.avpreserve.com/dvanalyzer/
- BWF MetaEdit: http://sourceforge.net/projects/bwfmetaedit/
- The public forum for the Visual Resources Association Embedded Metadata Working Group http://metadatadeluxe.pbworks.com/

Class 6 (Oct 9): Digital Preservation Workflows
*** EVERYONE BRING YOUR LAPTOP THIS WEEK. IF YOU CAN'T COMPLETE THE IN-CLASS ASSIGNMENT ON YOUR OWN THIS WEEK, YOU WILL HAVE TO DO IT FOR HOMEWORK ***

Topics
- Check in on Twitter, final project topics, upcoming due dates, outstanding questions, date for final presentations
- Group exercise: Command line digital preservation workflows

Required Readings
- 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. Also read the tool descriptions in the Reading Room section toward the end (p, 33-24). PDF available from: http://www.prestocentre.org/library/resources/av-insider-2-preservation-times-precarity
- 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
- Look over the websites for these tools (read the FAQ, README, and other documentation)
    - ○ Checksum: md5, fixi, md5deep/hashdeep
    - ○ Packaging: BagIt
    - ○ Transcoding: FFMPEG

NO CLASS OCTOBER 16 – FALL BREAK

Class 7 (Oct 23): Multi-institutional projects | Intellectual Property Issues

Due
- Assignment 1 – paper and presentation
- One-page final project proposals

<u>Topics</u>
- Intellectual Property Issues and Digital Rights Management
  - Howard Besser Guest Lecture (60 min)
    - Slides from Howard's talk http://besser.tsoa.nyu.edu/howard/Talks/12drm.pdf
- Student Presentations for Assignment 1 (120 min)

<u>Required Readings</u>
- **Multi-institutional Projects** (spend at least 10 minutes looking at each of these websites)
  - NDIIPP: National Digital Information Infrastructure and Preservation Program – Library of Congress http://www.digitalpreservation.gov
  - Digital Curation Center http://www.dcc.ac.uk/

- Intellectual Property Issues, and Digital Rights Management
  - Howard Besser's "Copyright and Digital Preservation" notes, 10/27/2011. http://www.nyu.edu/tisch/preservation/program/11fall/digital-copyright.html
  - "VIDEO AT RISK" DRAFT MEMORANDUM ON REFORMATTING CIRCULATING VIDEO COLLECTIONS (via email)
  - OCLC's Well-Intentioned Practice for Putting Digitized Collections of Unpublished Materials Online" (http://www.oclc.org/research/activities/rights/practice.pdf)
  - Reproduction, Copyright and Ownership in the context of digital preservation (http://www.nyu.edu/tisch/preservation/program/05spring/digital-copyright.html)
  - American Library Association, "Digital Rights Management and Libraries" on the web site of ALA. http://www.ala.org/ala/issuesadvocacy/copyright/digitalrights/index.cfm
  - Besser, Howard. The Challenge of Copyright & Difficult Materials -- besser.tsoa.nyu.edu/howard/Talks/09soima-copyright.pdf
  - Techdirt.com website, *Can a Contract Remove Fair Use Rights* http://www.techdirt.com/articles/20110209/03531413021/can-contract-remove-fair-use-rights.shtml

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

<u>Recommended Readings</u>
- "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)
- Duncan, Charles et al. "Digital Rights Management: Final Report". JISC. 2004. http://www.intrallect.com/index.php/intrallect/knowledge_base/general_articles/jisc_drm_study_2004__1

Class 8 (Nov 6): Web and New Media Preservation
BRING YOUR LAPTOP TO CLASS

Due

Topics
- Finalize make up dates
- Assignment 1 presentations
- Guest Lecture by Ben Fino-Radin, Rhizome Art Base (60 min)
- Hands-on exercise in web archiving (60 min)
- Group Discussion (60 min)

Required Readings
- Fino-Radin, Ben. "Sustainable Preservation Practices and the Rhizome ArtBase" http://media.rhizome.org/blog/8332/rhizome-digital-preservation-practices.pdf
- Watch entire Webinar video on using Archive-IT http://archive-it.org/public/webinar.html
- Archive-It Page:
  - https://webarchive.jira.com/wiki/display/ARIH/5+Challenges+of+Web+Archiving
  - https://webarchive.jira.com/wiki/display/ARIH/Archiving+Social+Networking+Sites+with+Archive-It
- Internet Archive's Glossary of Web Archiving Terms https://webarchive.jira.com/wiki/display/ARIH/Glossary+of+Web+Archiving+Terms
- Look over what the International Internet Preservation Consortium does, review recent activities, etc. (http://www.netpreserve.org/about/index.php) IIPC Project to Evaluate Emulation and Migration as Long-Term Preservation Solutions for Web Archives
- Listen to "Michael Nelson on Digital Preservation" on Surprisingly Free (stream or podcast) http://surprisinglyfree.com/2011/09/06/michael-nelson/
- Look over the Memento project website: http://www.mementoweb.org/
- ACT-IAC Collaboration & Transformation (C&T) Shared Interest Group (SIG), "Best Practices Study of Social Media Records Policies" http://www.actgov.org/knowledgebank/whitepapers/Documents/

Shared%20Interest%20Groups/Collaboration%20and%20Transformation%20SIG/ Best%20Practices%20of%20Social%20Media%20Records%20Policies%20-%20C T%20SIG%20-%2003-31-11%20%283%29.pdf
- Shiela A Brennan and T. Mills Kelly, "Why Collecting History Online is Web 1.5" http://chnm.gmu.edu/essays-on-history-new-media/essays/?essayid=47
- Video: *Web Archiving*, Library of Congress, 2009 (http://www.digitalpreservation.gov/ videos/webarch09/index.html)

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

Class 9 (Nov 13): Advanced Topics in Digital Preservation
BRING YOUR LAPTOP TO CLASS

Due
Assignment #2 due on Friday, November 16

Topics
- Microservices & SOA
- Emulation & Virtual Machines
- Forensics
- Software Preservation
- Group activity: Using Digital Preservation Repository Tools. Potential tools:
   ○ Preservica
   ○ Archivematica
   ○ Duraspace
   ○ BitCurator

Required Readings:
- 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=2Idbur1gR8I Slides: http://www.slideshare.net/ FuturePerfect_/jeff-rothenberg-digital-preservation-perspective
- Review and read a few papers from the keep Project http://www.keep-project.eu/ ezpub2/index.php

- 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
- 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
- "Digital Forensics and Born-Digital Content in Cultural Heritage Collections" by Matthew G. Kirschenbaum, Richard Ovenden, Gabriela Redwine (December 2010) CLIR Report http://www.clir.org/pubs/reports/pub149/pub149.pdf
- Digital Archaeology: Recovering Your Digital History http://www.nypl.org/blog/2012/07/23/digital-archaeology-recovering-your-digital-history
- California Digital Library "UC3 Curation Foundations" Rev 0.13 March 25, 2010 (https://confluence.ucop.edu/download/attachments/13860983/UC3-Foundations-latest.pdf)
- Watch this video of Mark Matienzo "Fiwalk with Me", starting at about 66:00. http://www.indiana.edu/~video/stream/launchflash.html?format=MP4&folder=vic&filename=C4L2011_session_3b_20110209.mp4

Class 10 (Nov 20): Storage, Distribution and Delivery
BRING YOUR LAPTOP TO CLASS

<u>Topics</u>
- Group 4 technology presentation
- 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
- Group exercise: Storage Planning using the IT-Innovation Modeling Tool

<u>Required Readings</u>
- 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_PreservationModellingTools_R0_v1.00.pdf
- November is "Lets solve the format problem" month. Read about it:
  - http://ascii.textfiles.com/archives/3645
  - http://www.archiveteam.org/index.php?title=Just_Solve_the_Problem_2012
  - Multimedia Wiki: http://wiki.multimedia.cx/index.php?title=Main_Page
  - http://www.openplanetsfoundation.org/blogs/2012-07-06-biodiversity-and-registry-ecosystem

- ○ http://registry.dcc.ac.uk:8080/RegistryWeb/Registry/
- John Footen and Mohan Ananthanarayanan "Service-Oriented Architecture and Cloud Computing" *SMPTE Motion Imaging Journal,* March 2012

Review
- Website for IT-Innovation PrestoPRIME Simulation Tool: http://prestoprime.it-innovation.soton.ac.uk/
  - ○ Download the tool for your OS

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

Class 11 (Nov 27): Sustainability
Topics
- Sustainability Issues lecture (60 min)
- Trustworthy Repositories Audit and Certification: Criteria and Checklist (60 min)
- Group exercise (90 min)

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
- 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)
- Beagrie, Neil. Keeping Research Data Safe Factsheet: Cost Issues in Digital Preservation of Research Data, Sept 2010 (http://www.beagrie.com/ KRDS_Factsheet_0910.pdf)
- Trustworthy Repositories Audit and Certification: Criteria and Checklist http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf
- Rosenthal, David, "Modeling the Economics of Long Term Storage" Lecture at Personal Digital Archiving 2012 http://archive.org/details/personaldigitalarchiving2012pt2

Recommended Readings
- Strategies for Sustainable Preservation of Born Digital Public Television -- http://www.thirteen.org/ptvdigitalarchive/

- Rusbridge, Chris (2006). "Excuse me... Some digital preservation fallacies?" *Ariadne* 46. (http://www.ariadne.ac.uk/issue46/rusbridge/?ref=cia-team.com)
- Lavoie, Brian. "The Fifth Blackbird: Some Thoughts on Economically Sustainable Digital Preservation." 2008. *D-Lib Magazine*. www.dlib.org/dlib/march08/lavoie/03lavoie.html.
- Academy of Motion Picture Arts & Sciences' *The Digital Dilemma* (http://www.oscars.org/science-technology/council/projects/digitaldilemma/)

NO CLASS DECEMBER 4 – AMIA CONFERENCE

Class 12 (Dec 11): Digital Asset Management, Requirements Gathering, Agile
Topics
- Lecture (90 min)
    - Assessing the Asset Management landscape
    - Types of systems – Content Management System, Collection Management System, Repository, Digital Asset Management System, Media Asset Management System, etc.
      Common misconceptions and issues: versioning, workflow, administration
    - Asset management system selection process
- Group Exercise: Use cases and requirements gathering for DAM and Digital Preservation (User requirements with Legos)

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

Class 13 (Dec 12): Presentations of Final Projects

ASSIGNMENTS

Assignment 1: Review of a digital preservation and access project (15%)

Due: Oct 23
Deliverables:
- 2-3 page paper (single spaced) + bibliography/webography
- 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. Be sure to look at the following: What need or problem is this project trying to solve? What collections and preservation 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
- plans for migration and/or emulation
- user interface (note: Perform sample queries to see what limitations you might face)
- query structure
- quality of files accessible (eg. resolution)
- intellectual property limitations, etc.
- sustainability planning

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)
- **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/)
- **AIMS Born-Digital Collections**: An Inter-Institutional Model for Stewardship
- **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)
- **PrestoCentre:** "The PrestoCentre Foundation is a membership-driven organisation that brings together a global community of stakeholders in audiovisual digitisation and digital preservation to share, work and learn" (http://www.prestocentre.eu/) and **PrestoPRIME** (http://www.prestoprime.org/)
- **SCAlable Preservation Environments (SCAPE) "**The **SCAPE** project will develop scalable services for planning and execution of institutional preservation strategies on an open source platform that orchestrates semi-automated workflows for large-scale, heterogeneous collections of complex digital objects" (http://www.scape-project.eu/)

- **Keeping Emulation Environments Portable (KEEP)** "KEEP (Keeping Emulation Environments Portable) is developing emulation services (KEEP Emulation Services) to enable accurate rendering of both static and dynamic digital objects" (http://www.keep-project.eu/ezpub2/index.php)
- **Blue Ribbon Task Force on Sustainable Digital Preservation and Access** (http://brtf.sdsc.edu/)
- **Archive-It** "a web archiving service to harvest and preserve digital collections" (http://www.archive-it.org/)
- **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)
- **UDFR** "a reliable, publicly accessible, and sustainable knowledge base of file format representation information for use by the digital preservation community" (http://www.udfr.org/) and **PRONOM**, an online registry for technical information about file formats (http://www.nationalarchives.gov.uk/PRONOM/Default.aspx)
- **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/)
- **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/)
- **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/)
- **Portico** "is a digital preservation service provided by ITHAKA, a not-for-profit organization with a mission to help the academic community use digital technologies to preserve the scholarly record and to advance research and teaching in sustainable ways. (http://www.portico.org/)
- **Federal Agencies Digitization Guidelines Initiative** (http://www.digitizationguidelines.gov)
- **Chronopolis at the San Diego Super Computer Center** "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: TOOLS FOR DIGITAL PRESERVATION (5%)

Due: November 13

Description:

Choose one of the two options below:

Option 1: Define two specific tasks that need to be performed in a preservation environment (example: validate checksum after transfer). Make sure the task is very specific. For each task, describe at least three functional requirements (what the tool is supposed to do/behave/accomplish) of a tool that would support/enable this task. Functional requirements should be specific and testable (e.g. User pushes a button to start checksum validation process), not vague (e.g. the tool needs to be user friendly)

For example, if your task is validate checksum, some functional requirements might be (there may be others):
1. Run the MD5 on all files immediately after they are received in the storage environment
2. Display checksum output on screen while processing
3. Compare the value generated with the content of a stored .md5 file that accompanies the content file
4. Report any differences between the calculated MD5 value and the stored value via email to systems administrator, including the full path to the file that didn't match and the date and time that the process occurred.

Option 2: Choose 2 command line tools, preferably from those we looked at in class, or from the tool registries. Using your virtual machine, install the tool if needed, then perform at least 2 tasks using that tool, in combination with others as needed (bonus: use pipe). Each task must include different arguments (options) available for that tool. Save your output to a text file. Write one paragraph for each tool or process that describes what the purposes of the tool and task executed for digital preservation.

HINT: For both options, you may want to look at the Issues and Solutions documented on the Open Planets Foundation wiki for some ideas.

Deliverables:
Option 1:
- 1 paragraph description of each task you define. Explain why the task is relevant to digital preservation. Provide the three functional requirements for each task using numbers or bullets formats. Keep the tasks and functional requirements very simple and focused.

Option 2:
- 1 paragraph explanation for each task executed, describing the role of the tool(s) in a digital preservation workflow, and why the tool was chosen. Include the actual command syntax in your paragraph (e.g. md5deep -r [filename])
- Text output from shell demonstrating the use of the tool/command

ASSIGNMENT 3: FINAL PROJECT (40%)

Due:
- One-page proposal due **Oct 23**
- Final paper due Dec 14 (by 5pm)
- Presentation on last day of class (Dec ??)

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 revolving around any topic we discuss in class. The instructor will provide a list of possible topics, before the end of September, or you may suggest your own. You should try to discuss your proposed topic with the instructor in September or October. You must turn in a formal project/paper proposal no later than Oct 23, and the instructor must approve your topic and scope before you start work on it.