

IS 241: Digital preservation, Winter 2016

Draft version, January 2016

Course information:

Number: INFS 241
Title: Digital Preservation
ID: 6282472000
Quarter: winter 2016
Location: room 121, GSE&IS Bldg.
Time: Wednesdays, 1h30pm-5.

Instructor information:

Instructor: Jean-François Blanchette
Office: room 218, GSE&IS Bldg.
Phone: 310 267 5137
Fax: 310 206 4460
Email: blanchette@ucla.edu
Web: <http://polaris.gseis.ucla.edu/blanchette>
Office hours: Tuesdays, 1h00-3h00

1. Course description

Cultural and scientific industries are today massively turning to digital media as the primary medium for the production and distribution of their products, either through digitization of cultural artifacts, creation of new forms of cultural expression and scientific experimentation (e.g., videogames, distributed simulations), or reliance on digital tools in the creation process itself (special effects, CAD). Yet, preserving digital objects over time has proved a complex problem on multiple fronts:

- While in the context of paper an information resource is typically conflated with its presentation, an electronic document is better characterized as an *aggregation of stored resources and a computed view*. In this context, only the stored resources (software code and data) have temporal persistence while the computed views (i.e., the intelligible content) are by essence evanescent – that is, the persistence of views is limited to their performance (i.e., the computation and rendering process);
- Because of the rapid obsolescence of software and hardware, the preservation and re-enactment of a digital document necessarily involves transformations, migrations, or emulations of the original. This is a fundamental departure from the traditional archival paradigm founded on the integrity of the documentary resource. Consequentially, there is not yet a stable consensus as to criteria for acceptable transformations;
- The definition of the “substance” of new digital forms has yet to stabilize through cultural and social conventions on authorship, authenticity, versions and performance. This lack of definitional stability is accompanied by a theoretical gap, an absence of concepts with which to manage this problem in archival science, information science, or system designs;

- Institutional models (economic, legal, and policy) appropriate for the long-term preservation of digital objects have yet to be developed.

Thus, the nature of digital media mandates reformulation of traditional concepts of authenticity, authorship, and originals, new kinds of information systems to manage the preservation process, along with new economic, legal, and policy tools with which to manage digital information over the long term. It's a full plate! This course will provide an introduction to this vibrant field, while at the same time helping participants develop some practical preservation skills.

2. Learning objectives

By the end of the course, class participants should:

- Be familiar with the technical literature relating to digital preservation and strategies for how to stay abreast of changes in this dynamic field;
- Be able to articulate the fundamental technical issues at the heart of digital preservation and the various trade-offs between cost, longevity, risk, and technical complexity entailed by solutions such as format migration, emulation, or bit preservation;
- Be able to develop and assess digital preservation policies and their appropriateness for specific types of content and user communities;
- Be able to identify and articulate the elements necessary to create and manage a trusted digital repository;
- Be able to assess and make recommendations with respect to the suitability of various storage options (including outsourcing to commercial providers, to institutional repositories, or developing in-house solutions) for meeting specific digital preservation needs;
- Further develop modes of communication, analysis, and presentation appropriate to a professional career in the LIS field, including conference posters and professional reports.

3. Requirements

The field of digital preservation is moving fast these days, so students are expected to read widely based on the readings and their own particular interests. No technical expertise is expected and assumed, beyond that expected of all entering students the MLIS regardless of specialization. Grades will be assigned based upon the level of critical and original thinking, depth of analysis, professional presentation of assignments, and class participation. In addition, class participants should:

- Come to class prepared to discuss the readings. See "How to Read a Book," (<http://pne.people.si.umich.edu/PDF/howtoread.pdf>).
- Forfeit the use of your laptop and other electronic devices during class lectures. In group exercises, one laptop per team is allowed.
- Participate in discussions. You are particularly encouraged to question the assumptions of the readings, the instructor, and your fellow students, as long as you do so respectfully. In doing so, you will sharpen your ability for critical thinking, innovation, debate, and public speaking, skills fundamental to your future professional life.
- Written work should be of high quality. If you have concerns about writing, address them early. A useful resource is UCLA's Graduate Writing Center (<http://gsrc.ucla.edu/gwc/>).

- Assignments must be turned in according to the scheduled due dates. In particular, no incompletes will be given.
- If you feel that you may need an accommodation for a disability or have any other special needs, make an appointment to discuss this with the instructor. I will best be able to address special circumstances if I know about them early in the term. The website for the UCLA Office for Students with Disabilities (www.osd.ucla.edu) contains a wealth of useful information as well as official policies about this issue.

4. Assignments

1. Complete required readings, participate in class discussions, and complete class assignments (20%)
2. **Case study analysis:** Working in groups of 2 or 3, you will identify an organization, group or individual with a digital preservation, curation, or recovery, or repatriation need. This can be as institutional as the Library of Congress' Twitter archive or as informal as your cousin's videogame collection from the 90s. You will gather information on the current environment, analyze the need for intervention as well as the existing preservation situation, present findings, and propose recommendations for future preservation action. The final report should include, at a minimum, an analytical discussion of the issues raised by (a) file formats, (b) migration strategies (if any), and (c) significant properties, a (d) cost/risk analysis of the proposed actions, and a (d) discussion of ethics, intellectual property rights, and other legal considerations. (50%). You will report back to the rest of the class on **Week 3** about the case study you have selected, and on **Week 6** on your progress.
3. **Poster creation and presentation:** Collaborate on creating a *poster* of your suggested plan suitable for presentation at a conference. Resources regarding the design of conference posters will be available on the course website. The posters will be displayed and judged in the IS Department lounge during Week 10 (30%).

5. Course schedule and readings

Week 1 (January 6): How bad is it?

Required readings:

Steward Brand, "Written on the Wind", in MacLean, M., & Davis, B. (eds.). (1999). *Times and bits: Managing digital continuity*. Los Angeles: Getty Research Institute.

<http://longnow.org/essays/written-wind/>

John M. Sarkissian, *The Search for the Apollo 11 SSTV Tapes*, May 2006.

http://www.parkes.atnf.csiro.au/news_events/apollo11/The_Apollo11_SSTV_Tapes_Search.pdf

Week 2 (January 13): Preserving bits

Every digital preservation strategy must solve the same problem: how to enable the *rendering mechanism* (software and hardware) to "perform" the *trace* (bitstrings on magnetic/optical media) when technological obsolescence has created a gap between the trace and its ability to be rendered. Approaches to bridging this gap consist of updating the trace (migration), the rendering mechanism (emulation), or perform various kinds of abstractions of the trace or the rendering mechanism (scoring, universal formats, UVC).

Required readings:

Thibodeau, K. (2002). Overview of technological approaches to digital preservation and challenges in coming years. In *The state of digital preservation: An international perspective* (pp. 4-31). Washington, DC: Council on Library and Information Resources. Retrieved from <http://www.clir.org/pubs/reports/pub107/pub107.pdf>

Commission on Preservation and Access and the Research Libraries Group. "Report on the Task Force on Archiving of Digital Information." (1995)
<http://www.clir.org/pubs/reports/pub63watersgarrett.pdf>

Galloway, P. (2004). Preservation of digital objects. *Annual Review of Information Science and Technology*, **38**:549-590.

Additional readings:

Yakel, E. (2001). Digital preservation. *Annual Review of Information Science and Technology*, **35**:337-378.

Rothenberg, Jeff. "Avoiding Technological Quicksand: Finding a Viable Technology for Digital Preservation." Washington, DC: The Council on Library and Information Resources, January 1999.

Kahle, Brewster. "Preserving the Internet: An archive of the Internet may prove to be a vital record for historians, businesses and governments." *Scientific American*. March 1997: 82-83.
<http://search.epnet.com/login.aspx?direct=true&db=buh&an=9704276050>

Rinehart, R. (2004). *A System of Formal Notation for Scoring Works of Digital and Variable Media Art*. In Proceedings of Annual Meeting of the American Institute for Conservation of Historic and Artistic Works, Portland, OR.

Lynch, Clifford. "Canonicalization: A Fundamental Tool to Facilitate Preservation and Management of Digital Information." *D-Lib Magazine* 5, no. 9 (September 1999).

Mellor, Phil, Paul Wheatley and Derek Sergeant. "Migration on Request, a Practical Technique for Preservation." *CAMiLEON Project*. Leeds, UK: The University of Leeds, 2002.

Van Wijngaarden, Hilde, and Erik Oltmans. "Digital Preservation and Permanent Access: The UVC for Images." Proceedings of the Imaging Science & Technology Archiving Conference. San Antonio, Texas. April 2004.

Reich, V. & Rosenthal, D.S.H. (2001). "LOCKSS: A permanent Web publishing and access system" *D-Lib Magazine*, 7(6). <http://www.dlib.org/dlib/june01/reich/06reich.html>

Moore, R. W. (2006). "Building preservation environments with data grid technology". *American Archivist* 69 (Spring/Summer): 139-158

Lee, K.-H. et al. (2002). The state of the art and practice in digital preservation. *Journal of Research of the National Institute of Standards and Technology* **107**: 93-106. Retrieved from <http://nvl.nist.gov/pub/nistpubs/jres/107/1/j71lee.pdf>

Hedstrom, M. (2002). The digital preservation research agenda. In *The state of digital preservation: an international perspective* (pp. 32-37). Washington, DC: Council on Library and

Information Resources. <http://www.clir.org/pubs/reports/pub107/pub107.pdf>

Week 3 (January 20): File Formats

The most widespread strategy for preservation today is the migration of file formats. This lecture will review what is currently understood about file formats, the dynamics of their evolution and obsolescence, and the modes of governance most appropriate for their longevity and stability.

Arms and Fleischhauer, "Digital Formats: Factors for Sustainability, Functionality, and Quality."

http://memory.loc.gov/ammem/techdocs/digform/Formats_IST05_paper.pdf

See <http://www.digitalpreservation.gov/formats/>

David Pearson and Colin Webb, "Defining File Format Obsolescence: A Risky Journey," *International Journal of Digital Curation* 4(3) (2009): 29-43

Lawrence, G.W., et al. (2002). *Risk management of digital information: A file format investigation*. Washington, DC: Council on Library and Information Resources.

<http://www.clir.org/pubs/reports/pub93/contents.html>

Week 4 (January 27): Authenticities/significant properties

When working with digital objects, the archivist is faced with a fundamental dilemma. On the one hand, to fulfill a professional mission founded on a classical notion of authenticity, she must preserve the bitwise integrity of the stored resource, but face losing any guarantees that content will remain accessible over the long term, because the software and hardware necessary to render the bits will be long obsolete. On the other hand, she can ensure that content will remain accessible through manipulation of the stored resources (e.g., the migration of data formats), which will entail foregoing traditional criteria of resource integrity. What to do?

Hedstrom, M. & Lee, C.A. (2002) "Significant properties of digital objects: Definitions, applications, implications" in *Proceedings of the DLM Forum 2002, Barcelona, 6-8 May 2002*. Luxembourg: Office for Official Publications of the European Communities, 2002.

Margaret Hedstrom, Cal Lee, Judy S. Olson, and Cliff Lampe, "The Old Version Flickers More': Digital Preservation from the User's Point of View," *American Archivist* 69/1: 159-187.

Representation Information Registries

http://www.planets-project.eu/docs/reports/Planets_PC3-D7_RepInformationRegistries.pdf

Additional readings:

Inspect project final report:

<http://www.significantproperties.org.uk/inspect-finalreport.pdf>

<http://www.jisc.ac.uk/whatwedo/programmes/preservation/2008sigprops.aspx>

Rothenberg, Jeff (2000), "Preserving Authentic Digital Information", in *Authenticity in a Digital Environment*, Washington, DC: The Council on Library and Information Resources.

Matthew Kirschenbaum, "An Old House with Many Rooms", Chapter 3 in *Mechanisms*, MIT Press, 2010.

Additional readings:

"Digital Preservation – Finding Balance", Special issue of *Library Trends*, Volume 54, Number 1, Summer 2005. http://muse.jhu.edu/journals/library_trends/toc/lib54.1.html

Depocas, Alain, Jon Ippolito, and Caitlin Jones. *Permanence Through Change: The Variable Media Approach*. New York: Guggenheim Museum, 2003.

David Phillips, "Judges in the Dock" chapter 5 of *Exhibiting Authenticity*, Manchester University Press, 1997.

Gilliland-Swetland, A. J. (2000). *Enduring Paradigm, New Opportunities : The Value of the Archival Perspective in the Digital Environment*. Washington, D.C.: Council on Library and Information Resources.

Week 5 (February 3): The repository movement

Guest speaker:

Shira Petzman, Digital Archivist, UCLA Library Special Collections

Readings:

Christopher Lee, "Open Archival Information System (OAIS) Reference Model"
<http://ils.unc.edu/caltee/p4020-lee.pdf>

Research Libraries Group/OCLC, Inc. (2002). *Trusted digital repositories: Attributes and responsibilities*. Mountain View, CA: Research Libraries Group. Retrieved from <http://www.oclc.org/programs/ourwork/past/trustedrep/repositories.pdf>

Trustworthy Repositories Audit & Certification: Criteria and Checklist
http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf

Katherine Skinner, Matt Schultz, "A Guide to Distributed Digital Preservation"
http://www.metaarchive.org/sites/metaarchive.org/files/GDDP_Educopia.pdf

Additional Readings:

Consultative Committee for Space Data Systems. (2002). *Reference Model for an Open Archival Information System (OAIS)*. Washington, DC.
<http://public.ccsds.org/publications/archive/650x0b1.PDF>

Massachusetts Institute of Technology Libraries. (2002). *DSpace: Durable digital repository*. MIT Services. Retrieved from <http://dspace.org/mit/services.html>

William G. LeFurgy, "Building Preservation Partnerships: The Library of Congress National Digital Information Infrastructure and Preservation Program", *Library Trends*, 54(1) Summer 2005. http://muse.jhu.edu/journals/library_trends/v054/54.1lefurgy.pdf

NDIPP, *Preserving our Digital Heritage: Plan for the National Digital Information Infrastructure and Preservation Program – A Collaborative Initiative of the Library of Congress*. Washington D.C., Library of Congress.

http://www.digitalpreservation.gov/about/ndiipp_plan.pdf

Browse:

DuraSpace

<http://www.duraspace.org/>

Center for Research Libraries

<http://www.crl.edu/archiving-preservation/digital-archives/certification-assessment>

arXiv.org

“e-prints in Physics, Mathematics, Computer Science and Quantitative Biology”

<http://arxiv.org>

The Berkeley Electronic Press

“The Premier Institutional Repository Platform”

<http://www.bepress.com/>

BioMed Central: Open Repository

“a service from BioMed Central to build, launch, host and maintain institutional repositories for organizations” - built on DSpace

<http://www.openrepository.com/>

DSpace

“captures, stores, indexes, preserves, and distributes digital research material”

<http://www.dspace.org/>

Fedora (Flexible Extensible Digital Object Repository Architecture)

“a general purpose repository system”

<http://fedora-commons.org/>

Internet Archive

“...is building a digital library of Internet sites and other cultural artifacts in digital form”

<http://www.archive.org/>

OCLC digital archive

“real-world solutions for the challenges of archiving and preservation in the virtual world”

<http://www.oclc.org/digitalarchive/>

UC Libraries Digital Preservation Repository

“a set of services that support the long-term retention of digital objects for ...UC libraries and their users”

<http://www.cdlib.org/services/uc3/dpr.html>

Open Preservation Foundation

<http://openpreservation.org>

Week 6 (February 10): Sustainability & Economics

Guest speaker:

Linda Tadic, CEO, Digital Bedrock

Readings:

Blue Ribbon Task Force on Sustainable Digital Preservation and Access, *Sustainable Economics for a Digital Planet*, (2010).

http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf

Richard Wright, Ant Miller, and Matthew Addis, "The Significance of Storage in the 'Cost of Risk' of Digital Preservation, *International Journal of Digital Curation* 4(3) (2009).

David S. Rosenthal, Daniel L. Vargas, "Distributed Digital Preservation in the Cloud," *International Journal of Digital Curation* 8(1) (2013): 107-11.

Lavoie, Brian F. "The Incentives to Preserve Digital Materials: Roles, Scenarios, and Economic Decision-Making." Dublin, Ohio: OCLC Research, 2003.

<http://www.oclc.org/research/projects/digipres/incentives-dp.pdf>

Additional readings:

Currall, James and Peter McKinney, "Investing in Value: A Perspective on Digital Preservation," *DLib Magazine*, April 2006.

<http://www.dlib.org/dlib/april06/mckinney/04mckinney.html>

ERPANET. 2003. *Cost Orientation Tool*,

<http://www.erpanet.org/guidance/docs/ERPANETCostingTool.pdf>

Currall, James, Claire Johnson, and Peter McKinney. 2005. 'The Organ Grinder and the Monkey. Making a business case for sustainable digital preservation', given at EU DLM Forum Conference 5-7 October 2005 Budapest, Hungary.

<https://dspace.gla.ac.uk/handle/1905/455>

Week 7 (February 17): Preserving the Web/Social Media

Guest speaker:

Scott Reed, formerly Web Archivist, Internet Archive

Readings:

NDSA, Web Archiving in the United States: A 2013 Survey.

http://www.digitalpreservation.gov/ndsa/working_groups/documents/NDSA_USWebArchivingSurvey_2013.pdf

Maureen Pennock, Web-Archiving, Digital Preservation Coalition, 2013.

Peter Buneman, Heiko Müller, and Chris Rudbrige, "Curating the CIA World Factbook," *International Journal of Digital Curation* 4(3) (2009):29-43.

Week 8 (February 24): Preservation metadata

Readings:

Gilliland-Swetland, Anne J. "Defining Metadata," in *Introduction to Metadata: Pathways to Digital Information*. Los Angeles: Getty Information Institute, 1998. 1-8.

Ross Harvey and Dave Thompson, "Automating the Appraisal of Digital Materials," *Library Hi Tech* 28(2) (2010): 313-322.

OCLC, & RLG Working Group on Preservation Metadata. (2002). *Preservation Metadata and the OAIS Information Model: A Metadata Framework to Support the Preservation of Digital Objects*.

Week 9 (March 2): Personal Archives

Kirk, D. S. and Banks, R. 2008. "On the design of technology heirlooms." In Proceedings of the International Workshop on Social Interaction and Mundane Technologies (SIMTech'08).

David S. Kirk and Abigail Sellen. 2010. "On human remains: Values and practice in the home archiving of cherished objects." *ACM Trans. Comput.-Hum. Interact.* 17, 3, Article 10 (July 2010), 43 pages.

Richard Banks , David Kirk , Abigail Sellen , A Design Perspective on Three Technology Heirlooms, *Human-Computer Interaction* Vol. 27, Iss. 1-2, 2012

Week 10 (March 9): Poster sessions

Posters session in IS Department lounge.

Final reports due.