

IS 289: Digital preservation, Fall 2006

Department of Information Studies, University of California, Los Angeles

(Last modified Monday October 2nd 2006)

Class hours: Tuesdays, 9-12h30, GSE&IS Bldg, Room 121.

Course homepage: <http://courses.gseis.ucla.edu/course/view.php?id=69>

Instructor: Jean-François Blanchette <blanchette@ucla.edu>

Web: <http://polaris.gseis.ucla.edu/blanchette/>

Office hours: Tuesdays 1h30-3h30, GSE&IS Bldg, Room 218

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, there are today no known solutions to the problem of preserving complex digital objects over time. We already face risk of significant loss to our cultural heritage – a situation that will only become worse if nothing is done.

The issue has been difficult to solve because of several interconnected issues:

- 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.
- As well, economic, legal, and policy models 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. This course will provide an introduction to the problems and possibilities of long-term preservation for digital objects ranging from video games and websites to 3-dimensional architectural and engineering drawings.

Readings

All readings will be available to students on the course website:

<http://courses.gseis.ucla.edu/course/view.php?id=69>

Readings are listed under the week for which they are due, i.e., the Galloway reading is to be read *prior* to attending the first class.

Expectations

The area of digital preservation is currently exploding, 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 in either the MLIS or MIAS programs. Students should be prepared to discuss and comment upon readings, lectures, and presentations. Grades will be assigned based upon the level of critical and original thinking, depth of analysis, professional presentation of assignments, and class participation. Assignments must be turned in according to the scheduled due dates. *No incompletes will be given.*

Requirements

1. Complete required readings and participate in class discussions (20%).
2. Write one **analytical essay** (4 pages, space-and-a-half), which synthesize a week's readings and suggest questions for discussion (20%). You decide which week you would like to cover. The essay covering a given week's readings is due by 5:00pm on the **day before** (that is, on the Monday before) that week's class, in the instructor's mailbox (no emails please). Each essay should include:
 - a. Summary introduction (context)
 - b. An analytical **synthesis** of major concepts and themes that emerge from the week's readings. You should **not** simply list the facts or arguments in all of the readings.
 - c. Suggested questions for the week's class discussion. The idea is to pose questions that might stimulate a reader who has already done the readings to think about them further (e.g. how the concepts or arguments relate to readings from earlier in the semester; specific ways that given readings for the week might complement or contradict each other; how ideas from the week's readings might be applied in particular professional situations). You should **not** generate questions that would simply prompt the reader to provide a factual response (e.g.). Your discussion questions should not constitute more than 1 page of your total essay for the week.
 - d. Conclusion
3. Collaborate on writing a grant proposal to the National Endowment for the Humanities' *Preservation Assistance Grants for Smaller Institutions* program, seeking funds to evaluate the preservation need of a digital collection (neh.gov/grants/guidelines/pag.html). The objective of the grant will be to secure funds to hire your group as consultants "to conduct a general preservation assessment, and to help draft a long-range plan for the care of humanities collections. The consultant visits the institution to assess policies, practices, and conditions affecting the care and preservation of humanities collections and prepares a report that summarizes the findings and contains prioritized recommendations for future preservation action." Working in groups of 4-5, you will collaborate with an assigned partner organization (e.g., the UCLA Ethnomusicology Archive) to gather information on

the current environment, analyze the preservation need and the existing preservation situation, present findings, and propose recommendations for a grant-funded project. Your final submission should follow the outlines of the call for proposal which includes a five-page (single-spaced) project narrative, an itemized budget for project expenses, and supporting documentation (30%).

4. Collaborate on creating a *poster* of your suggested plan suitable for presentation at a conference (follow the guidelines available at <http://www.euroia.org/callforpapers.html>, but posters should be a minimum of 3' by 6'). More 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%).

Course schedule

Week 1: Introduction to Digital Preservation (October 3)

Required readings:

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

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://www.longnow.org/views/essays/articles/writtenonwind.php>

Browse:

National Archives of Australia. (n.d.) *Preserving access to digital information* Website.

<http://www.nla.gov.au/padi>

Additional readings:

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

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.

http://www.dlmforum2002.org/download/margaret_hedstrom.pdf

"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

Week 2: (October 10)

Cancelled

Week 3: Authenticities (October 17)

Guest speaker: Bruno Bachimont, Université Technologique de Compiègne.

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?

Readings:

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.

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

Additional readings:

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 4: Preserving bits (October 24)

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).

For this week, read all assigned readings, plus two additional readings of your choice.

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>

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>

Additional readings:

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>

Lyman, Peter. "Archiving the World Wide Web" in *Building a National Strategy for Digital Preservation: Issues in Digital Media Archiving*. Washington, DC: Council on Library and Information Resources, 2002, pp. 38-51. <http://www.clir.org/pubs/reports/pub106/web.html>

AIIIM Industry White Paper on Records. "Conversion & Document Formats: Backfile conversion and format issues for information stored in digital archives." Hamburg: Project Cunsult, 2002. <http://whitepapers.zdnet.co.uk/0,39025945,60100435p-39000450q,00.htm>

Lawrence, G.W., et al. (2002). *Risk management of digital information: A file format investigation*. Washington, DC: Council on Library and Information Resources. Retrieved from <http://www.clir.org/pubs/reports/pub93/contents.html#about>

Browse:

Consultative Committee for Space Data Systems. (2002). *Reference Model for an Open Archival Information System (OAIS)*. Washington, DC.

National Library of Australia. "PADI: Preserving Access to Digital Information - Web Archiving." Available from <http://www.nla.gov.au/padi/topics/92.html>

Week 5: Institutions and Policies (October 31)

Guest speaker: Victoria McCargar, former Senior Editor for Technology, LA Times.

Readings:

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

http://www.parkes.atnf.csiro.au/apollo11/apollo11_sstv_search_report.html

Commission on Preservation and Access and the Research Libraries Group. "Report on the Task Force on Archiving of Digital Information." (1995)

http://www.rlg.org/en/page.php?Page_ID=20442

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>

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

Browse:

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

Archivists' Toolkit. <http://archiviststoolkit.org/>

National Archives of Australia, *DIRKS: A Strategic Approach to Managing Business Information*, 2001. <http://www.naa.gov.au/recordkeeping/dirks/dirksman/dirks.html>

CASPAR: Cultural, Artistic and Scientific Knowledge for Preservation, Access, and Retrieval.

<http://www.casparpreserves.eu/caspar-project>

Week 6: Law (November 7)

Guest speakers:

Stephen Davison, UCLA Library Digital Collections

Maureen Whalen, J. Paul Getty Trust

Readings:

Charlesworth, Andrew. *Legal issues relating to the archiving of Internet resources in the UK, EU, USA and Australia: A study undertaken for the JISC and Wellcome Trust*. JISC & The Wellcome Trust, February 2003.

http://www.jisc.ac.uk/uploaded_documents/archiving_legal.pdf

Iannella, R. (2001). Digital rights management (DRM) architectures. *D-Lib Magazine*, 7(6).

Retrieved from <http://www.dlib.org/dlib/june01/iannella/06iannella.html>

Additional readings:

Rosenblatt, B. et al. (2002). *Digital rights management: Business and technology*. New York: M&T Books.

Week 7: Economics (November 14)

Readings:

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>

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>>.

Additional readings:

DCC/DPC Workshop on Cost Models for preserving digital assets, British Library Conference Centre, 26th July 2005.

<http://www.dpconline.org/graphics/events/050726workshop.html>

Neale, Bill. "A New Approach to Cost Justifying ERM Systems." Richmond/Tidewater ARMA and & Old Dominion AIIM Meeting, November 2005.

<http://www.aiim.org/chapters/olddominion/Business%20case%20for%20ERMS%202011-15-2005.ppt>

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 8: Repositories (November 21)

Readings:

Lavoie, Brian F. "The Open Archival Information System Reference Model: Introductory Guide." OCLC and Digital Preservation Coalition, 2004.

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

Research Libraries Group/OCLC, Inc. (2002). *Trusted digital repositories: Attributes and responsibilities*. Mountain View, CA: Research Libraries Group. Retrieved from

<http://www.rlg.ac.uk/longterm/repositories.pdf>

Budapest Open Access Initiative: *A Guide to Institutional Repository Software v 3.0*, 2004

<http://www.soros.org/openaccess/software/>

E-Archiving: An Overview of Some Repository Management Software Tools, April 2005
<http://www.ariadne.ac.uk/issue43/prudlo/>

Browse:

ARNO (Academic Research in the Netherlands Online)

“university document servers to make available the scientific output of participating institutions”

<http://www.uba.uva.nl/arno>

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/>

CERN Document Server Software Consortium

“CDSoftware ...to run your own electronic preprint server, an online digital library catalogue or a document repository on the web”

<http://cdsoftware.cern.ch/>

Chronopolis

“Federated Digital Preservation Across Time and Space” NARA, University of Maryland, San Diego Supercomputer Center (SDSC). 2006 Internet2 Idea award winner...

<http://globalstor.org/pdf/presentations/Moore-chronopolis.pdf>

DIAS: Digital Information Archiving System

“provides a flexible and scalable open deposit library solution for storing and retrieving massive amounts of electronic documents and multimedia files”

<http://www-5.ibm.com/nl/dias/>

Digital Archive - “Digital Fridge”

Digital archive developed at The National Archives, UK

<http://www.nationalarchives.gov.uk/preservation/digitalarchive/default.htm>

Diva

“a software infrastructure for visualizing and interacting with dynamic information spaces”

<http://embedded.eecs.berkeley.edu/diva/>

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://www.fedora.info/>

FEZ

“Fedora-based Repository Management System”

<http://www.apsr.edu.au/currentprojects/fez06.htm>

Greenstone Digital Library Software

“a suite of software for building and distributing digital library collections”

<http://www.greenstone.org/cgi-bin/library>

Internet Archive

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

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

The Jakarta Slide project

“a content repository ... a basis for a content management system / framework ...”

<http://jakarta.apache.org/slide>

MyCoRe "Content Repositories"

“an Open Source project for the development of Digital Library and archive solutions”

<http://www.mycore.de/engl/index.html>

OCLC digital archive

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

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

Open Source Portfolio Initiative (OSPI)

Repository project from the e-learning community

<http://dkc.mse.jhu.edu/repository.html>

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/inside/projects/preservation/dpr/>

Week 9: Preservation metadata (November 28)

Guest speaker: TBA

Readings:

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*.

Gilliland-Swetland, Anne. "Social Science Data Archives in the New World?" *For the Record: Data Archives, Electronic Records, Access to Information, and the Needs of the Research Community*. Ed. Rena Lohan et al. Dublin: Institute of Public Administration, 1996. 54-63.

Karasti, Helena & Karen Baker. "Infrastructuring for the Long-Term: Ecological Information Management." Hawaii International Conference on System Sciences 2004 (HICSS'37), Hawaii, January 5-8 2004.

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

Additional readings:

National Science Board. *Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century* (draft). May 26, 2005.

Week 10: Poster sessions (December 5)

Posters session in IS Department lounge.

Rauber, Andreas et al. "Uncovering Information Hidden in Web Archives: A Glimpse at Web Analysis building on Data Warehouses." *D-Lib Magazine* 8, no. 12 (December 2002).
<http://www.dlib.org/ar/dlib/december02/rauber/12rauber.html>

Pardo, Theresa. (2005). *Building State Government Digital Preservation Partnerships: A Capability Assessment and Planning Toolkit*, Version 1.0. Available from
http://www.ctg.albany.edu/publications/guides/digital_preservation_partnerships

Iannella, R. (2001). Digital rights management (DRM) architectures. *D-Lib Magazine*, 7(6). Retrieved from
<http://www.dlib.org/dlib/june01/iannella/06iannella.html>

Lawrence, G.W., et al. (2002). Risk management of digital information: A file format investigation. Washington, DC: Council on Library and Information Resources. Retrieved from <http://www.clir.org/pubs/reports/pub93/contents.html#about>

Lyman, Peter. "Archiving the World Wide Web." In *Building a National Strategy for Digital Preservation: Issues in Digital Media Archiving*. Washington, DC: Council on Library and Information Resources, 2002. 38-51.

Rauber, Andreas et al. "Uncovering Information Hidden in Web Archives: A Glimpse at Web Analysis building on Data Warehouses." *D-Lib Magazine* 8, no. 12 (December 2002).

Boudrez, Filip and Sofie Van den Eynde. *DAVID: Archiving Websites*. Antwerp: Stadsarchief, 2003.

Lyman, P. (2002). *Environmental scan: Webpages*. Washington, DC: Library of Congress. National Digital Information and Infrastructure Preservation Program. Retrieved from
http://www.digitalpreservation.gov/ndiipp/repor/repor_back_web.html

Contracting Out For Digital Preservation Services: Information Leaflet And Checklist, 2004
<http://www.dpconline.org/docs/guides/outsourcing.pdf>

Lee, J. (2003). Extending storage resource management with active archiving to manage data lifecycle. *Information Management & Technology*, 36(4), 175-176.

Lee, J. (2005). Database archiving complements faster backup. *Information Management & Technology*, 38(4), 183-184.

Week 2: Preserving bits (October 10)

Additional Readings:

Young, J.R. (2002, July 5). "Superarchives" could hold all scholarly output. *The Chronicle of Higher Education*. Retrieved from <http://chronicle.com/free/v48/i43/43a02901.htm>