



PRESERVATION EDUCATION CURRICULUM



NORTHEAST
DOCUMENT
CONSERVATION
CENTER

CLASS 11 LESSON PLAN

Creating Sustainable Digital Collections, Part 2: Digital Preservation

The Lesson

Part I: Foundational Work (30 minutes)

This is an introduction to the foundational work that has driven most of the digital preservation development over the last ten years. Of key importance is for the class to be able to discuss and understand the categories of digital preservation challenges (roughly, organizational, legal, economic, technical), since digital preservation is far more than a technical problem. Another key distinction that should be introduced is the focus on the specific needs of electronic archives that is presented by the InterPARES work. Both Preserving Digital Information and the OAIS refer to high-level needs of all digital resources, but InterPARES strongly incorporates the issues of authenticity and integrity as they apply to managing digital information over time.

- A. CPA-RLG Task Force on Archiving of Digital Information, *Preserving Digital Information*
- B. *Reference Model for an Open Archival Information System (OAIS)*
- C. InterPARES

Part II: Research and Development (30 minutes)

- A. National Digital Information Infrastructure and Preservation Program (Library of Congress)

Discuss the relevant research and development activities of the organizations and initiatives listed below. Beginning with the broad US NDIIP program, discuss how research and development in digital preservation are being funded and carried out. [Note: the United States has a different model from most other nations actively engaged in digital preservation.] Discuss the areas of focus, the status of this loose “national program,” and prospects for the future.

- B. Joint Information Systems Committee (UK)
- C. Digital Curation Centre (UK)

To complement the view of the U.S. program, have students consider the projects and programs underway in the United Kingdom (especially through the Digital Curation Centre) and the large amounts of money and strong support they receive from the Joint Information Systems Committee. Compare the two systems and discuss why the United Kingdom might be ahead of the United States in terms of digital preservation.

D. ERPANET

Also consider the ERPANET project as an interesting and helpful set of digital preservation research and training activities that proved to be unsustainable. Discuss how research and development projects and programs—as well as digital preservation projects—are vulnerable to funding issues and the impact that this could have on the long-term preservation of digital materials.

Part III: Understanding Digital Objects (30 minutes)

This section is complemented by an activity. Its goal is to make students better understand digital objects and the associated metadata that serve a role in making them renderable over time. Key to this conversation is the understanding of the phrase “significant properties”—the ability to discern what aspects of the digital object and its functionality are important to preserve over time.

- A. Formats
- B. Preservation metadata
- C. Longevity

In-Class Activity

- Present to students a variety of digital objects. Ask them to identify the significant properties of the following named objects (objects to be provided):
 1. Encoded text document
 2. Photographic image (b/w)
 3. Photographic image (color) with color/grayscale/ruler included
 4. A video clip that at opening appears to be a photograph with overlaid text but after clicking on it, it becomes a motion clip with sound

Students should identify and discuss the significant properties of each and the effects on long-term access and functionality if the significant properties (metadata) are not collected and preserved with the digital object.

Part IV: Technical Strategies (30 minutes)

Following the discussion on understanding digital objects, this unit addresses the various technical strategies for preserving digital objects. Emphasize that no one strategy will be a valid answer for all digital objects; thus, this is not simply choosing migration versus emulation.

- A. Format migration
- B. Emulation
- C. Refreshing

Review the strategies and discuss which might be appropriate for a variety of file types. Discuss file formats documenting still images, moving images, encoded text, word-processed documents (e.g., a Microsoft Word 5.0 file), and so on. Be sure to discuss issues of mission/purpose, functionality/use, authenticity, and the like.

- D. Hybrid imperative—(It’s not “or” but “and”)—Briefly mention/discuss why some institutions choose to preserve digital objects using two different strategies. Why might that be important? Is this a good choice? Is it economically viable? Note: this “hybrid imperative” does not refer

to analog and digital preservation. It refers only to preserving digital objects utilizing two or more different technological strategies. The preservation of e-journals is a good choice to discuss here.

E. Web archiving

Because archiving the Web poses a vast and complex problem, discuss the ways various organizations are preserving Web-based digital information. Include the following in your discussions:

1. Pandora (Australia)
2. Internet Archive and European Archive
3. International Internet Preservation Consortium

Part V: Digital Repositories (30 minutes)

A. Defining and distinguishing digital repositories, digital archives, and institutional repositories

B. Open-source repository software (DSpace, Fedora, EPrints, BePress)

It is critically important for students to explicitly understand the differences among digital repositories, digital archives, and institutional repositories. The instructor may also want to briefly mention the additional concepts of content management systems and digital asset management systems to cover students who may wish to work in alternative environments like media preservation or museums. The readings for this section explain the capabilities and preservation services of which each kind of technological architecture may be capable. The key here is to underscore that most institutional repository software—e.g., DSpace, EPrints, etc.—do not yet have preservation functionality and were designed to facilitate local digital collections of born-digital content for *access purposes only!*

C. Standards and best practices (OAIS, etc.)

D. "Trusted repositories"

E. Certification for digital repositories and archives

The above grouping of topics is meant to underscore the existence of frameworks, papers, and projects that identify standardized approaches to digital preservation repositories capable of managing all kinds of digital objects over time. *None* provides a blueprint from which to build a system, but together the standards and reports identify attributes and metrics for assessing a repository's/archive's capabilities for long-term preservation of digital cultural heritage.

Part VI: Digital Registries (15 minutes)

A. Concept (shared metadata)

The concept of registries is not new (think union catalogs of bibliographic data), but in the new digital environment, people are beginning to think of new union databases of information that can be built to facilitate digital preservation, including provisions for more authentic data, as well as lowering the costs of producing necessary metadata. Briefly discuss the following initiatives to group this important metadata in a way that facilitates not only better access, but also long-term use and preservation.

B. Descriptive metadata registries (OAI, etc.)

C. Bibliographic "registries" of digitized materials (DLF Registry of digital masters)

These are both bibliographic in nature and perhaps are the ones with which students will be the most familiar. Discuss opportunities of harvesting/collecting these data so that they can be used

by others to promote discovery and use by all kinds of users.

D. Format registries (Global Digital Format Registry, PRONOM, Library of Congress Sustainability of Digital Formats site)

Moving the concept in the direction of digital preservation, discuss the need for and potential uses of format registries such as those mentioned above. Why should institutions want to use these kinds of registries? How do they facilitate digital preservation?

Suggested Graded Assignments

- Write a short paper defining and distinguishing digital repositories, digital archives, and institutional repositories.
- Write a short paper summarizing the various technical strategies for preserving digital objects.

Suggested Term Projects

- Write a research paper on a topic of the student's choice, such as the historical development of digital preservation initiatives, digital repositories, differences between U.S. and international digital preservation efforts, and so on.