

EDUCAUSE Center for Applied Research

Research Bulletin

Volume 2005, Issue 18

**August 30, 2005**

# Digital Preservation: A Campus-Wide Perspective

Richard Fyffe, University of Kansas

Deborah Ludwig, University of Kansas

Beth Forrest Warner, University of Kansas



Digital content permeates today's academic enterprise. It includes

- *Administrative content*—records of student learning (transcripts, enrollment data, and course descriptions and catalogs); financial records; employment records and personnel handbooks; and
- *Scholarly content*—research papers, experimental data sets, course syllabi, and so forth.

Preservation of digital content is an emerging aspect of the academy's stewardship responsibility. It requires more than simply extending traditional preservation practices to digital information or assuming that media backups are sufficient. Meeting universities' responsibilities for access to vital current content and stewardship of intellectual and cultural heritage challenges us to define a new information environment that promotes the preservation of fragile digital information.

Most current preservation practices developed out of a print-on-paper environment. Print endures with relatively modest upkeep and can be read without intermediary technology. We can set aside a paper document and reasonably expect to find it readable years later. As a consequence, formal preservation programs for print information can generally accommodate a wide range of user practices.<sup>1</sup> Digital information is not so forgiving. Digital objects cannot simply be encapsulated and set aside; they are never permanently preserved like pickles in a jar. Digital objects reside on physical media (tapes, disks) that are fragile and susceptible to corruption. More important, digital objects require intermediary technology (software and hardware) to create and render them—technology created in a marketplace that prizes proprietary control and innovation, resists standardization, and is often inattentive to backward compatibility. Digital objects can therefore easily fall out of synchronization with the tools needed to use them. Without a planning program for continued access to and usability of digital information, there can be no reasonable expectation that digital information created today will remain usable in a few years.

Notable efforts to preserve digital content are emerging from organizations such as the California Digital Library and the Library of Congress, but most of these efforts still represent early investigations and attempts at understanding. Moreover, they tend to focus on the technical aspects of digital preservation, whereas a fully successful program must also address the roles that staff throughout the organization will play in the preservation program as well as the kinds of policies, education, and training needed to support them in those roles. At the University of Kansas (KU) we are attempting to develop a holistic approach to digital preservation that will begin upstream in the headwaters of digital creation and change our day-to-day management of information, both academic and administrative, in such a way that preservation is a natural outcome of normal practice.

The focus of our approach at this point is current digital assets. In this research bulletin we do not address rescue of digital material in obsolete formats or on obsolete media, nor do we address digitization of analog material. This bulletin outlines the campus-wide issues relating to digital preservation. A follow-up bulletin details our proposed actions in creating an integrated digital preservation program.

## Highlights of Digital Preservation

Digital preservation is the ongoing process of managing data for continuing access and use. It is an outcome of the organization's successful day-to-day management of its digital assets,<sup>2</sup> not the condition of an object. Effective digital preservation requires that we understand and attend to the full life cycle of digital objects. This life cycle has two key dimensions:

- *Chronological*—highlights the various stages of an object's life from the point of creation forward.
- *Functional*—highlights the different roles that are occupied through an object's interaction with creators, editors, users, stewards, policymakers, and others over the course of its lifetime as shown in Figure 1.

**Figure 1. Functional Aspects of Digital Preservation**



No one step in an object's life cycle is “the preservation step” and no one role is the “preserving role.” Rather, digital preservation is the whole complex of roles and operations designed around management of information for long-term accessibility and usability. However, we give particular emphasis to one role and one point in the life cycle—the impact of choices made early in the object’s life by creators (and other agents close to the creator) on the likelihood that it will remain accessible and usable over time. There are four key areas in which those “upstream” choices can be exercised:

- *Appraisal*—early recognition that the object will (or will not) be used in the future will encourage creators and stewards to make preservation-oriented choices.
- *Choice of file format*—some file formats are more likely to remain usable over long periods than others. In general, proprietary formats are less likely to be usable in the future than “open” formats.
- *Choice of storage location*—in general, files placed in networked storage devices maintained by professional administrators adhering to best practices are more likely to remain usable than files held on isolated disks (including workstation hard drives) if other preservation steps are followed.
- *Description (metadata)*—good documentation of the content of an asset and the circumstances under which it was created (name of creator, date, data source, file format, and so forth) will enhance the likelihood that the asset can be discovered, rendered, and used effectively over time.

## Developing a Digital Preservation Program

A successful institutional digital preservation program requires the interaction of three elements: education, roles and policies, and an integrated technical architecture.

- *Education* relates to developing and implementing an educational program for faculty, staff, and administrators to address the basic concepts and challenges in digital preservation and information management practices. The goal is to encourage good stewardship of institutional digital assets by creators and users of information objects and to encourage support for digital preservation by administrators and resource allocators.
- *Roles and policies* refers to establishing a set of roles (functions exercised by staff) and institutional policies (to guide those roles) to ensure that appropriate digital preservation systems and services are implemented and maintained.
- *Integrated technical architecture* relates to foundation technologies that are designed around the complete life cycle of digital information, from creation forward. A technical infrastructure must be created that includes all the systems and services necessary to manage and access information over time.

For a detailed description of the elements, issues, and practical applications of a digital preservation program for higher education institutions, please see ECAR research bulletin #19, which will be available on September 13, 2005, (<http://www.educause.edu/LibraryDetailPage/666?ID=ERB0519>).

## Other Organizational Approaches

While a number of organizations have begun to explore the implications of preserving various types or groups of digital assets, it is critical for institutions of higher education to develop a more holistic approach and commitment to preserving both academic and administrative digital assets, and addressing both the technical aspects of preservation and the organizational behaviors that may result in successful preservation.

### Technical Approaches

Technological obsolescence is the result of the evolution of technology—as newer technologies appear, older ones cease to be used. Strategies for dealing with technological obsolescence include

- migration of digital information to technologies from which they are accessible,
- emulation of obsolete systems, and
- encapsulation—the preservation of obsolete technologies.

#### Migration<sup>3</sup>

The migration of digital information refers to the “periodic transfer of digital materials from one hardware/software configuration to another, or from one generation of computer technology to a subsequent generation.” Migration includes refreshing, specifically, copying digital information without changing it. However, although refreshing will overcome the problem of media instability, it usually isn’t enough to keep ahead of technological obsolescence. Migration to new operating environments often means that the copy is not exactly the same as the original object. Decisions will need to be made about which aspects (for example, functionality, presentation, and so forth) need to be preserved.

#### Emulation<sup>4</sup>

Emulation refers to the process of mimicking, in software, a piece of hardware or software so that other processes think the original equipment/function is still available in its original form. Emulation is essentially a way of preserving the functionality of and access to digital information which might otherwise be lost due to technological obsolescence.

One of the benefits of the emulation strategy compared with migration is that the original data need not be altered in any way. It is the emulation of the computer environment that will change with time. This should help maintain the integrity and “look and feel” of the material.

Encapsulation in the context of digital preservation is a technique of grouping together a digital object and anything else necessary to provide access to that object. This technique includes the details of how to interpret the digital bits in the object as part of the encapsulated information. Encapsulation can be achieved by using physical or logical structures called “containers” or “wrappers” to provide a relationship between the information components. The types of supporting information that should be included, apart from the digital object itself, are described by the *Reference Model for an Open Archival Information System (OAIS)*,<sup>6</sup> which is based on the report *Preserving Digital Information from the Task Force on Archiving of Digital Information*.

### OAIS Reference Model

An OAIS is an archive consisting of an organization of people and systems that has accepted the responsibility to preserve information and make it available for a Designated Community. The OAIS Reference Model has been designated as ISO standard 14721:2003 and is designed as a conceptual framework in which to discuss and compare archives. As such, it attempts to address all the major activities of an information-preserving archive in order to define a consistent and useful set of terms and concepts. This reference model does not specify a design or an implementation. Actual implementations may group or break out functionality differently.

### Examples of Other Efforts Under Way

- *National Digital Information Infrastructure and Preservation Program/Library of Congress*—as a pioneer in the field of digital information, the Library of Congress has been chosen to lead the development of a national collaborative digital strategy called the National Digital Information Infrastructure and Preservation Program. The Library will work closely with federal partners to assess considerations for shared responsibilities and will also seek participation from the non-governmental sector. The overall strategy will be executed in cooperation with the library, creative, publishing, technology, and copyright communities. The Digital Preservation Program will seek to provide a national focus on important policy, standards, and technical components necessary to preserve digital content. Investments in modeling and testing various options and technical solutions will take place over several years, resulting in recommendations to the U.S. Congress about the most viable and sustainable options for long-term preservation.
- *Digital Curation Centre*—working with other practitioners, the Digital Curation Centre will support U.K. institutions to store, manage, and preserve science and other research data to ensure their enhancement and their continuing long-term use. The purpose of the Centre is to provide a national focus for research into curation issues and to promote expertise and good practice, both national and international, for the management of all research outputs in digital format.

- *California Digital Library/UC Libraries Digital Preservation Repository*—as the foundation of the digital preservation program of the University of California system, the UC digital preservation repository is a set of services that support the long-term retention of digital objects for the benefit of the UC libraries and their users. The services and associated storage are based at the California Digital Library. The CDL developed the digital preservation repository to provide for the controlled, orderly deposit and delivery of objects between authorized users and for reliable, long-term storage.

## What It Means to Higher Education

Continuing access to administrative records is fundamental to colleges' and universities' fulfillment of their mission as administrative agencies. Such records are routinely requested 5, 10, and 20 years after they are created, and historians will expect them to be available much further into the future. Moreover, just as important as the ability to deliver such records when they are needed in the future is the academy's ability to demonstrate to funders and oversight agencies accountability and transparency in such areas as

- flow of data within the institution;
- positions and offices responsible for creating and maintaining mission-critical data;
- policies that govern access to and revision of data; and
- ongoing accessibility of primary research data created under grant sponsorship (for example, NIH and NSF requirements for data archiving).

Similarly, stewardship for intellectual heritage through continuing access to scholarly information is a vital part of the mission of higher education institutions. Books and journals published by independent scholarly publishers (commercial or not-for-profit) no longer represent the sole or even primary record of scholarship. The academy is increasingly directly responsible for disseminating work created within its schools, departments, and research centers, and therefore for assuring that such work remains accessible in an ongoing fashion. Examples of such academy-based scholarship include databases and other "new media" scholarship, institutional repositories, and individual faculty and departmental Web sites that host scholarly content.

A number of components of the technical infrastructure that we recommend to support preservation of digital information are already under development at many institutions. These systems support the new practice of scholarship and teaching in the digital academy—an academy where

- scholarly authors can readily share research reports and primary data under appropriate access controls;

- teachers can easily adapt research materials into a form suitable for beginning and advanced learners, and students can learn the practices of digital authorship as part of their curriculum; and
- business services support the long-term needs of its members, administrators, and stakeholders.

Those components include

- registry services and object repositories;
- metadata templates and repositories;
- identity, authentication, and authorization (IAA) services; and
- data management rules and processing.

In this respect, digital preservation—or life-cycle management—can be used as a framework for guiding the evolution of systems in higher education in the early years of the 21st century.

However, the development of technical systems will probably not be the greatest challenge for colleges and universities as they implement a digital preservation program. The model recommended here requires that digital information be treated, from the point of its creation, as a core institutional asset with stringent maintenance requirements, rather than as an ancillary product of other core activities. This shift, in turn, poses at least three related challenges for colleges and universities:

- Creating greater organizational integration and boundary-spanning than is often found in institutions of higher education. In particular, developing common systems and practices to address both academic and administrative data, as this Bulletin recommends, will require considerable coordination among technology, policy, and library units. This will also require developing a culture that treats administrative data as an institutional asset beyond its immediate departmental use.
- Greater collaboration between information creators and information professionals in the early phases of information creation.
- Developing a culture of stewardship among college and university faculty. The archiving function for scholarly research has traditionally been separated from the process of authorship, through the library's role as a collector of published books and journals. As scholars create and disseminate a wider variety of scholarly genres, they must take a more active role in the preservation process by choosing file formats that can be migrated over time, by documenting their work more carefully, and by storing their work in professionally managed file space.

## Key Questions to Ask

To assess their readiness to undertake a digital preservation program under the model proposed here, colleges and universities will want to ask the following key questions.

- How well is the flow of administrative and academic data understood in our institution? Do we know where data originates, who is responsible for creating and maintaining it, and how it flows from creators to users?
- Do we have well-formulated policies regarding data privacy and access-control systems required to implement those policies?
- What is the institution's stewardship responsibility for works hosted on its systems but not owned by the institution (for example, academic work owned by faculty)? Do local intellectual property policies clearly permit the institution to copy and migrate these works?
- Is the institution prepared to devote resources to instilling a culture of digital preservation and developing the requisite educational, staffing, and technical resources?

## Where to Learn More

- B. F. Lavoie, *The Open Archival Information System Reference Model: Introductory Guide* (Dublin, Ohio: OCLC and Digital Preservation Coalition, January 2004), <[http://www.dpconline.org/docs/lavoie\\_OAIS.pdf](http://www.dpconline.org/docs/lavoie_OAIS.pdf)>.
- National Library of Australia, *PADI: Preserving Access to Digital Information*, <<http://www.nla.gov.au/padi/index.html>>.
- *Preserving Digital Information: Report of the Task Force on Archiving of Digital Information*, Commission on Preservation and Access, and the Research Libraries Group, May 1996, <[http://www.rlg.org/en/page.php?Page\\_ID=114](http://www.rlg.org/en/page.php?Page_ID=114)>.
- RLG/OCLC Working Group on Digital Archive Attributes, *Trusted Digital Repositories: Attributes and Responsibilities* (Mountain View, Calif.: RLG, May 2002), <<http://www.rlg.org/longterm/repositories.pdf>>.
- Library of Congress, National Digital Information Infrastructure and Preservation Program, *Digital Preservation*, <<http://www.digitalpreservation.gov/>>.
- Digital Curation Centre, <<http://www.dcc.ac.uk/>>.
- UC Libraries Digital Preservation Repository, <<http://www.cdlib.org/inside/projects/preservation/dpr/>>.
- A. Smith, *New-Model Scholarship: How Will It Survive?* (Washington, D.C.: CLIR, March 2003), <<http://www.clir.org/pubs/reports/pub114/pub114.pdf>>.

- C. A. Lynch, "Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age," *ARL Bimonthly Report*, No. 226, February 2003, <<http://www.arl.org/newsltr/226/ir.html>>.

## Endnotes

1. Throughout this bulletin the preservation of tangible objects is presented in an oversimplified fashion. Effective preservation strategies for all information formats also require a life-cycle management approach and physical and organizational infrastructure.
2. A *digital asset* is an electronic object that has value for some purpose. It may have been created digitally, or it may have been digitized from a nondigital original source. Examples of digital assets include word-processing documents, databases, Web sites, organizational records, digital recordings of musical performances, and so forth.

To become part of a university's digital preservation program, the digital asset must support (directly or indirectly) the university's fundamental instructional, research, or public service missions. It can be academic or administrative in origin. A digital object should be considered a university digital asset if it satisfies one or more of the following criteria:

- It was created in fulfillment of the research, teaching, or creative work of university faculty, staff, or students.
- It is relevant to the planning, managing, operating, controlling, or auditing of administrative functions of an administrative or academic unit of the university.
- It was purchased or licensed by the university in fulfillment of an academic or administrative function under a contract that permits continuing use of the asset (for example, certain electronic journals licensed through the library).

Not all university digital assets will have equal priority for preservation. This working definition of *university digital asset* establishes minimum conditions for eligibility and does not determine any additional conditions that an object may need to meet to be included in the preservation program or to set priorities for preservation.

3. National Library of Australia (NLA), Migration, <<http://www.nla.gov.au/padi/topics/21.html>>.
4. NLA, Emulation, <<http://www.nla.gov.au/padi/topics/19.html>>.
5. NLA, Encapsulation, <<http://www.nla.gov.au/padi/topics/20.html>>.
6. CCSDS 650.0-B-1: *Reference Model for an Open Archival Information System (OAIS)*, Blue Book, Issue 1, January 2002, <[http://ssdoo.gsfc.nasa.gov/nost/isoas/ref\\_model.html](http://ssdoo.gsfc.nasa.gov/nost/isoas/ref_model.html)>.

## About the Authors

*At the University of Kansas, Richard Fyffe ([rfyffe@ku.edu](mailto:rfyffe@ku.edu)) is Assistant Dean of Libraries for Scholarly Communication; Deborah Ludwig ([dludwig@ku.edu](mailto:dludwig@ku.edu)) is Director, Enterprise Academic Systems; and Beth Forrest Warner ([bwarners@ku.edu](mailto:bwarners@ku.edu)) is Assistant Vice Provost for Information Services (Strategic Initiatives).*

Copyright 2005 Richard Fyffe, Deborah Ludwig, and Beth Forrest Warner. All rights reserved. This ECAR research bulletin is proprietary and intended for use only by subscribers. Reproduction, or distribution of ECAR research bulletins to those not formally affiliated with the subscribing organization, is strictly prohibited unless prior permission is granted by EDUCAUSE (through February 2007) and the authors.