

University of Louisville

ThinkIR: The University of Louisville's Institutional Repository

Faculty Scholarship

Summer 2008

Preserving digital information.

Rachel Howard

University of Louisville, rachel.howard@louisville.edu

Follow this and additional works at: <https://ir.library.louisville.edu/faculty>



Part of the [Archival Science Commons](#)

Original Publication Information

This was originally published in the "Preservation Perspectives" column of *Kentucky Libraries*, volume 72, issue 3, in Summer 2008.

This Article is brought to you for free and open access by ThinkIR: The University of Louisville's Institutional Repository. It has been accepted for inclusion in Faculty Scholarship by an authorized administrator of ThinkIR: The University of Louisville's Institutional Repository. For more information, please contact thinkir@louisville.edu.

PRESERVATION PERSPECTIVES: PRESERVING DIGITAL INFORMATION

BY RACHEL HOWARD

EKSTROM LIBRARY, UNIVERSITY OF LOUISVILLE

Library preservation strategies have long had to adapt to accommodate modifications in formats and materials. For example, in the mid-19th century the demand of an increasingly literate population for reading materials led to changes in paper-making, from high-quality linen or cotton rag stock to more abundant (and acidic) wood pulp, resulting in “brittle books” as the paper aged. Librarians and archivists mobilized behind the issue in the 1970s, which led to the dedication of resources to offset the problem of yellowing, fragile pages, including federal grants for preservation reformatting and climate and humidity controlled environments; the creation of standards for the permanence of paper, and of educational and job opportunities in the field of preservation; and the development of technology for mass deacidification.

Since that time, the environment in which we create and consume research and scholarship, personal, artistic, and journalistic documentation, and entertainment has become increasingly dependent on digital files, software, and equipment. So, we are now faced with the challenge of preserving digital information-only this time, we need to concern ourselves with “bit rot,” or the decay of information within a digital file; hardware, software, and format obsolescence; and server crashes and other disasters. Fortunately, librarians again have had the foresight to begin to create solutions before we enter what some doomsayers have called a “digital dark age” of information loss. The lifespan of born-digital and digitized materials can be extended through proper storage, backups, and documentation that will enable them to be migrated or their playback mechanisms to be emulated in the future. Once again, large-scale initiatives are providing much-needed funding and attention to preservation issues.

The Library of Congress initiated the National Digital Information and Infrastructure Preservation Program (NDIIPP) in 2000 in order to make digital information holding our history and heritage available and accessible for future generations.¹ The University of Louisville is participating in an NDIIPP-funded partnership called the MetaArchive Cooperative, where, along with five other Southeastern research libraries, we have developed a protocol for the distributed preservation of digital cultural heritage materials.²

The MetaArchive approach is based on the open-source Lots of Copies Keeps Stuff Safe (LOCKSS) software developed at Stanford University in the late 1990s and originally intended for the preservation of electronic journals.³ LOCKSS harnesses the power of librarians worldwide to jointly back up the materials to which they subscribe, as well as Web-based open access publications that might otherwise be lost. In the case of a natural or technical disaster, LOCKSS members can recover their electronic journal access from the backups of their peer institutions. Even if tightening budgets were to force the cancellation of an e-journal subscription, at least the back issues have been safely stored!

The MetaArchive is a private LOCKSS network (PLN) wherein the six geographically distributed partner institutions back up one another’s digital cultural heritage content in a “dark archive” not accessible to the public. (The Library of Congress serves as a seventh backup, although we do not back up its content.) Working together as a Cooperative, we set standards and policies for content selection, collection-level metadata, and technical specifications, but each partner is entrusted to make decisions about which formats to submit, how a collection should be arranged, and

whether or not to include item-level metadata. We each operate servers running LOCKSS software, and LOCKSS daemons which are computer programs running in the background, to locate and harvest authorized digital content from each partner's sites. Daemons continuously compare the files on each server and determine (by majority rule) whether a file has been damaged, restoring it from another server's cache if necessary.

This process guards against the "bit rot" of an individual digital file, while the multiple servers distributed in different locations guard against the loss of content due to hardware

failure or a natural disaster. Format, software, and hardware obsolescence remain inevitable, but we feel that by making a strategic commitment to digital preservation in the MetaArchive, we are prepared to cooperatively and proactively address migration and emulation when needed.

The MetaArchive Cooperative is accepting new partners. See <https://www.metaarchive.org/join.html> for details about joining the partnership.

Rachel Howard
rachel.howard@louisville.edu

WORKS CITED

- 1 Library of Congress. Digital Preservation. 9 June 2008 <<http://www.digitalpreservation.gov>>.
- 2 MetaArchive Cooperative. 9 June 2008 <<http://www.metaarchive.org>>.
- 3 Stanford University. Lots of Copies Keeps Stuff Safe (LOCKSS). 9 June 2008 <<http://www.lockss.org>>.