

Preservation of Scholarship

The Digital Dilemma

Deanna B. Marcum

DIGITAL TECHNOLOGY'S INFLUENCE on higher education has been a subject of great interest in recent years. Faculty are creating Web-based resources for classroom use; researchers gather data in massive amounts that can be manipulated and analyzed only in digital form; scholars are sharing their findings on Web sites and in e-journals; and students get assignments and do course work online.

We know from the recent study funded by the Pew Charitable Trust, *Counting on the Internet*,¹ that three-quarters of the more than 2,000 students surveyed from the full range of two- and four-year institutions use the Internet more frequently than the library. The Council on Library and Information Resource's (CLIR) study² of more than 3,000 faculty, graduate students, and undergraduates in universities and liberal arts colleges had similar findings. All groups, and undergraduates especially, depend primarily on the Internet for discovering information. Over half of the undergraduates surveyed³ in the CLIR study much prefer finding the information they need on the Internet

and go to the library shelves only when that is the only way to get the information. Anecdotal evidence from faculty tells the same story. If assignments can be completed by using the Web, that is the first resource undergraduates turn to.

It is not surprising, then, that faculty, librarians, courseware providers, and others are moving quickly to develop tools and content for the Web. The question such activity raises is, “Will these digital resources be preserved for future generations?” For the past two years, I have asked just that of several speakers at various forums who presented wonderful examples of digital content they had created. Generally, they looked away and said, “Good question. We’ll have to deal with that later.”

Why the Concern?

I believe we must address the problem of preserving digital information immediately—or the history of our academic achievements in the early twenty-first century will be incomplete, at best, or lost forever, at worst.

Jeff Rothenberg, a computer scientist at Rand, said of digital preservation,⁴ “Digital documents last forever—or five years, whichever comes first.” He further explained:

The physical lifetimes of digital storage media are often surprisingly short, requiring information to be “refreshed” by copying it onto new media with disturbing frequency. The technological obsolescence of these media (and the hard-

PRESERVATION OF SCHOLARSHIP

ware and software necessary to read them) poses a different and equally urgent threat. . . . As these programs (or the hardware/software environments in which they run) become obsolete, the digital documents that depend on them become unreadable—held hostage to their own encoding. This problem is paradoxical, given the fact that digital documents can be copied perfectly, which is often naively taken to mean that they are eternal.⁵

The problem is evident to any computer user who has stared at a screen full of gibberish when trying to open a document created in a different word-processing program, who has wondered how to save data from a “crashed” computer, or who has looked in vain for the slot on the new computer that will accept floppy disks from a previous era. Those who preserve library collections must deal with nontechnical questions as well, such as, Who is responsible for preserving an e-journal—the creating publisher or the access-providing library? For use by whom? For how long? And with what copyright provisions?

Faculty and administrators, for the most part, have expectations for long-term access to materials based on the paper model: Faculty conduct research and write journal articles or books, and libraries buy those books and journals for long-term retention. Administrators have been willing (usually) to fund these library purchases because the resulting collections are the long-term assets of the institution upon which new scholarship and knowledge creation depend.

The preservation problem first surfaced when journal pub-

lishers began distributing their articles in electronic form. Publishers stopped selling journals to libraries. Instead, they licensed the electronic content to libraries so that the information could be made available to members of the campus community. The journal was no longer a well-defined entity, but rather a database that could be configured at the point of use to display the article of interest to a reader. Librarians began to ask publishers about their plans for making the electronic materials available into the future, always worrying about the eventuality that the publishing house would be bought by another, would go out of business, or would drop a particular journal.

Electronic journals have now been with us for more than a decade, and although a small number of the large publishers have announced that they are taking responsibility for maintaining their electronic content, most of the smaller publishers are still in a quandary about preserving their materials. Preservation costs, traditionally borne by the library, are hard to transfer to the underfunded, struggling scholarly society or university press. JSTOR is a notable example of a nonprofit organization that began digitizing old issues of journals for purposes of access, while also building a digital repository for the purpose of preservation.

Funding agencies are eager to see the benefits of information access spread to all segments of society. Thus they are willing to fund large-scale conversion projects, allowing libraries to convert rare and special collections of audio, visual,

PRESERVATION OF SCHOLARSHIP

and manuscript collections to digital format for purposes of improved access. In recent years, as faculty members have undertaken more research-based and classroom-directed projects, concerns about preserving digital information resources have grown.

What Scholars Say

In April 2002, with funding from the Alfred P. Sloan Foundation, CLIR convened a small group of scholars who had created digital content for the Web. We recognized that these creators were interested in inventing tools that foster productive use of the Web as a medium of scholarship and teaching. We also knew that they were having trouble finding long-term stewards for the materials they created.

We acknowledged that it was impractical to simply tell the scholars to create something preservable, as much as some librarians would like to do so. We used the opportunity presented by this two-day meeting to ask creators what they are doing and learn more about what they want to have happen to their materials over time, and then to ask archivists, librarians, and technologists what has to be in place to enable them to act as stewards of the digital creations.

The kinds of projects that the scholars have been able to undertake—with Sloan funding—illustrate the preservation problem quite well.

Documenting Big Science

Babak Ashrafi of the Dibner Institute at the Massachusetts Institute of Technology is overseeing a project called the History of Recent Science and Technology. The project's aim is to document "big science." At the April 2002 meeting, Ashrafi explained that it is no longer sufficient for one person to choose one scientist and write a biography of the person and his or her ideas. Today, groups of historians collaborate on large projects, and scientists and engineers help them overcome the technical hurdles they encounter. One such project focuses on the physics of scales. (Since the early 1970s, physicists have attempted to understand cosmic phenomena by using methods that relate physical quantities at one scale to those of another scale.) Project team members are attempting to document the scientific developments in this area from World War II through the mid-1970s. The team of historians is interviewing scientists, mounting critical documents on the Web, collaborating with living scientists to annotate the documents, and engaging a few of the scientists to moderate Web-based forums. The scientists become "communication nodes" who help recruit other collaborators and students. Currently, three historians and two dozen scientists are working on the topic; eventually 60–80 scientists will be involved in building an archive about the fields they helped develop.

Ashrafi described the essential problem in creating interactive tools for scholarship as one of making tradeoffs between standardization, which is essential for digital preservation, and

flexibility, which allows scientists to customize the materials to make them more useful to their scholarly colleagues. He said, “The History of Recent Science and Technology has to be a site for and by historians. We do not want chat rooms. We want a heavily moderated archive that will be successful 150 years from now for people who want to look at the development of the theory of physics of scales. They will come to our archive. It will be there, and it will be authoritative.”⁶

Hypertext Scholarship on the Web

Roy Rosenzweig, director of the Center for History and New Media at George Mason University, saw the problem as a quite simple one: “We want to get things [digital content] out the door.”⁷ History faculty are not technologists. They are merely using technology as a medium for doing work they have not been able to do before. He praised Web materials for their ability to offer a more democratic and inclusive view of the past than typical historical documents. The Web can facilitate new modes of learning about the past that encourage student participation and engagement, and enable innovative scholarship that challenges traditional ways of doing history.

One project launched by Dr. Rosenzweig used the flagship journal in American Studies, *American Quarterly*, published by Johns Hopkins University Press, to present hypertext scholarship on the Web. He identified scholars whose work demanded hypertext presentation that went beyond what was typically included in a print-based article. He found working with the

publisher to be enormously difficult because it was unwilling to depart from its standard format. In the end, Dr. Rosenzweig worked with his group of authors and mounted the materials on the Web without respect to *American Quarterly's* standard formats. On the one hand, he regretted that decision because the materials probably will not be preserved. On the other hand, innovation in scholarly techniques must be disseminated quickly and boldly if they are to have an influence.

As part of another project of the Center for History and New Media, Dr. Rosenzweig created a Web site to which anyone and everyone is invited to contribute stories about their experiences of September 11, 2001. He has found that demanding requirements imposed on potential contributors simply scares them away. He emphasizes the value of the inclusiveness of the Web, and insists that this kind of history, which includes many segments of society, will work only if it is not too structured.

Rosenzweig acknowledges that librarians could be helpful to his team in creating formats and standards for materials so that they can be preserved, but emphasizes that time is of the essence in his projects. He cannot imagine working out the standards before a project is launched. His motto is, collect it first, then figure out how to keep it.

Recreating Victorian London

At the University of Virginia's Institute for Advanced Technology in the Humanities, John Unsworth and Michael Levenson

PRESERVATION OF SCHOLARSHIP

have collaborated on a project to recreate Victorian London. Their Web site includes a three-dimensional model of the Crystal Palace showing every nut, bolt, wire, and pane of glass. The 3D views are possible because of programming in Quick-Time. While this level of detail is enormously useful for architects, landscape designers, and historians, it will be challenging to maintain the model over time because there are no XML/SGML standards for 3D models. The University of Virginia library has agreed to make the material available as long as possible, but when changes are made in hardware and software, it is unlikely that a 3D model can be migrated successfully to a new system. Unsworth's plea is that publishers reconceive their role to include converting material that originates in digital form into a format that libraries can collect.

What Librarians Say

Librarians take seriously their role to collect materials that will be useful to current and subsequent communities of users. They have always faced difficulties in acquiring everything that all members of the campus community want. Nonetheless, as a result of thoroughly considered collection development policies and cooperative and consortial agreements, libraries for the most part meet the needs of their faculty and students, and have managed to effect preservation plans ensuring that the most important scholarly materials will be available for future generations.

With an increasing number of faculty creating Web-based scholarship and classroom materials, however, the focus of preservation responsibility has shifted from the library to the individual creator. Librarians are justifiably concerned that the result will be a serious loss of important materials. In some cases, faculty are simply experimenting—exploring what is possible with digital technology. They do not imagine that the work they are doing has lasting value or that digital media may become unreadable. Yet, the history of how the Web has altered and influenced teaching and learning will be known only through the preservation of a significant number of these experiments.

Bernie Hurley of the University of California, Berkeley's digital library calls for striking a balance between what we need to archive because of its intellectual value and what we can accomplish with the technology we have. His approach is to encourage faculty to use the commonly accepted standards for their projects so that their materials can migrate from one system or platform to another. At Berkeley, the library has created a digital repository, and librarians are soliciting materials from the faculty. Unless librarians reach faculty at the earliest stages of their projects, however, it is often the case that the materials have been built on proprietary standards and cannot be incorporated into the library's repository.

Dale Flecker, Associate Director for Planning and Systems at Harvard University, reports a quite different approach. At Harvard, individual curators and librarians are responsible for selecting material to be preserved. They treat digital material

PRESERVATION OF SCHOLARSHIP

in the same way they treat all other formats: the librarian/specialists determine which materials are most likely to be required by scholars and students of the future, and then proceed to do whatever is necessary to acquire that material and keep it alive over time. Digital material created outside the library system poses enormous problems because standards cannot be enforced at the creation stage. Harvard also has concentrated on making arrangements with publishers to archive their digital content—a process that has been fraught with tensions related to access and payment for use.

The Future of Scholarly Resources in the Balance

Scholarly communication in the twenty-first century is complicated. Today, content creators play a new role in the chain of scholarly communication—as do publishers, libraries and archives, and users. From the custodian's point of view, it is important to capture the attention of the content creators while they are in the process of creating digital content so as to affect the decisions they make, consciously or unconsciously, and help them become the stewards of their own intellectual property.

Ultimately, libraries and archives may find themselves in the business of negotiating with repositories that store digital bits so that they can make the necessary arrangements for providing access to digital materials to those who need them. It is likely that a few regional repositories with massive digital collections

may be established around the country, and local libraries will serve as intermediaries on behalf of local “customers.”

The most important question concerning the preservation of digital scholarship is, “How do scholars and librarians work together to ensure that resources created today will be available in the future?”

NOTES

1. *Counting on the Internet*, Internet and American Life Report, Pew Charitable Trust, December 29, 2002.

2. Amy Friedlander, *Dimensions and Use of the Scholarly Information Environment: Introduction to a Data Set Assembled by the Digital Library Federation and Outsell, Inc.* (Washington, D.C.: Digital Library Federation and Council on Library and Information Resources, November 2002).

3. Ibid.

4. Jeff Rothenberg, “Digital Information Lasts Forever—Or Five Years, Whichever Comes First,” RAND Video V-079, 1997.

5. Jeff Rothenberg, *Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation* (Washington, D.C.: Council on Library and Information Resources, 1998), p. 2.

6. Remarks at a workshop on “New Model Scholarship,” convened by CLIR with support from the Alfred P. Sloan Foundation, Washington, D.C., April 23–24, 2002.

7. Ibid.

Deanna B. Marcum is President of the Council on Library and Information Resources.