Digital documents are far from eternal. Anyone who has ever stared at a computer screen full of gibberish when trying to open a document created in an early word-processing program, or looked in vain for the slot on a computer that accepts floppy disks from a previous era, knows firsthand about the limited lifetimes of digital storage media. For the most part, higher education faculty and administrators have high expectations for long-term access to digital materials, based on their experience with the paper model. Yet today technology has moved us well beyond that model, giving rise to pressing issues and concerns about the preservation of digital scholarship. Deanna Marcum, president of the Council on Library and Information Resources, and Clifford Lynch, director of the Coalition for Networked Information, discuss the scope of these issues and some promising responses to them.

The Problem

Jeff Rothenberg, a computer scientist at Rand, has written, “Digital documents last forever—or five years, whichever comes first.” He explains that digital documents dependent on obsolete hardware or software to run are hostage to their own encoding. They become unreadable unless they are “refreshed”—copied onto new, updated media—with disturbing frequency.

Librarians started to recognize the digital preservation problem when journals began distributing their articles in electronic form a decade or so ago. 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 ensuring that these electronic materials would be available in 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 been with us for more than a decade now, and although a small number of the largest publishers (and a few of the most thoughtful and financially strongest of the scholarly societies) have announced plans to maintain their electronic content, most of the smaller publishers are still in a quandary about the question of preserving their materials. Preservation costs, traditionally borne by the li-
library, are hard to transfer to the underfunded, struggling scholarly society or university press.

Today, the problem extends well beyond the preservation of electronic journals. In recent years, faculty scholarship has incorporated more and more digital media for research and classroom-based projects, and concerns about preserving these newly created digital information resources have grown accordingly.

**Perspectives on the Problem**

Faculty working with new learning media are creating some of the most innovative scholarship today. But their concerns, rightly, are with how to use the new digital media most effectively to convey and document scholarship, and not with technical questions about our ability to preserve the works they create. Further, in too many cases, the pathbreaking work being done by these early adopters resides on their computers, which are being rather informally managed; they are not part of any institutional infrastructure or program to preserve and maintain access to content. If we were to lose the scholar—to a new interest, retirement, or the proverbial bus—we would lose the scholarship.

A few examples illustrate the range of the issues we face:

Babak Ashrafi, of the Dibner Institute at MIT, is overseeing a massive project aimed at documenting “big science.” He is collaborating with a group of historians who expect to eventually involve 60 to 80 scientists in creating an archive about the fields these scientists helped develop. The project includes conducting interviews, posting key documents on the World Wide Web, collaborating with the scientists to annotate the documents, and having some of them moderate Web-based forums. The result will be a rich interactive tool for scholarship. Ashrafi describes the essential problem as a trade-off between standardization, which is essential for long-term digital preservation, and flexibility, which allows the scientists to customize the materials to make them more useful for their scholarly colleagues.

Roy Rosenzweig, director of the Center for History and New Media at George Mason University, sees the problem as one of speed—or lack thereof. One of the center’s projects is a Web site to which anyone is invited to contribute stories about their experiences of September 11, 2001. Rosenzweig found that demanding requirements imposed on potential contributors simply scare people away. He notes that the kind of history that includes many segments of society cannot be too structured or it risks breaking down the Web’s inclusive approach, which he thinks is tremendously valuable. He acknowledges that librarians could be helpful in creating formats and standards for materials so they can be preserved, but time is of the essence in his projects. He cannot imagine working out the standards before the project is launched. At this point, his approach is to collect the information and figure out later how to keep it.

Finally, at the University of Virginia’s Institute for Advanced Technology in the Humanities, John Unsworth and Michael Levenson have created a digital three-dimensional model of Victorian London’s Crystal Palace. The model, which shows every nut, bolt, wire, and pane of glass, is enormously useful for architects, landscape designers, and historians. Yet because there are no technical standards for 3-D models, it will be a challenge to save this work when the hardware or software used to create it is changed. Unsworth and Levenson’s hope is that publishers will reconceive their role to include reshaping digital material into a form that libraries can collect.

At this stage, though, with an increasing number of faculty creating Web-based scholarship and classroom materials, de facto responsibility for preservation has shifted from the library to individual creators. Librarians are justifiably concerned that the result will be a serious loss of important materials.

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 commonly accepted standards for their projects so 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.

From the custodian’s point of view, it is important to capture the attention of content creators while they are in the
process of creating digital materials so as to affect the deci-
sions they make, consciously or unconsciously, and to help
them become stewards of their intellectual property. Unless
librarians reach faculty at the earliest stages of their projects,
it is often the case that their materials are built on proprietary
standards and therefore can be incorporated into the library’s
repository only with considerable difficulty.

Dale Flecker, associate director for planning and sys-
tems at Harvard University, reports a quite different ap-
proach. At Harvard, individual curators and librarians are
responsible for selecting material to be preserved. They
treat digital material in the same way that they treat all
other formats: the librarians/specialists determine which
material is most likely to be required by scholars and stu-
dents of the future and then proceed to do whatever is nec-
essary 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 has concentrated on making
arrangements with publishers to archive their digital con-
tent, a process that has been fraught with tensions related
to access and payment for use.

Organizational Responses
The stakes in addressing digital preservation problems are
high. Publishers today, particularly in the sciences, are
distributing the same materials in both electronic and
print formats, and in some cases libraries are still receiv-
ing both forms, using the print largely for preservation.
The financial savings would be substantial if publishers
could discontinue the printed versions, but this can’t be
done until there is a persuasive preservation strategy for
the digital materials. Abandoning the print format not
only would save money, but would also clear a path to-
ward the more extensive exploitation of the capabilities of
the digital medium by authors. The big questions here are
ones of economics: Who pays for archiving the digital
materials, for distributing the work among the appropriate
organizations (national libraries, research libraries, and
other players), or for developing appropriate licensing
and access models to ensure not only preservation but
community confidence in the system of preservation?

In some cases the primary reason to convert materials
into digital form is enhanced access, but these programs
often also help to ensure preservation of the materials.

JSTOR is a notable example of a nonprofit organization
that began digitizing old issues of journals for purposes of
access, at the same time building a digital repository for
preservation purposes. Other funding agencies, eager to see
the benefits of information access spread to all segments of
society, have been willing to support large-scale conversion
projects, allowing libraries to convert rare and special col-
cections of audio, visual, and manuscript collections to dig-
ital format for purposes of improved access.

The recent report of the National Science Founda-
tion’s Advisory Committee for Cyberinfrastructure,
chaired by Daniel Atkins, notes that as the sciences de-
pend more on digital data, the need for data management
services, including those for its preservation and curatorial
oversight, becomes more pressing. The need to store,
manage, and preserve digital databases and data sets to fa-
cilitate their use by scientists may lead toward some sort of
national system of stewardship, perhaps funded and or-
ganized by the scientific communities. In some cases fed-
eral science agencies such as the National Library of Med-
icine or NASA may play a direct role; in other cases or-
ganizations such as the NSF may fund centers within the
research and higher education community to curate data,
much as they have created supercomputer centers to sup-
port access to high-performance computing services.

Our focus here has been on the digital products of
scholars in the higher education community, be they part
of the traditional book and journal publishing system,
which is moving to digital distribution, or innovative new
works outside the system that are exploring the new capa-
bilities of digital media. Even in the case of for-profit commercial publishers, there’s a clear shared set of values that honor the importance of preservation, and we can expect that over time the scholarly publishing community will reach agreements to ensure that these materials are preserved.

But higher education doesn’t exist in a vacuum: higher education’s content is minimal compared with the vast amounts of digital content produced in the news, cultural, and entertainment worlds. These are the raw materials of future scholarship, and yet they are not subject to the shared values about preservation that help us in the scholarly domain. Forces and trends stemming from those realms—such as pay-per-view and restrictive copyright laws—will make it extremely difficult for traditional noncommercial organizations such as libraries and universities to archive the mass-market digital materials that may be crucial to the scholarship of future decades. The Library of Congress recently received funding for planning the development of a national digital preservation strategy intended to broadly address the American intellectual and cultural record (as opposed to the narrow scholarly sphere); it has already begun a series of what will undoubtedly be difficult negotiations with mass-market content owners.

Notwithstanding these difficulties, there is exciting news at the institutional level, where work is under way on the creation of repositories. In this context, a repository is an institutionally managed place for housing the intellectual assets produced by an organization. It deals with issues of access and of preservation. The technology for institutional repositories is the least of the hurdles that must be overcome. Rather, the notion of such repositories begs questions related to management of the intellectual assets, such as stewardship and organizational structure, institutional access and content policies, and funding.

Not every institution will need its own digital repository. Consortia offer perhaps the best solution for smaller institutions; another option would be to form groups based on geographic location rather than size. Further, institutions can collaborate on many levels: they could share just the technology necessary to house digital media, for example, or they could share managerial and curatorial roles.

**Conclusion**

The stakes in this discussion are high: the future of scholarly resources lies in the balance. Responsible preservation of digital scholarship—some of it the most innovative work happening on campuses today—will help to both encourage different and creative approaches to scholarship and legitimize such efforts. Moreover, the issue of preservation of the raw materials for future research and creation of knowledge is one that extends beyond the higher education community to society as a whole. At another level, a core responsibility of colleges and universities is to maintain their institutional records, now available nearly exclusively in digital form, and by so doing to preserve their institutional memories.

The difficult questions surrounding the preservation of digital scholarship and media demand careful thought and broad discussion within the higher education community. Absent that, these invaluable resources could simply evaporate and disappear forever as a result of inattention to the issue.

Clifford Lynch is director of the Coalition for Networked Information (CNI). CNI, jointly sponsored by the Association of Research Libraries and EDUCAUSE, includes about 200 member organizations concerned with the use of information technology and networked information to enhance scholarship and intellectual productivity.

Deanna Marcum is president of the Council on Library and Information Resources (CLIR), formed by the merger of the Commission on Preservation and Access and the Council on Library Resources. Prior to joining CLIR, she was the Director of Public Service and Collection Management at the Library of Congress.