E-Content

By Abby Smith

Digital Preservation: An Individual Responsibility for Communal Scholarship

Ask anyone to conjure an image of a scholar, and he or she will probably envision a figure in solitary concentration, frowning over a text: St. Jerome, perhaps, whom we see in widely reproduced images as a monk alone in his study pursuing knowledge of some divine subject. This is a common misperception, however. Scholarship is in fact a highly social activity in which communication—among peers and across generations, spanning time zones and great distances—is essential for knowledge creation. St. Jerome was not alone in that study; the colleagues with whom he was communing were present through their written works, if not in person.

To this day, scholarship depends on access to recorded information, access that digital technology is expanding in mind-boggling ways. Of course, we are also familiar with new types of barriers to access in the digital realm: bandwidth limits, unequal resources for technology, copyright restrictions, and so forth. But the greatest risk to present and future access to digital information is none of these; rather, it is inadequate preservation.

Preservation encompasses all of the actions that must be taken to ensure that information resources will be fit for use at some time in the future, including capture, curation, and storage. In the world of books, the burden of preservation has been shouldered almost entirely by libraries, which rely on a sturdy recording medium—acid-free paper—and on controlled storage conditions to ensure longevity. When those conditions are met, librarians rarely have to intervene. Digital preservation, on the other hand, requires regular—almost constant—intervention and cannot be accomplished by librarians alone. From the moment a file is created, it is crucially dependent on software, hardware, and fragile storage media, such as a PC hard drive or a magnetic tape. In some sense, the creator, knowingly or not, makes all the critical decisions about life expectancy in the first few hours after the file's conception: which software to encode the creation, which hardware to play it back, and how to name the files.

Scholars in all fields are creating all sorts of scholarly resources in digital form, from simple word-processing documents headed for publication in peer-reviewed journals and classroom Web pages to more complex digital objects such as scans of the neurological functioning of brains, computer simulations of seismic instabilities, and virtual-reality renderings of Roman forums. Yet few scholars are aware of their roles in the preservation of their own intellectual property. They have arrived in the digital realm rather suddenly, their work habits forged in the analog era when preserving (in addition to cataloging and collection building) was the exclusive concern of other professionals on campus.

Why does preservation matter to these scholars? Humanists, of course, have long depended on historical resources for their research and teaching. Scientists have, until recently at least, been less concerned about access to information from the past. But that is no longer the case. At the beginning of the twenty-first century, both humanists and scientists are threatened by the vulnerability of digital resources. Increasingly, scientists develop longitudinal data and construct digital maps—for example, of the sky or of the United States—that will take decades to complete and that must be maintained for decades more and longer. Similarly, the digital conversion of humanities texts—the works of Shakespeare or classical writings—demand a high investment of resources and should be preserved. Additionally, all scholars share the need for a scholarship and professional literature record that is reliable, readily accessible, and authentic.

The question remains, however: Who is responsible for preservation? For their part, libraries and publishers are beginning to step up to the challenge of preserving electronically published professional literature, as Dale Flecker previously noted in this department.[1] Although we are a long way from having agreed-upon procedures for archiving scholarly journals supported by robust business models for sustainability, we can, as Flecker points out, cite more than one example of libraries and publishers forging partnerships on behalf of scholars.

But where are the scholars themselves? Although many of them are creating digital objects worthy of access over time, these objects may not be destined for publication and therefore may not fall under the stewardship of libraries. What will happen to a Web resource built by a scholar once the scholar moves on and no longer supports it? And what about the numerous databases and information resources that have been built, usually with taxpayers' money, by scientists engaged in long-term research projects outside the purview of federal agencies such as the National Aeronautics and Space Administration or the National Oceanic and Atmospheric Administration? Such agencies are careful stewards of their own data resources, but what about the valuable resources that exist on campuses and in labs that lack preservation capabilities?

A blue-ribbon commission of the National Science Foundation recently concluded that the promise of transforming science and engineering through information technologies will be realized only with the building of an infrastructure that supports reliable information creation and transmission over time.[2] The sciences are not the only disciplines being

transformed by information technology, though they certainly are the best funded and are in the best position to manage their information resources responsibly. They are doing so in many cases by developing so-called informatics, or information-science specialties, which have prestige and are well rewarded.

In the humanities, however, where the practice of building common resources and mining them with sophisticated computer technologies in teams has not taken hold—at least not yet—the challenge is to help individual scholars develop the good working habits that will ensure the integrity of the resources until they are transferred elsewhere for long-term custody. Humanists may in time also develop informatics in their fields. For now, those who are working in new-model scholarship can do well by partnering with their libraries *from the beginning* to create digital resources that can be preserved.

New-model scholarship requires new models of collection development, cataloging, and preservation. Scholars must manage their own data and create their own libraries of working materials, and libraries must collaborate in maintaining those collections over time. MIT, Harvard, the California Digital Library, Stanford, and other research libraries are developing programs that promote such collaborations. Libraries, accustomed to focusing on preserving content, will need to attend to the software requirements of digital resources as well as the data in them. Often, digital research resources are most valuable for the tools and behaviors embedded in them. Librarians have been deft at preserving both the informational content of a book or map and the physical object itself when the artifact in question has research value. In the digital realm, they will likewise need to distinguish between the value of preserving data alone and saving both the data and the functionality embedded in the software.

We cannot expect any one academic institution, or even an association of them, no matter how well-funded and well-intended, to perform the preservation tasks demanded by the growing body of digital information. For this we must continue to rely on agencies and institutions that serve large communities. Examples of such entities include GenBank, the Inter-university Consortium for Political and Social Research, and JSTOR.

Fortunately, the problem of preserving digital resources has gained national attention. The Library of Congress has recently completed the planning phase of its National Digital Information Infrastructure and Preservation Program (NDIIPP), a program funded for up to $100 million to design and begin to build a national infrastructure to preserve library digital content (the sort of resource usually protected by copyright). The NDIIPP plan identifies a series of areas for action: selection; intellectual property; business models; standards and best practices; communication and outreach; and technology and infrastructure. This plan expects that the academy will take primary responsibility for the information resources it creates for its own use, especially those resources that are not published and deposited for copyright protection.

The academic community must act now to educate its members about their needs and to inculcate in them a sense of stewardship over the information assets on which scholarship depends. For without preservation, scholarship will become a truly solitary pursuit, solipsistic and cut off from its connections with both the past and the future.

Notes

1. Dale Flecker, "Digital Archiving: What Is Involved?" *EDUCAUSE Review* 38, no. 1 (January/February 2003): 10–11.
2. National Science Foundation, "Revolutionizing Science and Engineering through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure," January 2003, <http://www.communitytechnology.org/nsf_ci_report/> (accessed February 25, 2003).

Abby Smith is Director of Programs at the Council on Library and Information Resources.

Illustration by Philip Kaake, © 2003