Digital Preservation: An Individual Responsibility for Communal Scholarship

By Abby Smith

A sk anyone to conjure an image of a scholar, and he or she will probably envision a figure in solitary concentration: a scholar writing a book or manuscript, reading 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 his study; the colleagues with whom he was communicating were present through their written correspondence.

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, unique 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 remain usable over time. It is a matter of concern for 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 librarians, 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 records 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 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 emerging in the analog era when preserving the physical content of a book or map and the intellectual content of a scholarship and professional literature, as Dale Flecker previously noted in this department.

We know preservation three ways. Schol- larly records require regular—almost constant—inter- vention 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 records 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 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 emerging in the analog era when preserving the physical content of a book or map and the intellectual content of a scholarship and professional literature, as Dale Flecker previously noted in this department. Although we are a long way from having agreed-upon procedures for archiving scholarly journals supported by robust business models, we can do even better. We can do even better. We can do even better.

The question remains, however: Who is responsible for preservation? For their part, publishers and librarians are beginning to step up to the challenge of preserving digital research resources. Libraries, accustomed to managing their own data and creating their own digital content (the sort of resource we see in widely reproduced images as a monk alone in his study pursuing knowledge of some divine subject) are careful stewards of their own data but 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, the practice of building common resources and mining them with sophisticated computer technologies in teams 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 preservation. Librarians have come to realize that providing support in setting up such shared environments for managing and sharing data is a crucial task. The digital realm, they will likely 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 institutions, to be 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 STOR.

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), which was funded for up to $30 million in the fiscal year 2002 by a national initiative to preserve digital content (the sort of resource usually protected by copyright). The NDIIPP plan identifies a series of areas for action: selection; intellectual prop- erty; business models; standards and best practices; communication and outreach; technology and infrastructure. This plan expects that the academy will take primary responsibility for the informa- tion technologies that serve large communities. We know preservation three ways. Scholarly records require 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 records 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 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 emerging in the analog era when preserving the physical content of a book or map and the intellectual content of a scholarship and professional literature, as Dale Flecker previously noted in this department. Although we are a long way from having agreed-upon procedures for archiving scholarly journals supported by robust business models, we can do even better.

The question remains, however: Who is responsible for preservation? For their part, publishers and librarians are beginning to step up to the challenge of preserving digital research resources. Libraries, accustomed to managing their own data and creating their own digital content (the sort of resource we see in widely reproduced images as a monk alone in his study pursuing knowledge of some divine subject) are careful stewards of their own data but 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, the practice of building common resources and mining them with sophisticated computer technologies in teams 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 preservation. Librarians have come to realize that providing support in setting up such shared environments for managing and sharing data is a crucial task. The digital realm, they will likely need to distinguish between the value of preserving data alone and saving both the data and the functionality embedded in the software.

Notes

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

© 2003 Abby Smith

EDUCAUSE Review May/June 2003