Digital Archiving: Who, What, When, Where, Why, and How?

Panelists:
Michael A Keller
HighWire Press and Stanford University Press
Stanford, California

Kevin M Guthrie
JSTOR
New York, New York

Deanna Marcum
Council on Library and Information Resources
Washington, DC

Reporter:
Leon Jeter
Federation of Animal Science Societies
Savoy, Illinois

Michael A Keller delivered a wealth of information on this timely and complex topic. Answering the interrogative *why*, Keller explained that we archive materials because of the subject's intrinsic significance: If it has withstood the tests of editorial scrutiny or marketability, we had better save it. In addition, we save works so that we can chart social or technical progress. How long a work should be saved is difficult to determine. Intertwined with this part of Keller's presentation were answers to the interrogative *who*. Traditional archivists have been private collectors, libraries, and publishers. Today, a trusted third party might enter the scene to provide electronic archiving.

What form of archive? We are seeking to understand today's and tomorrow's electronic means of archiving. Keller mentioned our shared concerns about multiple formats at the multiple levels of production: From the author's desktop to the publisher's workstation and eventually to the user's terminal or printer, the problem of maintaining the usability (print, graphics, audio, and video) challenges all. As Kevin M Guthrie was to mention later, money must be allocated for future data "migrations" to accommodate inevitable hardware and software developments. The third speaker, Deanna Marcum, strongly emphasized that we in publishing must monitor the accessibility of our digitized materials. Whether we store the data ourselves or entrust the job to others, we must know how long our material will be usable.

When to archive? "Now!" So rapid is technologic development that the software and hardware applications used to present our archived works might exist for only 5 years.

Where is answered through the concept of distributed and replicated copies, a reasonably fail-safe system that we have enjoyed with traditional libraries. Again, the issue of data formats—their proliferation and their ultimate accessibility to readers—is paramount. Keller noted the wisdom of storing application, source, and production files at different locations to ensure safety.

How to archive our digital materials pertains directly to the cross-format issue. Magnetic media endure only a few years and CDs only a couple of decades; the layers of materials used in those media deteriorate at different rates, and that causes data loss. New and soon-to-be-tried media are those made with nonoxidizing substrates. The speakers emphasized the importance of having duplicate storage locations. Keller touched on data encapsulation but reminded us about the extremely high rate of change in the information technologies. All three speakers called for explicit standards and for best-practice examples to be studied as various institutions seek to become trusted third parties; explicit statements of intended practices by such organizations are necessary.

Marcum brought library and information-science interests to bear. Her extensive experience in discussion groups in the 1990s and this year has given her particular insights. She emphasized that publishers and librarians want some of the same things, such as preservation and access. She offered the Web address for a Council on Library and Information Resources (CLIR, pronounced "clear") publication called *CLIR issues*. Her article therein reports on minimal requirements for digital repositories.[1]

Guthrie described JSTOR, a digital repository that combines preservation with access. It enables readers at "subscriber" libraries to view millions of pages' worth of social science and humanities journal articles. JSTOR and HighWire Press are examples of trusted third-party digital archives. Guthrie contrasted paper and digital archiving and noted the inherent geographic limitations of the dissemination of paper copies compared with the freedom of Internet-based transmission. Guthrie emphasized the importance of copyright issues and their inclusion in contracts. He stressed the need to back up one's digitized holdings by giving the example of JSTOR's three mirror databases, residing in three cities, that are reduplicated nightly.

We must identify the future users and begin some reasonable method of archiving right away. As interested members of a community engaged in scholarly communication, we should experiment with ways of creating and operating our digital archives and share our results with our colleagues.

1. Marcum DB. Establishing minimal requirements for archival repositories. CLIR issues 2000;15:May-Jun, *www.clir.org/pubs/issues/issues15.html* (13 Sep 2000).