

# Digital archiving

## Whose responsibility is it?

by Deanna Marcum

In the world of paper, preservation has been the clear responsibility of libraries. While not every library has made it a high priority, many institutions—individually or consortially—have committed resources to preserving the materials they have acquired that have long-term value. However, this model of preserving acquired materials does not apply easily to electronic journals, which publishers typically license to libraries for use.

Because so much of the scholarly record is found in journal articles, and because libraries are relying more on providing access to the electronic versions than on acquiring the paper copies, there is growing interest in establishing archival repositories for electronic journals.

In 1996, the Research Libraries Group (RLG) and the Commission on Preservation and Access issued a report, "Preserving Digital Information," based on the work of the Task Force on Archiving Digital Information, co-chaired by Donald Waters and John Garrett. The task force recommended the establishment of certified digital archives—agencies that spell out how they are preserving digital objects and that can be held accountable for fulfilling their promise. The task force also called for a fail-safe legal mechanism that would allow these certified archives to rescue digital files that are in danger of being lost.

Perhaps the call for such a formal mechanism was too jarring to the library community four years ago for specific steps to be taken. However, as libraries rely increasingly on licenses for electronic journals, both librarians and publishers have become more aware of the potential problems. When negotiating licenses, librarians are quick to ask what the publisher has done to ensure long-term access to the material. Since libraries traditionally have financed preservation, publishers have been forced to consider new business models that include the cost of preserving digital information.

### Who is responsible for archiving?

Archiving should be accomplished through a set of standards and practices set out jointly by librarians and publishers. To be of lasting value, electronic journals must become part of the permanent intellectual record. Archiving efforts must take into consideration all the parties involved—readers, authors, publishers, libraries, and the scholarly community.

To be successful, an agreement must be reached that allows publishers to make a profit, libraries to provide ongoing access, and users to take advantage of intellectual content.

In much the same way that standard practices are in place for print sources, parallel

### About the author

Deanna Marcum is president of the Council on Library and Information Resources, e-mail: [dmarcum@clir.org](mailto:dmarcum@clir.org)

## Editors' introduction

Changes in scholarly communication and technology have made librarians, publishers, and researchers re-examine their roles in this complex process. Nowhere is this more important than in the area of archiving.

In the pre-electronic era, archiving was clearly the role of the library. Long-term preservation of knowledge was one of our most important professional values and was reflected in our practices and organizations. Paper publications were effectively archived in a distributed informal system. Most academic libraries participated in this system by preserving materials locally.

With the advent of electronic publications, the issue has become more complex, and old assumptions no longer necessarily hold. Libraries are often reluctant to take on the task of archiving electronic materials. The level of redundancy needed for archiving in the paper world seems unnecessary in the electronic environment.

Even for the largest academic and research libraries, the complexity of the technical knowledge needed is an impediment to tak-

ing action. The prospect of migrating massive amounts of information on a regular basis is daunting. The difficulties of archiving a more interactive and less static medium are enormous. The temptation is to eliminate all redundancy or to rely on publishers or aggregators to perform these functions.

But do these organizations have the same commitment to archiving as libraries did in the past? Are there opportunities for partnerships in this new environment? What role can aggregators, national libraries, and library organizations play?

We've asked Deanna Marcum of the Council on Library and Information Resources to address the issue of archiving in this column. Marcum is known to most of you for her years of work on important library and information issues for the Council. She has been deeply involved in recent discussions about archiving electronic publications. We're delighted that she has agreed to share of her perceptions with us in this issue.—*Ivy Anderson, Gail McMillan, and Ann Schaffner*

mechanisms must be implemented for electronic journals.

### Who is working on projects and models?

RLG, OCLC, and several international groups are deeply engaged in digital archiving. RLG, OCLC, and the Cedars Project (UK) are actively exploring metadata standards for digital objects. An international workshop on the subject is scheduled for late fall.

CLIR has commissioned several reports on approaches to digital archiving.<sup>1</sup> Individual institutions concerned about both "born digital" materials and their reformatted digital collections are setting requirements for archival repositories.

Despite all of the good work that these organizations have done to define the problems and suggest possible solutions, there has been a dearth of digital archival repository experience to guide us.

CLIR, in collaboration with the Coalition for Networked Information (CNI), convened a meeting of librarians and publishers in the fall of 1999 to answer the question, "What would be required to make electronic journals accessible for 100 years?" (We are confident that print materials that have been reformatted to microfilm will last a century.)

Following the initial meeting, CNI hosted a workshop on the subject for a larger number of libraries and publishers. The participants, working from the premise that it is

### About the editors

*Ivy Anderson is coordinator for Digital Acquisitions at Harvard University, e-mail: ivy\_anderson@harvard.edu; Gail McMillan is head of the Digital Library and Archives (formerly the Scholarly Communications Project) at Virginia Tech University, e-mail: gailmac@vt.edu; Ann Schaffner has been an academic librarian for more than 20 years and is currently a full time MBA student at Simmons College, e-mail: ann.schaffner@simmons.edu*

necessary to create a standard for digital repositories, discussed what that standard should entail.

At this workshop, the participants explored technical, practical, and theoretical questions. The group was asked to consider whether a statement regarding acceptable formats should be released, whether best practices should be defined, and what makes one journal more suitable for archiving than another. Other issues brought to the forefront included how the repository would function financially and how to establish a minimum level of public access.

The discussions were difficult because the stakes are high for all parties. To make progress, we concluded that it was important to consider the publishers' and librarians' views separately. To provide a point of departure for the discussions, Dan Greenstein, director of the Digital Library Federation, and I extracted minimum criteria for archival repositories of electronic journals from the Open Archival Information System (OAIS) reference model and modified them to reflect the specific needs of the library, publishing, and academic communities.<sup>2</sup>

The eight criteria set forth highlight the importance of working together with publishers to create an environment where the free flow of information will benefit all parties. Agreed-upon minimum criteria are essential to ensure the preservation of intellectual content.

### **Proposed minimum specifications**

The archival repository must exist as a trusted third party that conforms to minimum requirements that both scholarly publishers and libraries agree upon. This will provide a benchmark against which service can be measured, validated, and, above all, trusted by the libraries and publishers that rely upon them.

The repository will define its mission with regard to the needs of scholarly publishers and research libraries. Repositories will also be explicit about which publications they are willing to archive. A repository needs to spell out the scope and nature of materials it aims to collect, the strategy and methods it adopts for developing its collections (attracting deposits), and the community of li-

braries and other users it seeks to serve. Registries should also be developed that document what scholarly publications are archived and where.

Once an agreement is reached, the repository will have sufficient control of deposited information to ensure its long-term preservation. The repository will negotiate and accept appropriate deposits from scholarly publishers. It will negotiate the requisite perpetual licenses and rights, even as they change over time. The repository will follow documented policies and procedures that ensure the information is preserved against all reasonable contingencies and that enable the information to be disseminated as authenticated copies of the original or as traceable to the original. Because preservation practices are likely to vary across repositories, it may be useful to request that participants in any such coordinated effort agree to document the practices they adopt and make them available for community review and evaluation.

The repository will make preserved information available to libraries. Publishers seem willing to deposit electronic content in archival repositories only under certain conditions. Although repositories will need to support access at some level, they should not replace the normal operating services through which digital scholarly publications are typically made accessible to end users. The repository will ensure that data can be disseminated to libraries in a renderable form. At a minimum, libraries should be able to create end-user services appropriate to the disseminated data and to do so independently of any assistance from those who initially produced the data.

Repositories will work as part of a network. Libraries may benefit from common finding aids, access mechanisms, and registry services that are supported by a network and allow uniform identification and access to information. Publishers may benefit from having access to a single repository or group of repositories that specialize in publications of a particular type and the cost effectiveness that results from such a network.

### **Technical requirements**

Without an established minimum criterion for preserving electronic journals, multiple

overlapping versions in fragile, rapidly changing environments are at risk of loss. The solution is to create a system that allows for technical growth.

Everyone agrees that redundancy is essential. By requiring three separate, independent facilities to act as repositories, the documents can be protected from human error and natural disaster. If at least two of these repositories exist in different countries, governmental policies that may interfere with content are no longer a concern. Regardless of the technological implementation chosen, the materials must exist in a safe haven where they can be accessed over time.

How access should change or not change in the networked environment is central to the issue of technical development. Once standards are set, a number of events will occur. Publishers can make business decisions about the electronic format. Libraries will be forced to evaluate the archivability of electronic journals when making decisions about purchasing and collection development. Authors will have better information for establishing expectations about the long-term availability of potential publishing venues for their works. Scholarly

societies designing electronic publishing programs will have guidelines for addressing questions of archivability of their publications.

Once disseminated, these copies can potentially be maintained in perpetuity without further permission from or even interaction with the publisher. One major caveat is the need to ensure a limited level of permanent public access to these copies by the broad scholarly community while the work remains under copyright. Protocols must be set to recreate these mechanisms for electronic journals. Once copyright has expired, the materials must remain accessible.

### **How can individual librarians make a difference?**

With so much of the scholarly record now being distributed in electronic form, we risk losing an entire generation of accumulated knowledge. If no one is certain whose responsibility it is to preserve electronic journals, no one will be responsible. Publishers recognize that librarians are concerned about this issue, and they are, happily, beginning to issue their policies about digital archiving.

*(continued on page 807)*

## Tired of searching for the best reference & research Web sites?



### Get CHOICE Magazine's 4th Annual Web Review Issue!

CHOICE's "WEB IV" contains more than 600 reviews of research-related Internet sites, arranged by subject. Plus information on searching, forthcoming books about the Internet, and more. Order now!

\$24.00 a copy, 1 to 4 copies  
\$21.50 each, 5 to 9 copies  
\$19.00 each, 10 to 19 copies  
\$17.50 each, 20+ copies  
WEB III: limited number of copies still available. Same prices as above.

Get extra copies for Reference and for your Internet trainers!

### CHOICE

100 Riverview Center Middletown, CT 06457-3445

For more information: [www.ala.org/acrl/choice](http://www.ala.org/acrl/choice)

Phone: (860) 347-6933 x33

Fax: (860) 704-0465

CHOICE is a publication of the ACRL division of ALA.



# The gateway to a world of information

In a changing world, vision, experience and reliability are the qualities you value in a partner. Whether it be the management of your subscriptions or the delivery of an article to your desktop, Swets Blackwell provides quality services to libraries and information centers all over the world.

**Swets Blackwell**  
Enhancing the power  
of your knowledge



For further information

please contact:

Swets Blackwell

22 Cortlandt Street • New York, NY 10007-3107

T: 800-221-3306 • F: 212-233-0746

P.O. Box 1459 • Blackwood, NJ 08012

T: 800-458-3706 • F: 856-232-5397

440 Creamery Way, Suite A • Exton, PA 19341-2551

T: 800-447-9387 • F: 800-590-9850

[www.swetsblackwell.com](http://www.swetsblackwell.com)

 Swets Blackwell