

Urgent Action Needed to Preserve Scholarly Electronic Journals

Digital preservation represents one of the grand challenges facing higher education. In field after field, research and teaching are generating data, reports, publications, teaching materials, and other forms of scholarly communication in digital formats. Research and teaching are also increasingly dependent on data mining tools and other computer-based techniques that require the long-term persistence of these various forms of digital information to advance knowledge. Yet as the creation and use of digital information accelerate, responsibility for preservation is diffuse, and the responsible parties—scholars, university and college administrators, research and academic libraries, and publishers—have been slow to identify and invest in the necessary infrastructure to ensure that the published scholarly record represented in electronic formats remains intact over the long-term. Inaction puts the digital portion of the scholarly record—and the ability to use it in conjunction with other information that is necessary to advance knowledge—increasingly at risk, and solutions may require unique arrangements within the academy for sharing preservation responsibility.

The shift from print to electronic publication of scholarly journals is occurring at a particularly rapid pace. Researchers, teachers, students and other readers demand electronic formats because it provides so many advantages over print, especially for search and retrieval. Recognizing the greater capability of the digital medium, editors are treating the electronic versions of journals as the definitive versions of record. Scholarly publishers are shifting their business models accordingly and are pricing print and electronic formats separately so that they can survive as electronic subscriptions supplant print. And research and academic libraries are increasingly canceling print subscriptions in favor of electronic licenses both to satisfy user demands and to avoid the substantial costs associated with ordering, receiving, cataloging, binding, storing, and circulating paper volumes.

In the face of this shift, what makes preservation action so urgent for electronic scholarly journals—and the risk of failure so high for the academy—is the nature of the licensing regime under which these journals are now distributed. When research and academic libraries license electronic journals, they do not take local possession of a copy as they did with print. Rather, they use content stored on remote systems controlled by publishers, and economies of scale in electronic publishing are driving control of more and more journals into fewer and fewer hands. Although some – but certainly not all – licenses now recognize that libraries have permanent rights to use electronic journal content, these rights remain largely theoretical. If a publisher fails to maintain its archive, goes out of business or, for other reasons, stops making available the journal on which scholarship in a particular field depends, there are no practical means in place for libraries to exercise their permanent usage rights and the scholarly record represented by that journal would likely be lost. For electronic journals, the academy has as yet no functional equivalent in long-term maintenance and control over the scholarly record that “owning a copy” provided for printed journals. Unless and until it creates digital archiving services, the academy cannot fully shift to electronic-only journal publishing, and cannot fully achieve the system-wide savings and benefits associated with such a shift.

Urgent Action Needed to Preserve Scholarly Electronic Journals

Universities and colleges and their libraries have recently been working together to help scholars manage their copyrights and to craft alternatives to high priced forms of scholarly publishing. It is now equally and perhaps even more important that research and academic libraries work with scholars—and their publishers—to sustain future research and teaching by establishing trusted archives in which the published scholarly record in electronic form can persist outside of the exclusive control of publishers, and in the control of entities that value long-term persistence. Four key actions are essential.

First, research and academic libraries and associated academic institutions must recognize that *preservation of electronic journals is a kind of insurance*, and is not in and of itself a form of access. Preservation is a way of managing risk, first, against the permanent loss of electronic journals and, second, against having journal access disrupted for a protracted period following a publisher failure. Storing electronic journal files in trusted archives outside the control of the publisher addresses the first risk. Mitigating the second risk requires investment in substitute access systems, which may cost more or less to construct depending on the quality and duration of disruption that the academic community would be willing to tolerate in the event of a failure.

Second, in order to address these risk factors and provide insurance against loss, *qualified preservation archives would provide a minimal set of well-defined services*. Such archives are beginning to emerge and must at least:

- Receive files that constitute a journal publication in a standard form either from a participating library or directly from the publisher;
- Store the files in non-proprietary formats that could be easily transferred and used should the participating library decide to change its archives of record;
- Use a standard means of verifying the integrity of ingoing and outgoing files, and provide continuing integrity checks for files stored internally;
- Limit the processing of received files, in order to keep costs down, but provide sufficient processing so that the archives could locate and adequately render files for participating libraries in the event of loss;
- Restrict access by the participating libraries to archived files that are under copyright in order to protect the publisher's business interests, except when (a) the publisher goes out of business or is otherwise unable to provide consistent access; or (b) the content is no longer protected by copyright.
- Offer an open, transparent means of auditing these archival practices.

Certifying agencies might recognize qualified preservation archives that provide these services with a publicly visible symbol of compliance. Additional preservation services for specific purposes beyond this minimal set, and provision of data-mining and other value-added services based on the preserved materials, would depend on the funding available to the archives, the permission of the publisher or other rights holder, and possibly both.

Third, *libraries must invest in a qualified archiving solution*. A library may itself operate a qualified archive as defined above. Otherwise, research and academic libraries may collaborate in the form of an insurance collective, or mutual assurance society. Such an

Urgent Action Needed to Preserve Scholarly Electronic Journals

entity may be governed in a variety of ways, but libraries would exercise their preservation obligation, in part, by paying fees to support the archive. In the event of a loss of access to an archived journal through the publisher, only paying participants would be able to have access to lost content through the archive. The collective would institute financial and other measures to ensure that potential participants who might choose initially to withhold support would pay their full fair share should they eventually need access to preserved materials.

Finally, *research and academic libraries and associated academic institutions must effectively demand archival deposit by publishers as a condition of licensing electronic journals.* Standard form clauses need to be crafted and implemented that require publishers to transmit all files upon publication either directly to a qualified archive or to the licensing library for deposit in a qualified archive. To express demand via such a contractual mechanism, research and academic libraries may need to seek support from university administrators and faculty governing bodies. They may also need to mobilize membership organizations, such as the American Library Association, the Association of Research Libraries, and the International Federation of Library Associations; local, regional, national, and international consortia; and services such as LibLicense. In addition, libraries must both urge publishers to describe their archiving provisions publicly and prominently in their publications, and educate authors and readers to consider these archiving provisions in evaluating the suitability of journals as a durable record of scholarship.

These actions may not be easy, but in a scholarly environment that is increasingly dependent on information in digital form, preservation of electronic journals is necessary and urgent. It will provide critical infrastructure and serve as a model for the preservation of other forms of digital information. In the end, those institutions that invest in digital preservation to ensure the persistence of and enduring access to the scholarly record will secure widely shared values of scholarship, and scholars and the public who are committed to those values will recognize the competitive advantage that preservation efforts afford in their ability to advance knowledge in service of the public good.

This document represents a consensus of the following academic librarians, university administrators and others who participated in a meeting to discuss electronic journal preservation at The Andrew W. Mellon Foundation offices in New York on September 13, 2005: Paul Courant (University of Michigan), Sam Demas (Carlton College), Nancy Eaton (Pennsylvania State University), David Ferriero (New York Public Library), Daniel Greenstein (California Digital Library), James Hilton (University of Michigan), Deborah Jakubs (Duke University), Micheline Jedrey (Wellesley College), Paula Kaufman (University of Illinois at Urbana-Champaign), Robert Kieft (Haverford College), Clifford Lynch (Coalition for Networked Information), Carol Mandel (New York University), James Neal (Columbia University), Elliott Shore (Bryn Mawr College), Sarah Thomas (Cornell University), Karin Wittenborg (University of Virginia), and Ann Wolpert (Massachusetts Institute of Technology).

*Edited by Donald J. Waters
Program Officer
The Andrew W. Mellon Foundation
October 15, 2005*

<http://www.diglib.org/pubs/waters051015.htm>