E-Journal Archiving Metes and Bounds:
A Survey of the Landscape

by Anne R. Kenney, Richard Entlich,
Peter B. Hirtle, Nancy Y. McGovern, and
Ellie L. Buckley

September 2006

Council on Library and Information Resources
Washington, D.C.

ISBN 1-932326-26-X
ISBN 978-1-932326-26-0
CLIR Publication No. 138

Published by:

Council on Library and Information Resources
1755 Massachusetts Avenue, NW, Suite 500
Washington, DC 20036
Web site at http://www.clir.org

Additional copies are available for $30 each. Orders must be placed through CLIR's Web site.
This publication is also available online at no charge at http://www.clir.org/pubs/abstract/pub138abst.html.

The paper in this publication meets the minimum requirements of the American National Standard for Information Sciences—Permanence of Paper for Printed Library Materials ANSI Z39.48-1984.

Library of Congress Cataloging-in-Publication Data

Kenney, Anne R., 1950-
E-journal archiving metes and bounds : a survey of the landscape / by Anne R. Kenney ... [et al.].
p. cm. -- (CLIR publication ; no. 138)
Includes bibliographical references.
ISBN-13: 978-1-932326-26-0 (alk. paper)
1. Electronic journals--Conservation and restoration. 2. Electronic journals--Publishing. 3. Scholarly periodicals--Conservation and restoration. 4. Scholarly periodicals--Publishing. 5. Scholarly electronic publishing. 6. Libraries--Special collections--Electronic journals. 7. Digital preservation. 8. Libraries and electronic publishing--United States. 9. Libraries and electronic publishing--Canada. I. Council on Library and Information Resources. II. Title.
Z701.3.E44K46 2006
070.5'797--dc22

2006028961

Contents

About the Authors

Anne R. Kenney is senior associate university librarian for public services and assessment at Cornell University Library. For more than 15 years, she has led research focusing on digital imaging and digital preservation. She is the coauthor of the award-winning *Moving Theory into Practice: Digital Imaging for Libraries and Archives* (Research Libraries Group 2000) and *Digital Imaging for Libraries and Archives* (1996), as well as of numerous articles and reports. Ms. Kenney is a fellow and past president of the Society of American Archivists. She has served on the RLG/OCLC Working Group on the Attributes of a Trusted Digital Repository and on the National Science Foundation/European Union Working Group on Digital Preservation. Ms. Kenney currently serves on the Portico Advisory Board.

Richard Entlich is digital projects librarian in the Research and Assessment Services Division of Cornell University Library. He began work with electronic publishing in 1990 as project manager for Cornell's part in the Chemistry Online Retrieval Experiment (CORE), one of the earliest attempts to test the feasibility of online access to scholarly journals. He has subsequently played a role in several other scholarly e-publishing initiatives, including Core Historical Literature of Agriculture, Making of America I, TEEAL (The Essential Electronic Agricultural Library), Prism (Preservation, Reliability, Interoperability, Security, Metadata), and KMODDL (Kinematic Models for Design Digital Library). Richard is a lecturer in Cornell's Digital Preservation Management Workshop and coauthored the associated online tutorial. He also writes regularly on digital preservation topics for *RLG DigiNews*.

Peter B. Hirtle is the technology strategist for the Cornell University Library's Public Services and Assessment Division. He also serves as the intellectual property officer for the Cornell University Library and is the bibliographer for United States and General History. Previously, Mr. Hirtle served as director of the Cornell Institute for Digital Collections, where he explored the use of emerging technologies to expand access to cultural and scientific sources through the development and management of distinctive digital collections. He also served as the associate editor of *D-Lib Magazine*. He is a fellow and past president of the Society of American Archivists, and chairs its Working Group on Intellectual Property. He is currently a member of the Copyright Office's Section 108 Study Group and is a contributing author to the LibraryLaw.com blog.

Nancy Y. McGovern is the digital preservation officer for the Inter-University Consortium for Political and Social Research (ICPSR) at the University of Michigan. Until August 2006, she was the director of research and assessment services and digital preservation officer at Cornell University Library. She is a codeveloper of and a principal instructor for Cornell's Digital Preservation Management workshop series and tutorial. While at Cornell, she coedited *RLG DigiNews*. She serves on the Technical Policy Committee of the LOCKSS Alliance and is a member of the RLG/NARA Digital Archive Certification Task Force. She has focused on digital preservation research and practice since 1986 when she joined the staff of the Center for Electronic Records at the U.S. National Archives. She is completing her Ph.D. on a technology responsiveness model for digital preservation at University College London.

Ellie Buckley is a digital research specialist for Research and Assessment Services at Cornell University Library. Her current focus is digital preservation and information science. She comes from a medical science background and has experience with developing evidence-based health care initiatives on the Web and telemedicine research projects.

Acknowledgments

We would like to thank both the Council on Library and Information Resources (CLIR) and the Association of Research Libraries for the opportunity to conduct this study. Nancy Davenport, former president of CLIR, made the initial contact with us. She, along with Karla Hahn and Carol Mandel, helped shape the project's parameters. We especially value the support and critical review provided by Ann Okerson through all phases of our work. Kathlin Smith provided excellent editorial and design help. Thanks also go to The Andrew W. Mellon Foundation, which provided financial support.

Many individuals gave generously of their time throughout this project. We are particularly grateful to the counsel of the library directors and senior managers from across North America who shared their thoughts and concerns about e-journal archiving with us. They included Joe Branin, Jane Bryan, Sam Demas, Ray English, Lee Hisle, Wendy Lougee, Catherine Murray-Rust, Kathy Perry, Louis Pitschmann, Sarah Pritchard, Mara Saule, Martha Sites, Sarah Thomas, Leslie Weir, Betsy Wilson, and Karin Wittenborg.

Representatives from the 12 e-journal archiving programs surveyed in this study not only completed a lengthy survey form but also participated in phone discussions and patiently answered our questions throughout the course of the investigation. They included:

CISTI Csi: Beverly Brown, Lucie Molgat, Daping Tan
LOCKSS Alliance and CLOCKSS: Vicky Reich, David S. H. Rosenthal, Michael A. Keller
Koninklijke Bibliotheek e-Depot: Erik Oltmans
Kooperativer Aufbau eines Langzeitarchivs digitaler Informationen/ Die Deutsche Bibliothek (kopal/DDB): Thomas Wollschläger, Tobias Steinke
Los Alamos National Laboratory Research Library: Richard Luce, Miriam Blake
National Library of Australia's PANDORA: Margaret Phillips, Gerard Clifton, Paul Koerbin, Chris Foster, Julie Whiting
OCLC Electronic Collections Online: Tim Martin
OhioLINK Electronic Journal Center: Tom Sanville, Thomas Dowling, Anita Cook
Ontario Scholars Portal: Leslie Weir, Alan Darnell, Peter Clinton
Portico: Eileen Fenton, Evan Owens
PubMed Central: David J. Lipman, James Ostell, Betsy Humphreys, Ed Sequeira

We would also like to acknowledge the help and advice of our colleagues at Cornell University Library, especially William Kara, William Kehoe, Linda Miller, Kornelia Tancheva, Sarah Thomas, and Scott Wicks. Finally, thanks goes to those who offered suggestions throughout the project or reviewed the final draft, including Robin Dale, Eileen Fenton, Bernie Reilly, and Roger Schonfeld.

Anne R. Kenney
Richard Entlich
Peter B. Hirtle
Nancy Y. McGovern
Ellie L. Buckley

EXECUTIVE SUMMARY

This report summarizes a review of 12 e-journal archiving programs from the perspective of concerns expressed by directors of academic libraries in North America. It uses a methodology comparable to the art of surveying land by "metes and bounds" in the era before precise measures and calibrated instruments were available. It argues that current license arrangements are inadequate to protect a library's long-term interest in electronic journals, that individual libraries cannot address the preservation needs of e-journals on their own, that much scholarly e-literature is not covered by archiving arrangements, and that while e-journal archiving programs are becoming available, no comprehensive solution has emerged and large parts of e-literature go unprotected. Academic libraries of all sizes have both a responsibility and an opportunity to support the development of e-journal archiving programs to better meet the needs of students, faculty, and other researchers. Libraries that elect not to support such programs in the near future risk incurring costly and delayed access to essential resources if and when publishers cease to make content available.

This report makes the following recommendations to academic libraries, publishers, and e-journal archiving programs.

Recommendations: Academic Libraries and Organizations

1. Libraries and consortia should press publishers hard to enter into e-journal archiving relationships with bona fide programs and to convey all necessary rights and responsibilities for digital archiving as part of their license negotiations. Research libraries should collectively agree not to sign new licenses or renew old ones for access to electronic journals unless these conditions are met.

2. Libraries should share information with each other about what they are doing in e-journal archiving, including their internal assessment processes for decision making.

3. Institutions should become members of or participate in at least one e-journal archiving initiative. A broad range of academic and research libraries should be encouraged to affiliate with appropriate e-journal archiving programs.

4. Academic libraries of all sizes should act collectively to press for digital archiving programs that meet their needs. As a condition of support, they should request details on the programs' ability to meet base-level requirements for responsible stewardship of journal content and, ultimately, insist on some form of accreditation to ensure the development of full-fledged preservation programs.

5. Libraries should participate in developing a registry of archived scholarly publications that indicates which programs have preserved them. This registry can be used to identify gaps in publisher or content coverage. Models for such registries include the Registry of Open Access Repositories (ROAR) and ROARMAP.

6. Libraries should lobby e-journal archiving programs to participate in a network that shares information, codifies best practices, and promotes sufficient redundancy, and also shares responsibility for preserving peer-reviewed e-journals that are not currently included.

Recommendations: Publishers

1. Publishers should be overt about their digital archiving efforts and enter into archiving relationships with one or more e-journal archiving programs of the sort described in this report.

2. Publishers should provide enough information to e-journal archiving programs to ensure that the scope, content, date span, and title coverage are adequately recorded.

3. Publishers should extend liberal archiving rights in their licensing agreements with content aggregators and consortia. Digital archiving of e-journals should be a distributed responsibility.

Recommendations: E-Journal Archiving Programs

1. Archiving programs should present compelling public evidence that they offer at least the minimal level of services for well-managed collections. They should be open to audit, and when certification of trusted digital repositories is available, they should be certified.

2. Archiving programs should be overt about the publishers, titles, date spans, and content included in their programs. They should make this information easily accessible on their Web sites.

3. Archiving programs should ensure that, once content is ingested, it becomes the repository's property and cannot be removed or modified by a publisher or its successor. If there is an alleged breach of contract, there should be a process for mediating disputes to protect the longevity and integrity of the e-journal content.

4. A study should be conducted to examine the rights and responsibilities necessary to ensure adequate protection for digital archiving actions so that these rights are accurately reflected in contracts. Archiving programs should periodically review contracts, because changes in publishers, acquisitions, mergers, content creation and dissemination, and technology can affect archiving rights and responsibilities.

5. Archiving programs should consider that some content they store might eventually enter the public domain and should negotiate all agreements with publishers to take this possibility into account.

6. Archiving programs should form a network of support and mutual dependence to exchange information on content coverage, technical implementations, and best practices; to obtain the necessary contractual rights to preserve and eventually provide access to content; to create a safety net for one another for succession planning and secondary archival functions; and to share responsibility for identifying and preserving peer-reviewed e-journals that are not currently protected.

> A survey of a tract by metes and bounds is the oldest known manner of describing land and is the outgrowth of the art of surveying as practiced in the olden times.
>
> Frank Clark,
> *A Treatise on the Law of Surveying and Boundaries*

INTRODUCTION

"Digital preservation represents one of the grand challenges facing higher education," wrote a working group of influential academic administrators and librarians who participated in a special meeting convened at The Andrew W. Mellon Foundation in September 2005 (Waters 2005). Their statement, titled "Urgent Action Needed to Preserve Scholarly Electronic Journals," signaled the intensity of the broad concern that had been voiced for more than a decade, and summoned the educational community to action. The statement emphasized that preserving scholarly electronic publications has become a critical need as e-publication comes to dominate scholarly expression and as user communities increasingly depend on electronic publications much as they used to rely on paper.

Ten years earlier, an active thread of discussion on the issue of archiving e-journals appeared on the ARL-EJOURNAL list. Sharon McKay of Blackwell Publishing initiated the discussion by asking who should be responsible:

> Who should do it, then? Non-profit organizations? Do we need a new organization formed specifically to perform the function of e-journal archiving? Should there be one for each continent? Each nation? Each language? What are the implications for access, and what kind of economic model would work? Is it possible to have publishers to establish something like an escrow deposit for archival data that would be available to subscribers, regardless? What technical issues should be considered, and how do we prepare for unknown technological changes in the future which will affect storage techniques and access? (McKay 1996)

These questions remain relevant. Even as academic libraries recognize the growing concern over e-journal archiving, many are unclear on the dimensions of the problem, the alternatives for action, and what role they might play. In the past few years, several promising alternatives for addressing e-journal preservation have emerged. To help libraries better understand the emerging strategies and options and determine their best course of action, the Association of Research Libraries (ARL) and the Council on Library and Information Resources (CLIR) agreed that a survey of the e-journal archiving

landscape was needed. CLIR commissioned the Cornell University Library Research and Assessment Services Department to undertake this review.

Why Is E-Journal Archiving Such a Concern?

The "Urgent Action" statement listed several trends that raise concerns over the long-term viability of serial literature. These trends include publishers' shift to electronic distribution, users' preference for online resources, and libraries' ability to respond to these two trends, given constrained budgets.

The Shift to Electronic Publishing

Many have noted the difficultly of determining how many peer-reviewed journals are currently online, but all agree that growth has been dramatic over the past decade. In 1996, Stephen P. Harter and Hak Joon Kim identified 131 refereed or peer-reviewed e-journals. Ten years later, estimates range in the tens of thousands to the hundreds of thousands. In 2003, Carol Tenopir found inconsistencies in using *Ulrich's International Periodicals Directory* to determine the number of scholarly journals available online. A June 2006 search of that database for all active, online, and refereed journal titles confirmed these inconsistencies but did return 14,338 hits (1,429 of which are open-access titles), representing 62% of all 23,187 active and refereed titles listed. (The remaining 38% were print only). The *Directory of Open Access Journals*, by comparison, listed 2,044 peer-reviewed open-access journals in February 2006, up 600 from the year before (Tenopir 2004; Van Orsdel and Born 2006).

As online access grows, publishers are beginning to consider eliminating print runs, although the number of electronic-only titles is still a significant minority of all publications (Ware 2005, 194). The National Library of Medicine (NLM) is undertaking an effort to identify journals that have gone to an electronic-only format.[1] In 2003, the British Library commissioned Electronic Publishing Services Ltd. to project publishing trends to 2020. Among other things, the report looked at the migration from print to electronic formats for serial literature (including scholarly publications) in the United Kingdom and elsewhere. The authors concluded that by 2016, half of all serial publications will have migrated to electronic-only format. They predicted that science, technology, and medical (STM) titles would be the first to switch. Large publishers will start with less-profitable titles. Smaller publishers, especially scholarly societies, will switch on the basis of rising print and distribution costs (Powell 2004). In a thoughtful paper, Karen Hunter from Elsevier outlined four issues that will have to be resolved before publishers move to electronic-only formats; among these issues is "bullet-proof digital archiving of electronic journals" (Hunter 2006).

[1] Dianne McCutcheon, chief of technical services, National Library of Medicine, and Beth Weston, head of serial records section, National Library of Medicine. Telephone conversation with Anne Kenney, June 14, 2006.

User Preferences for Online Journals

Across all disciplines, faculty members and other users have come to value electronic access to scholarly literature, and use of such resources increases dramatically with their availability in electronic form (Guthrie and Schonfeld 2004; Tenopir 2003). A 2003 study by the Digital Library Federation (DLF) and CLIR reported that 75% of faculty members and graduate students surveyed use e-journals (Marcum and George 2003). A second study that year conducted for Ithaka confirmed these findings: of the 7,400 faculty members surveyed, 78% characterized electronic scholarly journals as "invaluable research tools."[2] Carol Tenopir found that for scientists, two-thirds of their reading now comes from e-resources, and in some fields, such as astronomy, the number is approaching 80% (Wolverton and Tenopir 2005). Although faculty members are concerned about how digital resources will be preserved, many accept the cancellation of print editions, especially in the sciences and social sciences, if a choice has to be made between retaining print or greater e-access (Salisbury, Vaughn, and Bajwa 2004).[3]

Library Response

Libraries have responded to changes in publishing and user behavior by increasing the percentage of their serials expenditures on licensing e-serials. Between 1995 and 2004, the median serial expenditure of ARL members rose from a little more than $3 million annually to just under $5.5 million—an increase of more than 80%. Meanwhile, the median amount devoted to e-journals increased from $156,754 to $2,348,463—nearly a 1,400% increase. E-serials represented 5% of total serials expenditures in 1995; by 2004, that portion had jumped to 42% (Kyrillidou and Young 2005). Cary Bruce of EBSCO estimates that for STM titles, online journal subscriptions will exceed print subscriptions by 2008 (Bruce 2005).

As libraries continue the shift to licensing e-journals, it is becoming more common to cancel the print equivalents in response to serials prices that have increased faster than inflation for the past two decades. In a 2004 Publishers Communication Group survey of 155 librarians from academic libraries worldwide, 84% of respondents said they cancel print when an electronic version is available. Forty percent of current subscription revenues for Elsevier's Science Direct come from electronic-only subscriptions (Hunter 2006). In a recent ARL member survey, research libraries reported that they had canceled print equivalents for bundled e-content in 153 out of 266

---

[2] "Electronic Research Resources" survey of 7,403 faculty members conducted in 2003 by Odyssey, a market research firm, on behalf of Ithaka (unpublished). A summary and additional information can be found at http://www.educause.com/ir/library/pdf/ERM0248.pdf (summary) and http://www.jstor.org/about/faculty.survey.ppt (PowerPoint with many charts and figures from the survey).

[3] The Institute for Museum and Library Services has funded a study by Carol Tenopir, Donald King, and others on how to maximize library investment in digital collections (including e-journals), through better data gathering and analysis of user preferences. See http://web.utk.edu/~tenopir/imls/.

contracts (58%) for 2006 (Hahn 2006). Print repositories are being developed at the regional and national levels to ensure that at least one paper copy remains accessible,[4] but increasingly institutions recognize that print is not an acceptable archival format for electronic content.[5]

Concern over reliance on leased, rather than owned, electronic content has led libraries to include "perpetual access" rights in their licenses. According to the 2005 ARL member survey, 98% of contracts included a provision for some form of backfile access if a library cancels its electronic subscription. In identifying technical requirements for e-journal licenses, the California Digital Library, among others, requires that vendors agree that the institution will retain "*use* of material to which it previously subscribed, and allow users to continue to access that data in the event the subscription is cancelled" (California Digital Library 2006).

There are two primary options for ensuring continued access to licensed content. The first is to rely on the publisher or distributor to provide perpetual access. OCLC's Electronic Collections Online (OCLC ECO), for example, stresses that the content it delivers from publishers will always be available to subscriber libraries, as long as they continue to pay access fees. Its policy states: "Your library retains the right to access all journals to which you have subscribed even after you discontinue subscriptions to any of them."

The question, of course, is whether one can trust the publisher or distributor to keep older content accessible and unchanged, especially after the publisher stops distributing a title or the library stops subscribing to it. Hence, the second option found in many licenses: the requirement that publishers will give libraries copies of the files that constitute an e-journal. The NorthEast Research Libraries Consortium's (NERL) Generic License provides a good example (NERL 2006). The agreement specifies that if the distributor discontinues any of the licensed materials or if either party terminates the agreement, the distributor must provide the library with one copy of subscribed materials in a mutually acceptable form. The license further stipulates that the library can make any copies needed into perpetuity "for purposes of archival preservation, refreshing, or migration." Of course, few libraries are equipped either to preserve or to provide access to a large number of e-journal files. The NERL Generic License, therefore, also authorizes libraries to contract with third-party trusted archives or to participate in collaborative archiving endeavors to fulfill the requirements of this provision.

The NERL Generic License distinguishes between perpetual access and e-journal archiving. The focus of the former is to maintain

---

[4] See, for example, Committee on Institutional Cooperation Libraries Pilot Cooperative Program to Archive Print Journals press release, http://www.cic.uiuc.edu/programs/JournalArchiving/archive/PressRelease/PrintJournalArchiving4-25-05.pdf, and the Center for Research Libraries Web site, at http://www.crl.edu/content.asp?l1=13&l2=19&l3=35&l4=64.

[5] See, for example, California Digital Library 2006.

access rights beyond subscription periods; the latter focuses on mitigating risk of permanent loss to ensure availability for future users. Nonetheless, e-journal archiving does not necessarily equate to ongoing access. One can have e-journal archiving without current access, but it is difficult to imagine how one could ensure perpetual access without having an e-journal archiving program. Jim Stemper and Susan Barribeau (2006) provide a thoughtful discussion on perpetual access in light of results from a survey of the University of Minnesota's contracts with publishers and aggregators. They discovered that 64% of publishers with which Minnesota has license arrangements grant perpetual access rights. Perhaps ironically, more commercial publishers (72%) than society publishers and aggregators (56%) granted these rights. In many cases "perpetual" covered a limited number of years after cancellation of subscriptions; almost half the publishers specify that there will be or may be a charge associated with such access. Stemper and Barribeau concluded that only 20% of large research libraries would consider the lack of perpetual access assurances a reason for not signing a license with a publisher.

In addition to demanding perpetual access rights, libraries and others are establishing institutional repositories, using systems such as DSpace, Fedora, and bepress, and are joining with faculty members and professional organizations to urge publishers to provide self-archiving rights to authors. SHERPA/RoMEO, funded by the Joint Information Systems Committee on Higher and Further Education Councils (JISC) and the University of Nottingham (U.K.), provides a list of 155 publishers' copyright conditions that relate to authors who are archiving their work online. The service categorizes publishers and their conditions as follows: *green publishers* allow self-archiving of both preprints and postprints (45% of publishers); *blue publishers* allow self-archiving of postprints but not preprints (24%); *yellow publishers* allow self-archiving of preprints but not postprints (10%); and *white publishers* do not allow self-archiving (22%) (SHERPA/RoMEO 2006).

Beyond perpetual access and self-archiving, institutions are beginning to ask that publishers establish preservation programs. In the 2005 ARL survey on large publisher bundles, most libraries reported that they had investigated the publishers' archiving plans (71% of contracts); of those who did, only 60% found the publishers' plans acceptable. This calls into question more than half of the archiving arrangements by publishers reported in that survey (Hahn 2006). As Mary Case commented, "No clause in a license guaranteeing perpetual access or any other user rights will help if the resource suddenly disappears for no matter what the reason" (Case 2004). Stemper and Barribeau (2006) hypothesize that

> … a research library's mandate to provide current access to journals for today's scholars can be at odds with the mandate to keep those journals available to be accessed by scholars in the future. Librarians still value their stewardship role in the digital realm, but they perhaps fear that pressing the issue contractually is commercially and financially unrealistic at this time.

A Gathering Momentum

By 2000, libraries' concerns over their e-journal vulnerabilities led many to press for trusted e-journal archiving programs that were independent of the publishers and did not rely on individual libraries' efforts.[6] The past several years have seen the following developments:

- publishers collaborating with cultural institutions to provide dark archives for their backfiles;
- in several countries, passage of legal deposit laws that mandate deposit of online publications, including e-journals;
- the National Institutes of Health's (NIH) effort to create a freely accessible archive of government-funded research publications and the corresponding protests from commercial and not-for-profit publishers and societies;
- the coupling of the open-access movement with preservation;
- national libraries establishing or financially supporting e-journal archiving programs and emerging standards;
- the launch of third-party and consortial efforts that focus on e-journals;
- development of the draft "Audit Checklist for the Certification of Trusted Digital Repositories" by RLG and the National Archives and Records Administration (NARA) (RLG 2005); and
- road testing of the RLG-NARA certification requirements by the Center for Research Libraries in several digital repositories, with a heavy focus on e-journal preservation and an eagerly awaited report on the results due later this year.

These efforts are beginning to bear fruit, and academic libraries are now being offered viable options for e-journal archiving. This report looks at 12 of the more promising options and provides a means for assessing their viability and suitability for academic libraries.

METES AND BOUNDS

A survey "by metes and bounds" is a highly descriptive delineation of a plot of land that relies on natural landmarks, such as trees, bodies of water, and large stones, and often-crude measurements of distance and direction. This was accepted practice before more precise instruments and methods were developed—indeed, the original 13 U.S. states were laid out by metes and bounds. More accurate means of measuring were established to overcome the method's serious shortcomings: streambeds move over time, witness trees are struck by lightning, compass needles do not point true north, and measur-

[6] See, for example, "Minimum Criteria for an Archival Repository of Digital Scholarly Journals," Digital Library Federation, May 15, 2000, http://www.diglib.org/preserve/criteria.htm. In 2001, The Mellon Foundation funded seven institutions to research archiving options. The results of these studies pointed to the need for collective action.

ing chains and surveyor strides can be of slightly differing lengths. However, the metes and bounds system is still used when it is impossible or impractical to make more precise measurements.

In undertaking our survey of the e-journal archiving landscape, we found that precise measurements and controlled data collection were not always possible. The e-publishing terrain is changing at time-lapse photography speed. Definitions and terms are widely interpreted, and standards are not yet established. These factors, along with our need to rely heavily on self-reporting by the programs, mean that direct comparisons between them may not always be valid. Despite this, we describe in this report the current lay of the land for scholarly e-journal archiving.

This study focuses on the "who, what, when, where, why, and how" of significant archiving programs operated by not-for-profit organizations in the domain of peer-reviewed journal literature published in digital form. Not included are preservation efforts covering digitized versions of print journals, such as JSTOR; library-led digital conversion projects; self-archiving efforts by publishers; and initiatives still being planned.

In preparing this report, our team focused on the following:

- soliciting library directors' concerns and perceptions about e-journals;
- compiling responses from e-journal archiving initiatives taken from written surveys and semistructured interviews; and
- analyzing the issues and current state of practice in e-journal archiving, and forming recommendations for the future.

Library Directors' Concerns

We began the study by developing a list of what library decision makers are likely to consider as they assess preservation strategies for e-archiving. The list was informed by our own research, discussions with colleagues, and comments made to staff members of the Center for Research Libraries (CRL) by member library directors.[7]

During March and April 2006, 15 North American library directors, representing a range of public and private institutions of various sizes as well as consortia, participated in telephone interviews designed to solicit their views on six key areas:

1. Library motivation (*Why should we be concerned about or invest in this?*)
2. Content coverage (*Are current approaches covering the subject areas, titles, and journal components in which we are most interested?*)
3. Access (*What will we gain access to? When and under what conditions?*)
4. Program viability (*What evidence is there that these efforts are sufficiently well-governed and financed to last?*)

[7] "Digital Repositories: Some Concerns and Interests Voiced in the CRL Directors' Conversation," January 21–22, 2006 [at ALA midwinter] as distributed on the CRL Member Directors' listserv, February 3, 2006, by CRL President Bernard F. Reilly. See also Digital Archives and Repositories Update, FOCUS 25(2). Available at http://www.crl.edu/PDF/pdfFocus/Winter2005-06.pdf.

5. Library responsibilities and resource requirements (*What will this cost our library in staff time, expertise, financial commitment? Would our support save the library money?*)
6. Technical approach (*How do we judge whether the approach is rigorous enough to meet its archiving objectives?*)

The interviews helped refine the issues to be covered in our survey. They also revealed some interesting opinions on the topic. Three common themes emerged in the interviews: the sense of urgency, resource commitment and competing priorities, and the need for collective response.

Sense of Urgency

These directors were all aware of digital preservation as a major concern, but they differed on whether it was a priority for support and action. Some felt the sense of urgency as a vague concern rather than as an immediate crisis, and several were willing to defer action until a crisis point is reached. Digital preservation is a "just-in-case scenario," commented one director, "and this is very much a just-in-time operation." Another noted, "Archiving is the last thing that gets taken care of because it's the farthest thing out." One director did assert that she would not want to gamble on what it would take to obtain access later if her institution did not invest now, likening that decision to not buying a book and waiting three years to see whether there was a demand for it. Several directors who have committed to supporting e-journal archiving do so because they have experienced loss. One acknowledged that her institution's willingness to support digital archiving stemmed from the losses caused by a devastating flood: "Natural disasters make people focus." Another director indicated that 9/11 raised his level of concern: "Prior to that, I had scoffed at the idea that the Internet would break down and I wouldn't have access to my journals restored in 24 hours."

One-third of the directors expressed more concern about the preservation of digital content other than e-journals. Virtually all expressed a lack of trust in publishers providing the solution, but many argued that publishers had to take on more responsibility. They pointed to efforts to include archiving clauses in licensing agreements. One questioned why she should have to pay additionally to support e-archiving initiatives: "We've pressured publishers to include archiving, and now we're giving up on this?" Several pointed to the role that some publishers were already undertaking in collaborating with libraries to share preservation responsibility. One suggested that as the number of publishers decreases because of mergers and acquisitions, those remaining are making money and are not as apt to go under in the short term. Can an effective case be made, some asked, without there being an actual disaster? Another wondered about the future of licensed content in general for reasons other than digital preservation: "If you can't get [e-journals] on the open public Internet, do they have much value anymore?" Several identified university records, Web sites, and digital content produced

within institutions as more immediate concerns and were committing resources to their protection. "How do we sustain our role as the university archives in the digital age?" one asked.

Interviewees from some of the larger ARL libraries expressed the most concern about preserving e-journals. Although they argued that publishers had to bear some responsibility for e-journal archiving, they do not necessarily trust them to do this over time. One put it bluntly: "We definitely can't wait this one out. I have a bias toward action and want to be involved. Until you explore it, you really don't know what's going on." This concern was compounded by a sense of frustration over the options available. Understanding the issues is not the real problem, one noted: a lack of clarity about the solutions is. To date, few have committed real resources to address e-journal archiving, in part because they are unclear about what needs to be done. All directors interviewed acknowledged that a perfect solution is still many years away, and those who were willing to commit resources now stated their goal was to support a "good enough" solution that would be viable until the desired solution came along. One director characterized the decision of whether to commit resources as particularly acute for medium-size libraries. "The large ones will do it and worry about whether they should be doing this for others," she argued, "and the smaller ones will say they don't have the money. The ones in the middle with some resources and some sense of obligation are the fence sitters." A director of an Oberlin Group library argued that leading liberal arts colleges would want to be involved as well.

Of the fifteen directors interviewed for this study, nine have committed or are prepared to commit resources to e-journal archiving, two are not, and four characterize themselves as fence sitters. The two who have decided to do nothing view their positions as managing risks and making hard decisions. Of the four who are undecided, one called himself a fence sitter only because he has not made up his mind about which initiative to support. Another characterized her institution as an "early follower, sitting on a fence by design, not because we wound up on one," and a third concluded at the end of our discussion "I'm starting to think as we talk that sitting on the fence isn't helping." When asked what would provide additional incentives for getting off the fence, several pointed to peer pressure and reaching the "tipping point" of enough institutions participating. One said that he wanted to know where the major ARL libraries were going to put their money and why. One cited the importance of pressure from funding agencies such as The Andrew W. Mellon Foundation or their professional organizations. Another said that she would decide to do something in response to pressure from the administration or faculty members. Another indicated that having transparency in what is being done would be important, as was whether her institution would have a say in future directions. Several wanted to know about the circumstances and effort involved in committing to e-journal archiving, and how long they would have to wait before their institutions could restore access to their users fol-

lowing loss of normal access channels. Others wanted to know the costs involved, including staff effort, and what they would get from their commitment. They wanted to support those whom they could trust the most, whom they would have to pay the least, and who covered the material they care most about. Incentives to be an early subscriber were a big carrot. Knowing the penalties for waiting to join later was a potential big stick.

Resource Commitment and Competing Priorities

A recurring concern among the library directors interviewed was finding resources to commit to e-journal archiving programs. They pointed to competing priorities and the difficulty of identifying ongoing funds to support the effort.[8] Many felt that while they might be able to provide resources for the next several years, support would eventually have to be found at the university or college level. Some were concerned that senior administrators would agree that the problem was real and that the library should address it, but that it would be difficult to get additional support. Digital archiving, one noted, is a new kind of expense, which is more difficult to argue for than increases to an existing expense. The directors requested sound bites to use with their provosts, presidents, and chancellors. (One mused that real horror stories would be better.) Several focused on the need to have faculty identify digital preservation as a major concern that directly affects them.

Almost all the directors rejected the argument that the savings in moving to electronic-only could cover the archiving costs. For most of them, that shift has already occurred as a result of lean budget years and dramatic increases in serials subscriptions, and the savings have already been reallocated to other purposes. "We couldn't wait for the safety net to cancel," said one. A director from the East Coast noted that many competing demands from new initiatives require ongoing financial support.

The greatest competition, however, lies in providing ongoing access to electronic resources. When a choice has to be made between the two, "broad and deep access at this point trumps more restricted access but a reliable archive," concluded one director. "I'd rather buy more titles now than pay for something I might never use," said another. Several directors from state institutions worried about justifying the use of state funds to purchase something "intangible" and questioned whether e-journal archiving could substitute for risk management measures locally. Others expressed more concern about guaranteeing perpetual access to e-journals than archiving them. One pointed out that his main worry was ensuring future access to content "below the trigger threshold" that would not be addressed by e-journal archiving. Another director questioned whether it was counter to his responsibilities to try to "preserve all e-journals when I can't even get access to many of them because I can't afford it." An-

[8] Small and medium-size libraries expressed this concern in a 2003 study on the state of preservation programs by Kenney and Stam (2002).

other commented, "It all comes down to money: present money versus future money." One even suggested that it would almost seem like throwing money away: "You don't have anything to show for it, and I'm not even sure that the solution would survive when you do need it."

Need for Collective Response

All the directors interviewed rejected the notion of creating their own institutional solution. A major finding of the seven e-journal archiving projects supported by The Andrew W. Mellon Foundation in 2001 was the difficulty of developing an institution-specific solution. At the end of that project, the Mellon Foundation decided to provide startup funds for both Portico and the LOCKSS Alliance (Bowen 2005). Several directors called for the creation of a national cooperative venture, saying, "We want to throw our lot in with other libraries." Some wanted to tie e-journal archiving to their consortial buying and licensing efforts. Others felt that publishers had to be at the table as well, noting that libraries are too prone to seek internal solutions. One mused that libraries can now do with e-journal archiving what they have wanted to do for 40 years with shared print repositories, and that the two could not be handled in isolation.

Although agreeing that a collective response is needed, several directors worried about having too many options. "I have heard others say we need lots of strategies to keep stuff safe," said one, "but I'm not sure that's true." Another worried about ending up with two or three competing models that would be difficult to sustain. He suggested not investing in any of the options until they get together to build "something we can all get behind." Keeping track of what is archived by whom raised the specter of major management overhead. One director mused that this might represent a new business for Serials Solutions. All agreed that while it was still early, it would be "nice if the market sorted itself out fast."

Another concern of the directors was the long-term viability of any e-journal archiving initiative. Several wanted reassurance that their investment would be secure for at least 10 to 20 years. Others argued that it was unrealistic to expect assurances up front, noting that all the options are still experimental and that there is no right solution. Several suggested that it was important for institutions to support different options because it is not clear "which model will win out." The right answer, one stated, "is that more people must participate in order to uncover the problems and workable solutions." One director argued that instead of focusing on the existing options, libraries should collectively define what the solution should look like.

Cornell Survey of 12 E-Journal Archiving Initiatives

The directors' concerns helped shape a questionnaire that our team used to survey e-journal archiving programs. The survey covered six areas: organizational issues, stakeholders and designated communi-

ties, content, access and triggers, technology, and resources. The form went through several iterations in response to reviewer feedback and was pilot-tested with one digital archiving entity before being finalized. A version of the final survey form is located in Appendix 1. Project staff sent surveys to 12 e-journal archiving programs in March and held hour-long interviews with key principals (and subsequent follow-up) between April and June 2006.

Several criteria guided the selection of electronic journal archiving initiatives to include in this study. First, each initiative had to have an explicit commitment to digital archiving for scholarly peer-reviewed electronic journals. Second, it had to maintain formal relationships with publishers that include the right to ingest and manage a significant number of journal titles over time. Third, work addressing long-term accessibility had to be under way. Fourth, the efforts had to be by not-for-profit organizations independent of the publishers. Finally, the work had to be of current or potential benefit to academic libraries that have a preservation mandate.

The following 12 e-journal archiving programs met these criteria. Appendix 2 includes longer descriptions of these programs.

Canada Institute for Scientific and Technical Information (CISTI Csi)
The National Research Council of Canada (NRC), Canada's governmental organization for research and development, was mandated by the National Research Council Act (August 1989) to establish, operate, and maintain a national science library. In that capacity, the NRC hosts CISTI to provide universal, seamless, and permanent access to information for Canadian research and innovation in all areas of science, engineering, and medicine for Canadians, the NRC, and researchers worldwide. To achieve its mission as Canada's national science library, CISTI has established a three-year program called Canada's scientific infostructure (Csi) and is partnering with Library and Archives Canada (LAC) to ensure business continuity. This program is creating a national information infrastructure in collaboration with partners to provide long-term access to digital content loaded at CISTI and to support research and educational activities. In 2003, CISTI began loading e-journal content from three publishers and now has loaded close to 5 million articles. Additional content from other publishers in the sciences is planned.

LOCKSS Alliance and CLOCKSS
The Lots of Copies Keep Stuff Safe (LOCKSS) program, based at Stanford University, launched the beta version of its open-source software between 2000 and 2002. LOCKSS intended the software to allow libraries to collect, store, preserve, and provide access to their own, local copies of authorized content. Some 100 participating institutions in more than 20 countries use the LOCKSS software to capture content. About 25 publishers of commercial and open-access content (including large aggregators) participate in the LOCKSS program. In 2005, the LOCKSS Alliance was launched as a membership organization built on the LOCKSS software. The purpose of the

alliance is to develop a governance structure and to address sustainability issues. The Controlled LOCKSS (CLOCKSS) initiative, added to the LOCKSS program in 2006, brings together six libraries and twelve publishers to establish a dark archive for e-journals.

Koninklijke Bibliotheek e-Depot (KB e-Depot)
As the national deposit library for the Netherlands, the Koninklijke Bibliotheek (KB) is responsible for preserving and providing long-term access to Dutch electronic publications. To meet that responsibility, the KB started planning for e-journal archiving in 1993 and began to implement an archiving system between 1998 and 2000. It was initially intended as a system in which Dutch publishers would voluntarily deposit their publications for archiving. The KB's current goal is to include journals from the 20 to 25 largest publishing companies, which produce almost 90% of the world's electronic STM literature. The KB e-Depot currently offers digital archiving services for eight major publishers.

Kooperativer Aufbau eines Langzeitarchivs Digitaler Informationen (kopal/DDB)
Funded by the German Federal Ministry of Education and Research, kopal/DDB is a cooperative project begun in July 2004. A main impetus for kopal has been the need for the national library of Germany, Die Deutsche Bibliothek (DDB), to manage the legal deposit of electronic publications. DDB had been experimenting with electronic journals since 2000; in 2006, Germany enacted legal deposit legislation for electronic publications, making the implementation of a system a priority. Through voluntary agreements with publishers, DDB has acquired a variety of electronic content, including e-journal titles from Springer, Wiley-VCH, and Thieme. Under legal deposit, DDB will start acquiring and adding to kopal all electronic journals published in Germany. In the future, kopal/DDB intends to offer other institutions data archiving services.

Los Alamos National Laboratory Research Library (LANL-RL)
Los Alamos National Laboratory is one of three U.S. national laboratories operated under the National Nuclear Security Administration of the U.S. Department of Energy. LANL-RL has been locally loading licensed backfiles from several commercial and society publishers since 1995. Focusing on titles in the physical sciences, the library maintains content from 10 publishers primarily for the use of the LANL-RL staff, but it also serves a group of external clients who pay for access (LANL charges on a cost-recovery basis). LANL-RL has done substantial research and development work on repository and digital object architecture for long-term maintenance of electronic journal contents. A major focus of the research and development work has been the creation of the aDORe repository.

National Library of Australia PANDORA (NLA PANDORA)
The NLA selects e-journals from its Australian Journals Online database for preservation in PANDORA (Preserving and Accessing Networked Documentary Resources of Australia), which was established in 1996. E-journals is one of six categories of online publications included in PANDORA, which lists 1,983 journals published in Australia. Of these, 150 are commercial titles. The NLA released the first version of the PANDORA Digital Archiving System (PANDAS) in 2001.

OCLC Electronic Collections Online (OCLC ECO)
OCLC launched ECO in June 1997 to support the efforts of libraries and consortia to acquire, circulate, and manage large collections of electronic academic and professional journals. It provides Web access through the OCLC FirstSearch interface to a growing collection of more than 5,000 titles in a wide range of subject areas from more than 40 publishers of academic and professional journals. Libraries, after paying an access fee to OCLC, can select the journals to which they would like to have electronic access. OCLC has negotiated with publishers to secure for subscribers perpetual rights to journal content. In addition, OCLC has reserved the right to migrate journal backfiles to new data formats as they become available.

OhioLINK Electronic Journal Center (OhioLINK EJC)
The Ohio Library and Information Network is a consortium of Ohio's college and university libraries, comprising 85 institutions of higher education and the State Library of Ohio. OhioLINK's electronic services include a multipublisher Electronic Journal Center (EJC), launched in 1998, which contains more than 6,900 scholarly journal titles from nearly 40 publishers across a wide range of disciplines. OhioLINK has declared its intention to maintain the EJC content as a permanent archive and has acquired perpetual archival rights in its licenses from all but one publisher.

Ontario Scholars Portal
Launched in 2001, the Ontario Scholars Portal serves the 20 university libraries in the Ontario Council of University Libraries (OCUL). The portal includes more than 6,900 e-journals from 13 publishers and metadata for the content of an additional 3 publishers. The primary purpose of the portal is access, but the consortium has made an explicit commitment to the long-term preservation of the e-journal content it loads locally. The initiative began with grant funding but as of 2006 became self-funded through tiered membership fees.

Portico
Publicly launched in 2006, Portico is a third-party electronic archiving service for e-journals, and serves as a permanent dark archive. E-journal availability (other than for verification purposes) is governed by specific "trigger events" resulting from substantial disruption to access from the publishers themselves. A membership organization, Portico is open to all libraries and scholarly publishers,

which support the effort through annual contributions. As of July 1, 2006, 13 publishers and 100 libraries participated in Portico.

PubMed Central

Launched in February 2000, PubMed Central is NIH's free digital archive of biomedical and life sciences journal literature, run by the National Center for Biotechnology Information of the National Library of Medicine (NLM). PubMed Central encompasses about 250 titles from more than 50 publishers. It prefers that the complete contents for participating titles be submitted, but it will accept at minimum the primary research content, and it allows publishers to delay deposit by a year or more after initial publication. PubMed Central retains perpetual rights to archive all submitted materials and has committed to maintaining the long-term integrity and accuracy of the archive's contents.

General Characteristics

Three organizational types are represented among the twelve programs, as presented in Figure 1. The largest category includes government-supported efforts, with five of the six sponsored by a national library (CISTI Csi, KB e-Depot, kopal/DDB, NLA PANDORA, PubMed Central). LANL-RL receives funding from the U.S. Department of Energy and the U.S. Department of Defense. Two (OhioLINK EJC and the Ontario Scholars Portal) represent consortia that aggregate content primarily for access but have assumed archiving responsibility. Four (CLOCKSS, LOCKSS Alliance, OCLC ECO, and Portico) are member or subscriber initiatives, with all except ECO launched specifically to address digital archiving issues.

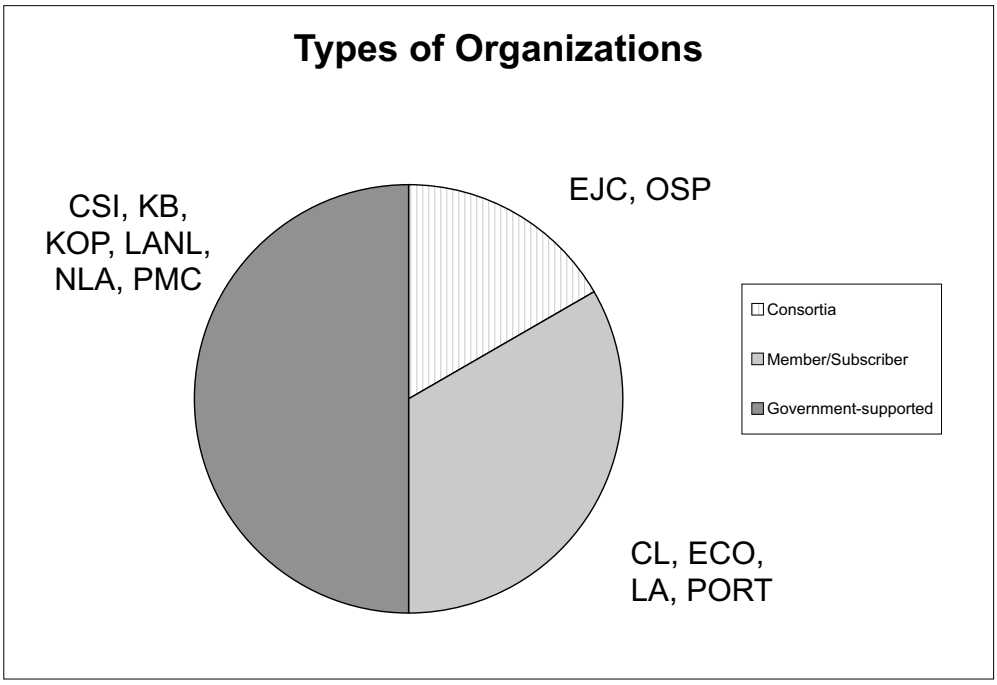

Fig. 1. Types of organizations included in survey

These programs are of recent origin. The oldest (LANL-RL) began in 1995, and four were launched within the past two years. Seven of the programs provide ongoing access to content and five limit access to current subscribers or members. Two (PubMed Central and NLA PANDORA) are open to all, but access to some material may not occur immediately following publication (this waiting period creates a "moving wall" for access). Five provide current access only for auditing purposes and for checking the integrity and security of systems and content; otherwise, access will be given after a trigger event occurs. A trigger event may occur, for example, when a publication ceases to be available online because of publisher failure or lack of support, a major disaster, or technological obsolescence.

Table 1 compares major attributes for the group, including year of inception, organizational type, access mechanisms, and designated users (i.e., those who receive access whenever it is provided).

Name	Start Year	Type of Entity	Principal Access Mechanism	Designated Users
CISTI Csi (CSI)	2000	National science library	Current online (partners)	Partners
CLOCKSS (CL)	2006	Limited membership / subscription	Trigger/audit	Members initially
LOCKSS Alliance (LA)	2005	Open membership	Trigger/audit	Local users served by library maintaining LOCKSS box
KB e-Depot (KB)	2002	National library	Trigger/audit (plus onsite), may provide online access (open) to open access content	Everyone in the Netherlands and beyond
kopal/DDB (KOP)	2004	National library	Trigger/audit (plus onsite), moving wall planned	Patrons of Die Deutsche Bibliothek and the Goettingen State and University Library
LANL-RL (LANL)	1995	Government/limited subscription	Current online, plus onsite to the general public	LANL staff and subscriber staff
NLA PANDORA (NLA)	1996	National library	Current online (open)	Australians and anyone with a research interest in Australia
OCLC ECO (ECO)	1997	Open membership/ subscription	Current online	Subscribers to the ECO collections
OhioLINK EJC (EJC)	1998	Limited Consortium	Current online	Supporting members of Ohio higher education community
Ontario Scholars Portal (OSP)	2001	Consortium	Current online (members) based on what they have licensed/purchased	Institutional members of the Ontario Council of University Libraries
Portico (PORT)	2005	Open membership/ subscription	Trigger/audit	Members, at least initially
PubMed Central (PMC)	2000	National medical library	Current online (open) and moving wall	Users of the NLM in the United States and beyond

Table 1. Major attributes of programs surveyed

Note: For the purposes of this report, the abbreviations listed in the left-hand column above will be used for all figures and tables. CLOCKKS was not considered as a separate entity from LOCKSS during the initial round of survey and interview and, therefore, will not be listed separately in many tables.

Assessing E-Journal Archiving Programs

Our team compiled and analyzed the survey responses in May and June 2006, freezing the addition of new information on July 1. A set of indicators for assessing the e-journal archiving programs was derived, in part, from two statements. The first is the *Minimum Criteria for an Archival Repository of Digital Scholarly Journals*, issued in May 2000 by the DLF. The second is the minimal set of services for an archiving program represented in the "Urgent Action" statement noted above.

As a result of this work, we identified seven indicators of a program's viability. In meeting its obligations to archive e-journals, the repository should

1. have both an explicit mission and the necessary mandate to perform long-term e-journal archiving;
2. negotiate all rights and responsibilities necessary to fulfill its obligations over long periods;
3. be explicit about which scholarly publications it is archiving and for whom;
4. offer a minimal set of well-defined archiving services;
5. make preserved information available to libraries under certain conditions;
6. be organizationally viable; and
7. work as part of a network.

Figure 2 shows our estimate of the current state of program viability for the 12 e-journal archives under review based on the seven indicators. These programs have secured their mandates, defined

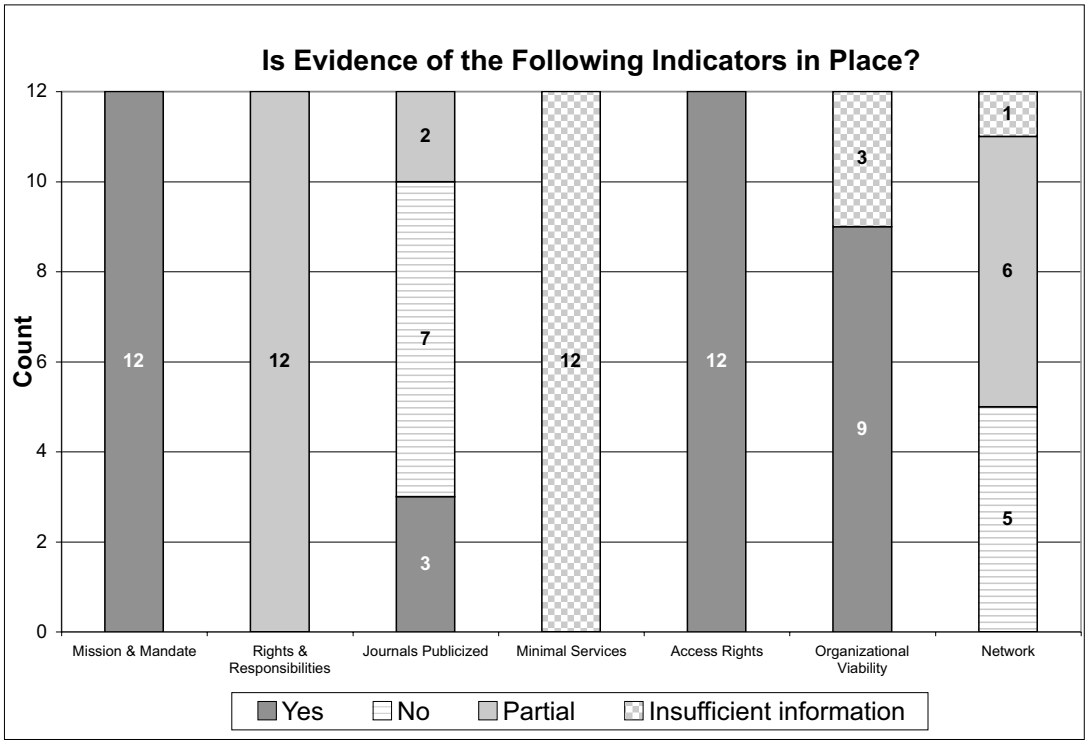

Fig. 2. Measuring e-journal archiving programs against seven indicators

access conditions, and are making good progress toward obtaining necessary rights and organizational viability, but room for improvement is apparent in three key areas: content coverage, meeting minimal services, and establishing a network of interdependency.

A discussion of the seven indicators follows.

Indicator 1: Mission and Mandate

The repository should have both an explicit mission and the necessary mandate to perform long-term e-journal archiving.

All 12 programs confirmed that their missions explicitly committed them to long-term e-journal archiving, and each has negotiated with publishers to secure the archival rights to manage journal content. Many publishers are willing to participate in these programs in part to protect their digital assets and in response to increasing demand from their principal customers. For example, the five largest STM publishers—Blackwell, Elsevier, Springer, Taylor & Francis, and Wiley—are all engaged in more than one of the e-journal archiving efforts reviewed in this report. Their participation, however, is voluntary, and at least one other publisher refused to grant OhioLINK EJC archival rights as part of its license agreement. E-journal archiving efforts could be strengthened considerably if publishers were required by legislative mandate or as a precondition in license arrangements to deposit their content in suitable e-journal archives.

The Role of Legal Deposit in E-Journal Archiving

More and more nations are requiring the deposit of electronic publications, including electronic journals, in their national libraries. Both the British Library and Library and Archives Canada, for example, are designing electronic-deposit repositories, and Germany recently passed a law that mandates the deposit of German publications, a move that will strengthen kopal/DDB's program.[9] Other nations are expected to follow suit.

While legal deposit is often implemented as a requirement for copyright protection, in practice it can also become an important component of a digital preservation program. Legal deposit laws provide the designated deposit libraries with both an explicit mission and a mandate to preserve a nation's publications. Once a journal has been deposited, the repository library is responsible for its preservation.

One question is whether legal deposit requirements will obviate the need to establish other e-journal archiving programs. We suggest that it will not, for at least four reasons. First, and most important, while most of the laws are intended to ensure that the journals will be preserved, there is less clarity as to how one can gain access to those journals. In almost all cases, one can visit the national library

[9] See the "*Gesetz über die Deutsche Nationalbibliothek (DBNG)*," signed into law June 22, 2006, and available at http://www.d-nb.de/wir/pdf/dnbg.pdf.

and consult an electronic publication onsite. It is unlikely, however, that the national libraries will be able to provide online access to remote users in the event of changes in subscription models, changed market environments, or possibly even publisher failure. The recently revised "Statement on the Development and Establishment of Voluntary Deposit Schemes for Electronic Publications," endorsed by both the Committee of the Federation of European Publishers (FEP) and the Conference of European National Librarians (CENL) and intended to serve as a model for national deposit initiatives, makes no mention of access beyond the confines of the national legal deposit library, leaving such issues to separate contractual arrangements with the publishers (CENL/FEP 2005). None of the national deposit programs we surveyed currently has the capability to serve as a distributor of otherwise unavailable archived journals.

Second, because legal deposit requirements are so new, the ability of the national libraries to preserve content is largely untested. Spurred by the requirements of legal mandates to acquire and preserve digital information, the national libraries have made tremendous strides in developing digital preservation programs. Many advances in our understanding of digital preservation have come through the work of the KB, the NLA, and other pioneering national libraries and archives working in this area. None of these libraries, however, would claim that it has developed the perfect, or only, solution to digital preservation. At this early stage in our knowledge, it is important to have competing digital preservation solutions that can, over time, help us develop a consensus as to what constitutes best practice.

Third, while the movement for national digital deposit legislation seems to be spreading, major gaps remain. In many cases, such as in the Netherlands, the deposit program is a voluntary agreement between the library and the publishers. Publishers are encouraged, but not required, to deposit electronic material. In other cases, most notably the United States, there is neither mandatory legal deposit for electronic publications nor clear evidence that the Copyright Office could demand the deposit of electronic publications (Besek 2003). At a minimum, the United States will need to adopt strong mandatory digital deposit legislation if legal deposit is ever to replace library-initiated preservation.

Finally, and somewhat paradoxically, the concept of national publications is becoming problematic, especially when dealing with electronic journals. Elsevier, for example, may be headquartered in the Netherlands, but does that make all its publications Dutch and subject to any future deposit laws in the Netherlands—even when those journals may have a primarily U.S.-based editorial board and may be delivered from servers based in a third country?

Although legal deposit may not be the silver-bullet solution to archiving e-journals, it is clearly an important component of the preservation matrix. If nothing else, a legal requirement that would force publishers to deposit e-journals in several national deposit sys-

tems (because of the international nature of publishing) would create pressure for standard submission formats and manifests for e-journal content. In addition, once material is preserved, it may be possible to revisit the trigger events that allow access to the content and even to permit remote access in narrow circumstances. The national libraries are also well positioned to develop technical expertise related to digital preservation and to share that expertise. For these reasons, we hope that efforts to develop more e-journal deposit laws will continue. It would be particularly beneficial if the U.S. Copyright Office started requiring deposit of electronic journals for copyright protection and the Library of Congress (LC) assumed responsibility for the preservation of those journals.

The Role of Open-Access Research Repositories in E-Journal Archiving

A development closely related to mandatory legal copyright deposit is the mandatory deposit of funded research into an open-access research repository, such as PubMed Central or arXiv. To date, participation in such repositories has been voluntary, and the results have been mixed. NIH, for example, estimates that only 4% of eligible research is making its way into the PubMed Central online digital archive as a result of the voluntary provisions of NIH's Policy on Enhancing Public Access to Archival Publications Resulting from NIH-Funded Research, implemented in May 2005 (DHHS 2006). Indeed, member publishers of the DC Principles Coalition fiercely contested the idea of a "mandated central government-run repository" (AAP, AMPA, DCPC 2004).

Several initiatives now under way could alter the voluntary nature of most agreements. In the United Kingdom, the Wellcome Trust and the Medical Research Council have ordered that the final copies of all research they fund be deposited in the UK PubMed Central, and the Biotechnology and Biological Sciences Research Council has mandated that publications from research it funds after October 1, 2006, will be deposited "in an appropriate e-print repository" (BBSRC 2006). Research Councils UK (RCUK) has encouraged the other U.K. research councils to consider deposit of funded research in an open-access repository.[10] In the United States, a recent NIH appropriations bill was modified in committee to mandate the deposit of copies of all NIH-funded research in an open-access repository within 12 months of publication (Russo 2006). In addition, Senators John Cornyn (R–TX) and Joe Lieberman (D–CT) have introduced the Federal Research Public Access Act of 2006 (FRPAA), which would require that research funded by the largest federal research agencies and published in peer-reviewed journals be deposited and made

[10] See "RCUK Position on Issue of Improved Access to Research Outputs" Web page at http://www.rcuk.ac.uk/access/.

openly accessible in digital repositories within six months of publication. Publishers oppose this proposed legislation.[11]

Given that more and more funded research is going to find its way into open-access repositories, an obvious question is whether libraries can rely on those repositories to preserve that information. There are at least two reasons why we would not recommend relying solely on open-access repositories for an archiving solution at this time.

First, while much research that appears in journals is funded by major U.S. or U.K. funding sources, many articles are not so funded. Consequently, much information will remain outside open-access repositories for the foreseeable future. Open-access article repositories are unlikely to function as substitutes for electronic journals.

Second, open-access repositories are not necessarily digital preservation solutions, although sometimes their names suggest otherwise. For example, one of the oldest open-access repositories, arXiv, suggests by its name that it is involved with preservation, yet there is nothing in the repository software that will ensure the preservation of deposited digital objects. Similarly, the protocol that links many preprint servers was named the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), suggesting that its activities are related to the Open Archival Information System (OAIS) framework. In reality, OAI and OAIS have nothing to do with each other (Hirtle 2001). Open "archives" are primarily concerned with providing open access to current information and not with long-term preservation of the contents.

In its draft position statement on access to research outputs, issued June 28, 2005, RCUK noted the distinction:

> RCUK recognises the distinction between (a) making published material quickly and easily available, free of charge to users at the point of use (which is the main purpose of open access repositories), and (b) long-term preservation and curation, which need not necessarily be in such repositories. . . . [I]t should not be presumed that every e-print repository through which published material is made available in the short or medium term should also take upon itself the responsibility for long-term preservation.

[11] See Van Orsdel and Born 2006; see also letter to Senators Cornyn, Lieberman, and Collins from signatories of the Washington D.C. Principles for Free Access to Science, June 7, 2006, available at http://www.dcprinciples.org/LiebermanLetter.pdf. The D.C. Principles, released on March 16, 2004 (see http://www.dcprinciples.org/), lay out seven principles constituting "commitment to innovative and independent publishing practices and to promoting the wide dissemination of information in our journals" by dozens of nonprofit scholarly journal publishers that oppose government-mandated public release of scholarly research articles. One of the seven principles is, "We will continue to work to develop long-term preservation solutions for online journals to ensure the ongoing availability of the scientific literature." As of August 1, 2006, only about half of the 75 scholarly society publishers who have signed the D.C. Principles had committed to one of the twelve e-journal archiving programs profiled in this report. Most are users of HighWire Press, which is in the process of including all its titles in LOCKSS.

RCUK's proposed solution was not to assume that the open-access repositories would perform preservation, but instead to work with the British Library and its partners to ensure the preservation of research publications and related data in digital formats.

Similarly, the Cornyn/Lieberman bill does not assume that institutional or subject-based repositories will be able to preserve research articles. Instead, it requires that their long-term preservation be done either in a "stable digital repository maintained by a Federal agency" or in a third-party repository that meets agency requirements for "free public access, interoperability, and long-term preservation."

In sum, the existing open-access research repositories (other than PubMed Central) are unlikely to qualify at this time as stable digital repositories. Libraries should therefore not presume that the scholarly record has been preserved just because it has been deposited in such a repository. At the same time, initiatives such as those from the RCUK and in FRPAA could be important to the development of digital preservation because they would force agencies either to develop digital preservation solutions themselves or define the requirements for third-party solutions.

Recommendations

1. More effort needs to go into extending the legal mandate for preserving e-journals through legal deposit of electronic publications around the world, to formalize preservation responsibility at the national level.
2. As part of their license negotiations, libraries and consortia should strongly urge publishers to enter into e-journal archiving relationships with bona fide programs.
3. Publishers should be overt about their digital archiving efforts and their relationships with various digital archiving programs. The five largest STM publishers are all engaged in more than one of the e-journal archiving efforts reviewed in this report, but only one (Elsevier) presents its digital archiving program on its Web site. Several others have announced their archiving policies in newsletters or press releases—which may still be included on their Web sites as part of a publicity archive—but it can be difficult to locate this information.[12]

---

[12] A study of publishers' archiving policies conducted in 2002 produced similarly disappointing results, indicating little progress in this area in the past four years. See Hughes 2002. Elsevier's home page offers a link to a set of resources for librarians that includes Elsevier's archiving policy: http://www.elsevier.com/wps/find/librariansinfo.librarians/libr_policies#sdarchiving. Publishers that have issued press releases announcing their participation in archiving programs have advertised only those most closely associated with archiving (Portico, LOCKSS, CLOCKSS, and KB e-Depot). If the others are noted (e.g., OhioLINK EJC and Ontario Scholars Portal), the announcements say nothing about archiving but focus on their roles in providing access. Other publisher sites checked were Oxford University Press, Kluwer, Sage, and Cambridge University Press. A few e-journal publishers and providers have provided prominent references to their archiving efforts, including Project MUSE, which has a link to Archiving and Preservation available at http://muse.jhu.edu/about/index.html, and the journals home page for the American Institute of Physics (http://journals.aip.org), which has a direct link to its archives and use policy at http://www.aip.org/journals/archive/arch&use.html.

4. Programs with responsibility to provide current access and archiving should publicize their digital archiving responsibilities both to publishers and to the research library community. Our discussions with library directors revealed that several of them were unaware of PubMed Central's archiving responsibility or that it could serve as part of their preservation safety net.
5. As the "Urgent Action" statement stipulates, research libraries should not sign licenses for access to electronic journals unless there are provisions for the effective archiving of those journals. The archiving program should offer at least the minimal level of services defined in the "Urgent Action" statement. In addition, the programs should be open to audit, and, when certification of trusted digital repositories is available, they should be certified. Unless e-journal content is preserved in such a repository, research libraries should not license access.

Indicator 2: Rights and Responsibilities

Rights and responsibilities associated with preserving e-journals should be clearly enumerated and remain viable over long periods.

Closely related to mission and mandate is the need for clarity of a repository's rights and responsibilities vis-à-vis publishers, distributors, and content creators. Although a publisher may grant archiving rights to a repository, the circumstances surrounding the exercise of these rights may not be uniform or clearly enumerated—or even fully understood when the contract is written. Including input from research libraries and publishers in the governance or operation of the repository would be a useful way to monitor policies as circumstances change (Table 2).

	CSI	ECO	EJC	KB	KOP	LA	LANL	NLA	OSP	PMC	PORT
Yes, in governance		L	L			P/L		L	L	P/L	P/L
Yes, in operation		L	L			P/L	P	P/L	P		
No, in neither	P/L	P	P	P/L	P/L		L				

Table 2. Responses to question: "Do publishers have any voice in the governance/operation of your e-journal archiving program?" (P = publishers; L = libraries)

The following three questions should be carefully considered in laying the foundation for digital archiving responsibility:

First, do the contracts consider all intellectual property rights held by publishers, creators, and technology companies that pertain to the content, and do they convey to the repository the right to perform necessary archiving functions to prolong the life of the content? Such rights can include basic permission to copy or reformat material, or both. They extend to bypassing copy and access restrictions, expiration, and other embedded technological controls. If not granted explicit permission, the repository may be unable to provide ongoing access through copying, migration, or reproduction.

Second, does the publisher or its successor reserve the right to remove or alter content from the archival institution under certain circumstances? If so, the archived content could be placed at risk. When asked whether agreements with publishers allow the repository to continue to archive content if the publisher is sold or merges with another company, seven programs answered "yes," one answered "no," and two were unsure. PubMed Central reported an instance when a publisher acquired one of the journals previously included and decided not to participate further, so new content has not been added. The content already in the repository remained. OhioLINK EJC's publisher agreements make no mention of exceptions caused by future changes in ownership. Could their rights under these conditions be only indirectly protected? The KB e-Depot and kopal/DDB recommend that publishers continue to ensure compliance with archiving agreements in the event of mergers, buyouts, or discontinuation of publishing operations, but these recommendations are not legally binding. Elsevier reserves the right to remove content from the KB e-Depot if there is a breach of contract; the LANL-RL indicated that material received could be kept indefinitely, "as long as previously agreed-upon usage restrictions are adhered to." CISTI Csi will seek to obtain a new agreement in the case of a merger or title transfer to a new publisher.[13]

Finally, are agreements with publishers regarding archival rights of limited duration? If so, the circumstances governing preservation responsibilities may be subject to change. Four of the twelve repositories reported that their contracts are of fixed, limited duration. They are reviewed regularly, at which time they may be renewed but also canceled. The remaining contracts are of indefinite duration or automatically renewable; all have cancellation options.

Recommendations

1. Once ingested into the digital archive repository, e-journal content should become the repository's property and not subject to removal or modification by a publisher or its successor.
2. In case of alleged breach of contract, there should be a process for dispute mediation to protect the longevity and integrity of the e-journal content.
3. Contracts need to be reviewed periodically, because changes in publishers, acquisitions, mergers, content creation and dissemination, and technology can affect archiving rights and responsibilities. Continuity of preservation responsibility is essential.

---

[13] An interesting glimpse at the perspective of publishers of journals for small scholarly societies regarding perpetual access responsibilities during title transfers appears in a publication of a British publisher's association. "If an unequivocal contractual commitment to provide 'perpetual' access was made by the transferring publisher, then strictly speaking it should bear the cost of whatever solution is adopted (be careful of this when drawing up your own journal licenses for journals you do not own!)." See ALPSP 2002.

4. A study should be conducted to identify all necessary rights and responsibilities to ensure adequate protection for digital archiving actions, so that these rights are accurately reflected in contracts and widely publicized.
5. Research libraries and consortia should pressure publishers to convey all necessary rights and responsibilities for digital archiving to e-journal archiving programs (i.e., the same rights should be conveyed in all archiving arrangements).

Indicator 3: Content Coverage

The repository should be explicit about which scholarly publications it is archiving and for whom.

Although this indicator seems to be straightforward, it is surprisingly difficult to identify what publications are being preserved and by whom. Six of the programs make public their list of publishers (OhioLINK EJC, PubMed Central, CLOCKSS, OCLC ECO, LOCKSS Alliance, Portico), three do so indirectly (KB e-Depot, CISTI Csi, Ontario Scholars Portal), and three do not (LANL-RL, NLA PANDORA, kopal/DDB). Even when the publishers are known, one should not assume that all journals owned by that publisher are included in the archiving programs. For instance, PubMed Central reported the largest number of publishers represented in its holdings, but the smallest number of titles of the 12 programs surveyed.

Locating a list of specific titles included is even more difficult. When asked whether they made an up-to-date, definitive list of titles available to the public, five responded "yes" (NLA PANDORA intersperses the list of journal titles with other content, with no ability to sort on e-journals only; the LOCKSS Alliance is building its list alphabetically by journal title). Five said "no," (the KB e-Depot and kopal/DDB indicated that they will archive all publications published in their respective countries). The remaining two programs plan to make such a list available. Further, even when the publications are listed, it is difficult to determine what date spans are included (only four repositories list this information) and how complete the contents of the publication are. For instance, the LANL-RL purchased backfiles of the Royal Chemistry Society journals from their inception to 2004, but is not receiving current content for local loading and archiving and does not intend to purchase it. Table 3 shows the availability of title lists and date spans by e-journal archiving repository. Maintaining content currency is a moving target; all repositories indicated they expect to add new titles and, indeed, during the course of our investigation new titles and publishers were being added frequently.

	CL	CSI	ECO	EJC	KB	KOP	LA	LANL	NLA	OSP	PMC	PORT
Title list			•	•			•		•	P	•	P
Date spans			•	•			P		•	P	•	P

Table 3. Responses to question "Do you make information about journal titles and date spans included in your program available to the public?" (• = yes; P = plan to within six months)

The pace of consolidation within scholarly publishing also creates dilemmas for those attempting to chronicle the state of the industry at any one time. Ownership of publishing houses, imprints, and individual titles is in constant flux, making it difficult to accurately associate large lists of titles with the correct publisher. In recent years, large companies with no name recognition as publishers have swallowed up a number of venerable publishing houses. Should these titles continue to be listed under the familiar, original publisher or by the new owner? Particularly complex are cases wherein a publisher has sold a portion of its titles or entire imprints but held on to others.

When evaluating data from e-journal archiving initiatives, it is sometimes impossible to tell whether lists of participating publishers or the names of publishers associated with particular titles reflect current status or are based on legacy metadata. For example, some initiatives still list Academic Press as a separate entity, while others have incorporated its titles under the current owner, Elsevier. When an initiative lists titles from Kluwer, is it referring to Kluwer Academic Publishers, which was purchased by Springer from Wolters Kluwer in 2004, or to Kluwer Health, which is still part of the original firm and includes labels such as Adis International and Lippincott, Williams & Wilkins? If complete title listings are available, it may be possible (though onerous) to make such a distinction, but lists are not always available.

Thus, the publisher listings presented here should be viewed as nothing more than a fuzzy snapshot of circumstances on July 1, 2006. The kind of precision that would allow us to determine the archived status of specific titles and publishers is not possible given the market's volatility and ambiguity in the current data.

Adding to the confusion about which titles and publishers are included in archiving initiatives is the fact that not all the "publishers" listed are truly publishers. Some are really aggregators—essentially republishers that provide electronic publication, marketing, and dissemination services for (usually) small scholarly societies that produce only one or a few titles and therefore benefit from aggregation to achieve visibility, critical mass, and state-of-the-art electronic publishing services.

Two prominent aggregators that turned up many times in our surveys are BioOne and Project MUSE. BioOne is a nonprofit aggregator that disseminates noncommercial titles in the biological, ecological, and environmental sciences. Most of the original publishers contracting with BioOne are scholarly societies and associations. As of July 1, 2006, BioOne handled 84 titles from 66 publishers. Even though none of the e-journal archiving initiatives we surveyed listed the American Association of Stratigraphic Palynologists as a publisher, its lone journal, *Palynology*, is included in LOCKSS Alliance, OhioLINK EJC, and Portico, by virtue of its contract with BioOne.

Project MUSE fills a similar niche for small publishers in the humanities, arts, and social sciences. Incorporating more than 300 journals from 62 publishers, predominantly university presses, as

of July 1, 2006, Project MUSE provides a portal and search facility that brings together many related titles. But MUSE also boasts that it provides a "stable archive." The overview on its Web site states the following:

> It is a MUSE policy that once content goes online, it stays online. As the back issues of journals increase annually, they remain electronically archived and accessible. We also have a permanent archiving and preservation strategy, including participation in LOCKSS, maintenance of several off-site mirror servers, and deposition of MUSE content into third-party archives.

MUSE participates in LOCKSS Alliance, OhioLINK EJC, and OCLC ECO. So, despite the absence of the George Washington University Institute for Ethnographic Research on the publisher listings of any of the e-journal archiving initiatives included here, its journal, *Anthropological Quarterly*, is being archived.

Other aggregators that are participating in at least one of the archives include HighWire Press (which hosts nearly 1,000 titles from large and small publishers and is affiliated with LOCKSS Alliance), the LOCKSS Humanities Project, the History Cooperative, and ScholarOne, Inc.

With all these caveats in mind, the number of titles included in these 12 programs is impressive, exceeding 34,000, as shown in Figure 3.

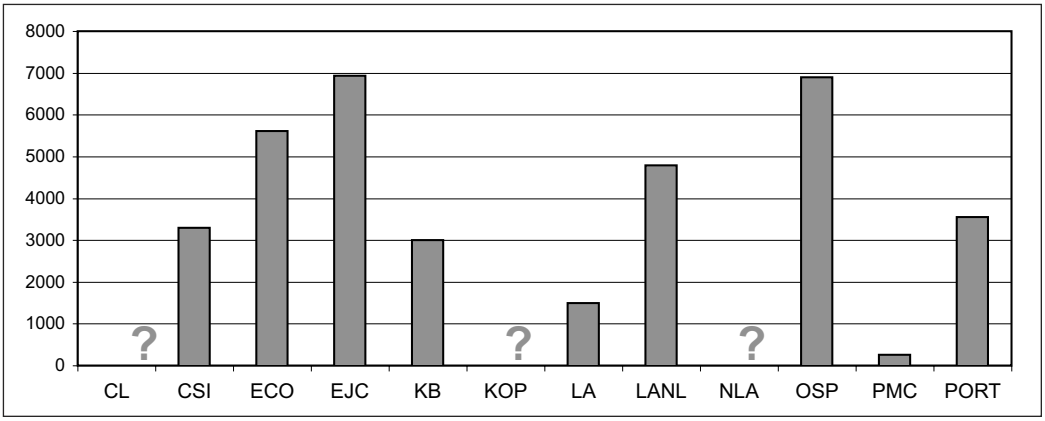

Fig. 3. Approximate number of titles included in e-journal archiving programs

Because there is no definitive list of titles covered in all these programs, the degree of overlap in content coverage is unknown. We were able to identify 220 publishers mentioned as participating in one or more of the e-journal archiving programs under review. We omitted PANDORA because the NLA preserves only Australian publications and does not maintain e-journal publisher data separately. Figure 4 provides the total publisher count for each e-journal archiving program. Appendix 3 lists the publishers in each archiving program.

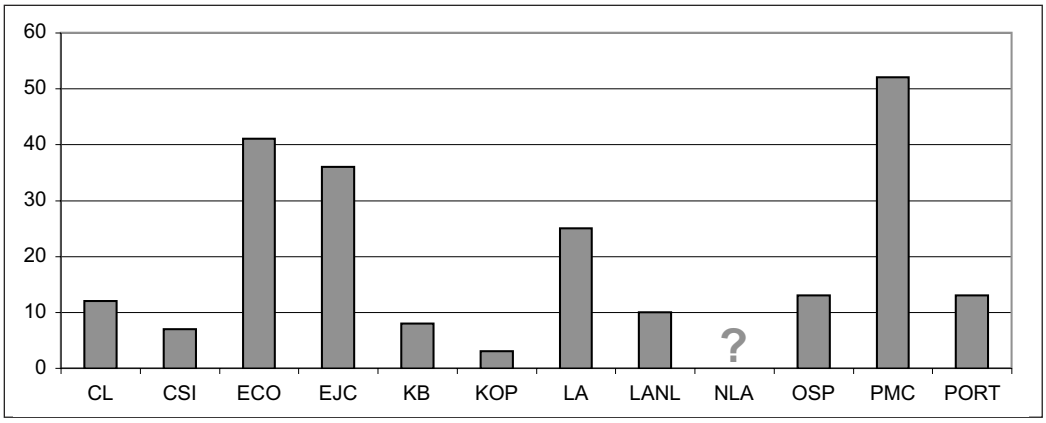

Fig. 4. Number of publishers included in the 12 e-journal archiving programs surveyed

The number of unique publishers in this pool is 128 (58% of the total). Of those, 91 (71%) are participating in only 1 program; 20 (16%) are involved in 2 programs. The major publishers are well represented in multiple arrangements. As Figure 5 reveals, 17 of them (13%) are involved in 3 or more programs and 6 of them (5%) are involved in 7 or more programs. Appendix 4 identifies the publishers included in more than one e-journal archiving arrangement.

Although there may not be complete overlap in content in each program, it appears that there is much redundancy for the major publishers of STM e-journals, especially those in English, many of which have their own archiving programs. Other disciplines, smaller publishers (especially independent Web publications of a dynamic nature), and most material published in non-Roman alphabets are less represented in general and particularly in multiple arrangements. They are also less likely to have developed a full-fledged archiving program in-house.

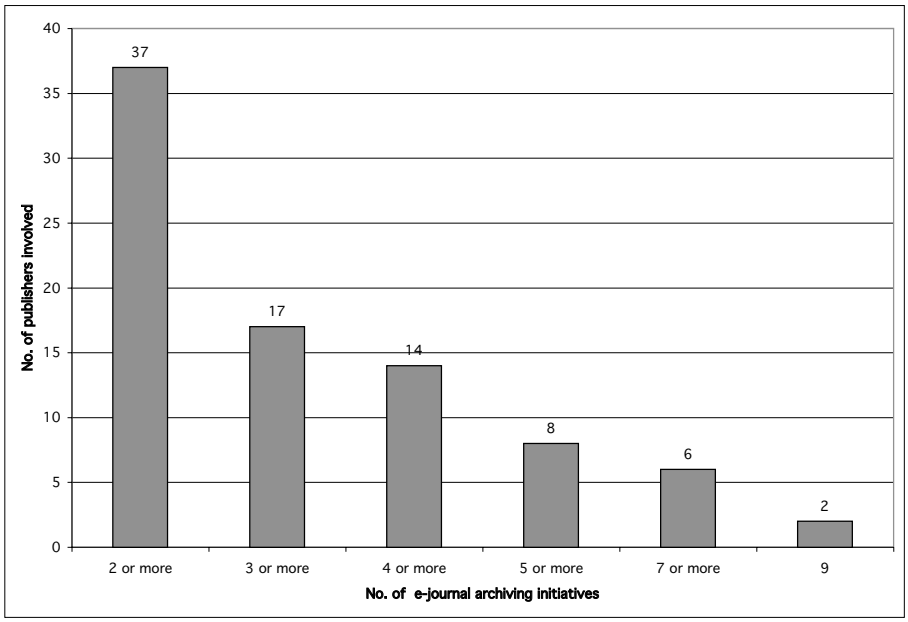

Fig. 5. Publisher overlap

It is unclear what the trend toward amalgamation of smaller presses into larger entities will mean for digital archiving, but it might prove beneficial. Recognizing the extent of at-risk e-journals in the humanities, LOCKSS launched its Humanities Project in 2004. Selectors at a dozen research libraries are participating in the project to identify significant content in the humanities for preservation, and programmers at those institutions are developing the plug-ins needed to capture the content, once the relevant publishers sign on.[14]

In addition to being transparent about the list of journals included and the date spans covered for each journal, archiving programs should be explicit about the content captured at the journal level (see next section). Content captured can vary by publisher as well as by journal. Given the differing archiving approaches used, it is likely that the extent of content captured for a particular journal held by more than one archive will vary among archives.

Recommendations

1. E-journal archive repositories need to be more overt about the publishers, titles, date spans, and content included in their programs. This information should be easily accessible from their respective Web sites.
2. A registry of archived scholarly publications should be developed that indicates which programs preserve them, following such models as the Registry of Open Access Repositories (ROAR), which lists 667 open-access e-print archives around the world, and ROARMAP, which tracks the growth of institutional self-archiving policies.
3. Research libraries should lobby smaller online publishers to participate in archiving programs and encourage e-journal programs to include the underrepresented presses; ideally, e-journal programs would cooperate to ensure that they share the responsibility to include these journals. (Only the LOCKSS Alliance allows a library to choose which publications to include.)

Indicator 4: Minimal Services

E-Journal archiving programs should be assessed on the basis of their ability to offer a minimal set of well-defined services.

This indicator is among the most elusive to assess because there is no universally agreed-on set of requirements for digital preservation, no mechanism to qualify (or disqualify) archiving services, and no organized community pressure to require it, although promising work is under way.

In 2003, RLG and NARA established the RLG-NARA Digital Repository Certification Task Force to develop the criteria and means for verifying that digital repositories are able to meet evolving digital preservation requirements effectively. The task force built on the earlier work of the OAIS working groups, especially the Archival

[14] http://www.lockss.org/lockss/Related_Projects.

Workshop on Ingest, Identification, and Certification Standards. In September 2005, RLG issued the task force's draft *Audit Checklist for Certifying Digital Repositories* for public comment. The checklist provides a four-part self-assessment tool for evaluating the digital preservation readiness of digital repositories. A revised version of the checklist is planned for release by the end of 2006.

To further the digital preservation community's certification efforts, The Andrew W. Mellon Foundation awarded a grant to fund the Certification of Digital Archives project at CRL. This project used the draft RLG audit checklist as a starting point for conducting test audits for four archival programs: Portico, LOCKSS Alliance, the Inter-University Consortium for Political and Social Research, and the KB e-Depot. The results of these test audits are informing the revision of the checklist. The project's final report, also scheduled for release by the end of 2006, will include recommendations for future developments in the audit and certification of digital repositories.

The Digital Curation Centre in the United Kingdom is conducting test audits of three digital repositories. It has a particular interest in and focus on the nature and characteristics of evidence to be provided by an organization during an audit to demonstrate compliance with the specified metrics. An interesting aspect of its approach is the value and use of evidence provided by observation and testimonials (Ross and McHugh 2005, 2006).

Germany is developing a two-track program for certification. DINI (Deutsche Initiative für Netzwerkinformation), a German coalition of libraries, computing centers, media centers, and scientists, encourages institutions to adopt good repository management practices without being overly prescriptive—steps that would lead to soft certification. The aim of soft certification is to motivate institutions to improve interoperability and gain a basic level of recognition and visibility for their repositories. The nestor project (Network of Expertise In Long-term STOrage of Digital Resources) is investigating the standards and methodologies for the evaluation and certification of trusted digital repositories and embodies rigorous adherence to requirements, leading to hard certification. The principles embraced by the nestor team include appropriate documentation, operational transparency, and adequate strategies to achieve the stated mission. DINI focuses on document and publication repositories at universities for scientific and scholarly communication and had issued 19 certifications as of July 2006. Nestor's scope goes beyond the realm of higher education and also targets repositories in national and state libraries and archives, museums, and data centers. Nestor is finalizing its certification criteria and has not yet issued any certificates (Dobratz and Schoger 2005; Dobratz, Schoger, and Strathmann 2006).[15]

It is not now possible for digital archiving programs to be certified, but when asked whether they would seek to become certified

[15] A list of institutions that have received DINI certificates is available at http://www.dini.de/dini/zertifikat/zertifiziert.php.

once such a process is in place, five of the e-journal archiving programs indicated they would, one indicated it would not, and five were uncertain or unaware of the certification effort. Table 4 reports their responses.

	CSI	ECO	EJC	KB	KOP	LA	LANL	NLA	OSP	PMC	PORT
Yes				•	•	•	•				•
No			•								
Not sure	•	•						•	•	•	

Table 4. Responses to question: "Will you seek to become a certified repository?" (• = yes)

In the absence of a certification process, adherence to digital preservation standards is a potential gauge to the technical viability of a program. Some existing digital preservation standards and best practices provide pieces of the puzzle.[16] We asked the surveyed repositories whether they were adhering to or planning to follow some of the key standards in the next six months. Table 5 lists these standards and best practices and provides the repositories' responses. Of interest is that only 5 of 11 programs report adherence to OAIS, an International Standards Organization standard that is gaining strong purchase in the digital preservation community. NLA PANDORA sees compliance to standards as a long-term goal and aligns with them as much as possible.

Despite the lack of a means to certify the operation of digital re-

Standard or Practice	CSI	ECO	EJC	KB	KOP	LA	LANL	NLA	OSP	PMC	PORT
OAIS	•			•	•	•					•
Trusted Digital Repositories				P		•	P				•
PREMIS	•			P		P	P				•
OAI-PMH	•			P		•	•			•	•
JAI DTD (NLM)	•									•	•
Audit Checklist				P		P					•
Open source software	•				•	•	•	•			•
Open file formats					•	•	•		•	•	•
Nonproprietary storage media						•	•				•
Other				P	•						

Table 5. Responses to question: "Do you follow any of the following standards and best community practices for archiving?" (• = yes; P = plan to within six months)

[16] Relevant standards include OAIS (Open Archival Information System), Reference Model, ISO 14721:2002; PREMIS (PREservation Metadata: Implementation Strategies); METS (Metadata Encoding and Transmission Standard); NISO MIX (NISO Metadata for Images in XML), NISO z39.87; MPEG-21; PDF/A-1 (Portable Document Format/Archival), ISO 19005-1:2005(E); OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting); Journal Archiving and Interchange DTD (Document Type Definition); and Journal Publishing DTD.

positories, enough conceptual work has been done to identify minimal expectations of best practices for a less rigorous standard—that of a well-managed collection. Measures such as an effective ingest process with minimal (even manual) quality control, acquiring or generating minimal metadata for digital objects in collections, maintaining secure storage with some level of redundancy, establishing protocols for monitoring and responding to changes in file format and media standards, and creating basic policies and procedural documentation—all acknowledge and address fundamental threats to digital document longevity.

There is widespread agreement about the nature of those threats—information technology (IT) infrastructure failure (hardware, media, software, and networking), built environment failures (plumbing, electricity, and heating, ventilation, and air conditioning), natural disaster, technological obsolescence, human-induced data loss (whether accidental or intentional, internal or external in origin), and various forms of organizational collapse (financial, legal, managerial, societal). There is far less uniformity of thought about the best means to confront each threat, or even which approaches should be considered effective to provide minimal protection.

Not surprisingly, therefore, the programs we surveyed, despite claiming a similar mandate, have chosen a variety of ways to carry it out. The diversity of approaches is healthy and useful, since only time and experience will tell us which techniques are effective. It is critical, however, that existing programs honestly and accurately document their successes and failures. The need for a risk-free mechanism to report negative results was noted in a previous CLIR report, which recommended "establishing a 'problems anonymous' database that allows institutions to share experiences and concerns without fear of reprisal or embarrassment" (Kenney and Stam 2002). The recommendation to establish such a system arose again in a more recent paper, which suggested the National Aeronautics and Space Administration's Aviation Safety Reporting System as a possible model (Rosenthal et al. 2005b). We heartily endorse these recommendations and believe that the community should place high priority on creating such a reporting system soon. The only way we will learn about the efficacy (or lack thereof) of various approaches is by having truthful reporting of experiences.

Short List of Minimal Services

As a starting point for documenting the digital preservation services being executed by the programs under review, we chose to assess them by five technical requirements laid out in the "Urgent Call to Action" statement, plus an additional requirement that we believe qualifies for the "short list" of minimal services:

- 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 nonproprietary 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 to contain costs, but provide enough processing so that the archives could locate and adequately render files for participating libraries in the event of loss;
- guard against loss from physical threats through redundant storage and other well-documented security measures; and
- offer an open, transparent means of auditing these practices.

Our discussion of these services presumes that programs should address not only what the services consist of but also how they intend to implement them.

Receive files that constitute a journal publication in a standard form, either from a participating library or directly from the publisher. This ingest-focused requirement encompasses at least two major elements. The first deals with the standard form that received files take. Before delving into specific standards, it is necessary to distinguish two basic approaches that e-journal archiving programs can use to receive the files that constitute a journal publication from the publisher. The most common approach is often referred to as "source-file archiving." In it, the archival agency receives from the publisher the files that constitute the electronic journal. These could be the standard generalized markup (SGML) files used to produce the printed volumes or the word processing or extensible markup language (XML) files used by the publisher to produce both printed and online products, such as portable document format (PDF) files. Graphic files and supporting material can also be included. In some cases, the files sent to an archival agency can be more complete than what is actually published. For example, a high-resolution image could be preserved even though a lower-resolution image is used on an online access site. PubMed Central and Portico are focused on preserving the source files received from the publishers.

A second approach is to receive the files that constitute the journal as published electronically. We call this approach "rendition archiving," since it focuses on preserving the journal in the form made available to the public. PDF files are the most common format for displaying journals as published, although some programs also receive the HTML and image files that are used to display a journal to readers. All the programs we surveyed welcome the submission of rendition files, and some, such as OCLC ECO, NLA PANDORA, and the LOCKSS Alliance, are based entirely on preserving and delivering the content as published. The LOCKSS Alliance and NLA PANDORA are special cases of rendition archiving. Rather than relying on rendition files provided by the publisher, they harvest (with the permission of the publishers) files from the publishers' Web sites.

Each of these approaches has advantages and disadvantages. With source archiving, the most complete version of the e-journal content is preserved. Furthermore, as is discussed in detail below,

source-file content is often either delivered in or converted to a few normalized formats, on the assumption that it will be easier to ensure the long-term accessibility of standardized and normalized files. One disadvantage to source archiving is that it requires a large upfront investment, with no assurance that the archive will ever actually be needed. In addition, the presentation of the e-journal content will almost certainly differ from that of the publisher; the "look and feel" of the journal will be lost.

Rendition archiving can maintain the look and feel of the journal, but it may be harder to preserve the content. No one knows, for example, what an effective migration strategy for PDF documents might be. In addition, it may be difficult to preserve the functionality of a dynamic e-journal if harvesting screen "scrapes" of static hypertext markup language (HTML) pages is the preferred ingest solution. On the plus side, the initial costs associated with preserving rendition files are likely to be lower (and, in the case of the harvesting projects, much lower). Migration, normalization, and other preservation activities need take place only when actually needed.

At this point, it is impossible to say which of these two approaches is the better solution to archiving. Those programs that solicit both source files and rendition copies of e-journal content (PubMed Central, Portico, KB e-Depot, kopal/DDB) probably are the safest archiving solution—but at a potentially greater cost.

Since text structure is the aspect of journal publishing that has been subject to the greatest standardization effort, source files are the type most commonly produced in a standard form. Several SGML and XML DTDs (document type definitions) have been devised specifically to support publishing of scholarly journal articles. One of the most popular is the NLM/NCBI (National Library of Medicine/National Center for Biotechnology Information) Journal Archiving and Interchange DTD. The full Journal Archiving and Interchange DTD Suite also includes modules that describe the graphical content of journal articles and certain nonarticle text, including letters, editorials, and book and product reviews. Acceptance of the Journal Archiving and Interchange DTD received a major boost in April 2006 when LC and the British Library announced support for the migration of electronic journal content to the NLM DTD standard, "where practicable" (Library of Congress 2006).[17] Four of the programs we surveyed currently use the NLM DTD.

Use of XML and SGML with DTDs designed for journal articles and other components has implications for "standard form" of structure and interchange capability at the lowest levels. The definition

[17] Even in the case of those programs that are using the NLM DTD, none requires the publisher to submit its material in that form. PubMed Central requires participating publishers to submit research articles in SGML or XML, based on an established journal article DTD. Although it does impose certain minimum coding requirements, it does not insist on use of the NLM DTD. More and more publishers are moving to XML-based production systems, and consider the XML version (not PDF or HTML) to be the official version. Nevertheless, there is still a considerable lack of publisher consistency regarding the "standard form" for journal articles.

of a character in the XML specification is based on the Unicode set. We queried the programs about the Unicode compatibility of their systems and found that at least some components of legacy systems (ScienceServer sites in particular) lacked it. With many publishers now supplying both journal content and metadata in XML, this has caused problems, particularly with the display of bibliographic data for some access-driven programs. We heard complaints that publishers had made the switch to Unicode compliance without giving the archive enough time to adjust its ingest procedures, resulting in incompatibilities. Two archives (PubMed Central and Portico) mentioned that despite being fully Unicode compliant, they could not support non-English metadata because of limitations in their ability to perform quality control and, in PubMed Central's case, because the search-and-retrieval system is based on English-language indexing and text matching.

Given that many of the programs profiled here are research driven, it is not surprising that they are trying to break new ground in repository development. Consequently, some of the "standard forms" used in the programs are unique to them. In LANL-RL's new aDORe repository, digital objects are represented using MPEG-21 DID (digital item declaration) and stored in an XML tape, while kopal/DDB has developed a Universal Object Format (Steinke 2006) for archiving and exchange of digital objects. Unfortunately, nothing yet qualifies as "universal" when it comes to digital objects. (As a cynic once said, "The nice thing about standards is that there are so many to choose from.") Until digital repository design matures and stabilizes, exchange of complex digital objects (i.e., archival information packages, or AIPs) among repositories will be less than transparent. However, proposals are emerging for facilitating the exchange of complex digital objects between repositories and archives.[18] Experimentation with a variety of approaches is appropriate at this stage of archive development. We also recommend that e-journal archives using different standards begin examining interoperability issues for digital objects and metadata, with an eye on maximizing compatibility.

There is as yet no standard form for source files. Although many programs prefer, and some require, files to be delivered as PDFs, no specific version of PDF is required. No program requires that PDFs adhere to ISO 19005-1 (PDF/A-1), and we are not aware of any major publishers that offer their files in that format.

Asked about the existence of file-format requirements (or preferences) for ingest, eight programs said they have such requirements, and half of them provided us with technical documentation describing them. Four do not (LOCKSS Alliance, Ontario Scholars Portal, NLA PANDORA, Portico). LOCKSS Alliance and NLA PANDORA harvest files from the Web and take whatever content can be delivered through Web protocols.

---

[18] See, for example, Bell and Lewis 2006, which examines interchange of electronic theses between a DSpace- and a Fedora-based repository; and Bekaert and Van de Sompel 2006.

The second major element of this minimal service is the receipt of "files that constitute a journal publication." Identifying the entirety of a journal publication in print is a straightforward matter, but the components of e-journals are more varied both in form and content and are far less tightly bound together. The lack of an established standard for what constitutes the essential parts of an e-journal was made abundantly clear by the nonuniform responses to our questions about which journal content types and features each archiving program includes (see Table 6).

All said they include research articles and errata, but beyond that there was no consistency. Athough most said they maintain "whatever the publisher sends," many do not include advertisements (which are often generated on-the-fly in a user-dependent manner) and certain other non-editorial content. Some do not capture supplemental materials, and even fewer are able to capture external features associated with publisher Web sites, such as discussion forums and other interactive content. Although it encourages the deposit of all journal components, PubMed Central, for example, requires only that research articles be provided; the presence of other kinds of content may vary among publishers, and even among titles.

The programs are aware that different publishers send different kinds and numbers of files for each title, but they seem less aware of what those components are. Survey comments made it clear that some responses to this question were guesses. Particularly for the access-driven programs, the focus is primarily research articles. Several respondents said that although they keep everything they receive, they are not necessarily able to provide access to all components.

Content/Feature	CSI	ECO	EJC	KB	KOP	LA	LANL	NLA	OSP	PMC	PORT
Research articles	•	•	•	•	•	•	•	•	•	•	•
Internal linking		•	•	•	•	•	•	•	•	•	•
Book reviews		•	•	•	•	•		•	•	•	•
Letters to the editor		•	•	•	•	•	•	•	•	•	•
Lists of editorial board members		•	•	•	•	•	•	•	•		•
Copyright statements		•	•	•	•	•	•	•	•	•	•
Journal descriptions		•	•	•		•		•			•
Advertisements		•	•	•		•		•	•		•
Reprint information		•	•	•		•		•	•		•
Editorials		•	•	•	•	•		•	•	•	•
News and announcements		•	•	•	•	•	•	•	•	•	•
Errata	•	•	•	•	•	•	•	•	•	•	•
Supplementary materials			•	•		•		•	•	•	•
Covers of corresponding print editions	•	•	•		•	•	•	•	•	•	•
Special features						•		•			
Other								•			

Table 6. Journal content types and features

There is likewise considerable variability within programs, because publishers have different definitions of what constitutes a complete e-journal. With no means to standardize journal components, and given that publishers are generally unable to provide manifests of how many files of what type the archive is supposed to be receiving, uncertainty at the receiving end is inevitable. Several programs noted that the lack of publisher manifests was a big problem. There is less ambiguity with programs that harvest content from publisher Web sites (NLA PANDORA and LOCKSS Alliance). Since the content is coming directly from the publisher's officially disseminated version, the only potential for missing components is if the harvesting itself is incomplete.

Users read and access the content of e-journals very differently than they do print journals (Olsen 1994). As more scholarly publishers eliminate print versions of their titles, it is possible that certain once-common features, such as advertisements or conference announcements, will be dropped or disseminated by different means (e.g., blogs or RSS feeds). The scholarly publishing landscape is not stable enough to prescribe what components (at minimum) constitute a journal publication in electronic form. But publishers need to do a better job of specifying exactly what they call a complete issue, and archiving programs need to pay more attention to exactly what they are receiving.

Store the files in nonproprietary formats that could be easily transferred and used should the participating library decide to change its archives of record. Use of nonproprietary formats has long been recognized as a strategy to fight obsolescence and improve the portability of digital objects. Depending on the ingest and archive approach of a particular program, the role of nonproprietary formats may be to

- take everything and store it in the supplied format (e.g., OhioLINK EJC, Ontario Scholars Portal, LOCKSS Alliance);
- take everything (or nearly so), preserve the original, but normalize it on ingest (e.g., Portico); or
- require use of a particular format or formats for deposit (e.g., PubMed Central, KB e-Depot, OCLC ECO).

The choice of preferred formats varies. Some require a form of XML (PubMed Central) or one that can be converted to XML (Portico), for articles, metadata, or both. Others accept PDF as the primary deposit format (OCLC ECO, KB e-Depot, OhioLINK EJC, CISTI Csi) or as an optional secondary format (PubMed Central). PDF is widely regarded as so open a specification that it is deemed nonproprietary. The lack of any credible competitor has made PDF seem a safe choice for long-term archiving, as evidenced by the work on PDF/A-1 and now PDF/A-2. However, the PDF specification is owned by Adobe, and recent events have slightly clouded the picture around it. Microsoft has announced the development of a competing product called XPS (XML paper specification), an XML-based document format

with many similarities to PDF. In June 2006, Microsoft reported that Adobe had threatened a lawsuit if plans to incorporate the ability to save as PDF into Office 2007 were carried out. Adobe denied making such a threat and said that its primary concern was that Microsoft would produce PDFs that strayed from its specification. Regardless of whom one believes, the bottom line is that no file format, no matter how open or popular, can be deemed permanently "safe."

The survey addressed the ability of programs to archive a variety of text, still image, and multimedia (sound and moving image) file formats (Tables 7–9). The gamut ranged from format-agnostic initiatives such as LOCKSS Alliance, which archives any format a publisher can make available through Web protocols, to prescriptive operations, such as PubMed Central, which requires submitted

Format Type	CSI	ECO	EJC	KB	KOP	LA	LANL	NLA	OSP	PMC	PORT
Plain text	•		•	•	•	•	•	•	•		•
HTML	•	•	•	•	•	•	•	•	•		•
SGML	•		•	•	•	P	•	•	•	•	•
XML	•	•	•	•	•	•	•	•	•	•	•
PDF	•	•	•	•	•	•	•	•	•	•	•
Postscript			•	•	•	•	•	•			•
TeX			•	•	•	•	•	•			•
Other				•		•					

Table 7. Text formats and page description languages accepted (P = plan to accept within six months)

Format Type	CSI	ECO	EJC	KB	KOP	LA	LANL	NLA	OSP	PMC	PORT
TIFF			•	•	•	•	•	•	•	•	•
JPEG	•	•	•	•	•	•	•	•	•	•	•
PNG			•	•	•	•	•	•		•	•
JPEG 2000			•	•	•	•	•	•		•	•
GIF	•	•	•	•	•	•	•	•		•	•
SVG		•	•	•	•	•	•	•		•	
Postscript			•	•	•	•	•	•			•
EPS			•	•	•	•	•	•		•	•
Other						•					

Table 8. Still-image formats accepted

Format Type	CSI	ECO	EJC	KB	KOP	LA	LANL	NLA	OSP	PMC	PORT
Sound file formats			•	•	•	•	•	•		•	
Moving image file formats			•	•	•	•	•	•	•	•	
Not sure								•			
Other											

Table 9. Other formats accepted

content to be in either XML or SGML. Just because a program says it accepts a format in its archive does not mean that it has the ability to provide access to it. For example, programs using an older version of ScienceServer software (three programs, at the time of our survey) are largely limited to displaying PDF, Tagged Image File Format (TIFF), and some XML files.

Effective transfer of archives content between programs requires more than simply using nonproprietary file formats. XML comes in many different flavors, with an external specification (the DTD) determining how the content should be interpreted. Metadata are moving toward standardization of both content and format, but metadata standards still vary widely among e-journal archives. Thus, even if we achieved universal adoption of nonproprietary file formats, easy transfer will be possible only with greater standardization of externalities and the containers that surround the basic digital objects.

Use a standard means of verifying the integrity of ingoing and outgoing files, and provide continuing integrity checks for files stored internally. This specification presumes that there is a standard means of determining and maintaining integrity, but our survey suggests that this area is ill-defined. Procedures for integrity testing differ greatly across the programs. Completeness testing can be automated or manual, and no two programs do it exactly the same way. Some test at the volume level, some at the issue level, and some at the article and article-component level. Some use byte counts while others use markup callouts. Only LOCKSS/CLOCKSS appears to have a system that incorporates a publisher's manifest for each transaction. Integrity testing at ingest is similarly nonstandard. Some programs use checksum comparisons or network transfer protocols that employ checksums (e.g., ftp). Others rely on random sampling with visual inspection or validation. LOCKSS boxes can do comparisons with both publisher sites and other LOCKSS boxes containing the same content.

Even though there are considerable differences in conducting completeness and integrity tests at ingest, ongoing integrity testing reveals the greatest divisions among the programs (see Table 10). Some lack any means for doing ongoing integrity testing. Several programs do periodic integrity checks using checksums. Although some access-driven programs conduct automated integrity checks, a prevailing view of those programs is that daily use by the constituency is the most effective way to uncover problems with individual files. At the same time, operators of access-driven programs are skep-

	CSI	ECO	EJC	KB	KOP	LA	LANL	NLA	OSP	PMC	PORT
Completeness upon ingest	•	•	•	•	•	•	•	•	•	•	•
Integrity upon ingest	•	•		•	•	•	•	•	•	•	•
Ongoing integrity		•		•	•	•	P	•	N/S	•	•

Table 10. Responses to question: "Do you conduct validation/testing?" (• = yes; N/S= not sure; P= plan to within six months)

tical that a dark archive can be properly maintained and ready for active use at any time simply by testing static properties of the content. They argue that usage patterns are ever-evolving and are themselves an essential part of curation. PubMed Central articulated this view most clearly:

> PMC operates on the philosophy that the best way to ensure the integrity of archived content is to use it directly, actively and continuously. Effective use of the content by humans and by automated processes proves the integrity and continued usability of the content. Therefore, the archive is made freely available to all users, encouraging repeated use—by between 50,000 and 90,000 different users each day and an estimated 1.5 million or more users a month. HTML views of articles are generated dynamically, directly from the archival XML copy, thus proving its integrity.
>
> Changing usage modalities reveal incremental problems in the data and allow them to be addressed before becoming massive and insurmountable. The bottom line is that there is a continuously ongoing process of archive curation.

Writing from a LOCKSS perspective, Rosenthal et al. (2005b) counter that relying on access alone as a means of integrity testing is inadequate because most items in an e-journal repository are infrequently used. The reliability of this approach is further called into question by the fact that one of the access-driven programs had a known problem (involving Unicode compatibility) that caused some bibliographic data to display as gibberish and yet logged no complaints from users. To obtain the greatest benefit from use testing, access systems should be designed to encourage and facilitate the reporting of integrity problems by users (Marty and Twidale 2000). Preservation-driven programs, however, can face resistance from publishers who can oppose regular use-based testing that does not derive from a trigger event (Honey 2005). Ultimately, both access-driven and preservation-driven programs need a combination of routine automated checks and regular review by a variety of users to maximize the benefits of integrity testing.

Limit the processing of received files to contain costs, but provide enough processing so that the archives could locate and adequately render files for participating libraries in the event of loss. Data are not yet widely available on the relative cost of file processing within digital repositories and the impact of various procedures on long-term renderability of files. Consequently, it is impossible to identify which programs have found the best balance between cost savings through minimizing file processing, and sufficient investment in metadata creation, integrity testing, and techniques to fight obsolescence. We can, however, look at examples of different approaches to limiting file processing and speculate about their

impact on efficiency of operations. Three approaches stand out:

- automating manual processes,
- offloading tasks to parties outside the archive, and
- making architectural decisions (e.g., about repository design, normalization, digital preservation strategy).

In operating and maintaining an e-journal archive, there are several steps with the potential to require large amounts of file processing. These include integrity and completeness validation at ingest, metadata creation at ingest, ongoing integrity testing, and responding to file-format obsolescence. The following paragraphs look at each of these activities in relation to the efficiency strategies mentioned above.

Integrity testing and completeness validation at ingest. These procedures are still conducted manually at many of the archives, even by programs with otherwise high levels of automation. Maintaining quality control at the point of ingest is sufficiently complex and important to warrant the time and expense of manual labor. If the completeness and integrity of content are not established at this point, the archive's ability to "locate and adequately render files for participating libraries" is substantially compromised. Tools for automating validation, such as JHOVE, are becoming available, and some archives are using them; Portico and the KB e-Depot both report using JHOVE in their workflows. However, there are limits to what automated validation can do, and a file deemed by JHOVE to be valid and well formed is not necessarily error-free.

Survey comments indicated that archives want more help from publishers in facilitating ingest. Archives would like publishers to provide a detailed manifest of the contents of each issue so that they have something against which to gauge completeness. The LOCKSS Alliance and CLOCKSS use an automated procedure to validate that everything the publisher made available has been collected. But that automated process would not be possible without the cooperation of the publisher (which creates a manifest page) and without the design of an architecture that supports this kind of testing as well as recovery from an error situation. So, LOCKSS/CLOCKSS combines all three approaches for maximizing the efficiency of completeness testing at ingest.

Metadata creation. Many see metadata creation as the most onerous step in digital repository management. There is a temptation to generate a lot of metadata (a tendency not discouraged by the size of the PREMIS data dictionary), on the presumption that "more is better" when it comes to managing digital files. However, there are significant costs in creating metadata, as well as ongoing costs for its maintenance and preservation. Some argue forcefully that hand-generated format and bibliographic metadata do not add enough value to merit the effort they require, relative to automated capture of the same class of data (Rosenthal et al. 2005b). LOCKSS uses completely automated metadata collection and believes that what it gets is good enough (although it notes that others disagree) and that the savings

from forgoing a more aggressive metadata-creation policy is better used in preserving additional content.

Automation is clearly an option for increasing the efficiency of metadata creation. Tools such as DROID, JHOVE, and the National Library of New Zealand Metadata Extraction Tool can aid in file-format identification as well as in extraction of deeper technical characteristics. Thus far, automated characterization is limited to a few popular file formats, but for most collections, that is probably adequate to deal with a distribution model in which 80% of the files are represented by a few common formats. Considerably more testing and experience with these tools are needed to improve their efficiency, learn their limitations, and develop best-practice guidelines for their deployment.

Since truly reliable automated means for extracting bibliographic and other forms of nontechnical metadata have yet to be perfected, such information should ideally be provided by the data submitter. If the publisher can be convinced to provide metadata in a standard format, so much the better.

Ongoing integrity testing. Several aspects of ongoing integrity testing, especially fixity verification, are routinely automated. KB e-Depot, Portico, kopal/DDB, and NLA PANDORA reported using checksums. The LOCKSS architecture uses a more robust system in which checksums are regularly generated and compared with newly generated checksums on peer LOCKSS boxes with the same content. If a discrepancy arises, a voting system is used to determine which box has the corrupted file and it is then replaced with a deemed "good" copy. The entire process is automated (Maniatis et al. 2003).

Some programs (OhioLINK EJC, Ontario Scholars Portal, CISTI Csi) have, in effect, offloaded the task of ongoing integrity testing to their users. Such an approach reduces costs by eliminating the programming and processing needed to implement and carry out automated checks, but it may leave large portions of a repository's content vulnerable to undetected corruption or loss. This is the case because standard usage patterns suggest that most articles will be infrequently accessed and because users tend to be unreliable at reporting data integrity problems unless empowered to do so (Marty 2005). Thus, opting to maintain data integrity by relying primarily on user feedback rather than other techniques may not be a good trade-off between cost savings and maintenance of long-term renderability.

Responding to file-format obsolescence. The role of repository architecture in streamlining operations comes to the fore in the design of procedures to respond to file format obsolescence. The options include the following:

- offloading some normalization responsibilities to the publisher (PubMed Central, KB e-Depot, OCLC ECO, OhioLINK EJC);
- normalization on ingest (Portico, PubMed Central, Ontario Scholars Portal);
- migration on-the-fly/just-in-time migration (LOCKSS Alliance, LANL-RL);

- batch migration/just-in-case migration (OhioLINK EJC, PubMed Central, OCLC ECO); and
- emulation (KB e-Depot, kopal/DDB, and NLA PANDORA).

The differences are even finer than these options suggest. For example, both PubMed Central and OhioLINK EJC request publisher normalization before ingest, but their strategies are very different. PubMed Central asks for partial normalization (publisher files delivered as XML or SGML based on an accepted journal publishing DTD), which it then fully normalizes to the NLM DTD. OhioLINK EJC, because its access software can handle only a limited range of file formats, requests that publishers normalize to one of those formats (typically PDF or XML) so that it can display the files to users. It does no internal normalization but assumes it will eventually have to do a batch migration of its currently used formats to more-modern formats. Thus, in the short term, PubMed Central has to process any file not already using the NLM DTD; later, it will have to batch-migrate its entire collection each time there is a significant change in the NLM DTD. OhioLINK EJC has essentially no up-front overhead for file-format management, but will eventually face multiple batch-migration operations when its prenormalized formats are no longer supported.

Strategies that envision doing on-the-fly migration also differ in implementation details. LOCKSS anticipates maintaining a suite of converters that will be called as needed, depending on whether an HTTP query indicates that the browser can handle the existing file format or not (Rosenthal et al. 2005a). LANL-RL, on the other hand, uses changes in the metadata envelope to indicate how a file should be decoded. Which technique will be judged more efficient and effective remains to be seen, since neither has had sufficient use in operational repositories to prove itself.

There are prospects for automating portions of the process of coping with file format obsolescence. XENA (XML Electronic Normalizing of Archives), a tool from the National Archives of Australia that facilitates normalization to XML-based formats, is now in its third postproduction release. None of the programs surveyed use XENA, which is not surprising since it is geared toward normalizing office-type documents rather than e-journal articles. However, one could imagine its utility for normalizing image files or supplemental data files that accompany some journal articles.

Another potential means for automation is the preservation-planning component of PRONOM 5b from the U.K. National Archives, slated for release in December 2006. According to the description, "The system will . . . focus on the development of migration pathways for the automatic conversion of electronic records to new formats as required for preservation or presentation purposes" (PRONOM 2006).

Three programs (KB e-Depot, kopal/DDB, and NLA PANDORA) said they would use emulation as a means of coping with file-format

obsolescence, though not to the exclusion of other techniques. A pair of studies published in *RLG DigiNews* deals directly with the competing interests represented by this minimal service: long-term usability versus cost of maintenance. Hedstrom and Lampe (2001) compared migration and emulation in terms of renderability; Oltmans and Kol (2005) compared them in terms of cost, providing some insight into the potential trade-offs between the two approaches.

Hedstrom and Lampe measured user satisfaction in response to both a migrated and an emulated form of a computer game. They found no statistical difference between users' perceptions of how well each approach preserved the game's look and feel. However, the authors concluded

> Further research on the effectiveness of emulation and migration needs to account for the quality of the emulator, the impact of specific approaches to migration on document attributes and behaviors, and on numerous aspects of the original computing environment that may affect authenticity and user experience.

Studies making similar comparisons between migrated and emulated components of scholarly e-journal articles, as well as user response to the repositories employing the different strategies, should help sort this out.

The Oltmans and Kol study, conducted as part of the KB e-Depot's research-and-development efforts, compared the projected costs of maintaining renderability of a large collection of digital objects over 50 years through either migration or emulation. The authors' model presumes higher up-front costs for emulation (mostly for emulator development), but cost savings from eliminating the need to periodically migrate every file soon thereafter tilt the advantage significantly toward emulation. At the end of 50 years, depending on the archive's size and other parameters, the authors predict that migration will be up to twice as expensive as emulation.

Regardless of the conclusions of these early studies, considerably more time and experience with large collections is needed before the relative merits of the different approaches to file-format obsolescence can be determined with any certainty. Most of the programs have only done small-scale testing or proof-of-concept exercises, particularly with regard to migration and emulation. Table 11 summarizes the programs' responses about the archiving strategies they use now or will adopt, when necessary.

Archiving Strategy	CSI	ECO	EJC	KB	KOP	LA	LANL	NLA	OSP	PMC	PORT
Migration	•	•	•	•	•	•	•	•		•	•
Emulation				•	•			•			
Normalization	•						•	•	•	•	•
Reliance on standards	•			•	•		•	•		•	•
Refreshing	•		•	•	•			•		•	•
Use of durable media				•	•				•	•	•

Table 11. Responses to question: "What type of archiving strategies do you use or plan to use?"

Whether we will learn which of these strategies best balances production efficiencies with protection of users' interests in the integrity of stored files depends heavily on how open the repositories are willing to be about their operations. Some archives are ingesting files that they currently have no means to render or disseminate or have no plan to migrate to more-manageable formats. Careful scrutiny and diligent reporting will be needed to ensure that such files are not forgotten or marginalized.

Guard against loss from physical threats through redundant storage and other well-documented security measures. Potential loss from physical threats is easily the best-understood and most widely appreciated aspect of digital preservation. Since the advent of digital-storage technology, IT professionals and casual computer users alike have maintained backup copies as a bulwark against the ephemeral nature of digital information and its vulnerability to a raft of destructive forces.

Redundancy provides an important hedge against immediate, large-scale data loss. In practice, redundancy can take many forms. Although local backups provide a convenient second source in cases of media or hardware failure, they are of limited value in cases of natural disaster, infrastructure failure, or any other widespread destruction. Awareness of the need for off-site storage (at a sufficient distance to preclude loss of primary and secondary copies in the same disaster) has noticeably increased in the aftermath of recent natural disasters (hurricanes, tsunamis, earthquakes) and political upheaval (Entlich 2005). An additional level of redundant security is the use of mirror sites, which not only hold an off-site copy of primary data (sometimes updated in real time) but also replicate the entire IT infrastructure so that they can substitute for the primary site should it become unavailable. Mirror sites are particularly important for those programs providing current access, since restoration of data from backup copies can be extremely time-consuming. Ontario Scholars Portal reported that it would take months to restore its terabyte-size primary online data store from backup tapes.

We asked each program about its use of local backups, off-site storage, and mirror sites, and about the total number of redundant copies of the journal data maintained (Table 12). Other than the LOCKSS Alliance, all programs currently maintain or shortly plan to implement both local backups and off-site storage. The preferred mechanism for backing up LOCKSS boxes is the LOCKSS system itself. LOCKSS boxes are designed to be "self-healing" and to detect and correct corruption on the basis of comparisons with and downloads from other LOCKSS boxes carrying the same content. However, for very large collections, rebuilding an entire LOCKSS box in that manner could be time-consuming and incur substantial network traffic charges. Nevertheless, even though it might be faster and cheaper in some cases to restore a LOCKSS box from a local, offline backup, most installations have opted to forgo their use. In fact, LOCKSS content licenses lack authorization to make such backups, so their

Redundancy Procedure	CSI	ECO	EJC	KB	KOP	LA	LANL	NLA	OSP	PMC	PORT
Local backups	•	•	•	•	P		•	•	•	•	•
Mirror sites	•	•		P						P	P
Off-site storage	•	•	•	P	•		•	•	•	•	•
Not sure											
Other						•					
Minimum number of copies (including master or production system)	2	4	3	2	4	See text	5	3	2	6	6

Table 12. Responses to questions: "Do you use any of the following redundancy procedures?" and "How many copies of your content do you maintain?" (• = yes; P = plan to within six months)

legality, at least under U.S. copyright law, is unclear. An alternative for institutions with very large storage caches would be to establish a second complete LOCKSS box within the same network domain.

Two initiatives—OCLC ECO and CISTI Csi—have established mirror sites. Portico, the KB e-Depot, and PubMed Central all have them in the planning stages. PubMed Central is in different stages of negotiation to establish mirrors in at least five countries; U.K. PubMed Central is expected to be the first to go live, possibly as early as January 2007 (UKPMC 2006). The concept of a mirror site has a different meaning in the context of LOCKSS; in a sense, all the content is mirrored, because every LOCKSS box has the complete LOCKSS software. Although no two LOCKSS boxes necessarily carry exactly the same content, any particular content should be available on a minimum number of other boxes.

There are not only different techniques for carrying out redundancy but also varying degrees of practice for each technique, as evidenced by differences in the number of redundant copies each program maintains. However, it is the operational details behind the numbers that determine the degree of protection provided. For example, a program that keeps five copies of only its data files, all on the same kind of media and in the same location, is more vulnerable to loss than is a program that maintains a single mirror site with both applications software and data that are in a geographically distinct location, on a different power grid, in a different network, and operated by different personnel. LOCKSS proponents claim that one strength of its architecture is that distinct systems personnel operate every site, increasing the protection of the content against loss by human error or deliberate attack from a determined insider. In fact, they assert that "unified system administration should be an unacceptable feature of digital preservation" (Rosenthal 2005b). We agree.

Different levels of redundancy may be appropriate for different types of archiving programs. Preservation-driven programs have less need for real-time mirroring, because they do not provide current access and typically do not promise immediate access to their subscribers or members in the case of a trigger event. Furthermore, the publisher can usually resupply content that has been processed,

Type of Threat	CSI	ECO	EJC	KB	KOP	LA	LANL	NLA	OSP	PMC	PORT
Malicious attack	•	•	•	•	•	•	•	•		•	•
Natural disaster	•	•	•	•	•	•		•		•	•
Infrastructure failure	•	•	•	•	•	•	•	•		•	•
Other						•					

Table 13. Responses to question: "Do you have written procedures and protocols to minimize vulnerability to various threats?"

but not yet backed up. However, over time, it can be expected that publisher failures, expiration of copyright, and other kinds of trigger events will eventually turn preservation-driven programs into content providers, thereby changing the nature of their responsibilities and, presumably, their redundancy planning.

Redundancy should be seen for what it is—a stopgap measure designed to restore data integrity or operations following a loss of primary systems. It is always preferable to prevent data loss in the first place. The need to rely on redundant storage, which can mean considerable expense and downtime, can be reduced through disaster planning. We asked each program whether it had established written procedures and protocols for dealing with three major classes of physical threats: malicious attacks, natural disasters, and infrastructure failure. As shown in Table 13, most programs have policies to address all three.

A written plan shows that a program takes its data-security obligations seriously. To be effective, disaster plans have to be comprehensive, detailed, widely disseminated to relevant personnel, and regularly tested and updated. Programs could enhance members' and subscribers' confidence in their preparedness for disasters by making disaster-planning documents public.[19] Public versions of these documents should be edited to exclude information that might compromise security, such as the precise location of off-site storage facilities, the identity of security personnel, and details about the operation of antihacking and anti-intrusion systems.

Offer an open, transparent means of auditing practices. This requirement addresses two questions: are practices audited and is the audit process open and transparent? At this early stage, there appears to be little agreement about the appropriate means and level of openness and transparency needed to gain the trust of potential participants. Our survey included a question about the conduct of technical audits. Seven programs indicated that they conduct technical audits (OhioLINK EJC, LANL-RL, LOCKSS, NLA PANDORA, Portico, OCLC ECO, CISTI Csi), two do not (Ontario Scholars Portal, kopal/DDB), and one (KB e-Depot) plans to conduct a technical audit within the next six months.

We also asked about the existence of written documentation covering many aspects of the programs' e-journal archiving functions.

[19] Some do so now, e.g., OhioLINK; see http://www.ohiolink.edu/ostaff/it/docs/DisasterPlan.doc.

There is as yet no standard expectation for a minimal set of documentation, and as Table 14 indicates, no one type of document that all programs have created. In most cases, only some of the documentation is publicly available.

Documentation Type	CSI	ECO	EJC	KB	KOP	LA	LANL	NLA	OSP	PMC	PORT
Mission statement			•	•	•					•	•
Publishers agreements	•		•	•	•	•	•	•		•	•
Membership agreements							•				•
Selection/Acquisition policies	•		•	P	•	•	•	•	•	•	•
Transfer requirements and deposit guidelines	•		•	•	•					•	
Ingest			•	•	•		•			•	•
Archival storage							•			•	•
Quality control							•			•	•
Auditing							•				•
Data management	•						•			•	•
Disaster planning/Recovery	•		•							•	P
Preservation planning					P						P
Metadata				•			•			•	•
Access and use policies							•	•			•
Financial reports											•
Annual reports				•							

Table 14. Responses to question: "Do you have the following written documentation that explicitly refers to e-journal archiving?" (• = yes; P = plan to within six months)

We believe that to earn the trust of the user community, archives must have written policies in all major areas of operations that are available for public review. Table 14 does not even address public availability, but it does point to an absence of written documentation in several critical areas, particularly quality control, disaster planning and recovery, and preservation planning.

During the thaw in relationships between the Soviet Union and the United States that took place in the 1980s, a number of Russian terms became well known to English speakers in the United States. These included *perestroika* (economic restructuring) and *glasnost* (openness), which referred to policy changes within the Soviet Union. On the U.S. side, the cautious response from then President Reagan often took the form of "*Doveryay, no proveryay*," usually translated as "Trust, but verify." That expression is especially appropriate for tentative relationships, where there is insufficient history and experience for trust to be automatic and unequivocal. Relationships between libraries and commercial publishers, in particular, have been strained, if not adversarial, for many years. Consequently, even with trusted nonprofit entities, including national libraries and university libraries playing a major role in facilitating e-journal archiving, there is much that libraries want to scrutinize and evaluate

Fig. 6. Examples of logos symbolizing compliance

before they can feel comfortable investing in a particular solution. Especially in these early stages, programs and initiatives should be prepared to demonstrate an extraordinary level of openness and transparency if they expect to gain the trust and support of the user community.

Recommendations

1. Publishers, research libraries, and archiving entities must all be involved in defining requirements and the processes associated with certification. Although it is important to consider what future requirements will be, it is equally important to do things now and to document what works and what does not.
2. Digital repositories should be overt about their ability to meet minimal requirements for well-managed collections and, ultimately, for certification. As the "Urgent Action" statement noted, "Certifying agencies might recognize qualified preservation archives that provide these services with a publicly visible symbol of compliance." Figure 6 shows examples of such symbols that are already in use: the NLA PANDORA's use of Safekept for materials on digital preservation that are preserved by Preserving Access to Digital Information (PADI), the National Archives of Australia's e-permanence program, and the server-certification program in Germany sponsored by DINI.
3. Research libraries should probe e-journal archiving programs for details on their ability to meet base-level requirements for responsible stewardship of journal content.
4. An anonymous reporting service should be established so that e-journal archiving programs and others in the community can share negative experiences with digital preservation procedures and tools without embarrassment or loss of credibility.
5. To achieve maximal feedback on the state of an archive's content, e-journal archiving programs should use a combination of automated integrity testing and active usage. Systems providing current access should be designed to encourage and facilitate reporting of data quality problems. Publishers should relax usage restrictions on dark archives to boost confidence that the content is "user ready" at all times.[20]

[20] Ken Orr proposes six data-quality "rules" of potential relevance to maintainers of and contributors to dark e-journal archives. Among these are (1) unused data cannot remain correct for very long; (2) data quality will, ultimately, be no better than its most stringent use; (3) data-quality problems tend to become worse as the system ages; and (4) laws of data quality apply equally to data and metadata (Orr 1998).

6. Programs should practice openness and transparency by making policy statements, model contracts, and technical procedure documentation publicly available.
7. E-journal archiving programs should begin examining interoperability issues for digital objects and metadata with an eye on maximinzing the ability to exchange data among them.
8. E-journal archiving programs should implement redundancy policies that maximize the survivability of data against the wide variety of potential threats. System administration responsibilities should be decentralized to reduce vulnerability to loss from a determined insider.

Indicator 5: Access Rights

A repository should negotiate with publishers to ensure that the digital archiving program has the right, and is expected, to make preserved information available to libraries under certain conditions.

The sine qua non of an effective e-journal digital archiving program is the ability to provide effective access to journals over time. If e-journals cannot be made available, there is little reason to preserve them. The conditions under which e-journal archiving programs can make preserved information available, and to whom, are two of the most important defining characteristics of the programs.

"Current Access" versus "Archiving"

One of the major distinctions in the surveyed initiatives is between those that provide immediate access to content, and promise to do so on a continuing basis, and programs whose primary responsibility is to ensure future availability of material, but which do not address current demand.

Tying digital preservation directly to current user access has pros and cons. On the plus side, it keeps preservation in the forefront. If a reader cannot currently access journals, either because of format changes or renderability problems, the provider will need to address the issue in relatively short order. Of the 12 initiatives we surveyed, 5 (CISTI Csi, OCLC ECO, LANL-RL, OhioLINK EJC, and the Ontario Scholars Portal) are focused primarily on making electronic journals available immediately to their authorized communities.

Two initiatives—PubMed Central and NLA PANDORA—offer online access to commercial publications after the expiration of a moving wall, normally six months to three years from date of publication.[21] In theory, one could substitute free access through PubMed Central or NLA PANDORA for a subscription, but in practice for most titles behind the moving wall, archival access is a supplement to, rather than a replacement for, current access from other sources.

---

[21] kopal/DDB hopes to negotiate with some publishers moving wall access to preserved content as well, but it cannot currently offer that service.

The drawback to programs that tie digital preservation to current user access is that they may be more motivated to perform functions supporting current, rather than future, access needs. One program providing immediate access commented on its use of standards and community practice: "As an access-oriented system, we struggled here. What we use is based on the current system for access. We would choose to use one or more [of these standards] if we were just archiving, or we may use them as we evolve to a new access system." Because proper preservation management embodies enough different and specialized responsibilities, the DLF *Minimum Criteria for an Archival Repository of Digital Scholarly Journals* document recommends against combining access and preservation in one system. Criterion six states that the limited-access services an archival repository provides "should not replace the normal operating services through which digital scholarly publications are typically made accessible to end users" (DLF 2000). Similarly, the authors of the "Urgent Call" statement suggested that digital archiving may best be viewed as a "kind of insurance" and not a form of access. They split archiving into two issues: mitigating risk of permanent loss and avoiding access disruptions for a protracted period.

The determination of whether a current e-journal access and delivery system can also effectively serve as an archival repository will ultimately rest upon a careful examination of all the program viability factors outlined in this report. Unlike the authors of the DLF *Minimum Criteria*, we do not reject out of hand the possibility that a program with a primary focus on current access could also serve as an archival repository.

"Dark Archive" versus "Light Archive"

A repository that preserves material for future use but does not provide current access is often referred to as a *dark archive* (Pearce-Moses 2005). In theory it might be possible to have a true dark archive that stores, maintains, and manages a sequence of bits without necessarily knowing what those bits contained. In reality, however, even the darkest of archives must permit some access by repository staff. The level of public access to the system can further distinguish dark archives. Some dark archives stress that they are dark because the system itself has no public interface and allows no public access. Only the person who deposits data into the dark archive can get it out, and it is the depositor's responsibility to provide access to the data. Other dark archives have public interfaces but allow no public access until a trigger event occurs. That trigger event could be negotiated with the content contributor (i.e., immediate onsite access to the files) or it could be related to an external event (such as the unavailability of the content owner's own Web site). People often refer to these archives as "dim," even "light," archives.

Librarians by and large have not been thrilled with the idea of pure dark archives. There are at least three reasons for this antipathy. The first is that for librarians, preservation and access have always intimately been linked. As Brian Lavoie and Lorcan Dempsey noted in

their 2004 article, "Thirteen Ways of Looking at . . . Digital Preservation":

> The notion of "dark archives," supporting little or no access to archived materials, has met with scant enthusiasm in the library community. This suggests that digital repositories will function not just as guarantors of the long-term viability of materials in their custody, but also as access gateways. Fulfilling this dual mission requires that preservation processes operate seamlessly alongside access services.

Don Waters made this same point in his paper "Good Archives Make Good Scholars: Reflections on Recent Steps Toward the Archiving of Digital Information":

> Access is the key. Over and over again, we have found that one special privilege that would likely induce investment in digital archiving would be for the archive to bundle specific and limited forms of access with its larger and primary responsibility for preservation (Waters 2002).

The second objection to dark archives concerns the funding mechanisms. As Sadie Honey (2005) noted:

> . . . the dark archive approach appears least likely to address long-term preservation needs. . . . The dark archive approach is weak in terms of equitable sharing of costs and long-term sustainability and does not score well against any of the criteria. The biggest obstacle for the dark archive approach is funding—who pays for it and how.

The third objection librarians have to dark archives is technical. It is far from certain that digital files stored in a system that is not accessible to the public can be safely managed. Don Waters, in the essay cited above, notes that, "User access in some form is needed in any case for an archive to certify that its content is viable." Harvard and others assert that they can safely audit and test a digital repository even when it is not open to public use, but this contention has not been proved. Cornell's experience with offline storage of digital masters has not been good and, in one case, a heroic rescue of digital files was necessary.

What librarians really want, in short, is at least a dim archive—though the level of dimness can vary. Fortunately, all the primarily preservation-oriented programs in our survey require staff access to content, with many assuming some level of public access. PubMed Central and NLA PANDORA, as noted above, are current publishers for some content and make other content available after a set period of time. The KB e-Depot and the kopal/DDB allow immediate onsite access to preserved content, with the possibility that online access can occur after certain trigger events. LOCKSS prefers that the publisher provide access to the reader, but when the publisher's copy is not available, the LOCKSS cached copy can be used for current access. To date, members of the LOCKSS Alliance have not experienced much need to initiate local access from their LOCKSS

boxes. Recently, however, when the journal *Communication Theory* moved from Oxford University Press to Blackwell Publishing, some LOCKSS Alliance libraries that do not subscribe through Blackwell began to provide local backfile access to their Oxford University Press content. As each institution's LOCKSS box serves only its own readers, the inexpensive machines used are more than adequate for a single institution's access load. Only Portico and CLOCKSS eschew some level of current access beyond audit, and both of them can become delivery mechanisms of choice under certain conditions. Portico plans to use the JSTOR access system to provide access in response to triggers or to secure perpetual access rights, if participating publishers choose to designate Portico as a provider of post-cancellation access. In addition, select librarians at participating libraries are granted password-controlled access for verification purposes.

Trigger Events

In a world of dim archives, the three key questions are who can have access to preserved content, how they can have access, and when they can have access. The conditions that can lead to a change in access to preserved content are usually called *trigger events* (Flecker 2001). A trigger event would occur when something goes wrong and a library could file a claim. We identified six trigger events that could change access conditions:

- a publisher ceases operation;
- a publisher no longer offers back issues;
- copyright in the journal expires;
- a journal ceases publication;
- the publisher or distributor experiences catastrophic system failure; or
- the publisher or distributor experiences temporary system failure.

Trigger events and the authorized community. We surveyed the archiving initiatives to see how a trigger event might change access for their authorized community. The results are presented in Table 15.

The programs that provide current access to content (OhioLINK EJC, LANL-RL, Ontario Scholars Portal, OCLC ECO, and CISTI Csi) would continue to provide such access even after a trigger event. As one of the providers noted, "Our partner model does not involve the idea of a 'trigger event.' Our repository is always available." Similarly, the moving-wall agreements that PubMed Central and NLA PANDORA have with publishers control access, regardless of trigger events. If either has received permission to make material available immediately or after a fixed period of time, that permission continues, regardless of the status of the publisher or the journal. LANL-RL is developing agreements with several scholarly societies, most notably the American Physical Society, to become a fallback provider if the primary servers fail completely.

Trigger events are more important for the other five repositories and can potentially alter the type and amount of access that each can provide. For example, if a publisher ceases operations, no longer of-

Trigger Event	CSI	ECO	EJC	KB	KOP	LA	LANL	NLA	OSP	PMC	PORT
Publisher ceases operation				•	•	•					•
Publisher no longer offers back issues				•	•	•					•
Copyright expires					•	•		•			
Journal ceases publication				•		•					•
Catastrophic failure				•		•	•				•
Temporary failure				•		•					
Other								•			•

Table 15. Trigger events that spark changes in access for the authorized community

fers access to back issues, ceases publication, or has a catastrophic failure of its delivery mechanism, LOCKSS and Portico would be able to make content available to authorized users. With LOCKSS, local access to the material preserved on a local LOCKSS box would be instantaneous, whereas with Portico it could take from 90 to 120 days to provide authorized user access to preserved material.[22]

In addition to the trigger events listed above, LOCKSS can provide access in the event of a temporary disruption in the publisher's distribution mechanism. Portico can in some cases provide ongoing access to subscribed content even after a library has terminated its license with the publisher. In these cases, the publisher will have decided that Portico, and not the publisher, will meet any perpetual access obligations of the original license.

Reactions to expiration of copyright as a trigger event were quite interesting. In theory, once copyright in a journal expires, the repository should be able to make it freely available to anyone. In practice, few repositories seem to have considered this possibility during their negotiations with publishers. If the negotiated agreements with the publishers limit access to a subset of users during the copyright term of the material, those restrictions would often still apply, even after the copyright has expired. As one interviewee somewhat sheepishly admitted, "Given the increasingly long duration of copyright terms, it is difficult to remember that copyright will eventually expire." Some of the initiatives (for example, PubMed Central, KB e-Depot, and kopal/DDB) are eager to make open-access material available to the world. Other initiatives appear to be concerned about the costs of giving nonmembers or nonsubscribers access to preserved open content. The benefit to society of providing ready access to public domain or otherwise open content can be great (Hamma 2005), and those programs providing current access to users should be urged to open access to the most material that the law, license agreements, and business plans allow.

[22] The other archiving initiatives (CLOCKSS, KB e-Depot, and kopal/DDB) would prefer to make content available to everyone after a trigger event, rather than manage authentication systems that control access to a select group of authorized users. These programs are discussed below.

Trigger events beyond the authorized community. The "Urgent Action" statement argued that access in response to a trigger event should be limited to designated member or subscriber communities. For those outside this group, access should come at a premium: "Potential participants who might choose initially to withhold support would pay their full fair share, should they eventually need access to preserved materials."

We therefore asked the e-journal archiving programs that restrict current or future access to a designated community whether, if one of the trigger events occurred, the repository would be able to provide access to those beyond their designated member or subscriber communities. Take, for example, an Elsevier journal that was no longer available electronically through the publisher. Would a library that subscribed to that journal and was not part of one of the archiving initiatives be able to turn to one of the e-journal archives to retain electronic access to the journal? And what about libraries that do not even have a current subscription? Would they ever be able to gain access to the preserved content?

Two of the initiatives—PubMed Central and NLA PANDORA—already make their content available to all after a publisher-specified waiting period. Of the remaining initiatives, only CLOCKSS said that it would be able to provide access to nonmembers in the event of a trigger event. A presumed trigger event would initiate collaboration among publishers, librarians, and representing societies to determine whether the trigger event had actually taken place and what the appropriate response should be: e.g., whether materials would be made generally available to all and whether such access would be for a limited or an indefinite period. Assuming general public access was authorized, the process of moving material from CLOCKSS's restricted storage environment into a public-access system would begin, and material would be available within six months.

The KB e-Depot, in principle, could also serve as a general delivery system for content in the event of a catastrophic collapse of the publisher's system, but some additional negotiations with publishers might be required, and the ramp-up time for the development of an online access system would likely be high, with no assurance that funding to develop such a system would be available. As yet, kopal/DDB has not negotiated the right to make material generally available after a catastrophic failure, though again this might be possible with the publishers' agreement and an appropriate ramp-up time.

Of the remaining seven initiatives, none opposed providing nonmembers access to preserved content at some time in the future, but all stressed that there would be myriad conditions and costs associated with doing so. As the respondent from the Ontario Scholars Portal noted, "Providing access outside the defined membership would be a problem financially and possibly ethically."

The reasons for the hesitation varied. In some cases, repositories did not know whether they would have the technical and financial resources necessary to make a general open portal to the preserved content. In other cases, agreements with publishers do not cover

such contingencies. In all cases, it was presumed that a nonmember would have to become a member to access the preserved content—presumably at a higher fee than if it had participated from the start. A library, for example, could join the LOCKSS Alliance, establish a LOCKSS box in the library, and then secure access to all content it had previously licensed or was freely available under a Creative Commons license. Alternatively, a library could join OCLC ECO or Portico to gain access to content to which it had once subscribed. The terms of the library's subscription and the archiving initiative's agreement with the publisher may limit what can be made available.

In short, it does not appear that there is a ready mechanism that can provide broad public access to currently access-restricted content should a triggering event occur. Subscribers to one of the current access services that also promise enduring access should be unaffected by any trigger event, assuming that the services can effectively preserve content. Participants in the LOCKSS Alliance and Portico should be able to "call in their insurance policy" and get ready access from these providers. The intention of CLOCKSS is to make its preserved content freely available to everyone in the event of a trigger event. The e-Depot at the KB and DDB's implementation of kopal would also like to provide worldwide, online access to content in the event of a publisher's failure, but for now the only certainty is that they will be able to continue to provide onsite access. Providers such as OCLC ECO and Portico may be willing to sign up new members when the need arises, but the costs are unclear.

The bottom line is this: unless electronic journals are available through the open-access portions of different repositories, the only certain method of access to preserved content for someone from outside a designated community is to fly to Amsterdam or Frankfurt to work with the preserved content onsite. The initiatives we examined have secured the necessary permissions to make material available to their designated community (e.g., subscribers, participants, onsite users). Few options, however, are available to users from outside the designated communities.

Recommendations

1. The only way a library can ensure that it will have continued access to subscribed (non-open access) content is through membership or participation in at least one of the e-journal archiving initiatives described in this report. This information should be conveyed to key library stakeholders to help them decide whether to support an e-journal archiving program at the local level.
2. National preservation projects should be encouraged to negotiate for broad access rights to copyrighted content in the event of a trigger event. Increased access may lead to increased preservation.
3. The preservation capabilities of any initiative whose primary purpose is the delivery of current journal literature should be carefully assessed. Access and preservation are not automatically at odds but focus on the former could be to the detriment of the latter.

4. All preservation initiatives should give more thought to the possibility that some of the content they store may eventually rise into the public domain and should negotiate all agreements with publishers accordingly.

Indicator 6: Organizational Viability

Repositories must be organizationally viable.
A digital preservation program exists within an organizational context and as such must fit the needs, priorities, and resources of the relevant stakeholders (e.g., publishers, the repository itself, members/subscribers/underwriters, users, and beneficiaries). *Trusted Digital Repositories: Attributes and Responsibilities*, produced by RLG and OCLC in 2002, defines the organizational context for a digital preservation program. Three attributes in particular relate to the viability of any e-journal archiving effort: administrative responsibility, organizational viability, and financial sustainability.

Administrative responsibility includes a commitment to implement community-agreed-upon standards and best practices, collect and share data measurements with depositors, regularly validate or certify processes and procedures, and maintain transparency and accountability in all actions. Organizational viability is reflected in a commitment to long-term retention and management in mission statements, legal status, business-practice transparency, staffing, the development and review of policies and procedures, testing, and contingency/escrow arrangements. Financial sustainability can be reflected in good business practices, business plans, annual reviews, standard accounting procedures, and short- and long-term financial-planning cycles.

What evidence exists that e-journal archiving programs are administratively responsible, organizationally viable, and financially sustainable? Our survey included questions on a range of issues, from organizational commitment, to documentation and standards adherence, to succession planning, to resources and cost models. The various programs' responses suggest that all have the potential for long-term viability. Each has an explicit mission committing it to long-term e-journal archiving and the legal right to do so. All have formal arrangements with publishers that spell out archiving and access requirements and show evidence of continued growth in publications covered. All are embedded in an organizational structure, and all except the government-supported programs have or plan to have a governance board that includes input from key stakeholders—libraries and publishers. Most make use of external advisers or are planning to do so within the next six months. All maintain Web sites and other publicity materials; many have contributed to the profession through participation in conferences, standards bodies, or digital preservation efforts, or through publication.

But these programs are of recent vintage and have limited track records in terms of digital preservation responsibility and practical experience. Except for the National Library of Australia, those with

a primary preservation focus are less than four years old; three have become operational since last year. Most are still building their digital preservation programs, and this is reflected in the fact that policies and practices are not as well documented as they might be. Well-defined service requirements are not fully met by all the repositories, and there appears to be little agreement regarding the appropriate means and level of openness and transparency needed to gain the trust of potential participants. Few have considered succession planning; none reported having a formal arrangement in place. That only half of them indicated a commitment to seek certification could also be a red flag for an institution that is relying on them for its preservation needs.

As shown in Table 16, only half of the programs reported that they have business and financial auditing processes in place or planned. However, the detailed comments accompanying these responses indicate that very few seem to conform to the standard set by the securities industry for a formal, externally conducted, and publicly released audit. Financial reports and publisher agreements, almost without exception, are not publicly available.

Economic issues related to digital preservation have been scrutinized in recent years, but the absence of any standard mechanism for accounting for all of the associated costs of e-journal archive management, and the early developmental stage of most of the programs, make meaningful comparisons of operating costs impossible—even if the programs surveyed had shared detailed budget documents with us. Perhaps the CRL report forthcoming by the end of 2006 will shed more light in this area.

We did look at two potential indicators of financial sustainability: sources of funding and stakeholder buy-in.

Sources of Funding

Programs with a government mandate may have an edge in terms of ongoing commitment and funding appropriations, although an exclusive dependence on government largesse could be detrimental in lean economic times. The KB, for example, has reallocated funding within its own budget to support e-Depot and since 2003, it has received an additional €1.1 million annually from the Ministry of Education, Culture, and Science for system maintenance and operations staff. In 2005, the ministry provided an additional €900,000 to be used exclusively in digital preservation research (Oltmans and van Wijngaarden 2006). Funding for PubMed Central is based on appropriations from the federal government for the NIH. In 2004, NLM's

Type of Process	CSI	ECO	EJC	KB	KOP	LA	LANL	NLA	OSP	PMC	PORT
Business	•		•	P		•					•
Financial	•		•	P		•					•

Table 16. Responses to question: "Do you have the following audit processes in place?" (• = yes; P = plan to within six months)

annual operating cost for PubMed Central was $2.3 million.[23] The *Bundesministerium für Bildung und Forschung* funded the three-year development of kopal/DDB with over €4 million in August 2004. To support the implementation of electronic legal deposit in Germany this year, kopal/DDB is getting a funding increase of about €2 million. Los Alamos National Laboratory receives appropriations from the U.S. Department of Energy, the U.S. Department of Defense, and elsewhere. The library receives funding from the institutional overhead in those appropriations or from grants and work for others that is done at the laboratory. The library charges external customers for access on a cost-recovery basis.

Programs with a primary mission to provide access may also be at a financial advantage, because the costs of archiving are tied directly to current use and subscriptions. Between 2001 and 2005, the Ontario Scholars Portal was supported by a grant and provincial matching funds as part of the Canadian National Site Licensing Program. The portal is now self-funded through a membership pricing model that adjusts for the varying size of consortium members and factors in usage, and includes tiered membership fees. Members have made a financial commitment through 2009–2010. OCLC ECO has been an online service provider for nearly 30 years and has the power of OCLC behind it. For OhioLINK EJC, all technical infrastructure costs, as well as about 20% of content-acquisition costs, are centrally funded though legislative appropriations. The remaining funding for content comes from member libraries, based on an institution's rate of expenditure on journals from publishers represented in EJC, including both print and electronic subscriptions. Most Ohio higher education institutions participate. Fluctuations in state appropriations, however, have resulted in discontinuation of some titles. EJC's contracts stipulate a nonpunitive approach to obtaining missing content if it resubscribes to a canceled title.

The three programs that are not funded by the government and are primarily intended for preservation may be the most vulnerable. All three have started within the past year or so; each has benefited from generous startup support from well-respected sources. The Andrew W. Mellon Foundation has supported both Portico and LOCKSS, and LC supports both Portico and CLOCKSS. In addition, LOCKSS received funding from the National Science Foundation, Sun Microsystems, and Stanford University libraries, and in-kind support from Sun, Intel Research Berkeley, HP Labs, and the computer science departments of Stanford and Harvard. Portico received heavy initial support from Ithaka and JSTOR, in addition to Mellon and LC.

[23] E-mail message from Ed Sequeira to Rich Entlich, April 14, 2006. "The last time we tallied the cost of PMC, in October 2004, we came up with an annual operating cost of $2.3 million."

Stakeholder Buy-in

Long-term sustainability for these efforts will depend on their ability to secure ongoing support from a number of quarters. The LOCKSS Alliance is an open-membership organization that began in 2005 to introduce governance for the program and to address sustainability issues. Its goal is self-sufficiency through membership fees, which are based on an institution's Carnegie Classification.[24] There is a 5% discount for consortia and library systems. Because some of the participating publishers make available for preservation only current content to current subscribers, the earlier a library joins the LOCKSS Alliance, the more complete its coverage is. Portico looks to a diversified revenue portfolio to fund ongoing operations, with major support coming from publishers and libraries. Publishers are asked to make annual contributions, which are tiered and vary according to the size of their annual revenue from journal subscriptions and advertising in addition to providing electronic journal source files. Libraries are asked to support the lion's share of expenses. Those that join pay an annual archive support payment, which is tiered according to a library's self-reported total library materials expenditure. Library systems and consortia are offered modest discounts. Published rates are available on the Portico Web site. To encourage early adoption, libraries that join in 2006 and 2007 will be designated "Portico Archive Founders." Those joining in 2006 receive a 25% savings in their payments for the next five years; those joining in 2007 will receive a 10% discount for the next five years.

CLOCKSS is in an initial two-year phase, and it is difficult to judge what will happen next. In the minds of many library directors, the e-journal–preservation issue comes down to two choices: LOCKSS Alliance or Portico. The long-term viability of these programs will be determined largely by how successful they are in signing up e-journal publishers as well as library members. The LOCKSS Alliance reported arrangements with more publishers than Portico, but Portico lists more titles covered. As of July 1, 2006, 13 publishers had committed more than 3,500 journals to Portico; 25 publishers had committed 1,500 titles to the LOCKSS Alliance.[25] Both continue to add new publishers and content.

More than 90 libraries worldwide joined the LOCKSS Alliance (157 institutions maintain LOCKSS boxes) in the first year it recruited members. In June 2006, the Alliance got a major boost when OCLC announced it had joined (OCLC 2006). According to the survey response from LOCKSS Alliance Director Vicky Reich, the LOCKSS Alliance "has reached an impressive level of sustainability." Eileen

---

[24] See http://www.lockss.org/lockysswiki/files/a/ad/AllianceInvoice.pdf. For a description of the Carnegie Classification system, see http://www.carnegiefoundation.org/classifications/. Equivalent measures are used for non–U.S. libraries.

[25] More publishers and titles are represented as being included in programs employing LOCKSS boxes, and the publishers' title listings on the Web site seem to be a work in progress. See http://www.lockss.org/lockss/Publishers_and_Titles.

Fenton, Portico's executive director, reported that as of July 1, 2006, 100 libraries had committed to supporting the archive. "Steadily growing participation from U.S. academic libraries and significant international expressions of interest suggest a broad base is building in support of Portico's efforts," she noted.

Both the LOCKSS Alliance and Portico have their supporters—and their detractors. Those who prefer to invest in an archiving solution by writing checks see Portico as the better choice and the annual fees a "bargain," especially given the early incentives and consortial discounts. The JSTOR imprimatur brings with it a sense of confidence in the approach. Some Portico supporters are also concerned by the technical requirements and staff time at the local level to participate in LOCKSS. Last February, the California Digital Library (CDL) estimated the impact of the Portico service on its systemwide e-journal preservation activities. They compared the journals then covered in Portico with CDL's 2005 journal packages, including nonprofit and for-profit publishers. The number of Tier 1 journals licensed was 4,593 for all 10 University of California (UC) campuses (9 campuses if the content is nonmedical and UC San Francisco is excluded). CDL negotiates the license, and all UC users have access to this material. It may be funded, in whole or in part, by CDL. CDL discovered that 45% of the journals were covered by Portico, representing 57% of the funds spent by CDL to license the journals.[26]

Those who favor the LOCKSS approach see it as the low-cost, technically proved, and organized way to go about archiving. "Any time someone asks us to write a check, we disappear," commented one director. They conceded that participating in the LOCKSS Alliance did require resources beyond the membership fee, but that the hardware and staff costs were negligible.[27] Others commented on the value of participating in collection development activities—choosing which publications to archive. They also valued the access to documentation, prerelease software, training, and involvement in planning efforts. Some expressed concern about the up-front efforts required by Portico to normalize data from the publishers, being one step removed from publishers by the participation of a third party, and the need to buy in before a full set of publishers was covered.

A few directors wondered whether the profession could financially sustain both the LOCKSS Alliance and Portico. Others valued the opportunity to participate in more than one program. As of July 1, 2006, 32 institutions had joined or were participating in both LOCKSS and Portico. Several members of OhioLINK EJC and the Ontario Scholars Portal are also participating in LOCKSS. Close to 300 institutions in the United States and Canada are covered by one or more e-journal archiving programs—a good beginning, but representing only a fraction of all higher education institutions in the country.

---

[26] E-mail, Patricia Cruse, Director, Digital Preservation Program, California Digital Library, to Anne R. Kenney, July 11, 2006.

[27] Libraries buying new hardware to support the LOCKSS box can be expected to spend approximately $1,000. Total staff costs, including technical support and collection development, average several hours per month.

Cornell University Library is participating in both Portico and the LOCKSS Alliance. Approximately 2,200 titles licensed by Cornell are covered in Portico (about 63% of Portico's total). As a LOCKSS Alliance member, Cornell's coverage includes 188 journals, 66 of which are also represented in Portico. Beyond the Alliance itself, Cornell subscribes to 618 titles from publishers in the LOCKSS program. Of these, 442 are also being archived through Portico.[28] It was surprisingly hard to determine the number of scholarly e-journals Cornell maintains that are not covered by these two options.[29] The cost to Cornell of participating in both Portico and the LOCKSS Alliance in 2006 is about $24,000, of which membership in the LOCKSS Alliance is $10,800 and participation in Portico is $13,125 (after the 25% early adopter discount). The LOCKSS box is running on a five-year-old Dell machine whose memory was upgraded twice, for a total of $125. The programmer responsible for managing the box estimates it took less than a day to set up the system and that he spends about 15 minutes a month to keep it running. With a three-year effort to move to electronic-only subscriptions in the sciences, social sciences, and the humanities, where possible, Cornell considers this money well spent, averaging approximately $10 per title and a little over one-tenth of 1% of total library materials expenditures. The money to support the memberships is coming from an account previously used for preservation microfilming.

Recommendations

1. Academic libraries should assess how much of their licensed content is protected in one of the e-journal archiving programs as a measure of the value of participation.
2. Academic libraries should share information with each other about what they are doing in terms of e-journal archiving, including their internal assessment process for decision making.
3. Mainstreaming commitment in terms of requisite resources and organizational support is essential. Participation in more than one program can ensure that different approaches and strategies are tried and assessed.
4. Academic libraries should press e-journal archiving programs for particulars on their business plans but not expect them to offer absolute guarantees of economic viability. Support should be viewed as an investment in developing viable models and an interim means for protecting vulnerable content.

---

[28] Information supplied by William Kara, e-resources and serials librarian, to Ellie Buckley, July 14, 2006.

[29] Cornell has about 42,000 unique bibliographic IDs for e-journals, so a little over 5% of the e-journal content Cornell makes available is covered in LOCKSS and Portico.

Indicator 7: Network

Repositories will work as part of a network.
The DLF *Minimum Criteria* lay out the advantages to creating a network: establishing a "satisfactory" degree of redundancy of their holdings; developing common finding aids, access mechanisms and registry services; and potentially reducing costs. In response to an evaluation by outside experts last year, the KB agreed that e-Depot should become part of a "larger international programme for preserving scientific literature." Yet what evidence exists that repositories are working toward this goal? Certainly they are holding common, often redundant, content and have common problems.

We asked the group whether they had any relationships with other archiving organizations in a number of categories. Table 17 summarizes their responses. Good collaboration is occurring in exchanging ideas and strategies (75%), sharing software (75%), and sharing planning documents (58%). The LANL-RL has shared its customized version of access software with both OhioLINK EJC and the Ontario Scholars Portal, and kopal/DDB and KB e-Depot are collaborating on the further implementation of IBM's DIAS software. Kopal is part of nestor, the alliance for Germany's digital memory; Portico and JSTOR have an agreement to use JSTORS's content-delivery infrastructure. The LOCKSS Alliance and CLOCKSS are using the same software. CISTI Csi and the Ontario Scholars Portal are having informal conversations on ways to collaborate. CISTI Csi has implemented business continuity facilities with Library and Archives Canada. OCLC ECO plans to work with OCLC's digital archives program in the future. And, as noted earlier, LC and the British Library intend to support the migration of electronic journal content to the NLM DTD standard.

Coordinating content selection and providing secondary archiving responsibilities is an under-represented form of collaboration. Only two repositories indicated that they coordinate content selection, but both are doing it in the context of their own consortial arrangements rather than with the other digital archiving programs.

Archiving Activity	CSI	ECO	EJC	KB	KOP	LA	LANL	NLA	OSP	PMC	PORT
Exchange ideas and strategies	P		•	•	•	•	•	•		•	•
Share planning documents				•	•	•	•	•			•
Share software			•	•	•	•	•	•		•	•
Coordinate content selection						•		•			
Reciprocal archiving/off-site storage/mirroring	•			P		•				•	P
Secondary archiving responsibility						•					P
Shared facilities/resources	•					•		•			•
Other	•										

Table 17. Responses to question: "Do you have any relationships with other archiving organizations involving the following activities?" (• = yes; P = plan to within six months)

Very few respondents have or are even thinking about succession plans or dependencies, as indicated by Tables 18 and 19, and only Portico has the contractual rights to pass on content and rights to another nonprofit organization. What may be more disturbing is that some may not even see the need to consider this option. One respondent wrote, "As a national library, we do not envisage that we would not continue." Another responded, "As a legal deposit repository, the need for succession is unlikely (if not unthinkable)." Although several respondents expressed a willingness to consider serving as a successor archive if another archive failed, in reality little formal commitment has occurred.

	CSI	ECO	EJC	KB	KOP	LA	LANL	NLA	OSP	PMC	PORT
Yes						•					•
No	•	•	•	•	•		•	•	•	•	

Table 18. Responses to question: "Do you have a succession plan in the event you are not able to continue your program?" (• = yes)

	CSI	ECO	EJC	KB	KOP	LA	LANL	NLA	OSP	PMC	PORT
Yes	•	•		•		•	•				•
No											
Not sure			•		•			•	•	•	

Table 19. Responses to question: "Do you or would you be willing/able to serve as a successor archive if another archive failed?" (• = yes)

Recommendations

1. Agree on the need for common rights to protect digital content and facilitate collaboration.
2. Investigate models for collaborative digital preservation action, such as Data-PASS (Data Preservation Alliance for the Social Sciences), a broad-based partnership of leading data repositories in the United States, to ensure the preservation of materials within and beyond current repository holdings. Supported by an award from LC through its National Digital Information Infrastructure and Preservation Program, Data-PASS is working in such areas as selection, appraisal, acquisition, and metadata and has developed the concept of partner-to-partner protocols for conveying content if an archive fails.
3. Fund a meeting of these programs' principals to identify areas of collaboration.

Getting and Keeping Informed

At a time when there is a great deal of activity related to e-journal archiving, there is unfortunately no comprehensive clearinghouse or gateway to all the relevant developments. The sources listed here cover at least a portion of the landscape.

Bibliographies

- PADI electronic journals topic, updated periodically: http://www.nla.gov.au/padi/topics/473.html
- Charles Bailey's Scholarly Electronic Publishing Bibliography "Library Issues: Information Integrity and Preservation," updated periodically: http://epress.lib.uh.edu/sepb/lbinteg.htm
- Drexel University Library E-Journal Archiving Issues, historical: http://www.library.drexel.edu/about/imls/preservation.html

Discussion Forums

- ARL-EJOURNAL: http://www.cni.org/hforums/arl-ejournal/about.html
- Yale LIBLICENSE-L: http://www.library.yale.edu/~llicense/mailing-list.shtml

Blogs

- Ten Thousand Year Blog, updated regularly: http://www.davidmattison.ca/wordpress/
- Peter Suber's Open Access News, updated regularly: http://www.earlham.edu/~peters/fos/fosblog.html
- Charles Bailey's Scholarly Electronic Publishing Weblog, updated once or twice per month: http://epress.lib.uh.edu/sepb/sepw.htm

What's New and News Listings

- Digital Preservation Coalition's What's New, updated 2–3 times per year: http://www.dpconline.org/graphics/whatsnew/
- Nestor's What's New: http://nestor.sub.uni-goettingen.de/aktuell/index.php?show=archiv&lang=en

Online Journals and Newsletters

- D-Lib, monthly: http://www.dlib.org
- RLG DigiNews, bimonthly: http://www.rlg.org/en/page.php?Page_ID=12081
- Ariadne, quarterly: http://www.ariadne.ac.uk/
- SPARC Open Access Newsletter, monthly: http://www.earlham.edu/~peters/fos/newsletter/archive.htm

Web Sites

- The Academic Publishing Industry: A Story of Merger and Acquisition, by Mary H. Munroe, last updated 9/30/2005: http://www.niulib.niu.edu/publishers/
- Washington DC Principles for Free Access to Science: http://www.dcprinciples.org/
- LIBLICENSE: Licensing Digital Information, A Resource for Librarians: http://www.library.yale.edu/~llicense/index.shtml
- University of California, Davis, University Library Scholarly Electronic Publishing Initiatives, historical: http://www.lib.ucdavis.edu/ul/about/sepi/index.php
- Electronic Journals: A Selected Resource Guide from Harrassowitz, historical: http://www.harrassowitz.de/top_resources/ejresguide.html

Promising E-Journal Archiving Programs Not Included in This Report

The 12 programs discussed in this report were selected on the basis of criteria presented earlier. One of those criteria was that the program had to already be archiving content. In the course of our research, we encountered references to additional programs that are still being planned or tested, or that have not yet devised a preservation strategy. Some of these programs are noteworthy because they will be archiving content that is not included in any of the 12 programs reviewed in this report, particularly e-journals using non-Roman alphabets. National libraries, through their legal deposit frameworks, are coordinating almost all this activity.

British Library (BL)

Subsequent to the passage of new legal deposit legislation in 2003, the British Library had been working with the Joint Committee on Legal Deposit to establish guidelines and procedures for deposit of materials not authorized for legal deposit in prior legislation (The British Library n.d.). To facilitate this work, three subcommittees were formed, including one to address issues relating to deposit of e-journals. The e-journals subcommittee has formed a working group that is conducting a pilot deposit project at the BL with more than 20 commercial, university, society, and small presses participating, representing more than 200 titles (Joint Committee on Legal Deposit 2004). The working group's first report, issued in June 2005, emphasizes technical issues, especially file formats and metadata (Inger 2005).

Det Kongelige Bibliotek (The Royal Library, Denmark)

Legal deposit legislation in Denmark that went into effect July 1, 2005, includes a new section that covers "materials made public via electronic communication network." It permits harvesting of public content on Danish Internet domains, as well as of materials intended for a Danish audience but made public on non-Danish Internet domains. A repository with preservation and access functions is being designed with the Royal Library's partner, the *Statsbiblioteket* (State and University Library), and the two locations will provide reciprocal backup capability. Danish law allows online access to content provided under legal deposit only for material that is not commercially available and, even then, only to meet strictly defined research needs. Most e-journals will be available only onsite at the Royal Library.

Library and Archives Canada (LAC)

The bulk of scholarly journal publishing in Canada is from university presses, trade associations, and individual academic departments. The National Research Council Research Press is the largest publisher of electronic journals in Canada, with 15 titles. Other e-journal publishers of note are the University of Toronto Press and the Canadian Medical Association (McDonald and Shearer 2006).

The most recent change to Canada's legal deposit laws, passed in 2004, includes a mandate for deposit of electronic publications that goes into effect in January 2007. According to its 2005–2006 Report on Plans and Priorities (Frulla n.d.), LAC is planning to develop a system to "facilitate the acquisition, management, preservation and accessibility" of Canadian digital content, in concert with the new legal deposit requirements.

National Diet Library (Japan)

Though amended in 2000 to include CD-ROM and other packaged digital publications, Japan's legal deposit legislation still does not cover online publications. Research preparatory to further amendments governing online publications has been under way at the National Diet Library, and revised legislation is expected soon. As part of its Digital Library Medium-Term Plan for 2004 (Mutoh 2005), NDL is conducting a digital library initiative that includes among its objectives the construction of a digital repository, Web archiving, and digital deposit for e-journals.

Since 2002, NDL has been pursuing an experiment called the "Web Archiving Project" (WARP), to preserve Japanese Web sites, including digital editions of periodicals on the Internet and born digital periodicals (NDL n.d.). By 2004, WARP had made available 1,496 e-journals harvested from the Japanese Web, although it is unknown how many of these are scholarly (Mutoh 2005). Mechanisms for long-term preservation are being discussed.

National Library of China (NLC)

The NLC is developing a digital repository that includes both access and long-term preservation as part of its mission. NLC recognizes the importance of e-journals and is working on a strategy for their preservation, with an emphasis on STM titles (Zhang, Zhang, and Wan 2005). The current NLC digital collection includes e-journals in Chinese and in Western languages. In May 2005, NLC launched a portal to its digital collections, including 16,000 periodicals in Chinese and other languages. Because of copyright restrictions, the portal is available only within the NLC building. It is not clear how many of the 16,000 periodicals are scholarly titles. Preservation activities are still in the planning stages.

Others

A recent report by the International Federation of Library Associations and Institutions describes the digital preservation activities and plans of 15 national libraries (Verheul 2006). Besides those mentioned above, several others are working on repositories that are expected to incorporate e-journals and will merit attention over the next few years.

CONCLUSION

> 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.
>
> *Urgent Call to Action*

Academic libraries have been slow to respond to the vulnerability of e-journal literature, because competing demands have taken precedence, because they have not fully embraced collective and shared responsibility for the safety of digital content, and because few options presented themselves. The landscape is changing and, as this report shows, several viable choices for exercising good digital stewardship for e-journals are emerging. Are these perfect solutions? No. Do they address preservation needs? Sort of. Do they adequately cover the domain of peer-reviewed e-journal literature? Somewhat. Are they worthy of support? Yes. Could they benefit from academic library input? Absolutely.

As we consider recommendations for the future, let us start with some givens:

1. It is a matter of when, not whether, e-journal publishing programs will suffer significant trigger events that put at risk ongoing access to vital scholarly resources.
2. Academic libraries cannot address all e-journal archiving needs at the local level. The requisite resources are simply not there.
3. Current guarantees included in e-journal licenses are inadequate: a perpetual-access clause does not equate to digital preservation, and the requirement to receive copies of the digital files on disk or tape is tantamount to buying pork bellies short on the commodities market and having them delivered to one's front door.
4. For the first time, viable options are emerging that address academic library needs and interests.
5. No single program can assume full responsibility for all e-journal preservation. Multiple programs are necessary, but they should cooperate with each other as part of a larger network.
6. Academic libraries have an opportunity to influence how these programs operate and whether they will succeed.
7. Academic libraries that do not support e-journal archiving programs in the near future risk incurring costly and delayed access to essential resources. E-journal archiving is not just a problem for large libraries
8. Current laws are inadequate to support digital archiving. Each country should enact legal deposit laws to provide a much-needed national safety net.

9. Coverage of scholarly literature is uneven across disciplines. STM journals are more heavily represented than are those in the humanities and social sciences; large commercial publishers are well represented; smaller, independent publishers are not.
10. Publishers and e-journal archiving programs alike need greater transparency of support, coverage, technical approaches, business practice, and contractual relations.

Our scan of the landscape highlights the need for action to address e-journal archiving challenges by three key players—publishers, archiving entities, and libraries. Looking ahead, what would progress look like? Publisher Web sites and other communication vehicles would highlight, even tout, their archiving arrangements, partners, and developments. Publishers would provide specific, comprehensive, and current information about archival strategies that is targeted at stakeholders beyond the library community and compliant with archival trends. Archiving would be a central and visible component of their digital asset management. The strategies and practices of e-journal archiving programs would be well known through publicly available and comprehensive documentation; the extent of their holdings in terms of publishers, titles, content included, and date spans would be current and readily accessible. A core group of archiving programs would be routinely audited and certified as adhering to prevailing standards and practice and would provide digital preservation models. Archiving programs would share information and collaborate to ensure that the main goal for preserving e-journal content and its scholarly successors is achieved. Libraries of all sizes and types would include explicit references in their mission statements to their ongoing investment and participation in e-journal archiving initiatives that both contribute to archiving programs and target specific categories of at-risk content. The extent and progress of e-journal archiving participation would be mainstreamed and would be a measure of success for libraries. In our ideal future scenario, key players would work together to codify standards and practice governing e-journal archiving. We have in mind something similar to COUNTER, a collaborative effort of publishers, libraries, consortia, intermediaries, and industry to measure the use of online resources through an agreed-on set of international standards and protocols governing the recording and exchange of online usage data. The COUNTER Codes of Practice provide these standards and protocols and are published in full on its Web site, as is a list of compliant vendors.

Recommendations: Academic Libraries and Organizations

1. Libraries and consortia should press publishers hard to enter into e-journal archiving relationships with bona fide programs and to convey all necessary rights and responsibilities for digital archiving to them as part of their license negotiations. There should be community agreement that the same rights are conveyed in all archiving arrangements. Research libraries should collectively agree not to sign new licenses or renew old ones for access to electronic journals unless these conditions are met.

2. Libraries should share information with each other about what they are doing in e-journal archiving, including their internal assessment process for decision making.

3. Institutions should become members of or participate in at least one e-journal archiving initiative; it is the only way a library can ensure it will have continued access to journal content. The institution must be prepared to commit the resources and organizational support needed. Participation in more than one program can ensure that different approaches and strategies are tried and assessed. A broad range of academic and research libraries should be encouraged to affiliate with appropriate e-journal archiving programs.

4. Academic libraries of all sizes should act collectively to press for digital archiving programs that meet their needs. As a condition of support, they should request details on the program's ability to meet base-level requirements for responsible stewardship of journal content and, ultimately, some form of accreditation. A first step would be to require each program to complete the audit checklist being developed by RLG and NARA, and to report the results. An archival program should also be able to provide a definitive list of titles and date spans covered, the level of content completeness, a description of institutional obligations, and a list of prevailing standards and best practices used to protect materials; it should specify the circumstances under which access to content is provided, and the timing of such access. Any initiative whose primary purpose is to deliver current journal literature should be carefully assessed for its preservation capabilities. Those that focus mainly on preservation should be examined for their ability to provide access in a timely and cost-effective manner following a trigger event. Access and preservation are not automatically at odds—but there is the danger that focusing on one could be to the detriment of the other.

5. Much of the e-journal literature remains outside the protection of the archiving programs. Libraries should participate in developing a registry of archived scholarly publications that indicates which programs have preserved them, following such models as

the Registry of Open Access Repositories (ROAR), which lists 667 open-access e-print archives around the world, and ROARMAP, which tracks the growth of institutional self-archiving policies. This registry could then be used to identify gaps in publisher or content coverage.

6. Libraries should lobby e-journal archiving programs to participate in a network that shares information, codifies best practices, and promotes sufficient redundancy but also shares responsibility for preserving peer-reviewed e-journals that are not currently included.

Recommendations: Publishers

1. Publishers should be overt about their digital archiving efforts and enter into archiving relationships with one or more e-journal archiving programs of the sort described in this report or their equivalents. Smaller presses appear to be at most risk.

2. Publishers should provide enough information to e-journal archiving programs to ensure that the scope, content, date span, and title coverage are adequately recorded.

3. Publishers should extend liberal archiving rights in their licensing agreements with content aggregators and consortia. Digital archiving of e-journals should be a distributed responsibility.

Recommendations: E-Journal Archiving Programs

1. Archiving programs should present compelling public evidence that they offer at least the minimal level of services for well-managed collections. They should be open to audit, and when certification of trusted digital repositories is available, they should be certified.

2. Archiving programs should be overt about the publishers, titles, date spans, and content coverage included in their programs. They should make this information easily accessible on their Web sites.

3. Archiving programs should ensure that once content is ingested it becomes the repository's property and cannot be removed or modified by a publisher or its successor. If there is an alleged breach of contract, there should be a process for dispute mediation to protect the longevity and integrity of the e-journal content.

4. A study should be conducted to examine rights and responsibilities necessary to ensure adequate protection for digital archiving actions so that these rights are accurately reflected in contracts. Archiving programs should periodically review contracts, be-

cause changes in publishers, acquisitions, mergers, content creation and dissemination, and technology can affect archiving rights and responsibilities. Continuity of preservation responsibility is essential.

5. Archiving programs should consider that some content they store might eventually enter the public domain and negotiate all agreements with publishers to take this possibility into account.

6. Archiving programs should form a network of support and mutual dependence to exchange information on content coverage, technical implementations, and best practices; to obtain the necessary contractual rights to preserve and eventually provide access to content; to create a safety net for one another for succession planning and secondary archival functions; and to share responsibility for identifying and preserving peer-reviewed e-journals that are not currently protected. As a first step, we recommend funding a meeting of the principals of these programs to identify areas of collaboration.

References

All URLs were valid as of September 15, 2006

AAP, AMPA, DCPC (American Association of Publishers, American Medical Publishers Association, DC Principles Coalition). "Open Letter to NIH Director Zerhouni, August 23, 2004." Available at http://www.dcprinciples.org/zerhouni.pdf.

ALPSP (Association of Learned and Professional Society Publishers). 2002. When a Society Journal Changes Publisher: ALPSP Guidelines for Good Practice (November). Available at http://www.alpsp.org/socjourn1.pdf.

BBSRC (Biotechnology and Biological Sciences Research Council). 2006. "BBSRC's Position on Deposit of Publications," news release, June 28, 2006. Available at http://www.bbsrc.ac.uk/news/articles/28_june_research_access.html.

Bekaert, Jeroen, and Herbert Van de Sompel. 2006. "Augmenting Interoperability across Scholarly Repositories." Paper presented at a meeting sponsored and supported by Microsoft, The Andrew W. Mellon Foundation, the Coalition for Networked Information, the Digital Library Federation, and the Joint Information Systems Committee, New York, N.Y., April 20–21, 2006. Available at http://msc.mellon.org/Meetings/Interop/FinalReport.

Bekaert, Jeroen, Xiaoming Liu, and Herbert Van de Sompel. 2005. "Using Standards in Digital Library Design & Development." Tutorial presented at the Joint Conference on Digital Libraries, Denver, Colo., June 7, 2005. Available at http://library.lanl.gov/cgi-bin/getfile?LA-UR-05-5219.pdf.

Bell, Jonathan, and Stuart Lewis. 2006. Using OAI-PMH and METS for Exporting Metadata and Digital Objects between Repositories. Available at http://cadair.aber.ac.uk/dspace/bitstream/2160/203/1/Using+OAI-PMH+and+METS.pdf.

Besek, June M. 2003. *Copyright Issues Relevant to the Creation of a Digital Archive: A Preliminary Assessment*. Washington, D.C.: Council on Library and Information Resources. Available at http://www.clir.org/pubs/reports/pub112/contents.html.

Bowen, William G. 2005. New Times Always; Old Time We Cannot Keep. *ARL Bimonthly Report* 243 (December). Available at http://www.arl.org/newsltr/243/bowen.html.

The British Library. [n.d.]. Legal Deposit in the British Library, Interim Arrangements for the Legal Deposit of Electronic Publications. Available at http://www.bl.uk/about/policies/legaldeposit.html#elec.

Bruce, Cary. 2005. "Going from Print to Online and the Role of Agents. EBSCO Information Services." Paper presented at Informatio Medicata, Semmelweiss University, Budapest, November 9–10, 2005. Available at http://www.lib.sote.hu/files/moksz/2005/going_from_print_to_online_and_the_role_of_agents.ppt.

California Digital Library. 2006. Technical Requirements for E-Journal Vendors, May 16, 2003; latest revision, April 21, 2006. Available at http://www.cdlib.org/vendors/CDL_ejournal_Vendor_Req.rtf.

Case, Mary M. 2004. A Snapshot in Time: ARL Libraries and Electronic Journal Resources. *ARL Bimonthly Report* 235. Available at http://www.arl.org/newsltr/235/snapshot.html.

Clark, Frank. 1922. *A Treatise on the Law of Surveying and Boundaries*. Indianapolis, Ind.: Bobbs-Merrill.

CENL/FEP (Conference of European National Librarians/Federation of European Publishers). 2005. Statement on the Development and Establishment of Voluntary Deposit Schemes for Electronic Publications. Approved by the Annual Conference of European National Librarians, Luxembourg, September 28–30, 2005. Available at http://www.nlib.ee/cenl/docs/05-11CENLFEP_Draft_Statement050822_02.pdf.

DHHS (Department of Health and Human Services), National Institutes of Health. 2006. Report on the NIH Public Access Policy. Available at http://publicaccess.nih.gov/Final_Report_20060201.pdf.

DLF (Digital Library Federation). 2000. Minimum Criteria for an Archival Repository of Digital Scholarly Journals (version 1.2). Available at http://www.diglib.org/preserve/criteria.htm.

Dobratz, Susanne, Astrid Schoger, and Stefan Strathmann. 2006. "The nestor Catalogue of Criteria for Trusted Digital Repository Evaluation and Certification." Paper presented at the Joint Conference on Digital Libraries (JCDL) Workshop: Digital Curation and Trusted Repositories: Seeking Success, University of North Carolina at Chapel Hill, June 15, 2006. Forthcoming in JCDL 2006 Proceedings. Available at http://sils.unc.edu/events/2006jcdl/digitalcuration/Dobratz-JCDLWorkshop2006.pdf.

Dobratz, Susanne, and Astrid Schoger. 2005. Digital Repository Certification: A Report from Germany. *RLG DigiNews* 9(5). Available at http://www.rlg.org/en/page.php?Page_ID=20793&Printable=1&Article_ID=1779.

Entlich, Richard. 2005. Too Close for Comfort? The Case for Off-site Storage. *RLG DigiNews* 9(6). Available at http://www.rlg.org/en/page.php?Page_ID=20865#article4.

Flecker, Dale. 2001. Preserving Scholarly E-Journals. *D-Lib Magazine* 7(9). Available at http://www.dlib.org/dlib/september01/flecker/09flecker.html.

Frulla, Liza. [n.d.] Library and Archives Canada 2005–2006 Report on Plans and Priorities. Available at http://www.tbs-sct.gc.ca/est-pre/20052006/LAC-BAC/pdf/LAC-BAC_e.pdf.

Guthrie, Kevin, and Roger C. Schonfeld. 2004. "What Do Faculty Think of Electronic Resources? Findings from the 2003 Academic Research Resources Study." Paper presented at Coalition for Networked Information (CNI) Task Force Meeting, Alexandria, Va., April 16, 2004.

Hahn, Karla. 2006. The State of the Large Publisher Bundle: Findings from an ARL Member Survey. *ARL Bimonthly Report* 245: 5. Available at http://www.arl.org/newsltr/245/bundle.html.

Hamma, Kenneth. 2005. Public Domain Art in an Age of Easier Mechanical Reproducibility. *D-Lib Magazine* 11(11). Available at http://www.dlib.org/dlib/november05/hamma/11hamma.html.

Harter, Stephen P., and Hak Joon Kim. 1996. Electronic Journals and Scholarly Communication: A Citation and Reference Study. Paper presented at the Midyear Meeting of the American Society for Information Science, San Diego, Calif., May 20-22, 1996. Published in *Journal of Electronic Publishing* 3(2) [December 1997]. Available at http://www.press.umich.edu/jep/archive/harter.html.

Hedstrom, Margaret, and Clifford Lampe. 2001. Emulation vs. Migration: Do Users Care? *RLG DigiNews* 5(6). Available at http://www.rlg.org/preserv/diginews/diginews5-6.html#feature1.

Hirtle, Peter B. 2001. OAI and OAIS: What's in a Name? *D-Lib Magazine* 7(4). http://www.dlib.org/dlib/april01/04editorial.html.

Honey, Sadie. 2005. Preservation of Electronic Scholarly Publishing: An Analysis of Three Approaches. *portal: Libraries and the Academy* 5(1): 59-75.

Hughes, Janet A. 2002. Issues and Concerns with the Archiving of Electronic Journals. *Science & Technology Libraries* 22(3/4): 113-136.

Hunter, Karen. 2006. "The End of Print Journals: (In)Frequently Asked Questions." Paper presented at the University of Oklahoma Libraries' Conference on Printed Resources and Digital Information: The Future of Coexistence, Oklahoma City, Okla., March 2–3, 2006. Forthcoming in *Journal of Library Administration*.

Inger, Simon. 2005. JCLD Pilot Project in Anticipation of E-Journals. June 2005 Project Report. Available at http://www.bl.uk/about/policies/pdf/jcldejpjunereport.pdf.

Joint Committee on Legal Deposit Working Group on E-Journals. 2004 (June). Terms of Reference. Available at http://www.bl.uk/about/policies/ejournals.html.

Kenney, Anne R., and Deirdre C. Stam. 2002. *The State of Preservation Programs in American College and Research Libraries: Building a Common Understanding and Action Agenda*. Washington, D.C.: Council on Library and Information Resources. Available at http://www.clir.org/pubs/reports/pub111/pub111.pdf.

Kiernan, Vincent. 2006. Libraries and Publishers Create Archive for Backup Copies of Electronic Journals. *Chronicle of Higher Education* 52(22): A37. Available at http://chronicle.com/weekly/v52/i22/22a03703.htm.

Kyrillidou, Martha, and Mark Young. 2005. ARL Library Trends. Washington, D.C.: Association of Research Libraries. Available at http://www.arl.org/stats/arlstat/04pub/04intro.html.

Lavoie, Brian, and Lorcan Dempsey. 2004. Thirteen Ways of Looking at . . . Digital Preservation. *D-Lib Magazine* 10(7-8). Available at http://www.dlib.org/dlib/july04/lavoie/07lavoie.html.

Library of Congress. 2006. "British Library to Support Common Archiving Standard for Electronic Journals," news release, April 19, 2006. Available at http://www.loc.gov/today/pr/2006/06-097.html.

Maniatis, Petros, Mema Roussopoulos, T. J. Giuli, David S. H. Rosenthal, Mary Baker, and Yanto Muliadi. 2003. "Preserving Peer Replicas By Rate-Limited Sampled Voting." Paper presented at the 19th ACM Symposium on Operating Systems Principles, Bolton Landing, N.Y., October 19–22, 2003. Available at http://www.eecs.harvard.edu/~mema/publications/SOSP2003-long.pdf.

Marcum, Deanna B., and Gerald George. 2003. Who Uses What? Report on a National Survey of Information Users in Colleges and Universities. *D-Lib Magazine* 9(10). Available at http://www.dlib.org/dlib/october03/george/10george.html.

Marty, Paul F. 2005. Factors Influencing Error Recovery in Collections Databases: A Museum Case Study. *Library Quarterly* 75(3): 295–328.

Marty, Paul F., and Michael B. Twidale. 2000. "Unexpected Help with Your Web-Based Collections: Encouraging Data Quality Feedback from Your Online Visitors." Paper presented at Museums and the Web 2000, Minneapolis, Minn., April 16–19, 2000. Available at http://www.archimuse.com/mw2000/papers/marty/marty.html.

McDonald, John, and Kathleen Shearer. 2006. Toward a Canadian Digital Information Strategy: Mapping the Current Situation in Canada, Version 2.0 (January 31, 2006). Available at http://www.collectionscanada.ca/obj/012033/f2/012033-700-e.pdf.

McKay, Sharon Cline. 1996. "Kickoff Discussion: Archiving." Message to ARL E-JOURNAL Listserv, November 19, 1996. Available at https://mx2.arl.org/Lists/ARL-EJOURNAL/Message/2331.html.

Mutoh, Toshiyuki. 2005. National Diet Library Digital Library Medium-Term Plan for 2004. Available at http://iadlc.nul.nagoya-u.ac.jp/archives/IADLC2005/mutoh.pdf. (Note: Pages 1, 5, and 8 will appear blank on systems lacking Japanese-language support.)

NDL (National Diet Library). n.d. The Internet—Intelligence for the Next Generation. Available at http://web.archive.org/web/20050331202749/http://warp.ndl.go.jp/pamph-e.pdf.

NERL (NorthEast Research Libraries Consortium) Generic License Agreement. January 4, 2006 (revised). Available at http://www.library.yale.edu/NERLpublic/NERLGenericLicRev010406.pdf.

OCLC. 2006. "OCLC Joins LOCKSS Alliance," news release, June 22, 2006. Available at http://www.oclc.org/news/releases/200629.htm.

OhioLINK Governing Board. 2006. "OhioLINK Library Community Recommendations on Retention of Intellectual Property Rights For Works Produced by Ohio Faculty and Students." May 2006, revised August 18, 2006. Available at http://www.ohiolink.edu/journalcrisis/intellproprecsaug06.pdf.

Olsen, Jan. 1994. *Electronic Journal Literature: Implications for Scholars*. Westport, Conn.: Mecklermedia.

Oltmans, Erik, and Hilde van Wijngaarden. 2006 (forthcoming). The KB e-Depot Digital Archiving Policy. *Library Hi Tech* 34(4) [December].

Oltmans, Erik, and Nanda Kol. 2005. A Comparison Between Migration and Emulation in Terms of Costs. *RLG DigiNews* 9(2). Available at http://www.rlg.org/en/page.php?Page_ID=20571#article0.

Orr, Ken. 1998. Data Quality and Systems. *Communications of the ACM* 41(2): 66–71.

Pearce-Moses, Richard. 2005. *A Glossary of Archival and Records Terminology*. Chicago, Ill.: Society of American Archivists. Available at http://www.archivists.org/glossary/term_details.asp?DefinitionKey=231.

Powell, David J. 2004. Publishing Output to 2020. In *The Future of Print and Electronic Publishing Output Worldwide*, a report by Electronic Publishing Services Ltd. to the British Library, January 29, 2004. Extract available at http://www.bl.uk/about/articles/pdf/epsreport.pdf.

PRONOM. 2006. PRONOM Future: Preservation Planning. Available at http://www.nationalarchives.gov.uk/aboutapps/pronom/default.htm.

Publishers Communication Group, Inc. 2004. Global Electronic Collection Trends in Academic Libraries: 2004. Available at http://www.pcgplus.com/resnresource/research/GlobalElTr.pdf.

RLG. 2005. *An Audit Checklist for the Certification of Trusted Digital Repositories, Draft for Public Comment, August 2005*. Available at http://www.rlg.org/en/pdfs/rlgnara-repositorieschecklist.pdf.

RLG. 2002. *Trusted Digital Repositories: Attributes and Responsibilities, An RLG-OCLC Report*. Available at http://www.rlg.org/longterm/repositories.pdf.

Rosenthal, David S. H., Thomas Lipkis, Thomas S. Robertson, and Seth Morabito. 2005a. Transparent Format Migration of Preserved Web Content. *D-Lib Magazine* 11(1). Available at http://www.dlib.org/dlib/january05/rosenthal/01rosenthal.html.

Rosenthal, David S. H., Thomas Robertson, Tom Lipkis, Vicky Reich, and Seth Morabito. 2005b. Requirements for Digital Preservation Systems: A Bottom-Up Approach. *D-Lib Magazine* 11(11). Available at http://www.dlib.org/dlib/november05/rosenthal/11rosenthal.html.

Ross, Seamus, and Andrew McHugh. 2006. "The Role of Evidence in Establishing Trust in Repositories." Paper presented at the Joint Conference on Digital Libraries Workshop: Digital Curation and Institutional Repositories, Chapel Hill, N.C., June 15, 2006. Available at http://sils.unc.edu/events/2006jcdl/digitalcuration/Ross_McHugh-JCDLWorkshop2006.pdf.

Ross, Seamus, and Andrew McHugh. 2005. Audit and Certification of Digital Repositories: Creating a Mandate for the Digital Curation

Centre. *RLG DigiNews* 9(5). Available at http://www.rlg.org/en/page.php?Page_ID=20793#article1.

Russo, Gene. 2006. Congress Pushes Plan to Make Papers Free. *Nature* (June 21).

Salisbury, Lutishoor, Ty Vaughn, and Vikramjit Bajwa. 2004. Evidenced-Based Services at the University of Arkansas Libraries: Results of a Faculty Survey to Assess the Usefulness of Electronic Resources. *Quarterly Bulletin of the International Association of Agricultural Information Specialists* 49(1-2): 36.

SHERPA/RoMEO. 2006. Publisher Copyright Policies and Self-Archiving. Available at http://www.sherpa.ac.uk/projects/sherparomeo.html.

Steinke, Tobias. 2006. Universal Object Format: An Archiving and Exchange Format for Digital Objects. Available at http://kopal.langzeitarchivierung.de/downloads/kopal_Universal_Object_Format.pdf.

Stemper, Jim, and Susan Barribeau. 2006. Perpetual Access to Electronic Journals: A Survey of One Academic Research Library's Licenses. *Library Resources and Technical Services* 50(2): 91-109.

Tenopir, Carol. 2004. Online Scholarly Journals: How Many? *Library Journal* (Feb. 1). Available at http://www.libraryjournal.com/index.asp?layout=articlePrint&articleid+CA374956.html.

Tenopir, Carol. 2003. *Use and Users of Electronic Library Resources: An Overview and Analysis of Recent Research Studies*. Washington, D.C.: Council on Library and Information Resources. Available at http://www.clir.org/pubs/abstract/pub120abst.html.

UKPMC (United Kingdom PubMED Central). 2006. "British Library-led Partnership Chosen to Run UK PubMed Central," news release, July 31, 2006. Available at http://www.bl.uk/news/2006/pressrelease20060731a.html.

Van Orsdel, Lee C., and Kathleen Born. 2006. Journals in the Time of Google. *Library Journal* (April 15). Available at http://www.libraryjournal.com/article/CA6321722.html.

Verheul, Ingeborg. 2006. *Networking for Digital Preservation: Current Practice in 15 National Libraries*. IFLA Publication 119. Koninklijke Bibliotheek/IFLA/Saur. Available at http://www.ifla.org/VI/7/pub/IFLAPublication-No119.pdf.

Ware, Mark. 2005. E-only Journals: Is It Time to Drop Print? *Learned Publishing* 18(3): 193–199.

Waters, Donald, ed. 2005. Urgent Action Needed to Preserve Scholarly Journals. Available at http://www.diglib.org/pubs/waters051015.htm.

Waters, Donald. 2002. Good Archives Make Good Scholars: Reflections on Recent Steps Toward the Archiving of Digital Information, in *The State of Digital Preservation: An International Perspective*. Conference Proceedings: Documentation Abstracts, Inc. Institutes for Information Science, Washington, D.C., April 24–25, 2002. Available at http://www.clir.org/pubs/abstract/pub107abst.html.

Wolverton, Robert E. Jr., and Carol Tenopir. 2005. Conference Report: Discovering the Magic: Faculty and Student Use of Electronic Journals. *The Serials Librarian* 49(3): 159.

Zhang Xiaolin, Zhixiong Zhang, and Ling Wan. 2005. "Developing a National Preservation System for STM e-Journals." Paper presented at the International Conference on Preservation of Digital Objects, Göttingen, Germany, September 15–16, 2005. Available at http://rdd.sub.uni-goettingen.de/conferences/ipres05/download/Developing%20A%20National%20Preservation%20Network%20For%20STM%20E-Journals%20-%20Zhixiong%20Zhang.pdf.

URLs of e-Journal Archiving Programs Reviewed

CISTI Csi. http://cisti-icist.nrc-cnrc.gc.ca/main_e.html

CLOCKSS. http://www.lockss.org/clockss/Home

KB e-Depot. http://www.kb.nl/dnp/e-depot/e-depot-en.html

kopal/DDB. http://kopal.langzeitarchivierung.de/index.php.en

LANL-RL. http://library.lanl.gov/

LOCKSS Alliance. http://www.lockss.org/lockss/LOCKSS_Alliance

NLA PANDORA. http://pandora.nla.gov.au/index.html

OCLC ECO. http://www.oclc.org/electroniccollections/

OhioLINK EJC. http://www.ohiolink.edu/

Ontario Scholars Portal. http://www.scholarsportal.info/

Portico. http://www.portico.org/

PubMed Central. http://www.pubmedcentral.nih.gov/

URLs of Other Resources Noted in This Report

aDORe Archive. Overview at http://african.lanl.gov/aDORe/projects/adoreArchive/index.html

Archival Workshop on Ingest, Identification, and Certification Standards (AWIICS). http://nost.gsfc.nasa.gov/isoas/awiics/

Archiving and Interchange DTD. National Center for Biotechnology Information. National Library of Medicine. http://dtd.nlm.nih.gov

Certification of Digital Archives. http://www.crl.edu/content.asp?l1=13&l2=58&l3=142

Cornell University Library Research and Assessment Services Department. http://www.library.cornell.edu/iris/research/index.html

COUNTER Codes of Practice. http://www.projectcounter.org/code_practice.html

Counting Online Usage of NeTworked Electronic Resources. http://www.projectcounter.org/.

Data Preservation Alliance for the Social Sciences (Data-PASS). http://www.icpsr.umich.edu/DATAPASS/

Digital Resource Commons. http://drc-dev.ohiolink.edu/

DINI. http://www.dini.de/

DROID (Digital Record Object Identification). http://www.nationalarchives.gov.uk/aboutapps/pronom/droid.htm

e-permanence. http://www.naa.gov.au/recordkeeping/default.html.

JHOVE-JSTOR/Harvard Object Validation Environment. http://hul.harvard.edu/jhove/

NASA's Aviation Safety Reporting System. http://asrs.arc.nasa.gov

National Library of New Zealand Metadata Extraction Tool Version 1.0. http://www.natlib.govt.nz/en/whatsnew/4initiatives.html#extraction

RLG-NARA Digital Repository Certification Task Force. http://www.rlg.org/en/page.php?Page_ID=367

Safekept. http://www.nla.gov.au/padi/safekeeping/safekeeping.html

XENA (XML Electronic Normalizing of Archives). http://xena.sourceforge.net

APPENDIX 1

Survey on E-Journal Archiving Programs

Council on Library and Information Resources (CLIR)
Association of Research Libraries (ARL)
Survey on E-Journal Archiving Programs
Spring 2006
Conducted by Cornell University Library
Research and Assessment Services

Thank you for agreeing to participate in an assessment of the leading e-journal archiving initiatives. We have endeavored to complete as much of the information below from publicly available resources. Please review this information to determine its accuracy. The survey form will also help you prepare for the phone interview. If you have the time, you may wish to correct and complete the rest of the form prior to our discussion and return it to us; if not, we'll use the interview to do so. This information will be used to prepare our final report for ARL and CLIR. However, if there is particular information you do not wish to share with a broad audience, please so note it and we'll omit it from the final report. Please feel free to attach any documentation that addresses the questions posed, so noting the attachments on the form itself.

Organization:

Web Site:

Date at which e-journal archiving program responsibility began:

Principal(s) Interviewed:

Name:

Title:

E-mail:

Phone:

Name of Interviewer:

Date of Interview:

A. Organization

A.1. ***Organizational Commitment***

A.1.a. Does your mission explicitly commit you to long-term e-journal archiving?

( ) Yes
( ) No
( ) Not sure

Comments:

A.1.b. In 2 to 3 sentences, how do you define e-journal archiving?

A.2. ***Governance and Legal Issues***

A.2.a. How would you characterize your organization? (check all that apply)

[ ] National library
[ ] Open membership/subscription organization
[ ] Limited membership/subscription organization
[ ] Consortium
[ ] Third-party organization
[ ] Other:

Comments:

A.2.b. Do you have a governance board?

( ) Yes
( ) No
( ) No, but plan to within next 6 months
( ) Not sure

Comments:

If ***yes***, how are the members selected?

A.2.c. Do you have the following kinds of external advisers? (Check all that apply; if you plan to within the next 6 months, indicate this with a **P.**)

[ ] Technical advisory committee
[ ] Selection/content advisory committee
[ ] Publishers group
[ ] Member/user groups
[ ] Consultants
[ ] Other:

Comments:

A.2.d. Is your program part of a national legal deposit requirement?

( ) Yes
( ) No
( ) Not sure
( ) Not applicable

Comments:

A.2.e. Do you address intellectual property issues as they relate to your e-journal archiving program?

( ) Yes
( ) No
( ) Not sure

Comments:

A.2.f. Do you have the following audit processes in place? (Check all that apply; if you plan to within the next 6 months, indicate this with a **P.**)

[ ] Technical
[ ] Business practice
[ ] Financial
[ ] Other:

Comments:

A.2.g. As the community completes defining requirements for certification, will you seek to become a certified repository?

( ) Yes
( ) No
( ) Not sure

Comments:

A.3. ***Documentation***

A.3.a. Do you have the following written documentation that ***explicitly refers to e-journal archiving***? (Check all that apply; if you plan to within the next 6 months, indicate this with a **P.**)

[ ] Mission Statement
[ ] Publishers Agreements (producer/archive agreements)
[ ] Membership Agreements
[ ] Selection/Acquisition Policies
[ ] Transfer Requirements and Deposit Guidelines
Technical Policies that explicitly refer to e-journal archiving for:
  [ ] Ingest
  [ ] Archival Storage
  [ ] Quality Control
  [ ] Auditing
  [ ] Data Management
  [ ] Disaster Planning/Recovery
  [ ] Preservation Planning
  [ ] Metadata
[ ] Access and Use Policies
[ ] Financial Reports
[ ] Annual Reports
[ ] Other:

Comments:

A.3.b. Are you willing to share with us copies of existing documentation?
( ) Yes
( ) No

Comments:

A.4. ***Standards and Community Practice***
A.4.a. Do you follow any of the following standards and best community practices for archiving? (Check all that apply; if you plan to within the next 6 months, indicate this with a **P.**)
[ ] Open Archival Information System (OAIS) Reference Model
[ ] Trusted Digital Repositories: Attributes and Responsibilities
[ ] Preservation Metadata Implementation Strategies (PREMIS) or any other metadata standard
[ ] OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting)
[ ] Journal Archiving and Interchange DTD
[ ] An Audit Checklist for the Certification of Trusted Digital Repositories
[ ] Open-source software
[ ] Open file formats
[ ] Non-proprietary storage technology
[ ] Other:

Comments:

A.5. ***Succession Planning***
A.5.a. Do you have a succession plan in the event you are not able to continue your program?
( ) Yes
( ) No
( ) Not sure

Comments:

If ***yes,*** what arrangements have you made?

B. Stakeholders/Designated Communities

B.1. ***General***
B.1.a. On whose behalf is your content being archived?

B.2. ***Publishers***
B.2.a. Do you have a published list of publishers?
( ) Yes
( ) No
( ) Not sure

Comments:

If ***yes,*** where is this list?

If ***no,*** which publishers are included in your program?

B.2.b. Are you expecting to add more publishers?

( ) Yes
( ) No
( ) Not sure

Comments:

If ***yes***, is this information made public?

( ) Yes
( ) No
( ) Not sure

Comments:

B.2.c. Do publishers have any voice in the governance/operation of your e-journal archiving program? (Check all that apply.)

[ ] Yes, in governance
[ ] Yes, in operation
[ ] No, in neither
[ ] Not sure
[ ] Not applicable

Comments:

B.2.d. Are your agreements with publishers:

( ) Fixed duration
( ) Open-ended/indefinite
( ) Not sure

Comments:

If ***fixed***, what is the duration?

B.2.e. Do your agreements with publishers allow you to continue to archive content even if the publisher is sold or merges with another company?

( ) Yes
( ) No
( ) Not sure

Comments:

B.2.f. Are the publishers relying on you as their sole archiving solution?

( ) Yes
( ) No
( ) Not sure

Comments:

If ***no*** (i.e., if publishers are *not* relying on you as their *sole* archiving solution), describe what else they are doing for preservation:

B.2.g. What are the arrangements in the event a publisher chooses to terminate its participation in your e-journal archiving program?

B.3. ***Designated User Communities***
B.3.a. How do you define your designated communities?

B.3.b. In 2-3 sentences, what benefits do your designated communities receive?

B.3.c. Do libraries or other subscribers have any voice in the ongoing governance/operation of the e-journal archiving program? (Check all that apply.)

[ ] Yes, in governance
[ ] Yes, in operation
[ ] No, in neither
[ ] Not sure
[ ] Not applicable

Comments:

B.4. ***Others***
B.4.a. Do you have any relationships with other archiving organizations involving the following activities? (Check all that apply; if you plan to within the next 6 months, indicate this with a **P.**)

[ ] Exchange ideas and strategies
[ ] Share planning documents
[ ] Share software
[ ] Coordinate content selection
[ ] Reciprocal archiving/off-site storage/mirroring
[ ] Secondary archiving responsibility
[ ] Shared facilities/resources
[ ] Other:

Comments:

B.4.b. Do you or would you be willing/able to serve as a successor archive if another archive failed?

( ) Yes
( ) No
( ) Not sure

Comments:

B.4.c. Are those beyond your designated communities able to have access to content in the event of publisher failure or some other trigger event?

( ) Yes, now
( ) Maybe in the future
( ) No
( ) Not sure
( ) Not applicable

Comments:

B.4.d. Do you have relationships with any of the following third parties that relate to your e-journal archiving program? (Check all that apply; if you plan to within the next 6 months, indicate this with a **P.**)

[ ] Abstracting/indexing services
[ ] Consolidators
[ ] Aggregators
[ ] Standards bodies
[ ] Government agencies
[ ] National libraries
[ ] Software developers—open source
[ ] Software developers—commercial
[ ] Nonprofit/professional organizations
[ ] Access infrastructure provider
[ ] Other:

Comments:

C. Content

C.1. *E-journal Content*

C.1.a. Do you have an up-to-date list of the *journal titles* and the *date spans* included in your program that is available to the public?

( ) Yes
( ) No
( ) No, but plan to within next 6 months
( ) Not sure

Comments:

If ***yes***, where is this list?

If ***no***, which titles and dates are included?

C.1.b. How many titles are represented?

C.1.c. Do you expect to add more journals?

( ) Yes
( ) No
( ) Not sure

Comments:

If ***yes***, do you have a list of *potential* journals that is publicly available?

C.1.d. Which of the following journal content types and features does your e-journal archiving program include? (Check all that apply.)

[ ] Research articles
[ ] Internal linking
[ ] Book reviews
[ ] Letters to the editor
[ ] Lists of editorial board members

[ ] Copyright statements
[ ] Journal descriptions
[ ] Advertisements
[ ] Reprint information
[ ] Editorials
[ ] News and announcements
[ ] Errata
[ ] Supplementary materials, including supporting data
[ ] Covers of corresponding print editions, if applicable
[ ] Special features associated with publisher websites, like discussion forums
[ ] Other:

Comments:

D. Access and Triggers

D.1. ***General Access***

D.1.a. What access is currently provided to your users? (Check all that apply.)
[ ] Online access is provided
[ ] On-site access is provided
[ ] Moving wall
[ ] Access for auditing/verification
[ ] Dark archive, but metadata access is provided
[ ] Dark archive (no access until a trigger event occurs)
[ ] Other:

Comments:

If ***moving wall***, describe:

D.2. ***Trigger Events***

D.2.a. Does access to content change in the event of a trigger?
( ) Yes
( ) No
( ) Not sure

Comments:

If ***yes***, under what circumstances? (Check all that apply.)
[ ] Publisher ceases operation or fails and content is no longer available from another source
[ ] Publisher no longer offers back issues to everyone and content is no longer available from another source
[ ] Expiration of copyright
[ ] Journal ceases publication and content is no longer available from another source
[ ] Catastrophic failure of publisher's traditional access mechanism
[ ] Temporary failure of publisher's traditional access mechanism
[ ] Other:

Comments:

D.2.b. When an access trigger event occurs, describe the nature of access provided? (Check all that apply.)

[ ] Not applicable
[ ] Online viewing access
[ ] Full download capability
[ ] Keyword or subject search
[ ] Full-text search
[ ] Access by journal title, volume, issue, and page
[ ] Other:

Comments:

D.2.c. In the event of a trigger event requiring access provision, which of the following would occur? (Check all that apply.)

[ ] Not applicable
[ ] Archive will provide direct online access
[ ] Subscriber access will automatically be switched from publisher/aggregator to archive
[ ] Subscriber must manually switch access from publisher/aggregator to archive
[ ] Online access responsibilities will be transferred to another entity
[ ] Subscribers/members will have the option of taking over stewardship of the content
[ ] Other:

Comments:

D.2.d. What is the anticipated ramp-up period from the time the trigger occurs until access becomes available? (If not applicable, answer "N/A.")

D.2.e. Do you have (or do you have a contingency plan to rapidly provide) the capacity, including network bandwidth, to supply online access to a large number of users in the event of a major publisher failure?

( ) Not applicable
( ) Yes
( ) No
( ) Not sure

Comments:

D.2.f. Have you tested your systems to ensure that proposed access mechanisms will function to expected usage levels when a trigger event occurs?

( ) Not applicable
( ) Yes
( ) No
( ) Not sure

Comments:

E. Technology

E.1. ***Ingest Policy and Practice***

E.1.a. How do you obtain content for the e-journal archive? (Check all that apply.)

[ ] Source files provided by publisher
[ ] Rendition/access files provided by publisher
[ ] Content harvested via the Internet (Web or ftp site)
[ ] Other:

Comments:

E.1.b. Do you validate the *completeness* of content upon ingest?

( ) Yes
( ) No
( ) Not sure

Comments:

If ***yes,*** how?

E.1.c. Do you validate the *integrity* of content upon ingest?

( ) Yes
( ) No
( ) Not sure

Comments:

If ***yes,*** how?

E.1.d. Do you have ingest requirements for file formats and other content received from publishers?

( ) Yes
( ) No
( ) Not sure

Comments:

If ***yes,*** what are they?

E.1.e. What file formats are you able archive? (Check all that apply.)

Text formats and page description languages

[ ] Plain text
[ ] HTML
[ ] SGML
[ ] XML
[ ] PDF
[ ] Postscript
[ ] TeX
[ ] Other:

*Still image*s
[ ] TIFF
[ ] JPEG
[ ] PNG
[ ] JPEG 2000
[ ] GIF
[ ] SVG
[ ] Postscript
[ ] EPS (Encapsulated Postscript)
[ ] Other:

Others
[ ] Sound file formats (e.g., aiff, mp3, ogg vorbis, MIDI, QuickTime, Realmedia)
[ ] Moving-image file formats (e.g., AVI, MPEG-4, Flash, QuickTime, Realmedia)
[ ] Not sure
[ ] Other:

Comments:

E.1.f. Do you test for and document the *ongoing integrity* of the archived content?
( ) Yes
( ) No
( ) No, but plan to within next 6 months
( ) Not sure

Comments:

If ***yes***, how?

E.1.g. Do you intend to retain the appearance ("look and feel") as well as the content of ingested material?
( ) Yes, for all formats
( ) Yes, for some formats
( ) No
( ) Not sure

Comments:

If ***yes***, describe:

E.1.h. Do you intend to retain the functionality of interactive content (e.g., spreadsheets, multimedia)?
( ) Yes, for all formats
( ) Yes, for some formats
( ) No
( ) Not sure

Comments:

If ***yes***, describe:

E.1.i. Are your systems Unicode compliant?
( ) Yes, completely
( ) Yes, partially
( ) No
( ) Not sure

Comments:

If ***partially***, describe:

E.1.j. What specialized content encodings are you able to archive? (Check all that apply.)
[ ] Table markup
[ ] Mathematical symbols
[ ] Chemical formulas
[ ] Archaic scripts or ideographs (e.g., Egyptian or Mayan hieroglyphs)
[ ] Musical notation
[ ] Other:

Comments:

E.2. ***Archiving and Disaster Planning***
E.2.a. Which of the follow e-journal archiving strategies do you/will you use? (Check all that apply.)
[ ] Migration
[ ] Emulation
[ ] Normalization
[ ] Reliance on standards
[ ] Refreshing
[ ] Use of durable media
[ ] Not sure
[ ] Other:

Comments:

E.2.b. Which of the follow e-journal archiving strategies have you *conducted*? (Check all that apply.)
[ ] Migration
[ ] Emulation
[ ] Normalization
[ ] Reliance on standards
[ ] Refreshing
[ ] Use of durable media
[ ] Not sure
[ ] Other:

Comments:

E.2.c. Do you use any of the following redundancy procedures? (Check all that apply; if you plan to within the next 6 months, indicate this with a **P**)
[ ] Local backups
[ ] Mirror sites
[ ] Off-site storage

[ ] Not sure
[ ] Other:

Comments:

E.2.d. How many copies of your content do you maintain?

E.2.e. Do you have *written* procedures and protocols designed to minimize vulnerability to the following threats? (Check all that apply.)
[ ] Malicious attacks (e.g., hacking, malware, vandalism, physical intrusion)
[ ] Natural disasters (e.g., flood, fire, earthquake)
[ ] Infrastructure failure (e.g., power outage, HVAC failure, water-containment failure, building fire)
[ ] Other:

Comments:

F. Resources

F.1. ***Funding Sources and Cost Model***
F.1.a. What is your source of funding to ensure long-term access to e-journals?
(Check all that apply.)
[ ] Appropriations
[ ] Endowment
[ ] Member/subscriber fees
[ ] Publisher support
Grants/contributions from:
[ ] Individuals
[ ] Organizations
[ ] Government agencies or departments
[ ] Private foundations
[ ] Professional associations
[ ] Corporations (other than publishers)
[ ] Other:

Comments:

F.1.b. Is financial information about your archiving effort publicly available?
( ) Yes
( ) No
( ) Not sure

Comments:

F.2. ***Resource Allocation***
F.2.a. In the event of a trigger event requiring provision of access, do you also have resources to cover the costs of providing general online access to the designated community?
( ) Not applicable
( ) Yes
( ) No
( ) Not sure

If ***no,*** how will additional funding be acquired? (Check all that apply.)
[ ] Additional fee borne by designated community
[ ] Additional appropriations or one-time monies
[ ] Solicitations from other entities (name):
[ ] Other:

Comments:

F.3. ***Membership/Subscriber Policies and Fees***
F.3.a. What are members'/subscribers' responsibilities? (Check all that apply.)
[ ] Pay the fees
[ ] Provide hardware/software
[ ] Provide technical expertise
[ ] Manage some aspect of content/functionality
[ ] Manage some aspect of access
[ ] No responsibilities
[ ] Not sure
[ ] Not applicable
[ ] Other:

Comments:

F.3.b. How do you determine the cost of joining up?
( ) Not applicable
( ) Fixed price
( ) Sliding scale, describe:
( ) Other model:

Comments:

F.3.c. Is there a discounted price for consortia or library systems?
( ) Not applicable
( ) Yes
( ) No
( ) Not sure

Comments:

If ***yes,*** describe:

F.3.d. Do membership/subscription fees cover specific titles or the full service?
( ) Not applicable
( ) Specific titles
( ) Full service
( ) Other:

Comments:

F.3.e. Is the cost of growth in content and new publications included in subscriptions or is it additional?

( ) Not applicable
( ) Included
( ) Additional
( ) Not sure

Comments:

F.3.f. Do you have incentives for early joiners?

( ) Not applicable
( ) Yes
( ) No
( ) Not sure

Comments:

If ***yes***, describe:

F.3.g. Do you have disincentives for joining later?

( ) Not applicable
( ) Yes
( ) No
( ) Not sure

Comments:

If ***yes***, describe:

F.3.h. Do you have a mechanism to ensure fair distribution of costs amongst all institutions that will benefit (i.e., dissuade "free riding")?

( ) Not applicable
( ) Yes
( ) No
( ) Not sure

Comments:

If ***yes***, describe:

F.3.i. Have you or would you lower the cost of membership/subscription if costs dropped or additional income was received (e.g., from more members or outside parties)?

( ) Not applicable
( ) Yes
( ) No
( ) Not sure

Comments:

If ***yes***, describe:

F.3.j. If two member institutions merged, could it roll over its archived content to the new entity?
( ) Not applicable
( ) Yes
( ) No
( ) Not sure

Comments:

If ***yes***, describe:

F.3.k. If a member withdraws, does it receive anything from the archive?
( ) Not applicable
( ) Yes
( ) No
( ) Not sure

Comments:

If ***yes***, describe:

Thank you very much! If you have completed this form, please return to:

APPENDIX 2

Profiles of the 12 E-Journal Archiving Initiatives

All data in the following summaries were current as of July 1, 2006

Canada Institute for Scientific and Technical Information

The National Research Council of Canada (NRC), Canada's governmental organization for research and development, hosts the Canada Institute for Scientific and Technical Information (CISTI), a major source for information in all areas of science, technology, engineering, and medicine. CISTI became the National Science Library in 1957.

CISTI has a key role as leader and catalyst in building universal, seamless, and permanent access to information for Canadian research and innovation. To help achieve this vision for Canada, CISTI has established a three-year program called Canada's scientific infostructure (Csi). This program will create a national information infrastructure and opportunities for collaborations with partners to support research and educational activities.

Using a leading-edge architectural approach, CISTI has built a reliable technology platform with expandable storage capacity that ensures long-term access to digital content loaded at CISTI. CISTI is partnering with Library and Archives Canada (LAC) to ensure business continuity for the infrastructure. With the infrastructure in place, CISTI has loaded close to 5 million articles from publishers NRC Research Press, Springer, and Elsevier. New content from the Institute of Physics, Oxford University Press, the American Society for Microbiology, Mary Ann Liebert, and Emerald will be added to increase the depth and breadth of the repository.

As part of the Csi program, CISTI is negotiating with publishers for rights to make content accessible to customers and partners. To ensure that access is as seamless as possible, CISTI is implementing SFX to support bibliographic linking and is investigating best options to support authentication and authorization in a digital environment. CISTI is also conducting research in the areas of text and data mining and text analyses for future implementation.

LOCKSS Alliance

The Lots of Copies Keep Stuff Safe (LOCKSS) program began in 1999 as a research project based at Stanford University Library. LOCKSS launched the beta version of its open-source software to 50 libraries

between 2000 and 2002. LOCKSS developed its software to allow libraries to collect, store, preserve, and provide access to their own, local copies of authorized content they purchase. The LOCKSS Web site[1] lists about 100 participating institutions in more than 20 countries that are using the LOCKSS appliance to capture content. About 25 publishers of commercial and open access are participating in LOCKSS not counting the individual publishers represented by aggregators such as HighWire Press and Project MUSE, and LOCKSS's own Humanities Project.[2]

In 2005, LOCKSS launched the LOCKSS Alliance as a membership organization that is built on the LOCKSS software to introduce governance for the program and to address sustainability issues. The LOCKSS Alliance is an open membership organization. Members have equal rights and responsibilities, though membership fees are based on an institution's Carnegie Classification. LOCKSS Alliance membership benefits include participation in collection-development activities (including publisher briefings); early access to LOCKSS documents, documentation, and prerelease software; access to implementation collection and technology workshops; involvement in community planning efforts; and access to the LOCKSS program staff.

The LOCKSS Alliance assures its members of access to participating publisher content, if the member has licensed or purchased that content. Libraries manage their LOCKSS boxes to include all the licensed content to which they wish to ensure long-term access. Libraries can also negotiate with publishers that are not participating in LOCKSS. Participating publishers may choose to prevent the collection of new content, but they cannot withdraw content that was previously ingested.

The LOCKSS appliance, an open-source software application, is the core of the LOCKSS program and the foundation for the LOCKSS Alliance. The appliance uses Web harvesting to capture content from participating publisher websites. To participate in LOCKSS, a publisher grants access to libraries to collect, preserve, and provide access to the content and grants access to the LOCKSS software to crawl, collect, and preserve the content by adding a Web page called a LOCKSS publisher manifest. The LOCKSS appliance has rules for monitoring, mediating, and repairing on the basis of the results of this continuous polling of the content.

CLOCKSS

The CLOCKSS (Controlled LOCKSS) initiative is a 2006 addition to the LOCKSS program that brings together 6 libraries (Edinburgh University, Indiana University, New York Public Library, Rice University, Stanford University, and University of Virginia) and 12 publishers and learned societies (American Chemical Society, American

[1] http://lockss.stanford.edu/about/users.htm.

[2] http://lockss.stanford.edu/about/titles.htm.

Medical Association, American Physiological Society, Blackwell Publishing, Elsevier, Institute of Physics, Nature Publishing Group, Oxford University Press, Sage Publications, Springer, Taylor & Francis, John Wiley & Sons, Inc.) to establish a large-scale, dark archive for e-journals. The libraries participating in CLOCKSS are also participants in the LOCKSS Alliance. Each library will host two servers, creating a network of 12 dark repositories.

CLOCKSS is a limited-membership organization that is holding assets on behalf of the broader community. CLOCKSS systems will harvest content by Web crawling and ingest source files provided by publishers. Access to CLOCKSS content will be made available to the community following an access trigger event. The CLOCKSS system will automatically detect the cessation of online access from the publisher and, if the content remains unavailable for six months, the governing board (made up of libraries and publishers) will work collaboratively to determine whether content will be made available to the community for a limited or indefinite time. "It's like a barn raising," Gordon Tibbitts, president of Blackwell Publishing's American division, said of CLOCKSS. "We all know we have to have the barn, so we're calling everyone together to build it" (Kiernan 2006).

During the two-year developmental phase, the CLOCKSS initiative will also test the responsiveness of this distributed test bed of content to various potential disasters and share the results of these tests to contribute to the development of global strategies for preservation.

Koninklijke Bibliotheek e-Depot

As the national deposit library for the Netherlands, the Koninklijke Bibliotheek (KB) has the responsibility for preserving and providing long-term access to Dutch electronic publications. At first, the KB focused on Dutch publishers, but more recently it has come to recognize that multinational publishers produce academic literature, and, as a consequence, there is often no longer a national library that is the natural repository for the content the publishers produce. The KB, therefore, has assumed the responsibility to acquire and preserve, in conjunction with other repositories, the published scientific output of the world, regardless of where it was formally published.

To meet that responsibility, the KB began planning for e-journal archiving in 1993, started experimenting with e-journal archiving systems in 1995, and conducted research and implementation of an e-journal archiving system as part of the NEDLIB project from 1998 to 2000. The current e-Depot was delivered in 2002 and is now fully operational: a fully automated system, dedicated to long-term storage and large-scale archiving. The e-Depot system has been made part of the general budget of the KB. In addition, since at least 2003, the KB has been receiving earmarked funds for the operation of the e-Depot system as well as monies for research and development in long-term preservation. Currently, those funds amount to €2 million a year.

The growth of content in e-Depot has been dramatic. As of March 2006, the e-Depot contained more than 6 million digital objects in about 6 terabytes of storage space. More than 3,500 e-journal titles are represented in the repository. Among the prominent publishers that have signed archiving agreements with the KB are

- Elsevier (1996, 2002)
- BioMed Central (2003)
- Kluwer Academic Publishers (now part of Springer) (2003)
- Blackwell Publishing (2004)
- Taylor & Francis (2004)
- Oxford University Press (2004)
- Sage Publications (2005)
- Brill Academic Publishers (2005)
- Springer (2005)

The KB's goal is to include in the e-Depot the journals from the 20 to 25 largest publishing companies, which produce almost 90% of the world's electronic STM literature.

Because there is no legal deposit requirement in the Netherlands, the deposit of material into e-Depot is managed through negotiations between the KB and individual publishers. At a minimum, the KB stipulates that there must be on-site access to all authorized library users. The archiving agreement with BioMed Central allows the KB to provide free remote access to more than 100 open-access journals. For non-open-access journals, the agreement with publishers stipulates that in the event that a publisher cannot deliver content for a long period of time, the KB could deliver the journals on an interim basis to subscribers. If a publisher should decide to stop providing electronic access, the KB could, if it so chooses, provide access to the world. Thus, while the e-Depot system is not primarily an access system, in an emergency the e-Depot could in theory provide access to users around the world—assuming sufficient funds to do so were available.

After receipt, ingest, and storage of electronic files from the publishers, the KB follows two technical approaches to long-term digital preservation. The first is migration: the KB plans to transform digital objects to keep them readable. The KB is also interested in emulation and has several projects under way to see whether it can be used both to lower the cost of preservation and to preserve the look and feel of the original object. The KB continues to work with IBM, the vendor for the e-Depot system, as well as partners from around the world, to create the technical tools required for digital preservation.

Perhaps the most important component of the KB's approach to digital preservation, however, has been the articulation of the need for what it has called the "Safe Places Network." The Safe Places Network will consist of a limited number of places that make a substantial investment in the equipment, skills, and expertise necessary to manage digital archiving programs. Sharing the risks inherent in a digital archiving system with a limited number of committed partners, it is hoped, will reduce the cost of digital preservation.

kopal/ Die Deutsche Bibliothek

The Kooperativer Aufbau eines Langzeitarchivs digitaler Informationen (kopal), is a cooperative project funded by the German Federal Ministry of Education and Research. It began in July 2004. Its goal is to develop an innovative technical solution to the problem of long-term accessibility of digital documents. Project partners Die Deutsche Bibliothek (DDB—the National Library of Germany) and the Lower Saxon State and University Library (SUB Göttingen) are storing a variety of digital materials in a repository based on DIAS, the Digital Information and Archiving System, developed by IBM and the National Library of the Netherlands, the Koninklijke Bibliotheek, in The Hague. The Gesellschaft für wissenschaftliche Datenvaraberitung mbH Göttingen (GWDG) is in charge of the archive's technical operation, with software support provided by IBM Deutschland GmbH.

One of the driving forces behind kopal has been the need of DDB for a system for managing the legal deposit of electronic publications. DDB had been experimenting with electronic journals since 2000; in 2006, legal deposit legislation for electronic publications was enacted in Germany, making the implementation of a system a priority. Fortunately, as part of the initiation of electronic legal deposit, DDB is getting a funding raise of about €2 million to implement it.

As part of its preliminary investigations, DDB had, through voluntary agreements with publishers, acquired a variety of electronic content, including 455 e-journal titles from Springer and many other e-journals from Wiley-VCH and Thieme. Under legal deposit, DDB will start acquiring and adding to kopal all electronic journals published in Germany.

DDB requires that publishers send to it compressed archive files that contain the journal contents plus some rudimentary metadata. At present, the intention is to maintain the readability of the archived file; when necessary, the content will be migrated into new formats. DDB has used emulation for some preservation activities and will continue to do so.

Voluntary agreements with publishers in the past have allowed for public access to the e-journals in the event of publisher failure. This "access of last resort" may also be possible with journals received via legal deposit. As yet, kopal has not built public-access systems, and so it is likely that there would be a significant delay between the collapse of a publisher's delivery system and remote access to content in kopal. Nevertheless, kopal/DDB is likely to serve as an important guarantor of the long-term availability of e-journals published in Germany.

Los Alamos National Laboratory Research Library

Los Alamos National Laboratory (LANL) is one of three U.S. national laboratories (the other two being Sandia and Lawrence Livermore) operated under the National Nuclear Security Administration of the U.S. Department of Energy. The Research Library at Los Alamos National Laboratory (LANL-RL) has been locally loading licensed backfiles from several commercial and society publishers since 1995. Focusing on titles in the physical sciences, the library maintains the content primarily for the use of LANL staff, but it also serves a group of external cost-recovery clients. These include five U.S. Department of Energy laboratories, nine members of the U.S. Air Force Library Consortium, Sandia National Laboratories, Santa Fe Institute, and five universities located in the western United States. LANL-RL's locally loaded e-journals are also available to members of the public who are on-site at the library during its regular hours. The titles come from the following publishers:

- American Chemical Society
- American Institute of Physics
- American Physical Society
- Elsevier
- Institution of Electrical Engineers
- Institution of Electrical and Electronics Engineers
- Institute of Physics
- John Wiley & Sons, Inc.
- Royal Society of Chemistry (backfiles through 2004 only)
- Springer

Through its digital library initiative, the Library Without Walls, LANL-RL has done substantial research and development work on repository and digital object architecture for long-term maintenance of electronic journal contents. In November 2004, LANL-RL received a $750,000 grant from the U.S. Library of Congress's National Digital Information Infrastructure and Preservation Program "to support research and development of tools that will help address complex problems related to collecting, storing and accessing digital materials."

A major focus of the research-and-development (R&D) work at LANL-RL has been the aDORe repository. aDORe uses a modular architecture, and is based on the following standards (Bekaert, Liu, and Van de Sompel 2005):

- MPEG-21 DID (Digital Item Declaration) to represent digital objects
- MPEG-21 DII (Digital Item Identification) to identify digital objects
- XMLtapes and Internet Archive ARC files to store digital objects and constituent data streams
- OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting) to harvest resources
- The OpenURL Framework to convey context-sensitive dissemination requests

- Info URI to facilitate the referencing of information assets under the URI allocation

LANL-RL is moving its main e-journal repository from ScienceServer to aDORe and expects to complete the transfer by the first quarter of 2007. Until then, it has to live with some of the limitations of ScienceServer, including the inability to display certain formats and partial lack of Unicode compliance. The new architecture will be considerably more flexible and was built with long-term preservation of digital objects in mind. In particular, it provides an application-neutral, XML-based means to store a wide variety of file formats while maintaining a record of the infrastructure and tools needed to decode the files through evolving digital environments.

Despite the emphasis on preservation in its R&D work, LANL-RL does not offer e-journal archiving services to its external cost-recovery clients. The fees paid by clients cover only the cost of current access and do not provide for subsequent access, even to backfiles, in the event of termination. However, even beyond its digital repository development contributions, LANL-RL's e-journal preservation efforts have important implications, both for the LANL community and for the scholarly community at-large.

First, LANL-RL has insured through contractual negotiation that all acquired e-journal content can be perpetually archived. Second, it has extended its R&D work into the area of trustworthy and high-integrity transfer of e-journal content from publishers. Since 2003, LANL-RL has been working with the American Physical Society (APS) on a multiphase project that may lead to the establishment of a fully synchronized dark-mirror site for all APS publications wherein LANL-RL would become the worldwide source for APS content in the event of catastrophic failure of APS's primary servers. LANL is in various stages of negotiation with other publishers to offer similar mirror and fallback services.

LANL receives appropriations from the U.S. Departments of Energy and of Defense, among other sources. The Research Library receives funding out of the institutional overhead in those appropriations. Researchers receiving grants are taxed for institutional support, and a portion of those funds go to support of the RL. Therefore, part of the RL's funding comes indirectly from appropriations, though there is no explicit budget line for RL operations, let alone for e-journal archiving or other specific tasks.

This creates a certain amount of uncertainty regarding ongoing commitments to e-journal archiving. LANL-RL's primary concern is that the scholarly journal literature needed by its staff continue to be available via an affordable and trustworthy mechanism. If another source that provided sufficient functionality emerged, it could decide to contract for the services instead. On the other hand, LANL-RL was one of the earliest local loaders of e-journals, and as a result of ongoing R&D, has continued to offer LANL staff functionality not available elsewhere.

Another potential source of uncertainty is that LANL is under-

going a major restructuring that could affect priorities and funding. LANL is currently managed by the University of California (UC) under contract to the U.S. Department of Energy, but over the next year, operation of the laboratory will shift to a limited liability corporation called Los Alamos National Security that includes UC along with Bechtel National, Inc., BWX Technologies, Inc., and the Washington Group International, Inc. How the shift in management will affect the RL's operation is not yet known.

National Library of Australia PANDORA

The National Library of Australia (NLA) established PANDORA in 1996. PANDORA is an acronym for Preserving and Accessing Networked Documentary Resources of Australia. PANDORA serves "all Australians, present and future, and anyone with a research interest in Australia." In addition to the NLA, the PANDORA program includes nine national- and state-collecting agencies across Australia that partner to populate and maintain PANDORA. The NLA covers the infrastructure, and support costs for PANDORA through appropriations.

PANDORA contains six priority categories of online publications, including Commonwealth and Australian Capital Territory government publications, publications of tertiary education institutions, conference proceedings, e-journals, titles referred by indexing and abstracting agencies, and topical Web sites. There are 1,983 journals represented in PANDORA, although not all are scholarly or peer reviewed. The PANDORA Web site groups the content into a broad range of subjects covering academic, cultural, social, political, and technical topics. Apart from approximately 150 commercial titles, PANDORA contains publicly accessible content. The commercial content of PANDORA is typically restricted for one to three years.

The first version of the PANDORA Archiving System (PANDAS) was released in 2001. The members of PANDORA use PANDAS to gather content, which is stored on NLA servers using proprietary storage software called DOSS. The NLA developed the PANDAS software to support these workflows: identifying, selecting, and registering candidate titles; seeking and recording permission to archive titles; setting harvest regimes appropriate to the content; gathering (harvesting) files; undertaking quality assurance checking; initiating archiving processes; and organizing access, display, and discovery routes to, and metadata for, the archived resources. The PANDAS software manages administrative metadata about titles that have been selected for archiving, rejected, or are being monitored pending a decision; manages access restrictions; schedules and initiates the harvesting of titles; manages the quality checking and assurance process; prepares and organizes harvested content for public display through title entry pages and title and subject listings; and provides operational reports. The PANDAS software that the NLA developed to gather content will be made available as open-source software soon.

OCLC Electronic Collections Online

OCLC launched Electronic Collections Online (ECO) in June 1997 to support the efforts of libraries and consortia to acquire, circulate, and manage large collections of electronic academic and professional journals. It provides Web access via the OCLC FirstSearch interface to a growing collection of more than 5,000 titles in a wide range of subject areas, from more than 40 publishers of academic and professional journals. Libraries, after paying an access fee to OCLC, can select the journals to which they would like to have electronic access.

An important component of the ECO offering is its promise of long-term accessibility to subscribed content. OCLC's agreement with publishers ensures that it can continue to provide libraries with access to any content to which the libraries may have subscribed as long as the library continues to pay the access fee. Even if a user discontinues an ECO access account, OCLC will maintain the user's subscription profile for five years, and if a user renews an access account before five years have passed, the user can regain access to all the journals covered by the previous subscription.

Although ECO has not established the "minimal set of well-defined services" that would make it a "qualified preservation archives" (Waters 2005), it has undertaken a number of steps that increase the likelihood that it will be able to provide continued access to the content it offers. For example, OCLC maintains a copy of all journal content and the associated abstract and index data in an off-site storage facility. It has also secured the right to migrate journal backfiles to new data formats as current formats such as PDF, which form the vast bulk of ECO content, become outmoded. (OCLC has not as yet, however, had to migrate any file formats.) ECO is not part of OCLC's Digital Archive service and has no immediate plans to take advantage of OCLC's "real-world solutions for the challenges of archiving and preservation in the virtual world."

In the event of publisher failure or some other trigger event that would prevent a publisher from delivering content to subscribers, it is possible that subscribers might be able to shift their subscriptions to ECO in order to secure access. This would have to be worked out in negotiations with the publishers. Should OCLC decide to stop offering the ECO service, it can provide to libraries on tape or CD/DVD copies of any content to which the library had subscribed. It would then be the library's responsibility to mount that material and make it available.

OhioLINK Electronic Journal Center

The Ohio Library and Information Network (OhioLINK) is a consortium of Ohio's college and university libraries, comprising 85 institutions of higher education and the State Library of Ohio. OhioLINK's electronic services include a multipublisher Electronic Journal Center (EJC), launched in 1998, which contains more than 6,900 scholarly

journal titles from close to 40 publishers across a wide range of disciplines. Although several OhioLINK resources are available to all Ohio residents (with some open to all on the Internet), the content of EJC is available only to students, faculty, and staff members at OhioLINK-affiliated institutions. At this time, OhioLINK has neither the resources nor the legal right to make the contents of EJC available outside of the state of Ohio.

EJC is an optional service of OhioLINK, though the vast majority of Ohio higher education institutions have chosen to participate. The cost of joining EJC is determined by the institution's current spending on journals from the publishers who are represented in EJC, including print and electronic subscriptions. Most institutions wind up getting electronic access to far more titles than they previously were subscribing to for a similar outlay of funds. The access mechanism is shifted from a campus-based one through publishers and aggregators to one based on EJC.

EJC accepts most content as it is supplied by the publisher, but is limited in the formats that can be displayed by its main repository software, ScienceServer. The current version of ScienceServer can display only PDF, TIFF, and some types of XML. EJC intends shortly either to upgrade to a new version of ScienceServer or move to different repository software. Goals for the new software include expansion of the range of file formats that can be displayed and resolving existing display limitations caused by the lack of Unicode compliance in the old ScienceServer.

OhioLINK has declared its intention to maintain the EJC content as a permanent archive and has acquired perpetual archival rights in its licenses from all publishers but one (the American Chemical Society). Furthermore, in May 2006 the OhioLINK Governing Board approved a series of recommendations that included a commitment to seek the addition of a clause to all EJC contracts that would extend liberal self-archiving and access rights to all personnel of Ohio higher education institutions.

EJC relies on regular and heavy use by subscribers to help maintain the integrity of its archive and reveal problems. Though it anticipates having to perform file migrations in the future, it has not done any yet. It does not normalize incoming files. Instead, EJC relies on publishers to supply files in one of the standard formats that ScienceServer is capable of displaying. Content received from publishers in other formats is retained, but will not be displayable until the next-generation repository software is in place.

All technical infrastructure costs, as well as about 20% of content-acquisition costs, are centrally funded though legislative appropriations. The remaining funding for content comes from member libraries. Fluctuations in state appropriations have resulted in discontinuation of some titles. EJC's contracts stipulate a nonpunitive approach to obtaining missing content if EJC resubscribes to a canceled title.

EJC has been extremely popular and continues to experience growth in usage. OhioLINK would like to expand EJC to include

publishers such as Sage, Taylor & Francis, Cell Press, the Institute of Electrical and Electronics Engineers, GeoScienceWorld, and titles from a number of scholarly societies. Some of these acquisitions would fill gaps in disciplines such as nursing and the biosciences that OhioLINK officials feel are currently underserved. If funding can be found, OhioLINK also wants to purchase backfiles for many titles as a means to increase access and save member libraries money by reducing the need to store print copies at multiple sites.

Plans include development of a Digital Resource Commons (DRC),[3] with which OhioLINK hopes to accomplish with a shared repository environment what EJC and other OhioLINK components have done with shared content. Instead of member institutions investing the resources to create and manage their own repositories, DRC would provide a centrally managed repository (based on Fedora) with locally controlled infrastructure for ingest, and a sophisticated, multilevel access rights management system. According to OhioLINK, DRC "ingests, preserves, presents, and mediates administration of the educational and research materials of participating institutions." Capabilities envisioned include an institutional repository for research portfolios such as preprints, postprints, and working papers, electronic thesis and dissertation management, and Web-mediated peer-reviewed electronic journals with open access, self-archiving, and publishing.

Ontario Scholars Portal

Launched in 2001, the Ontario Scholars Portal (OSP) serves all 20 university libraries in the Ontario Council of University Libraries (OCUL) consortium.[4] The Portal includes more than 6,900 e-journals from 13 publishers and metadata for the content of an additional 3 publishers. The publishers currently represented include Elsevier, John Wiley & Sons, Inc., Springer, Kluwer Law International, Blackwell, Oxford University Press, Cambridge University Press, American Psychological Association, Emerald, Berkeley Electronic Press, Sage, Institute of Electrical and Electronics Engineers, and the Royal Society of Chemistry.

The Portal uses a combination of "push and pull" to gather content: publishers provide source files, and the Portal harvests content from publisher Web sites. The Portal stores all the content from publishers, but the current system cannot render all the formats that have been stored, e.g., video files and numeric data. Most of the content is in PDF or XML format.

The primary purpose of the Portal is access, but the consortium has made an explicit commitment to the long-term preservation of the e-journal content that it loads locally. The Portal provides online

[3] About the Digital Resource Commons, http://drc-dev.ohiolink.edu/.

[4] http://www.ocul.on.ca/.

access to the content that consortium members have licensed or purchased. Members of the consortium are required to pay membership fees and are represented on the executive board of the Portal. Preservation is included in the e-journal service to members.

Between 2001 and 2005, OSP was supported by a grant and provincial matching funds as part of the Canadian National Site Licensing Program.[5] Ongoing support for OSP relies upon a membership cost model that adjusts for the varying size of consortium members and usage factors and that includes tiered membership fees.

Portico

Portico is one of the newest of the archiving programs, having just gone "live" in 2006 (although planning began in 2004, and the preservation obligation was assumed in 2005). The mission of Portico is to "preserve scholarly literature published in electronic form and to ensure that these materials remain accessible to future scholars, researchers, and students." Specifically designed as a third-party electronic-preservation service, Portico serves as a permanent dark archives. E-journal availability (other than for verification purposes) is governed by trigger events resulting from substantial disruption to access via the publishers themselves.

The program's archival approach begins with the receipt of source files, which comprise the intellectual content of electronic scholarly journals directly from the publishers, and features transformation or normalization of these diverse files to a standard archival format that can be managed over time through the preservation strategy of migration.

Portico boasts a strong pedigree, with startup funding provided by The Andrew W. Mellon Foundation, Ithaka, JSTOR, and the Library of Congress. A membership organization, it is open to all libraries and scholarly publishers, both of which are asked to support the effort through annual contributions.

Thirteen publishers are participating in Portico:

- American Anthropological Association
- American Mathematical Society
- Annual Reviews
- Berkeley Electronic Press
- BioOne
- Elsevier
  - *Cell* Press
  - *The Lancet*
- John Wiley & Sons
- Oxford University Press
- Sage Publications, Inc.
- Society for Industrial and Applied Mathematics (SIAM)

[5] http://library.queensu.ca/libdocs/news/2001apr09.htm.

- Symposium Journals (Oxford UK)
- United Kingdom Serials Group
- University of Chicago Press

Recently announced library fees, ranging from $1,500 to $24,000 per year, are based on the total library materials expenditures for an individual institution. To encourage early adopters, libraries that subscribe to this service in 2006 and 2007 will be designated "Portico Archive Founders" and will receive substantial savings on their annual archive support payment for five years. Library systems and consortia that facilitate support for the archive among their member institutions will be offered modest savings in their annual payments. According to Eileen Fenton, executive director, Portico is aiming to attract additional libraries from across the Carnegie Classification of Institutions of Higher Education.[6]

PubMed Central

PubMed Central (PMC) is a free, publicly accessible digital archive of English language biomedical and life sciences journal literature, run by the National Center for Biotechnology Information (NCBI) of the U.S. National Library of Medicine (NLM). Launched in February 2000 with content from the *Proceedings of the National Academy of Sciences* and *Molecular Biology of the Cell*, PMC has grown to include hundreds of thousands of articles from about 250 titles and 50 publishers.

Like the similarly named PubMed, PMC is an integral component of NCBI's Entrez life sciences search engine. While PubMed contains citations, abstracts, and links to full-text articles, PMC consists of full-text research articles and other content from peer-reviewed life sciences journals. The two services are separate and not entirely complementary. PubMed points to numerous articles that are not in PMC, while some content in PMC (mostly nonarticle journal content) is not indexed in PubMed.

PMC's mandate to preserve the journal literature of biomedicine comes from the Congressional act that created NLM, which authorizes it to "acquire, organize, disseminate and preserve books, periodicals, . . . and other library materials pertinent to medicine." At the moment, NLM cannot compel researchers to deposit their publications in PMC, but authors of life science research sponsored by U.S. National Institutes of Health are requested to voluntarily deposit final manuscripts of articles into PMC within a year of publication.

That situation may change, however. Legislation entitled the Federal Research Public Access Act of 2006 (introduced in the U.S. Senate on May 2, 2006) would require that U.S. government agencies with annual extramural research expenditures of more than $100 million make journal articles based on research funded by that agency publicly available via the Internet within six months. If the bill is

[6] http://www.carnegiefoundation.org/classifications/.

passed, agencies in the U.S. Department of Health and Human Services, e.g., NIH and the Centers for Disease Control and Prevention, would presumably use PMC, since the law requires that manuscripts be preserved in a digital archive that supports free public access, interoperability, and long-term preservation.

Other content comes into PubMed Central by a variety of mechanisms. Some open-access journal publications (most notably the entire set of BioMed Central journals) use PMC as their archiving solution. Some commercial publishers that do not otherwise have agreements with PMC allow authors to designate their articles as open access and to deposit these articles in PMC. Finally, a growing number of publishers have reached contractual agreements with PMC to deposit all their journal contents with PMC.

To participate in PMC, a publication must be covered by a major abstracting/indexing service, or have three editorial board members with current grants from major nonprofit funding agencies. Publishers are required to supply source files (via FTP or on CD/DVD or tape) in either SGML or XML, conforming to the NLM Journal Archiving XML DTD or another full-text article DTD that is widely used in the life sciences. The original high-resolution digital image files must be provided for all figures. PMC prefers (but does not require) that publishers also include a PDF version of their articles in the archive. Publishers are encouraged to deposit the entire contents of their journals for archiving, but must at minimum provide all research articles. For display purposes, PMC performs an on-the-fly conversion of stored XML to HTML.

PMC has a flexible deposit policy designed to accommodate the desire of many publishers to delay appearance of journal content in PMC for a period of time following publication. Although publishers are encouraged to make content available via PMC as soon as possible after publication, they may request a delay of up to one year for research articles, and up to three years for other content, such as letters and reviews.

NLM is committed to long-term stewardship of the content in PMC. All contracts must include a clause granting PMC perpetual archiving rights for any deposited material. Two operational policies dominate PMC's approach to content longevity. One is an emphasis on standardized XML, which is portable, maintains document structure, and lends itself to intelligent processing without sacrificing human readability. NLM is continuing its work on the Journal Archiving and Interchange DTD from which the Journal Publishing DTD was derived and for which the Library of Congress and the British Library recently announced support. The other is free, open access to all content, which, in concert with automated processes, helps ensure the integrity of archived content through direct, active, and continuous use.

NLM is also committed to expanding PMC. New publishers and titles are being added regularly, and NLM has embarked on a program of back-issue digitization for the titles that are routinely depositing current content in PMC.

PMC is not identified specifically as a line item in the NIH or NLM budgets. In October 2004, a review of personnel, contract, and system (hardware/software) costs noted an annual cost of $2.3 million. This included most operating costs for staff, contract work, equipment, and software other than the cost of digitization of journal back issues.

APPENDIX 3

Publishers Included in Each Archiving Program

(except NLA PANDORA)

CISTI Csi (7)

American Society for Microbiology
Elsevier
Institute of Physics
Mary Ann Liebert, Inc.
NRC Research Press
Oxford University Press
Springer

CLOCKSS (12)

American Chemical Society
American Medical Association
American Physiological Society
Blackwell Publishing
Elsevier
Institute of Physics
John Wiley & Sons, Inc.
Nature Publishing Group
Oxford University Press
Sage Publications
Springer
Taylor & Francis

KB e-Depot (8)

BioMed Central
Blackwell Publishing
Brill Academic Publishers
Elsevier
Oxford University Press
Sage Publications
Springer
Taylor & Francis

kopal/DDB (3)

John Wiley & Sons, Inc.
Springer
Thieme

LANL-RL (10)

American Chemical Society
American Institute of Physics
American Physical Society
Elsevier
Institute of Electrical and Electronics Engineers
Institute of Physics
Institution of Electrical Engineers
John Wiley & Sons, Inc.
Royal Society of Chemistry
Springer

LOCKSS Alliance (25)

American Historical Association
American Meteorological Society
American Society of Hematology
American Society of Plant Biologists
Berkeley Electronic Press (bepress)
BioMed Central
BioOne
Blackwell Publishing
Cambridge University Press
Emerald Group Publishing Ltd.
HighWire Press
Histochemical Society
History Cooperative
Indiana University Press
Institute of Physics
Lippincott, Williams & Wilkins
Nature Publishing Group
Open Access Titles (LOCKSS Humanities Project)
Oxford University Press
Project MUSE (Johns Hopkins University Press)
Royal Society of Chemistry
Sage Publications
Society for Industrial and Applied Mathematics (SIAM)
Springer
Usenix Association

OCLC ECO (41)

Adis International Ltd.
Alliance Communications Group
American Meteorological Society
Blackwell Publishing
Brill Academic Publishers
The British Psychological Society
Brookings Institution Press
Cambridge University Press
Edward Arnold Publishers Ltd.
Emerald Group Publishing Ltd.
Guilford Publications
The Haworth Press, Inc.
Humana Press
IOS Press
Idea Group Inc.
Institution of Mechanical Engineers
International Union of Pure and Applied Chemistry
John Wiley & Sons, Inc.
Karger Publishers
Lawrence Erlbaum Associates
MIT Press
M. E. Sharpe, Inc.
Mary Ann Liebert, Inc.
National Research Council of Canada
Nature Publishing Group
The Ohio State University Press
Organisation for Economic Co-Operation and Development
Oxford University Press
Palgrave Macmillan
Pan American Health Organization
Project MUSE (Johns Hopkins University Press)
RAND
Radcliffe Publishing Ltd.
Royal Society of Chemistry
Sage Publications
Springer
Taylor & Francis
Transaction Publishers
University of California Press
Walter de Gruyter GmbH & Co.
World Scientific Publishing Co.

OhioLINK EJC (36)

Adis International Ltd.
American Institute of Physics
American Physical Society
American Psychological Association
Association for Computing Machinery
Berkeley Electronic Press (bepress)
BioMed Central
BioOne
Blackwell Publishing
Brill Academic Publishers
The British Psychological Society
Bulletin of the Atomic Scientists
CABI Publishing
Cambridge University Press
Duke University Press
Edward Arnold Publishers Ltd.
Elsevier
Emerald Group Publishing Ltd.
Humana Press
Institute of Physics
John Wiley & Sons, Inc.
Kluwer Law International
Lawrence Erlbaum Associates
Multi-Science Publishing Co Ltd.
Optical Society of America
Oxford University Press
Project MUSE (Johns Hopkins University Press)
Rodopi
Royal Society
Royal Society of Chemistry
Royal Society of Medicine Press
Sage Publications
Springer
Taylor & Francis
Thieme
Transaction Publishers

Ontario Scholars Portal (13)

American Psychological Association
Berkeley Electronic Press (bepress)
Blackwell Publishing
Cambridge University Press
Elsevier
Emerald Group Publishing Ltd.
Institute of Electrical and Electronics Engineers
John Wiley & Sons, Inc.
Kluwer Law International
Oxford University Press
Royal Society of Chemistry
Sage Publications
Springer

Portico (13)

American Anthropological Association
American Mathematical Society
Annual Reviews
Berkeley Electronic Press (bepress)
BioOne
Elsevier
John Wiley & Sons, Inc.
Oxford University Press
Sage Publications
Society for Industrial and Applied Mathematics
Symposium Journals (Oxford UK)
United Kingdom Serials Group
University of Chicago Press

PubMed Central (52)

American Academy of Family Physicians
American Clinical and Climatological Association
American Epilepsy Society
American Hospital Association
American Medical Informatics Association
American Ophthalmological Society
American Society for Cell Biology
American Society for Clinical Investigation
American Society for Experimental NeuroTherapeutics
American Society for Microbiology
American Society of Human Genetics
American Society of Plant Biologists
Amphibian and Reptile Conservation
Australian Association of Clinical Biochemists
Baylor Health Care System
BioMed Central
Biological Procedures Online
Biophysical Society
Blackwell Publishing
British Medical Journal Publishing Group
Canadian Medical Association
Canadian Veterinary Medical Association
Cell Stress Society International
Centers for Disease Control
Cold Spring Harbor Laboratory Press
College of Family Physicians of Canada
Hindawi Publishing Corporation
Igitur Publishing and Archiving Services
Ivyspring International Publisher
Lippincott, Williams & Wilkins
Marshfield Clinic
Masson
Medical Library Association
National Academy of Sciences
National Athletic Trainers Association
National Institute of Environmental Health Science
Nature Publishing Group
North American Skull Base Society
Nuclear Receptor Signaling Atlas
Oxford University Press
Physicians Postgraduate Press, Inc.
Portland Press Ltd.
Public Library of Science
Royal College of General Practicioners
Royal Society of Medicine Press
ScholarOne, Inc.
Society for the Experimental Analysis of Behavior
Texas Heart Institute
University of Arizona Library
WebMD/Medscape Health Network
Wellcome Institute of the History of Medicine
Zhejiang University Press

APPENDIX 4 Multiprogram Publishers

I. Publishers Represented in More Than One of the E-Journal Archiving Programs (except NLA PANDORA), Listed in Descending Order, from Most to Least Commonly Represented

Publisher/Imprint	#	CL	CSI	ECO	EJC	KB	KOP	LA	LANL	OSP	PMC	PORT
Oxford University Press	9	•	•	•	•	•		•		•	•	•
Springer	9	•	•	•	•	•	•	•	•	•		
Blackwell Publishing	7	•		•	•	•		•		•	•	
Elsevier	7	•	•		•	•			•	•		•
John Wiley & Sons, Inc	7	•		•	•		•		•	•		•
Sage Publications	7	•		•	•	•		•		•		•
Institute of Physics	5	•	•		•			•	•			
Royal Society of Chemistry	5			•	•			•	•	•		
Berkeley Electronic Press (bepress)	4				•			•		•		•
BioMed Central	4				•	•		•			•	
Cambridge University Press	4			•	•			•		•		
Emerald Group Publishing Ltd	4			•	•			•		•		
Nature Publishing Group	4	•		•				•			•	
Taylor & Francis	4	•		•	•	•						
BioOne	3				•			•				•
Brill Academic Publishers	3			•	•	•						
Project MUSE (Johns Hopkins University Press)	3			•	•			•				
Adis International Ltd	2			•	•							
American Chemical Society	2	•							•			
American Institute of Physics	2				•				•			
American Meteorological Society	2			•				•				
American Physical Society	2				•				•			
American Psychological Association	2				•					•		
American Society for Microbiology	2		•								•	
American Society of Plant Biologists	2							•			•	
British Psychological Society	2			•	•							
Edward Arnold Publishers Ltd	2			•	•							
Humana Press	2			•	•							
Institute of Electrical and Electronics Engineers	2								•	•		
Kluwer Law International	2				•					•		
Lawrence Erlbaum Associates	2			•	•							
Lippincott, Williams & Wilkins	2							•			•	
Mary Ann Liebert, Inc	2		•	•								
Royal Society of Medicine Press	2				•						•	
Society for Industrial and Applied Mathematics	2							•				•
Thieme	2				•		•					
Transaction Publishers	2			•	•							

Multiprogram Publishers

II. Publishers Represented in More Than One of the E-Journal Archiving Programs (except NLA PANDORA), Listed Alphabetically

Publisher/Imprint	#	CL	CSI	ECO	EJC	KB	KOP	LA	LANL	OSP	PMC	PORT
Adis International Ltd	2			•	•							
American Chemical Society	2	•							•			
American Institute of Physics	2				•				•			
American Meteorological Society	2			•				•				
American Physical Society	2				•				•			
American Psychological Association	2				•					•		
American Society for Microbiology	2		•								•	
American Society of Plant Biologists	2							•			•	
Berkeley Electronic Press (bepress)	4				•			•		•		•
BioMed Central	4				•	•		•			•	
BioOne	3				•			•				•
Blackwell Publishing	7	•		•	•	•		•		•	•	
Brill Academic Publishers	3			•	•	•						
British Psychological Society	2			•	•							
Cambridge University Press	4			•	•			•		•		
Edward Arnold Publishers Ltd	2			•	•							
Elsevier	7	•	•		•	•			•	•		•
Emerald Group Publishing Ltd	4			•	•			•		•		
Humana Press	2			•	•							
Institute of Electrical and Electronics Engineers	2								•	•		
Institute of Physics	5	•	•		•			•	•			
John Wiley & Sons, Inc	7	•		•	•		•		•	•		•
Kluwer Law International	2				•					•		
Lawrence Erlbaum Associates	2			•	•							
Lippincott, Williams & Wilkins	2							•			•	
Mary Ann Liebert, Inc	2		•	•								
Nature Publishing Group	4	•		•				•			•	
Oxford University Press	9	•	•	•	•	•		•		•	•	•
Project MUSE (Johns Hopkins University Press)	3			•	•			•				
Royal Society of Chemistry	5			•	•			•	•	•		
Royal Society of Medicine Press	2				•						•	
Sage Publications	7	•		•	•	•		•		•		•
Society for Industrial and Applied Mathematics	2							•				•
Springer	9	•	•	•	•	•	•	•	•	•		
Taylor & Francis	4	•		•	•	•						
Thieme	2				•		•					
Transaction Publishers	2			•	•							