Preserving e-Journals: A review of the digital preservation landscape

Ken DiFiore, MLS, MS
Director, Library Relations

DELTA
SERIALS
CONFERENCE
ARKANSAS STATE UNIVERSITY | DEAN B. ELLIS LIBRARY

PORTICO

Evolution of Digital Content: Global Perspective

1996

2006

2016

Increase in the availability and usage of data and analytics

adam vincenzini [dot] com

DELTA
SERIALS
CONFERENCE
ARKANSAS STATE UNIVERSITY | DEAN B. ELLIS LIBRARY
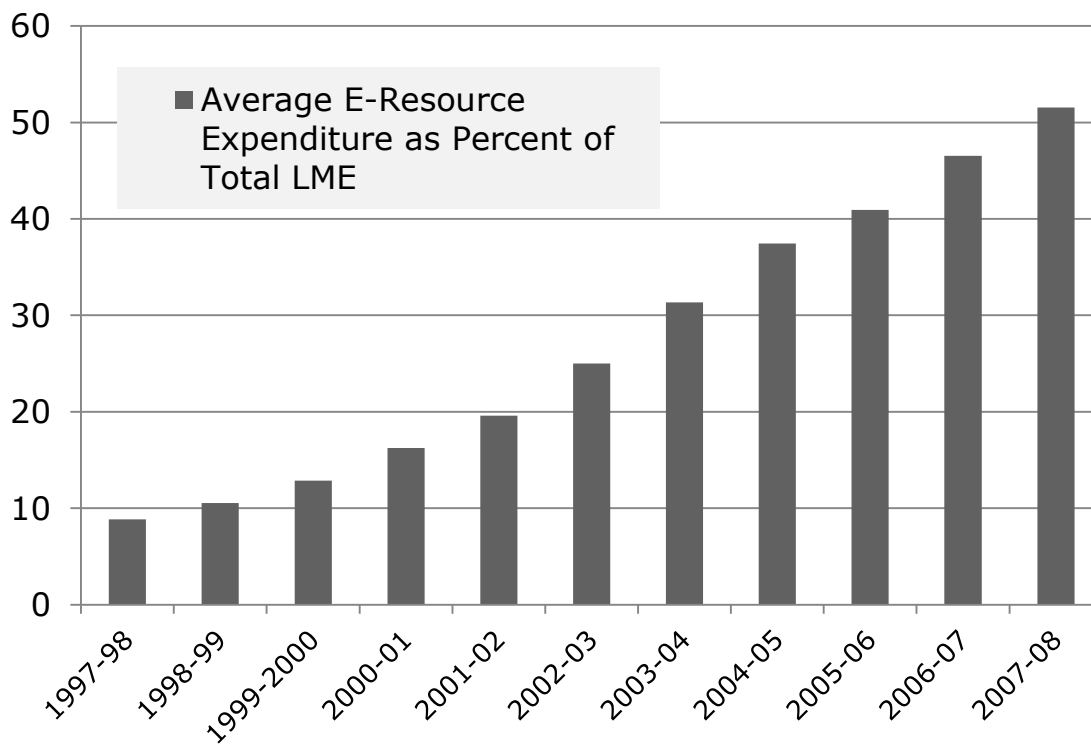
Evolution of Digital Content: Library Perspective

ARL STATISTICS
2007–2008

Compiled and Edited by

MARTHA KYRILLIDOU
LES BLAND

ASSOCIATION OF RESEARCH LIBRARIES
WASHINGTON, DC
2009

Average E-Resource Expenditure as Percent of Total LME
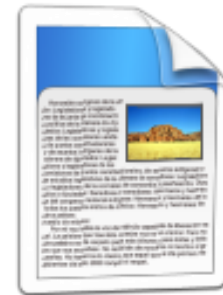
Evolution of Digital Content: Library Perspective

❖ Accommodate patron's demand for the convenience and accessibility of e-resources.

❖ Print processing and storage costs becoming prohibitive.

❖ Require floor space for new collaborative work environments.

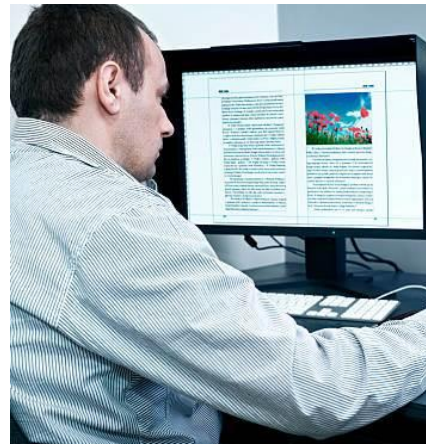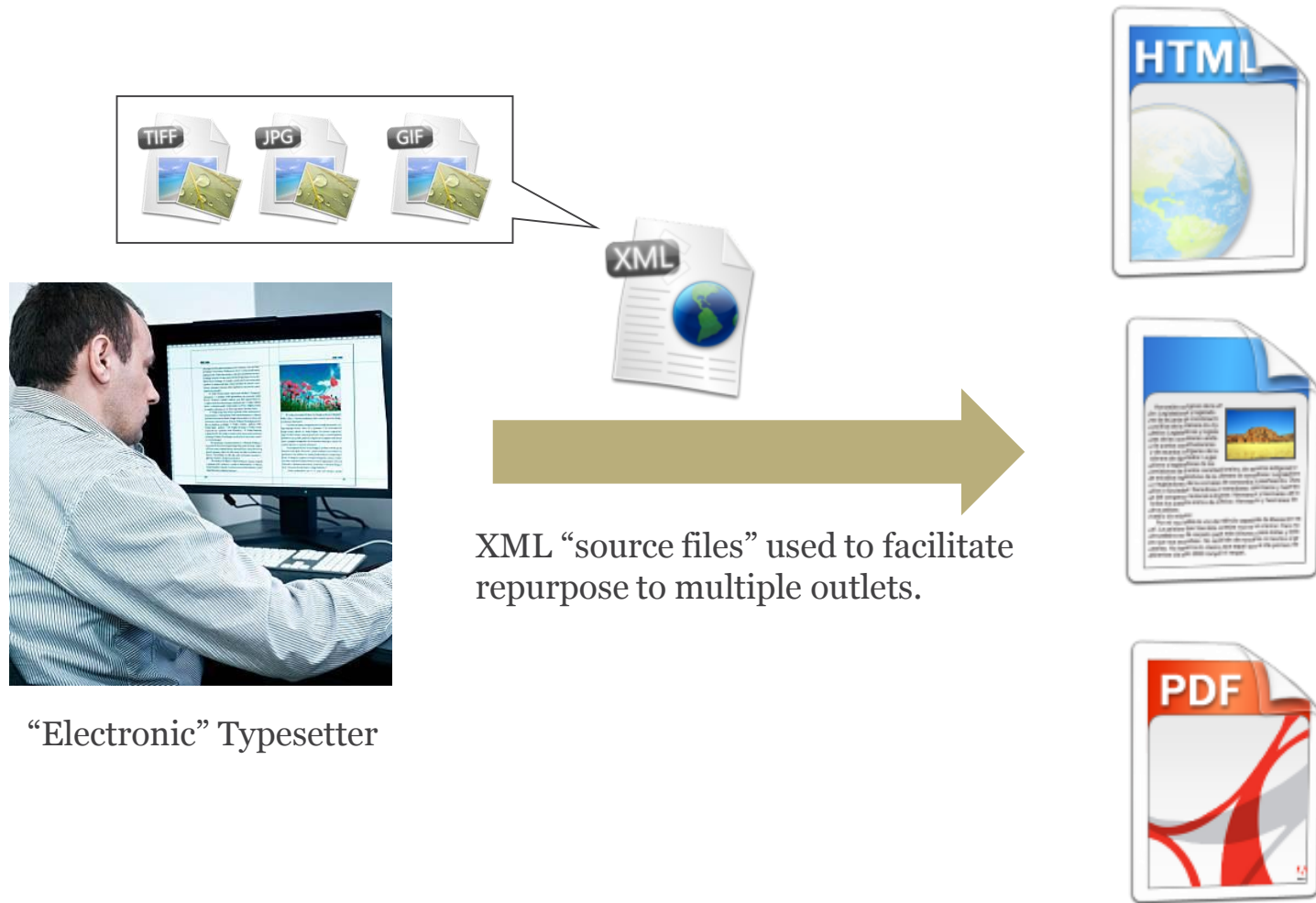❖ Maintain relevance: in the digital environment, networks and context are everything.

Evolution of Digital Content: Publisher Perspective

Print Typesetter

Evolution of Digital Content: Publisher Perspective

"Electronic" Typesetter

XML "source files" used to facilitate repurpose to multiple outlets.

Why Digital Preservation?

❖ Digital resources are now generally licensed rather than owned. Preservation and access are separated and the ownership of the preservation task is less clear.

❖ Multitude of electronic formats.

❖ Digital resources are surprisingly fragile and increasingly complex and innovative raising significant preservation challenges.

Influential Reactions from Library Community

The E-only Tipping Point for Journals

What's Ahead in the
Print-to-Electronic Transition Zone

Richard K. Johnson and Judy Luther

Association of Research Libraries

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.

Influential Reactions from Library Community

1. Research and academic libraries and associated academic institutions must recognize that preservation of electronic journals is a kind of insurance

2. Qualified preservation archives must provide a minimal set of well-defined services.

3. Libraries must invest in a qualified archiving solution

Urgent Action Needed to Preserve Scholarly Electronic Journals

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

The shift from print to electronic scholarly journals is occurring at a particularly rapid pace. Researchers, teachers and other readers demand electronic formats because it provides so many retrieval. Recognizing the great electronic versions of journals a shifting their business models separately so that they can surv and academic libraries are incre licenses both to satisfy user de ordering, receiving, cataloging,

Endorsed by:
ICOLC — International Coalition of Library Consortia
ALA — American Library Association
ASSOCIATION OF RESEARCH LIBRARIES

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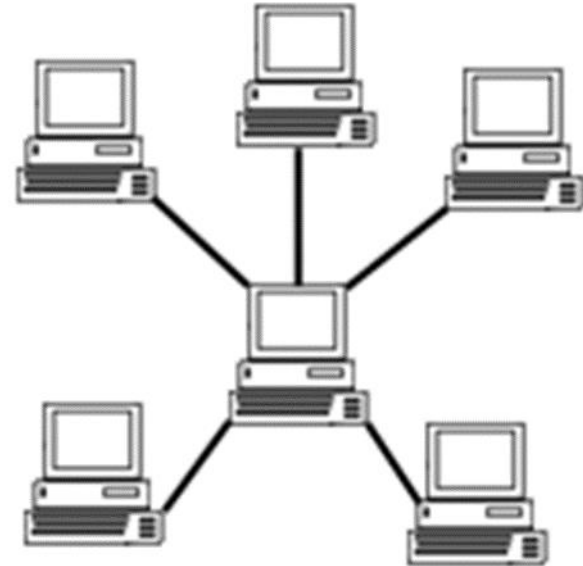
Collaborative Solutions: LOCKSS

❖ The LOCKSS Program is an open-source, library-led digital preservation system built on the principle that "lots of copies keep stuff safe."

❖ Allows libraries to take custody of and preserve access to the journals to which they subscribe.

❖ Analogous to libraries' using their own buildings, shelves and staff to obtain, preserve and provide access to paper content.

❖ Initiated and managed at Sanford University Libraries, with a non-profit business model.

LOCKSS
LOTS OF COPIES
KEEP STUFF SAFE

DELTA
SERIALS
CONFERENCE
ARKANSAS STATE UNIVERSITY | DEAN B. ELLIS LIBRARY

Collaborative Solutions: LOCKSS

❖ LOCKSS ingests content from publishers websites using a web crawler similar to those used by search engines.

❖ Preserves content by continually comparing the content it has collected with the same content collected by other LOCKSS Boxes, and repairing any differences.

❖ Delivers authoritative content to readers by acting as a web proxy, cache or via Metadata resolvers when the publisher's website is not available.

Collaborative Solutions: CLOCKSS

❖ CLOCKSS is a geographically distributed dark archive with which to ensure the long-term survival of Web-based scholarly publications for the benefit of the greater global research community.

❖ Built on low-cost, open-source, award-winning LOCKSS technology, the CLOCKSS archive comprises a network of redundant nodes located at 12 major research libraries, into which e-content is ingested, copied, and preserved. CLOCKSS's decentralized, geographically disparate preservation model ensures that the digital assets of the community will survive intact.

❖ Operates as a non-profit business.
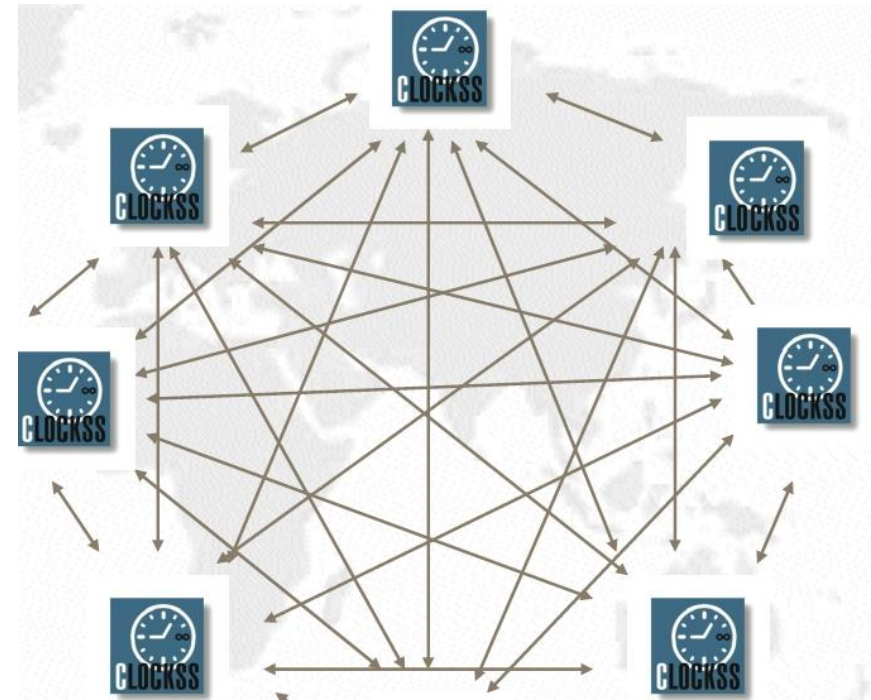
Collaborative Solutions: Portico

❖ Initiated by JSTOR, managed by ITHAKA.

❖ Portico is a centralized, replicated, and "dark" repository, utilizing a managed preservation methodology.

❖ Access is provided to participating libraries following a trigger event or post-cancellation access claim.

❖ Establishes and executes a preservation plan for each journal, which may include an initial migration from publisher specific article XML to the NLM archival standard.

PORTICO

Files Arrive

Files Are Processed
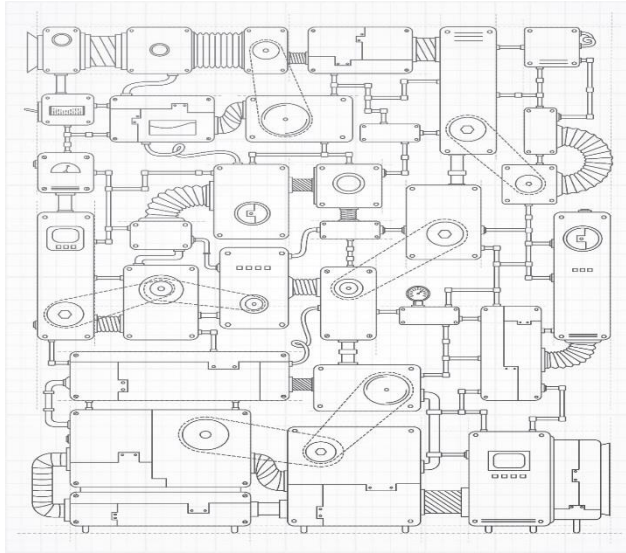
Files Preserved & Replicated
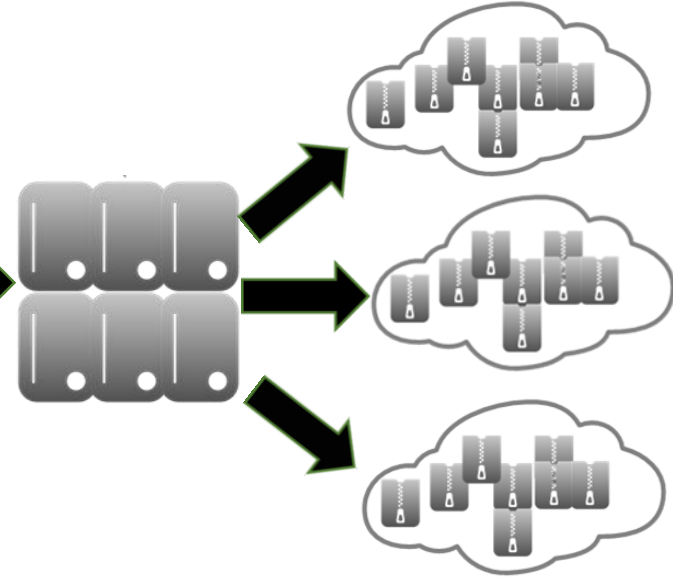
- Validating files against their format specifications
- Verifying checksums
- Extracting key technical and descriptive metadata to record in the preservation metadata
- Capturing all actions taken on the content and recording each as an event in the preservation metadata
- Creating a checksum for each file and recording it in the preservation metadata
- Creating a Portico archival information package
- Analyzing the structure of the content to identify all the files and determine whether all files received were expected.
- Transforming descriptive metadata or structured full-text to an archival format.

Collaborative Solutions: JSTOR

❖ Launched in 1997, JSTOR helps libraries provide adequate space for an ever-increasing amount of published scholarship.

❖ Digitizes and preserves printed scholarly journals and provides access through research and teaching platform.

❖ Collaborates with Harvard University and the University of California to support maintenance of print journals in under JSTOR's digital stewardship.

Collaborative Solutions: JSTOR

ITHAKA

ITHAKA is a not-for-profit organization that helps the academic community use digital technologies to preserve the scholarly record and to advance research and teaching in sustainable ways.

JSTOR   PORTICO

❖ Journal content under the stewardship of JSTOR and Portico is treated with similar technological tools, techniques, and best practices.

"Light" vs "Dark" Archive

Dark Archive

Light Archive

❖ "Trigger Events" dictate access, with the exception of Portico, which in addition provides post-cancellation and audit access:

    A. Publisher ceases business
    B. Journal abandoned
    C. Back issues eliminated

❖ JSTOR: Participation in archive collection ensures access

❖ LOCKSS: Publisher site no longer provides access.

Scholarly Publisher Participation

❖ There would be no digital preservation initiatives without publisher participation!

❖ Demonstrated ongoing willingness to collaborate with LOCKSS, CLOCKSS, and Portico.

Other Digital Preservation Service Providers

Non-profit / Academic

Commercial

DPN — THE DIGITAL PRESERVATION NETWORK

ACADEMIC PRESERVATION TRUST

CHRONOPOLIS

HathiTrust

DURACLOUD™

ExLibris
a ProQuest Company

Preservica®
Digital Preservation

Arkivum
Every bit archived

❖ No specificity for electronic journals. Intellectual property rights guide the scope of preservation targets.

Evaluating Digital Preservation Service Providers

Evaluating Initiatives: Trustworthiness

- ❖ In 2006, CRL began certifying trustworthiness of repositories.

- ❖ Metrics based on Trusted Repository Archiving Checklist (TRAC):
    - A. Organizational Infrastructure
    - B. Digital Object Management
    - C. Technologies, Technical Infrastructure and Security

- ❖ CRL consults a panel of advisers who represent the various sectors of its membership, ensuring that the certification process addresses the interests of the entire community.

Center *for* Research Libraries
GLOBAL RESOURCES NETWORK

Evaluating Initiatives: Trustworthiness

Portico Audit Report 2010

Release Date Friday, January 1, 2010

Downloads 📄 **Portico Audit Report 2010**

The Center for Research Libraries (CRL) conducted a preservation audit of Portico (**www.portico.org**) between April and October 2009 and, based on that audit, has certified Portico as a trustworthy digital repository. CRL found that Portico's services and operations basically conform to the requirements for a trusted digital repository. The CRL Certification Advisory Panel concluded that the practices and services described in Portico's public communications and published documentation are generally sound and appropriate to both the content being archived and the needs of the CRL community. Moreover, the CRL Certification Advisory Panel expects that in the future, Portico will continue to be able to deliver content that is understandable and usable by its designated user community.

This finding is based upon a site visit and sampling of archives content, and upon the review of information gathered by CRL and its Certification Advisory Panel and documents and documentation provided by Portico. CRL's analysis was guided by the criteria included in the *Trustworthy Repositories Audit and Certification* checklist, and other metrics developed by CRL on the basis of its analyses of digital repositories.

CRL conducted its audit with reference to generally accepted best practices in the management of digital systems; the interests of its community of research libraries; and the practices and needs of scholarly researchers in the humanities, sciences and social sciences in the United States and Canada. The purpose of the audit was to obtain reasonable assurance that Portico provides, and is likely to continue to provide, services adequate to those needs without material flaws or defects and as described in Portico's public disclosures. The CRL audit provides a reasonable basis for these findings.

CRL has assigned Portico the following levels of certification (the numeric rating is based on a scale of 1 through 5, with 5 being the highest level, and 1 being the minimum certifiable level): [1]

Category	Portico Score
Organizational Infrastructure	3
Digital Object Management	4
Technologies, Technical Infrastructure, Security	4

Center for Research Libraries
GLOBAL RESOURCES NETWORK

Report on Portico Audit Findings

Evaluating Initiatives: Trustworthiness

CLOCKSS Audit Report 2014

Release Date Tuesday, July 1, 2014

Downloads 📄 CLOCKSS Report 2014

The Center for Research Libraries (CRL) conducted a preservation audit of CLOCKSS (www.clockss.org/) between September 2013 and May 2014, and on the basis of that audit certified CLOCKSS as a trustworthy digital repository of e-journal content. The CRL Certification Advisory Panel concluded that the practices and services described in CLOCKSS' public communications and published documentation generally correspond to the operations of CLOCKSS and are appropriate to the e-journal content being archived and to the expressed needs of the CLOCKSS designated community. Moreover the panel indicated its expectation that in the future, CLOCKSS will be able to deliver the content it preserves to appropriate third parties who are equipped to make it available for use by the designated community. CRL certification applies to the repository's ability to preserve and manage digital content deposited by participating e-journal publishers as of May 2014.

The present report is based upon review, by CRL and the members of its Certification Advisory Panel, of extensive documentation gathered by CRL independently from open sources and from third parties as well as data and documentation provided by CLOCKSS. The review also included a site visit by CRL audit personnel to the offices of the LOCKSS team in Redwood City, California. CRL's evaluation of CLOCKSS and the information provided in this report reflect the policies, systems and procedures that were in place at CLOCKSS to manage e-journal content as of June 1, 2014.

On the basis of this evidence, the certification panel concluded that overall CLOCKSS can be recognized by its designated community as a trustworthy repository. However, in the course of the audit, the Certification Advisory Panel identified one issue that CLOCKSS will need to address to more fully satisfy the concerns of its research library constituents: the lack of a formal succession plan. In addition, two aspects of CLOCKSS operations became apparent that should be understood by stakeholders, as they may have a bearing on future CLOCKSS services. Those issues are described in the section titled, *Detailed Audit Findings*, with reference to the corresponding criteria in the TRAC checklist. CLOCKSS has agreed to address the succession plan issue and also to make certain disclosures to CRL periodically, as a condition of continued certification. Those ongoing requirements are outlined in Section C of this report.

CRL assessed CLOCKSS on each of the three categories of criteria specified in TRAC, and has assigned the level of certification below for each. The numeric rating used is based on a scale of 1 through 5, with 5 being the highest level, and 1 being the minimum certifiable level.

TRAC Category	CLOCKSS rating	Optimum rating
Organizational Infrastructure	4	5
Digital Object Management	4	5
Technologies, Technical Infrastructure, Security	5	5
TOTAL	**13**	**15**

The basis for assigning these ratings is provided *in Detailed Audit Findings*.

Center for Research Libraries
GLOBAL RESOURCES NETWORK
Certification Report on CLOCKSS

Comparing & Contrasting Initiatives

Comparing Initiatives: What's best for your library?

- ✓ Trustworthiness
- ❑ Journal Coverage
- ❑ Intellectual Property Rights
- ❑ Transparency
- ❑ Technological approach
- ❑ Sustainability
- ❑ Cost

Comparisons from Global Library Community

E-JOURNAL ARCHIVING FOR

UK HE LIBRARIES

A White Paper

Version 4.0 Final Draft 10/01/2011

Prepared by:

Charles Beagrie Limited

www.beagrie.com

funded by

JISC

Charles Beagrie

ENSURING PERPETUAL ACCESS:

ESTABLISHING A FEDERATED STRATEGY ON

PERPETUAL ACCESS AND HOSTING OF

ELECTRONIC RESOURCES FOR GERMANY

Edited Final Report – February 2010

Prepared by:

Charles Beagrie Limited in association with Globale Informationstechnik GmbH

www.beagrie.com

A study funded by

The Alliance of German Science Organisations

Final Report of the 2CUL LOCKSS Assessment Team
Cornell University Library & Columbia University Library

Report Completed: March 2011
Public Release: October 2011

Introduction

Although LOCKSS is considered a successful digital preservation initiative, neither of the CULs feels that they fully understand the potential of the system for their own settings and collections. There is a range of practical issues that need to be explored in order to leverage this preservation system. In support of this goal, a joint team was established in November 2010 to investigate a range of questions to assess how LOCKSS is being deployed and the implications of local practices for both CUL's preservation frameworks. This study was seen as a high-level investigation to characterize the general landscape and identify further research questions.

The team worked with a condensed timeline, November 2010-December 2011, and investigated the following questions:

1) To build a collection of preserved journals beyond the journals already preserved via the LOCKSS network, a library needs to select the titles it wants to preserve (subscription or open access). Working with the Stanford LOCKSS team, the next step is to obtain or confirm the publishers' permission to preserve the titles in the system. The Stanford team is responsible for building and testing the required publisher-specific "LOCKSS plugin." The size of the LOCKSS Alliance ensures each title chosen for preservation has a critical mass of preserving institutions. How does this process work for Columbia and Cornell? Who is involved in overseeing this process and tracking such collection decisions?

2) What needs to happen when a journal is canceled to have access to back issues? What kind of a mechanism needs to be put in place between the ERM license record for journal subscriptions (library management systems) and the local LOCKSS box to support uninterrupted access to digital content? What is the internal monitoring mechanism - all manual or can a part of it be automated?

3) Neither of the institutions has chosen to participate in CLOCKSS. Do we have a sufficient understanding of the difference between these two strategies? LOCKSS provides a community approach to long term preservation of a library's local content while CLOCKSS aims to provide a long-term global archiving solution that will serve the joint library and publisher communities in the event of a long-term business interruption or in making orphaned or abandoned works readily available to the scholarly community.

4) How do we keep track of which e-subscriptions are represented in LOCKSS to understand their preservation status?

2CUL LOCKSS Assessment Study, October 2011, p. 1

BIRMINGHAM
City University

**Evaluation of the JISC UK
LOCKSS Pilot**

May 2008

Pete Dalton and Dr Angela Conyers

**Evidence Base
Research and Evaluation Services
Library and Learning Resources
Birmingham City University**

DELTA SERIALS CONFERENCE
ARKANSAS STATE UNIVERSITY | DEAN B. ELLIS LIBRARY

Comparisons from Global Library Community

The Keepers Registry
Supporting long-term access to journal content

Home | Journals | Publishers | Archiving Agencies | Member Services | FAQ | About | Help

Discover who is looking after your e-journals

Search 🔍

Input a search term (e.g. free text, or a title or ISSN(s))

[   ] [Search]

➕ Show search examples
📘 Glossary and definitions

The following organisations are the Keepers of digital content, working on your behalf to ensure long-term access to the scholarly and cultural record. They provide the registry with information on their archival holdings, ordered by most recent update (date of which is shown):

- Archaeology Data Service (09 Jul 2017)
- Global LOCKSS Network (08 Jul 2017)
- CLOCKSS Archive (07 Jul 2017)
- Cariniana Network (06 Jul 2017)
- HathiTrust (05 Jul 2017)
- Portico (05 Jul 2017)
- Scholars Portal (04 Jul 2017)
- British Library (03 Jul 2017)
- Library of Congress (03 Jul 2017)
- PKP PLN (01 Jul 2017)
- Swiss National Library (22 May 2017)
- National Science Library, Chinese Academy of Sciences (13 Feb 2017)
- e-Depot (24 Apr 2012)

The most up-to-date information may be available on the Keeper's website.

DELTA SERIALS CONFERENCE
ARKANSAS STATE UNIVERSITY | DEAN B. ELLIS LIBRARY

Snapshot of Portico's Preservation Progress

80million ARCHIVAL UNITS PRESERVED

11% growth IN OVERALL CONTENT IN THE ARCHIVE IN 2017

3% INCREASE IN E-JOURNALS PRESERVED

3% INCREASE IN E-BOOKS PRESERVED

1% INCREASE IN D-CCOLLECTION ITEMS PRESERVED

CONTENT COMMITTED TO THE ARCHIVE
TOTAL NUMBER OF E-BOOKS AND JOURNALS COMMITTED

802,000+ E-BOOKS

26,000+ E-JOURNALS

47 NEW PUBLISHERS JOINED YTD

447 TOTAL PARTICIPATING PUBLISHERS

PORTICO

Portico's Audit Site

Portico's Audit Site

Portico's Holdings Comparison Tool

ACTA HAEMATOLOGICA

Publisher Karger

Identifier Content Set ID ISSN_00015792

Completeness (□ as of Aug 12, 2016)

Years Preserved: 1983, 1998-2016 (v.1-135)

Portico Holdings: 2010 - v. 1 (1-4), 2009/2010/2012 - v. 2 (1-6), 2009/2010 - v. 3 (1-6), 2009 - v. 4 (1-6), 2009 - v. 5 (1-6), 2009 - v. 6 (1-6), 2009 - v. 7 (1-6), 2009 - v. 8 (1-6), 2009 - v. 9 (1-6), 2009 - v. 10 (1-6), 2009 - v. 11 (1-6), 2009 - v. 12 (1-6), 2009 - v. 13 (1-6), 2009 - v. 14 (1-6), 2009 - v. 15 (1-6), 2009 - v. 16 (1-6), 2009 - v. 17 (1-6), 2009 - v. 18 (1-6), 2009 - v. 19 (1-6), 2009 - v. 20 (1-6), 2009 - v. 21 (1-6), 2009 - v. 22 (1-6), 2009 - v. 23 (1-6), 2009 - v. 24 (1-6), 2009 - v. 25 (1-6), 2009 - v. 26 (1-6), 2009 - v. 27 (1-6), 2009 - v. 28 (1-6), 2009 - v. 29

Status Journal incomplete

	1	2	3	4	5	6
v.1 (2010)	n.1: Complete Pub Date: 2010-05-25 Articles: 8	n.2: Complete Pub Date: 2010-05-25 Articles: 8	n.3: Complete Pub Date: 2010-05-25 Articles: 5	n.4: Complete Pub Date: 2010-05-25 Articles: 8	n.5: Missing Pub Date: 2010-05-25 Articles: 8	n.6: Missing Pub Date: 2010-05-25
v.2 (2009/2010/2012)	n.1: Complete Pub Date: 2010-05-25 Articles: 5	n.2: Complete Pub Date: 2010-05-25 Articles: 6	n.3: Complete Pub Date: 2012-10-22 Articles: 8	n.4: Complete Pub Date: 2009-02-18 Articles: 7	n.5: Complete Pub Date: 2009-02-18 Articles: 5	n.6: Complete Pub Date: 2009-02-18 Articles: 8
v.3 (2009/2010)	n.1: Complete Pub Date: 2010-12-15 Articles: 7	n.2: Complete Pub Date: 2009-02-18 Articles: 8	n.3-4: Complete Pub Date: 2009-02-18 Articles: 12		n.5: Complete Pub Date: 2009-02-18 Articles: 7	n.6: Complete Pub Date: 2009-02-18 Articles: 6
v.4 (2009)	n.1: Co...					
v.5 (2009)	n.1: Co...					
v.6 (2009)	n.1: Co...					
v.7 (2009)	n.1: Co...					
v.8 (2009)	n.1-2: C...					
v.9 (2009)	n.1: Co...					

	1	2	3	4	
v.131 (2014)	n.1: Complete Pub Date: 2013-09-24 Articles: 23	n.2: Complete Pub Date: 2013-10-22 Articles: 21	n.3: Complete Pub Date: 2013-11-14 Articles: 27	n.4: Complete Pub Date: 2013-12-13 Articles: 24	No more issues
v.132 (2014)	n.1: Complete Pub Date: 2014-01-14 Articles: 35	n.2: Complete Pub Date: 2014-01-01 Articles: 25	n.3-4: Complete Pub Date: 2014-01-01 Articles: 23		No more issues
v.133 (2014/2015)	n.1: Complete Pub Date: 2015-01-01 Articles: 26	n.2: Complete Pub Date: 2015-01-01 Articles: 23	n.3: Unknown Pub Date: 2015-01-01 Articles: 11	n.4: Complete Pub Date: 2015-01-01 Articles: 10	No more issues
v.134 (2015)	n.1: Complete Pub Date: 2015-01-01 Articles: 11	n.2: Complete Pub Date: 2015-01-01 Articles: 12	n.3: Complete Pub Date: 2015-01-01 Articles: 14	n.4: Complete Pub Date: 2015-01-01 Articles: 12	No more issues
v.135 (2015/2016)	n.1: Complete Pub Date: 2016-01-01 Articles: 12	n.2: Complete Pub Date: 2016-01-01 Articles: 14	n.3: Unknown Pub Date: 2016-01-01 Articles: 9	n.4: Missing Pub Date: 2016-01-01	No more issues

PORTICO

Portico's Comparison with C|LOCKSS

	Portico	CLOCKSS	LOCKSS
Total Titles	26,066	20,163	12,897
Titles Only Here	8,428	2,095	844
% of CrossRef	48%	38%	25%
Total Years (Volumes)	488,998	212,347	111,219
Participating Libraries	1,010	285	u/k

Portico's OA Preservation

Portico is now actively engaging publishers to update their preservation agreements to allow access to their OA content to be opened, in case of a trigger event, not only to Portico members but to anyone around the world.

- John Wiley and Sons
- Elsevier
- Sage
- Cappelen Damm Akademisk
- American Astronomical Society
- Science Publishing Corporation
- Csurgeries
- EDP Sciences
- Scientific Research Publishing
- InTech
- Methaodos Revista de Ciencias Sociales
- PeerJ
- Exercise Fitness Health Alliance
- Academic and Business Research Institute
- The Sax Institute
- Elmer Press Inc.
- African Online Scientific Information Systems Ltd.

And more…

Why Participate in Digital Preservation Initiatives?

Value of Investment in e-Journal Digital Preservation

❖ To ensure your library's secure and reliable transition from print to a reliance on electronic journals.

❖ Protects library's investment in electronic journals.

❖ Facilitates deaccessioning of print journals, thereby freeing up space in the library for collaborative work areas.

What if we don't participate…

Q&A

Ken DiFiore
Outreach Director

ken.difiore@portico.org
www.portico.org