Digital Preservation of Academic Content: A Global Overview

Randy S. Kiefer
Executive Director
The CLOCKSS Archive
randy.kiefer@clockss.org

I SINPRED
Brasilia Brazil 7-9 May 2014

Preservation of Digital Content

- **Long-term preservation** refers to the processes and procedures required to ensure content remains accessible well into the future.

- In other words, continuing or perpetual access is an attempt to replicate the situation with paper journals where a library receives, makes available and preserves the material for ongoing reference, regardless of whether or not the subscription is continued. Long term preservation, on the other hand, can be viewed as an issue, not just for the subscribing library, but for society as a whole, ensuring that the scholarly record continues to be accessible to future generations.

(after Morrow et al 2008, provided by Neil Beagrie)

Community Moving to e-Resource Environment

- Patron’s Demand for the convenience and accessibility of e-resources
- High cost of maintaining dual formats (print & online)
- Definitive versions are now the online renditions
- Online becoming preferred option

Characteristics of e-Resources Differ from Print

- Digital resources are now licensed, not owned
- Usability of digital content is technology dependent
- Inherent fragility due to the pace of technological change
- Multitude of electronic formats (PDF, XML, DOC, etc.)

Preservation Reasoning

- Preservationists become keepers of the content in case of:
  - Publisher failure and no pick up of their assets
  - Discontinuation of a journal and publisher removes access
  - Disaster disrupts publisher's availability for extended period of time

`(continued)

- Market demand by libraries that want to be assured there is an independent third-party preservation of electronic content.
- Transition from physical content on shelves to “access” to electronic content.
- Publishers want to be good stewards of their content

The Value of Preservation to Libraries

- An “insurance policy” for e-resources
- Provides libraries with access to archived content when it becomes lost, orphaned, or abandoned (Publisher ceases operation, discontinues title, or drops back file)

What is Not Preservation: Commercial Hosting

- Includes aggregation databases
  - Collection of multiple titles and multiple publishers that are available for a subscription fee
- Includes journal hosting platforms
  - Commercial sites for the publishers where libraries can find their subscription content
- Includes distribution platforms (ebooks)
  - Commercial sites for access
- These are not preservation, not an archive
  - Content is only available behind a subscription wall and generally for the duration of the agreement

What is Not Preservation: Aggregators

- Not preservation archives – some examples
  - Gale
  - EBSCOhost
  - ProQuest

Content is only available to subscribers. If publisher content disappears or is no longer available from publisher, content usually remains available to subscribers.

What Is Not Preservation: Journal Hosting Platforms

- Commercial arrangement where publisher's journals are hosted for access by subscribers
- If publisher does not pay due to closure, then the site and access is removed

- Atypon, Ingenta, HighWire, Silverchair, Metapress, Torrossa, Semantico, Impelsys and others

Two Types of Digital Preservation Archives

- **Global Archives**
  - CLOCKSS Archive, LOCKSS GLN
  - Portico
- **Regional Archives**
  - British Library
  - Dutch KB

Regional Archives

- Digital Preservation
- Generally the content is only available in the host country and often only within their buildings

The CLOCKSS Archive – A global Archive

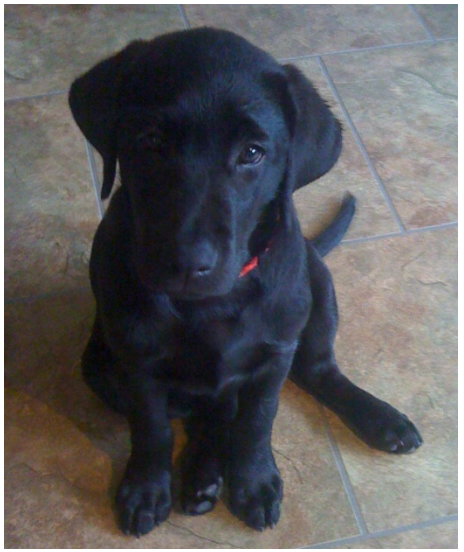
The beginning

In 2006, several of the world's largest scholarly publishers and top research libraries, challenged by the responsibility to preserve digital scholarly assets for the good of the entire community, joined forces to build a global archive for the very long term. Their unique collaboration emphasized community governance and a commitment to open access. Today, CLOCKSS (Controlled Lots of Copies Keep Stuff Safe) is a not-for-profit organization governed transparently and democratically by participants around the world.

CLOCKSS Statistics

- The CLOCKSS Archive as of December, 2010
- Number of Library Supporters = 86
- Number of Participating Publishers = 33
- Number of Archive Nodes installed = 7

- February 2011
  - First fulltime Executive Director
- My first office assistant →
- Charlie (our Labrador)
- 4 kilos



Principles of the CLOCKSS Archive

- Community Governed
- Global Approach: Decentralized Preservation
- Proven Technology using the open-source software LOCKSS
- Commitment to Open Access

Community Governed

CLOCKSS is designed to share the responsibility of archiving across the global academic community. Publishers and librarians have equal say in deciding procedures, priorities, and when to trigger content. CLOCKSS is committed to a transparent community-based governance structure, to be most sustainable in the future as technologies and priorities change.

Community Governed (continued)

Governing Libraries:

Roxanne Missingham, Australian National University
Peter Schirmbacher, Humboldt University
Brenda Johnson, Indiana University
Jun Adachi, National Institute of Informatics
Chip Nilges, OCLC [Treasurer]
Kerry Keck, Rice University
Michael Keller, Stanford University [Co-Chair]
Ellis Sada, Università Cattolica del Sacro Cuore
Geoff Harder, University of Alberta [Secretary]
Peter Burnhill, University of Edinburgh
Peter Sidorko, University of Hong Kong
Carla Lee, University of Virginia

Community Governed (continued)

Governing Publishers:

Vida Damijonaitis, American Medical Association
Rita Scheman, The American Physiological Society
Alicia Wise, Elsevier [Co-Chair]
Graham McCann, IOP Publishing
John Carroll, Nature Publishing Group
Mark Heaver, Oxford University Press
Carol Richman, SAGE Publications
David K. Marshall, SIAM
Wim van der Stelt, Springer
Ian Bannerman, Taylor & Francis
Craig Van Dyck, Wiley-Blackwell

Global Approach: Decentralized Preservation

Stewardship and Preservation
Reinforcing Social Value as Memory Organizations
Insuring against Geo-social and Geophysical Risks

Australian National University * *Australia*
Indiana University * *USA*
OCLC * *USA*
Stanford University * *USA*
University of Alberta * *Canada*
University of Hong Kong * *Hong Kong*
Humboldt University – Berlin * *Germany*
National Institute of Informatics * *Japan*
Rice University * *USA*
Università Cattolica del Sacro Cuore * *Italy*
University of Edinburgh * *United Kingdom*
University of Virginia * *USA*

Proven Open-Source Technology: LOCKSS

CLOCKSS runs on a proven digital preservation technology.

The LOCKSS technology has been safely and securely preserving web-published content for over 15+ years, and has evolved with web advances to preserve new content types.

Commitment to Open Access

Trigger Events include situations of non-availability of archived content in which:

Publisher No Longer in Business

The publisher is no longer in business or is no longer in the business of publishing content or providing access to previously published content and there are no successor interests or reversions or transfers of rights;

Title No Longer Offered

The publisher has stopped publishing and is no longer providing access to the content and there are no successor interests or reversion or transfer of rights;

Back Issues No Longer Available

The publisher has stopped offering or providing access to some or all of the back issues of the content and there are no successor interests or reversion or transfer of rights; or

Catastrophic Failure

While still publishing content, the publisher is not able to provide access to the content electronically due to technical or similar catastrophic and permanent failure.

In these cases, the Board of Libraries and Publishers can vote to trigger (release) the content.

CLOCKSS defines its designated community in three parts:

1. The scholars, students and readers of electronic academic content.

2. The libraries who purchase and manage this content on behalf of the scholars, students and readers.

3. The publishers of this content.

Services provided to the community

The CLOCKSS Archive delivers services to each component of its designated community by providing a sustainable, geographically distributed dark archive that ensures the long-term survival of Web-based, scholarly publications.

Scholars, students and readers are provided with free, open access to content that would otherwise have become inaccessible.

Librarians are reassured that the content which they purchase will remain accessible to their readers.

Publishers are relieved of the responsibility of providing for access to their content in the event that they no longer do so. The dark archive also does not allow access until triggered.

CLOCKSS Statistics today

- Supporting Libraries at 700+
  - 29 Countries
  - including Brazil through CAPES
- Participating Publishers = 200+
  - 34 Countries
  - SciELO
- Number of Archive Nodes installed = 12
  - Brazil is completing its application process to become our 13th node

2014

Now I am in my 4th year as Executive Director. Charlie is still my office assistant and is three years old and 41 kilos.

Thank You!

Randy S. Kiefer
Executive Director
The CLOCKSS Archive
randy.kiefer@clockss.org