

Digital Preservation of Academic Content: The Importance of Archiving

Kim Smilay
Director, Publisher Relations
The CLOCKSS Archive
pub-director@clockss.org

**CrossRef European Workshop
Vilnius
June 11, 2015**



The Importance of Archiving

Digital preservation combines policies, strategies and actions that ensure access to digital content over time.

<http://www.ala.org/alcts/resources/preserv/defdigpres0408>

The Importance of Archiving

Biggest Threats to Content:

- Format Obsolescence
- Media, Hardware, Software Obsolescence/Failure
- Economic Failure
- Natural Disaster
- Humans

Operator Error	Overwriting
Insider Attack	Indifference

The Importance of Archiving

- Market demand by libraries that want to be assured there is an independent third-party preservation of electronic content
- Centrally managed preservation of National collections preserved on national soil for safe-keeping
- Publishers want to be good stewards of content

The Importance of Archiving

- Preservationists become keepers of content in case a Trigger Event is needed:
 - Publisher failure and no pickup of their assets
 - Discontinuation of a journal and publisher removes access
 - Disaster disrupts publisher's availability for extended period of time

The Value of Preservation to Libraries

- An “insurance policy” for e-resources
- Provides all 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 - Content available behind a subscription wall, generally for the duration of the license agreement.

- **Aggregated Databases**

(examples: EBSCO, ProQuest, Gale)

Collection of multiple titles and multiple publishers that are available for a subscription

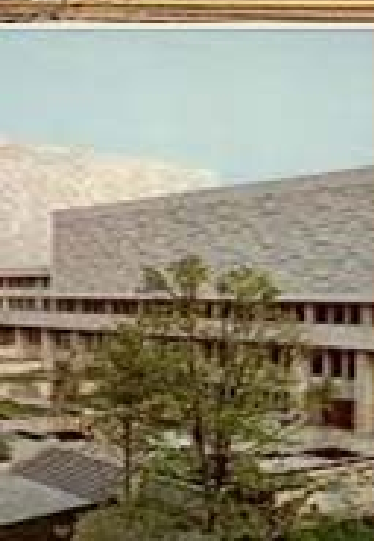
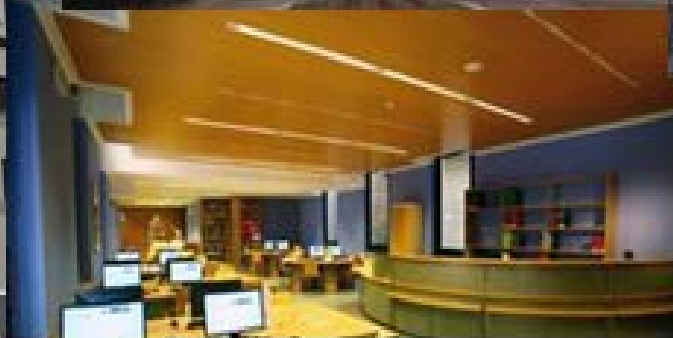
What is Not Preservation

Commercial Hosting - Content available behind a subscription wall, generally for the duration of the license agreement.

- **Journal-Hosting Platforms**
(examples: Atypon, HighWire, Ingenta, others)
- **e-book Platforms**

Two Types of Digital Preservation Archives

- Global Archives
 - CLOCKSS Archive
 - LOCKSS GLN
 - Portico
- Regional Archives - Generally the content is only available in the host country and often only within their buildings
 - British Library
 - Dutch KB



LIBRARY SUPPORTER



CLOCKSS



The CLOCKSS 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, **dark** 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.

Principles of the CLOCKSS Archive

- Community Governed
- Global Approach - Decentralized Preservation
- Proven Technology Using the Open-Source Software LOCKSS
- Commitment to Open Access

Principles of the CLOCKSS Archive

Community Governed

CLOCKSS is designed to share the responsibility of archiving across the global academic community

Publishers and Librarians have equal say:

- Procedures

- Priorities

- Trigger events

CLOCKSS is committed to a transparent community-based governance

Principles of the CLOCKSS Archive

Community Governed

Governing Libraries:

Roxanne Missingham, Australian National University

Peter Schirnbacher, 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

Principles of the CLOCKSS Archive

Community Governed

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

Principles of the CLOCKSS Archive

Global Approach – Decentralized Preservation

Nodes:

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

Principles of the CLOCKSS Archive

Proven Open-Source Technology - LOCKSS

CLOCKSS runs on proven, award-winning digital preservation technology

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

The LOCKSS technology has been adapted for dark archive functionality and for use by the CLOCKSS Archive

Scored a perfect 5 in Technologies, Technical Infrastructure and Security by Center for Research Libraries (CRL) in an independent audit

A Dark Archive is used as a repository of safekeeping of content that does not grant access. This limit on access protects the content from damage.

Principles of the CLOCKSS Archive

Commitment to Open Access

Trigger Events includes non-availability of archived content in which:

Publisher is

- No Longer in Business

- No Longer Provides Access

- No Successor Interests, Reversions or Transfer of Rights

Title is No Longer Offered

- No Longer Published

- No Longer Provides Access

- No Successor Interests, Reversions or Transfer of Rights

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

Principles of the CLOCKSS Archive

Commitment to Open Access

Trigger Events includes non-availability of archived content in which:

Back Issues

- No Longer Available

- Publisher No Longer Provides Access

- No Successor Interests, Reversions or Transfer of Rights

Catastrophic Failure

- Publisher is Not Able to Provide Access

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

CLOCKSS Community in Three Parts:

- Scholars, Students and Readers of electronic academic content
- Librarians who purchase and manage content on behalf of scholars, students and readers
- 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:

- **2010:**

- Number of Library Supporters = 86
- Number of Participating Publishers = 33
- Number of Archive Nodes installed = 7

- **2015:**

- Number of Library Supporters = 725+
- Number of Participating Publishers = 200
- Number of Archive Nodes Installed = 12

Thank You!