

# Digital Preservation of Academic Content: A Global Overview

Randy S. Kiefer  
Executive Director  
The CLOCKSS Archive  
[randy.kiefer@clockss.org](mailto:randy.kiefer@clockss.org)



REINVENTING UNIVERSITY PUBLISHING  
CAUL Library Publishing Advisory Committee  
**ANU 17<sup>th</sup>-18 March 2015**

# 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)

# Preservation of Digital Content

- 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 their content

# Preservation of Digital Content

- Preservationists become keepers of the 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

- 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 a license agreement

# What Is Not Preservation:

## Aggregators

Not preservation archives – some examples

- Gale
- EBSCOhost
- ProQuest
- Reveal Digital

Content is only available to subscribers. If publisher content disappears or is no longer available from the 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**
- **HighWire**
- **Impelsys**
- **Ingenta**
- **Silverchair**
- **Semantico**
- **Torrossa 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.
- Often include digitized collections of local/regional interest.



**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 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.



# Principles of the CLOCKSS Archive: Community Governed

## Governing Libraries:

Roxanne Missingham, Australian National University

Peter Schirmbacher, Humboldt University

Carolyn Walters, 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

## Stewardship and Preservation

## Reinforcing Social Value as Memory Organizations

## Insuring Against Geosocial and Geophysical Risks

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



# Principles of the CLOCKSS Archive: Proven Open-Source Technology - LOCKSS

CLOCKSS runs on a proven, award-winning, 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.

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

**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** include situations of nonavailability 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;

**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** include situations of nonavailability of archived content in which:

### **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: News

- Brazil is completing its application process to become our 13th node.
  - **Number of current Archive Nodes installed = 12**
  - **CLOCKSS Board has authorized 15 archive nodes.**
- CAPES negotiated a countrywide deal with CLOCKSS covering all major institutions.
- Opportunity for another archive node in Europe.
- TRAC Audit results: CLOCKSS Archive is a **Trusted Digital Repository**.
  - **Perfect score in Security and Technology**

# CLOCKSS Statistics: 2010



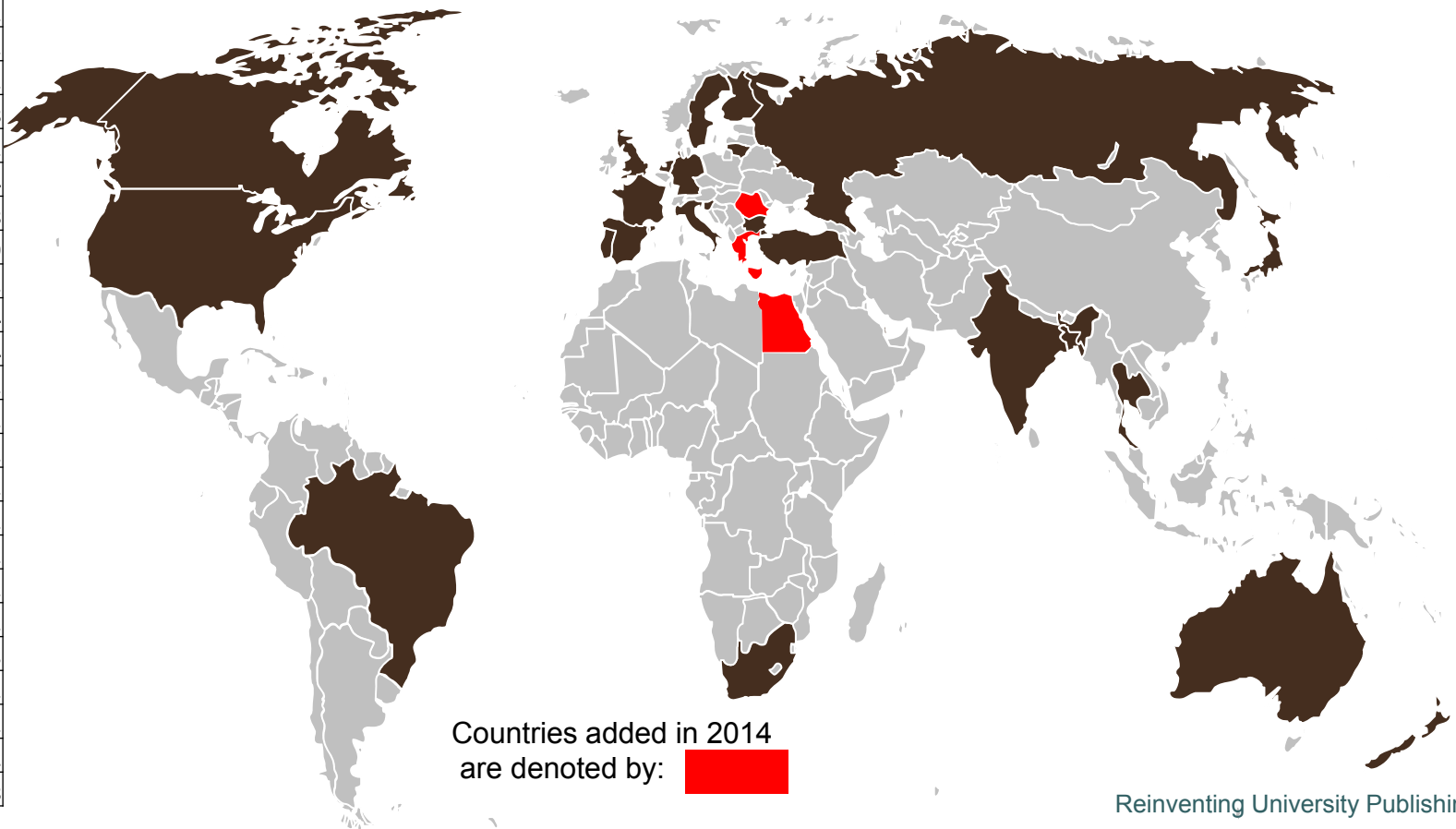
- The CLOCKSS Archive as of December, 2010
  - Number of Library Supporters = 86
  - Number of Participating Publishers = 33
  - Number of Archive Nodes installed = 7
- 
- February 2011
  - First full-time Executive Director
  - First office assistant →
    - Charlie (my Labrador)
    - 4 kilos



# CLOCKSS Statistics: 2014

CLOCKSS has **200** participating publishers in **29** countries, adding Egypt, Greece & Romania in 2014.

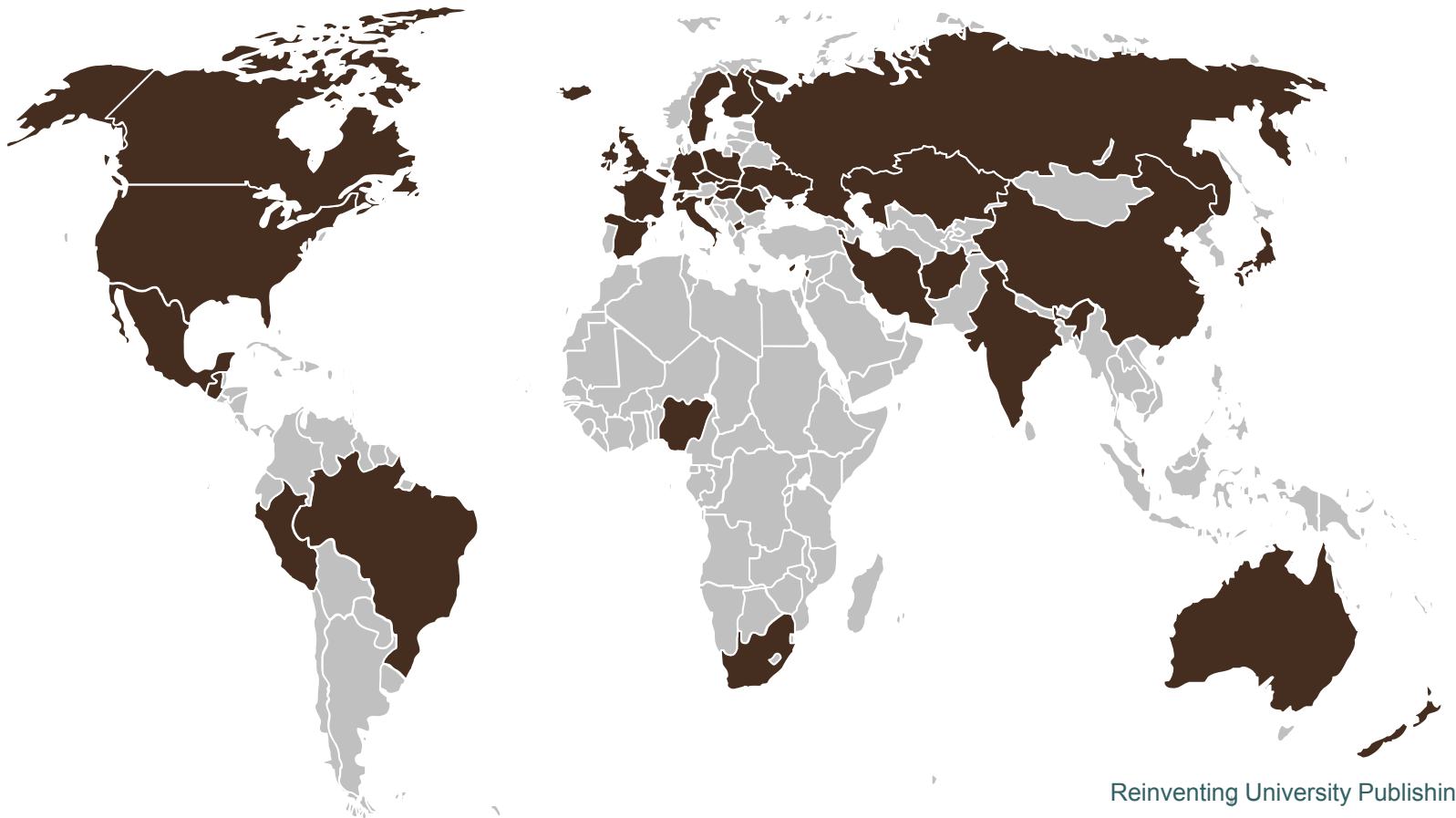
Australia	4
Bangladesh	2
Brazil	1
Bulgaria	1
Canada	7
Egypt	1
Finland	1
France	1
Germany	8
Greece	1
India	4
Italy	8
Japan	9
Lithuania	1
Netherlands	4
New Zealand	4
Portugal	1
Qatar	1
Romania	1
Russia	1
Slovenia	1
South Africa	1
Spain	1
Sweden	1
Switzerland	3
Thailand	1
Turkey	1
UK	42
USA	88



# CLOCKSS Statistics: 2014

CLOCKSS has **725** supporting libraries in **42** countries.

Afghanistan	1
Armenia	1
Australia	20
Bahrain	1
Belgium	1
Brazil	386
Canada	25
China	1
Cyprus	3
Czech Republic	3
Finland	1
France	4
Germany	2
Guatemala	1
Hong Kong	9
Hungary	2
Iceland	1
India	3
Iran	1
Ireland	1
Italy	10
Japan	86
Kazakhstan	3
Lebanon	1
Macedonia	2
Mexico	1
New Zealand	5
Nigeria	1
Peru	1
Poland	1
Romania	1
Russia	21
Singapore	3
Slovakia	1
Slovenia	5
South Africa	1
Spain	1
Sweden	1
Switzerland	8
UK	20
Ukraine	2
USA	83



## Discussion

1. Collection Development Policy
2. Formats – HTML5 and AJAX
3. Databases, datasets, supplementary materials
  - figshare
  - Reveal Digital
  - RUP, eLife photo images

Key issues are space (12x) and value to the community

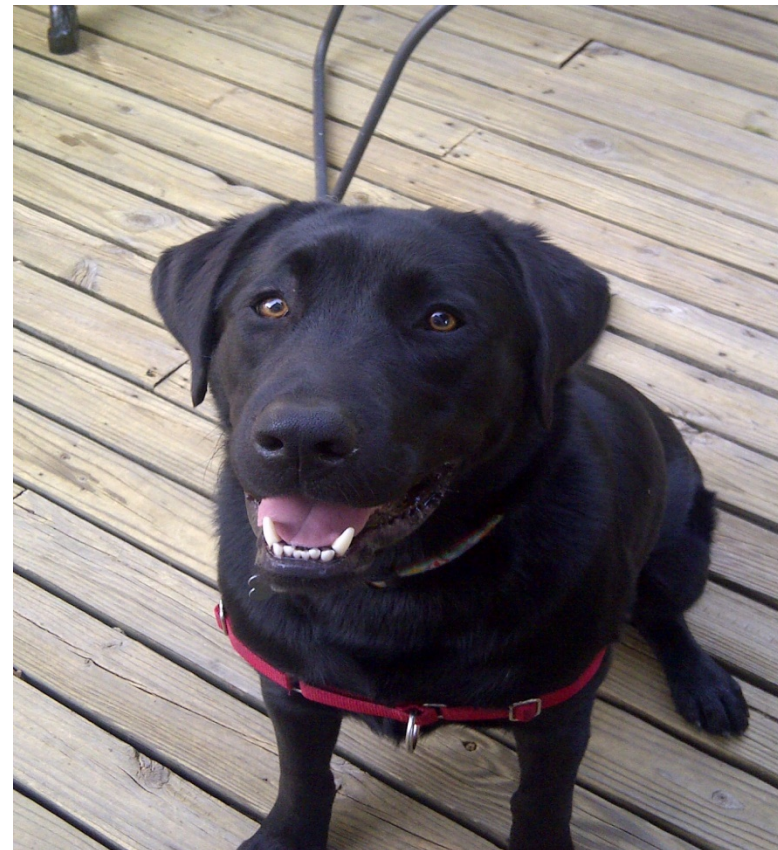
4. Funding

## Today

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

## Thank You!

Randy S. Kiefer  
Executive Director  
The CLOCKSS Archive  
[randy.kiefer@clockss.org](mailto:randy.kiefer@clockss.org)



# Further Reading

- From 2011
- <http://blog.dshr.org/2011/08/moonalice-plays-palo-alto.html>
- From 2012
- <http://blog.dshr.org/2012/05/harvesting-and-preserving-future-web.html>
- From 2013
- <http://blog.dshr.org/2013/02/rothenberg-still-wrong.html>
- and
- <http://www.cni.org/topics/digital-preservation/not-your-grandfathers-web-any-more/>
- From 2014
- <http://blog.dshr.org/2014/03/the-half-empty-archive.html>
- <http://blog.dshr.org/2014/08/trac-audit-process.html>
- <http://blog.dshr.org/2014/08/trac-audit-lessons.html>