Long-term preservation of digital scholarly publications

Library Publishing Forum May 2021
Craig Van Dyck / Alicia Wise, outgoing/incoming Executive Director, CLOCKSS
With print, many copies exist in many different libraries around the world
With digital, normally libraries do not have a copy; users access the content on the publishers’ platforms
How can libraries and users be sure that digital content will not disappear?
Preservation requires:
- active management, to ensure that the content data is healthy
- diversity in organization, governance, geography
- security, forward migration, and legacy
Levels of Preservation from the National Digital Stewardship Alliance
The Importance of digital preservation

- Scholars rely upon ongoing access to materials
- Scholarly literature is valuable over the long term
- Digital content can disappear from the Web
- Libraries are primary champions for preservation
- CLOCKSS ensures that the content will be available
Preservation options for library publishers

- CLOCKSS
- Portico
- Public Knowledge Project Preservation Network (PKP PN) – if using OJS
- Your own library
LOCKSS software and people support a myriad of digital preservation services and use cases:

- **Content type services** – for example Archive-IT, CLOCKSS and the PKP Preservation Network
- **Institutional services** – for example the Global LOCKSS network for local post-cancellation access
- **National services** – for example in Brazil, Canada, and the US
Things to know about CLOCKSS

- Community owned and governed
- Dark archive, keeping scholarship safe
- Triggered content made available Open Access
- World class preservation technology
- Contains:
  - 46 million journal articles
  - 260,000 books
  - Essential metadata, e.g. CrossRef DOIs
  - New content types, e.g. enhanced ebooks
- Content submission via file transfer or web harvest
CLOCKSS Costs – Library Publishers

Annual fee:
$242 if annual revenue <$250,000

One-time set-up fee:
$1000 minimum

Ingest fees:
$0.25/article, first 100/year free
$3/book, first 10/year free

When university library is a supporter, we offer a discount to publishers at the same university. Discounts begin at $100.00 for revenues under $250,000.00 and scale up.
CLOCKSS – Governance

- Community governance: Board of Directors equally comprised by libraries and publishers
- Funded by publisher fees and voluntary library contributions; publisher agreements
- Free-standing 501(c)(3) US-based non-profit
- Succession Plan announced November 2018
- Financially stable
2021 Priorities

- Adding journal backfiles and books
- Broader range of scholarly outputs
- Engaging with libraries
- Asia Pacific, South America, and the Global South
- Disappearing journals
Preserving at-risk open access journals

Project partners:

- CLOCKSS
- PKP
- DOAJ
- Internet Archive
Awareness among journal owners, editors and publishers about why archiving and preservation are important is low.

Problems to solve

1. Journals vanish from the Internet taking with them the research they published, leading to “reference rot” and broken URLs.[1]

2. Some archiving and preservation solutions that exist today come with financial and technical barriers that prevent some journals from taking part.

3. There is an initial investment in time and tech required by each service to set up a new journal before content is ingested. This can be expensive.

4. Communications with thousands of small publishers and a lack of standard practices makes archiving difficult.

5. Awareness among journal owners, editors and publishers about why archiving and preservation are important is low.
Phase 1

There are 7500 DOAJ-indexed, “diamond” open access journals not currently archived.

Only journals in DOAJ are accepted:
- ensures that they have an ISSN
- greatly reduces the possibility that they are predatory.

Over 50% of the journals use PKP’s OJS journal hosting platform.
Approach

An initial survey will be sent to 7500 journals to gather information on how ready the journal is to take part in one of three solutions. Responses will divide the journals into 3 groups:

01 **Those using an OJS version compatible with the PKP PN (3.2+)** will be directed to resources that tell them how to take part.

02 **Those that can produce and export metadata and full text** will be encouraged to participate in CLOCKSS.

03 **Those that cannot produce and export metadata** will be encouraged to allow Internet Archive to crawl their sites.
Deliverables

A sustainable solution that facilitates archiving for unfunded open access journals.

Aggregated archiving data available in Keepers Registry.

Goal is at least a 50% reduction in the number of “unarchived” journals in DOAJ.

Better documentation and training tools for raising awareness on why archiving is good for journals.
CLOCKSS holdings are publicly reported in the Keepers Registry:  https://keepers.issn.org

https://clockss.org
cvandyck@clockss.org
awise@clockss.org
@wisealici
Appendix
<table>
<thead>
<tr>
<th>Libraries Materials Budgets</th>
<th>Annual Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Under $1 million</td>
<td>$485</td>
</tr>
<tr>
<td>$1 - 2 million</td>
<td>$645</td>
</tr>
<tr>
<td>$2 - 3 million</td>
<td>$1,290</td>
</tr>
<tr>
<td>$3 - 4 million</td>
<td>$1,940</td>
</tr>
<tr>
<td>$4 - 5 million</td>
<td>$2,585</td>
</tr>
<tr>
<td>$5 - 7 million</td>
<td>$3,230</td>
</tr>
<tr>
<td>$7 - 9 million</td>
<td>$4,515</td>
</tr>
<tr>
<td>$9 - 11 million</td>
<td>$5,810</td>
</tr>
<tr>
<td>$11 - 13 million</td>
<td>$7,100</td>
</tr>
<tr>
<td>$13 - 15 million</td>
<td>$8,395</td>
</tr>
<tr>
<td>$15 - 20 million</td>
<td>$9,680</td>
</tr>
<tr>
<td>$20 - 25 million</td>
<td>$12,905</td>
</tr>
<tr>
<td>Over $25 million</td>
<td>$16,140</td>
</tr>
</tbody>
</table>