Long-term preservation of digital scholarly literature

LOCKSS Community webinar
Craig Van Dyck
26 June 2018
Contents

- Introduction
- Priorities
- Challenges
- Technology
- Processes
- Governance
- Statistics
- Triggers
CLOCKSS = Controlled LOCKSS*

- Began operations in 2006
- Ensuring long-term access to scholarly literature for researchers
- Part of a diverse, robust ecosystem of preservation solutions
- And part of the broader infrastructure of scholarly communications e.g. Crossref, ORCID, Datacite, CHORUS
- Libraries want publishers to archive their content
- CLOCKSS preserves and archives on behalf of libraries

*Lots of Copies Keep Stuff Safe
2018 Priorities

• Investing in hardware and software: capacity, timeliness
• Adding more large backfiles, and more books
• New content types e.g.
  Software – Code Ocean
  Digitized archival collections – “Divided Society”
  Databases – Crossref
  Video
  Data sets
Scholarly literature is diversifying

- Not only journal articles
- CLOCKSS scope is evolving
- Supplementary material, preprints, blogs
- Scale presents a challenge
- Considering fewer copies for preservation
Priorities, cont’d

- Stronger transparency
- Increased outreach
- Strengthened succession plan
### Challenges

1. **Preserving the “Long Tail”**
   - Long tail journals are the most at-risk, and the hardest to find and work with
   - We need libraries’ priorities for what to archive that is not yet archived

2. **Financial support for a sustainable, diverse, and robust digital preservation environment**
   - Libraries provide one third of our funding
CLOCKSS - Technology

Uses open-source LOCKSS technology, with 12 library server nodes:

North America
- Indiana, OCLC, Rice, Stanford, Virginia, Alberta

Europe
- Edinburgh, Humboldt/Germany, Universita Cattolica/ Italy

Asia/Pacific
- Hong Kong U, NII/Japan, Australia National U
CLOCKSS – Technology, cont’d

- CLOCKSS is certified as a *Trusted Digital Repository* by the Center for Research Libraries
- TRAC audit perfect score for technology; see David Rosenthal’s blog
- Spectrum of preservation solutions
- Archiving ≠ preservation
- LOCKSS at the high end
CLOCKSS - Processes

- Content submission via file transfer or web harvest
- Web harvest is particularly useful with “long tail” publishers
- The LOCKSS team at Stanford is our technology and operations partner
CLOCKSS is a “dark” archive

Triggered content is made available as open access

What does “trigger” mean?
- When digital content ceases to be available to end users
- Access must be ensured, to support scholarship
Community governance: equal number of libraries and publishers on the Board of Directors
Funded by publisher fees and voluntary library contributions
Free-standing 501(c)(3) US-based non-profit
Financially stable, with $1.7 million revenue/year
Trustee Committee owns succession plans
CLOCKSS - Statistics

- 250 publisher participants, 350 library supporters
- 30 million journal articles and books, adding ~4 million/year; over 20,000 journal titles
- 5 largest publishers = 70% of the content
- “long tail” publishers = 65% of the publishers
Triggering Content for Access

- Rigorous rules and practices
- Bylaws require 75% Board vote in favor, with no more than 2 voting against a trigger
- 53 triggered journals; ~1 million article downloads per year
- Triggered journals are open access via CLOCKSS, hosted at Stanford and Edinburgh Universities
- Creative Commons Attribution-Noncommercial-No Derivative Works License
CLOCKSS holdings are publicly reported in the Keepers Registry: https://thekEEPERS.ORG/

https://clockss.org
cvandyck@clockss.org
973-600-7397