



# LOCKSS

## Recentering Community for Decentralized Digital Preservation: LOCKSS Community and Technical Strategy

Nicholas Taylor ([@nullhandle](#))  
Program Manager, [LOCKSS](#) and [Web Archiving](#)  
[Stanford Libraries](#)

[DLF Forum](#)  
[Digitization: Pedagogy and Practice](#)  
16 October 2018



# overview

- LOCKSS  
**background**
- **community role** in  
LOCKSS  
preservation
- LOCKSS technical +  
community  
**strategy**





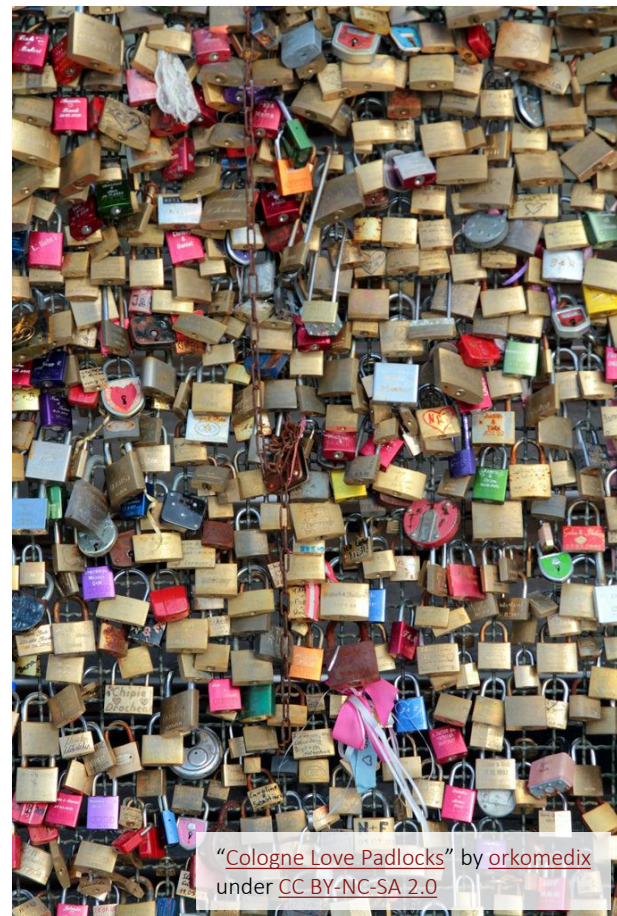
# LOCKSS Background

"Padlocks I" by Domiriel under CC BY-NC 2.0



# what is LOCKSS?

- a widely-accepted **principle** for the persistence of digital info
- a digital library-focused **program** of [Stanford Libraries](#)
- a research-informed **software** app for p2p digital preservation
- an international **community** of institutions + networks





# more than lots of copies

- lots of copies is **necessary but not sufficient**
- central points of failure **can undermine all copies at once**
- LOCKSS provides:
  - continual **integrity checking + repair**
  - b/t mutually-distrusting, **independent peers**
  - on a network that **your community controls**





# understand + mitigate threats

- long-term **data integrity** is hard
- needs architecture informed by **actual leading threats** to data
- don't underestimate:
  - people making **mistakes**
  - **attacks** on information
  - organizational **failure**





# Community-Based



# communities for preservation

- **shared** preservation responsibility
- **direct custody + stewardship** by community
- minimize risks to content from **external dependencies**
- **community governance** over their own preservation initiatives





# Global LOCKSS Network (GLN)

- libraries **collectively ensured** preservation of print journals, via independent acquisition
  - distributed
  - decentralized
  - irrevocable
  - tamper-evident
  - publisher-independent
- GLN re-implemented for e-journals, as **open network**





# LOCKSS Networks

- tens of **networks**
- hundreds of **institutions**
- **all types** of content
- preservation via **diversity** of:
  - technologies
  - institutions
  - networks



CLOCKSS

perma.cc



MetaArchive  
COOPERATIVE



COPPUL

Council of Prairie and  
Pacific University Libraries

PKP  
PUBLIC  
KNOWLEDGE  
PROJECT



Cariniana

Rede Brasileira de Serviços de  
Preservação Digital



THE ALABAMA DIGITAL PRESERVATION NETWORK  
PRESERVING ALABAMA'S DIGITAL RESOURCES

SAFE PLN  
SAFE Archiving FEderation



LOCKSS





# software re-architecture

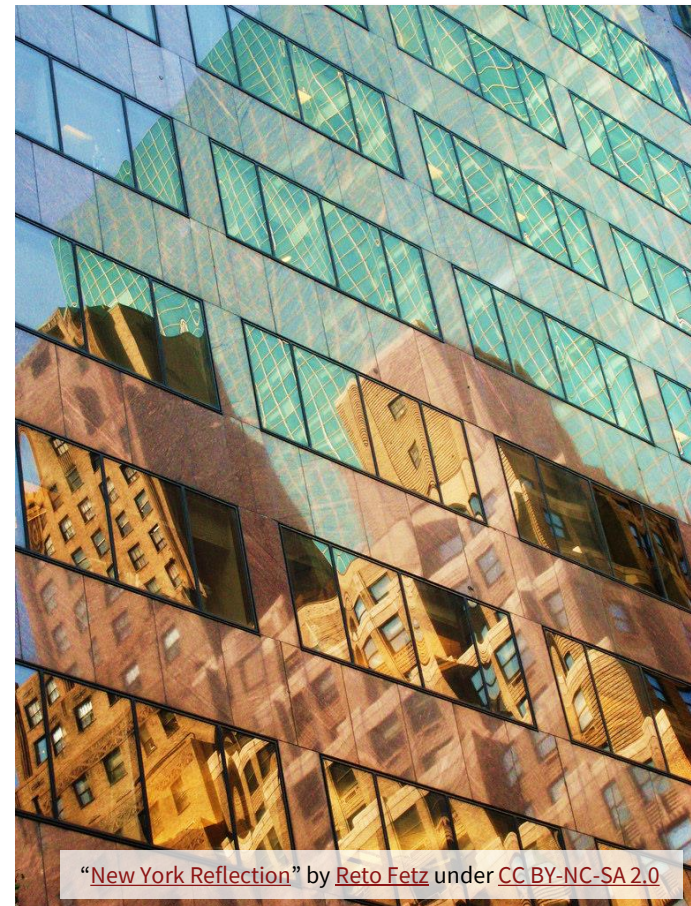
- **monolithic** Java application
- only deployable as **end-to-end solution**
- lacking **modern APIs**
- maintaining functionality on our own that others **increasingly address as a community**
- undertook re-architecture 2017-2019 w/ **Mellon Foundation funding**





# software re-architecture goals

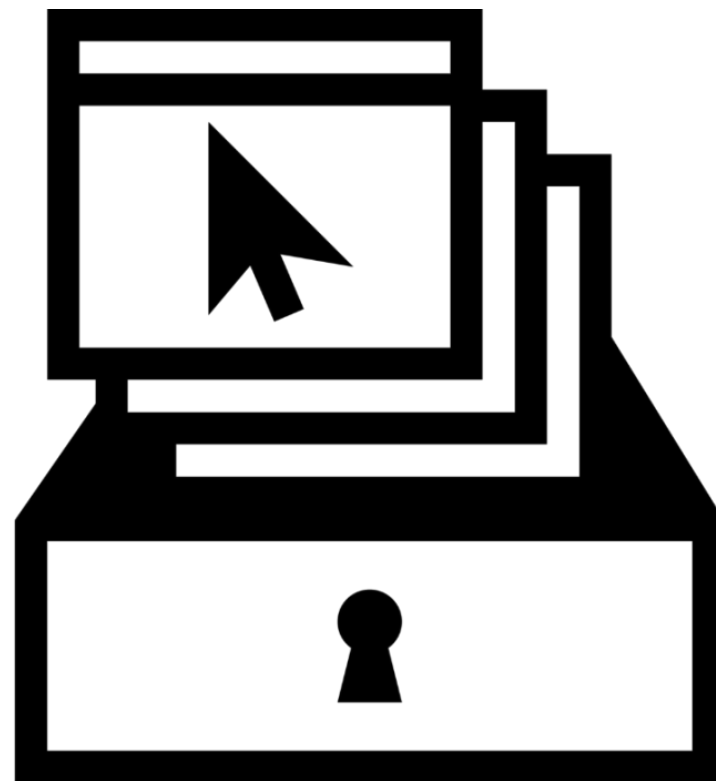
- capitalize on **work of broader communities**
- de-silo + enable **external integrations**
- empower community of practice w/ **better documentation + well-defined APIs**
- evolve w/ web + **digital preservation ecosystem**





# aligning w/ web archiving

- **common challenges:**
  - JavaScript link obfuscation
  - AJAX
  - high-fidelity replay
- **common approaches:**
  - JavaScript pre-execution
  - multi-crawler framework
- **common technologies:**
  - Open Wayback / PyWb
  - Web Archiving Proxy
  - WARC data format





# community-based service providers



← Digital Preservation Declarati...

## Digital Preservation Declaration of Shared Values

April 12, 2018  
Version 2

*Issued by representatives of Academic Preservation Trust (APTrust), Chronopolis, CLOCKSS, Coalition for Networked Information (CNI), Digital Preservation Network (DPN), DuraSpace, Educopia/MetaArchive Cooperative, HathiTrust, Stanford University - LOCKSS, Texas Digital Library (TDL), Council of Prairie and Pacific University Libraries (COPPUL).*

Digital preservation combines policies, strategies, and actions that ensure access to digital content over time.<sup>[1]</sup> Digital preservation is essential to the cumulative record of human activity and memory.

We are a group of collaborating organizations united in our commitment to preserve the cultural, intellectual, scientific and academic record for current and future generations. We believe that preservation should be sustainable, affordable, practical, and available to all. We believe we can accomplish these goals better together rather than separately. We provide services for the cultural heritage, research, and academic communities, and all kinds of scholars.

We embrace participation from communities reflective of the full range of human experience and strive to preserve the histories, cultures, experiences and perspectives of



# co-standardization opportunities

- web capture agent API standardization ([WASAPI](#))
- preservation reporting APIs ([Beyond the Repository](#))
- ingest package targets ([Beyond the Repository](#))





# user + developer documentation

- Documentation Portal  
([lockss.github.io](https://lockss.github.io))
  - access preserved content
  - discovery system integration
  - network setup
  - node management
  - node setup
  - ingest content
  - technical architecture
- Developers Portal
  - LAAWS API specs





# integrations

- **Data Life-Cycle Management (DLCM)**
  - Swiss universities collaborative research data management
- **Software Preservation Network (SPN)**
  - promoting best practices + piloting distributed emulation infrastructure
- **Webrecorder**
  - high-fidelity web capture + replay software





# anticipated outcomes

- collaborate to **build new hybrid solutions**
- align better w/  
**community workflows + interfaces**
- **simplify adaptation** of LOCKSS to local needs
- support LOCKSS technical **community of practice**
- **expand contexts** where LOCKSS can contribute to digital preservation





# questions + feedback

- other logical **collaborators + communities?**
- how can we make LOCKSS a more **robust OSS community?**
- how can we better support **community capacity** for digital preservation?

