



LOTS OF COPIES KEEP STUFF SAFE

Unlocking LOCKSS with APIs

Nicholas Taylor ([@nullhandle](#))

Program Manager, [LOCKSS](#) and [Web Archiving](#)
[Stanford University Libraries](#)

[LDCX](#)

29 March 2017

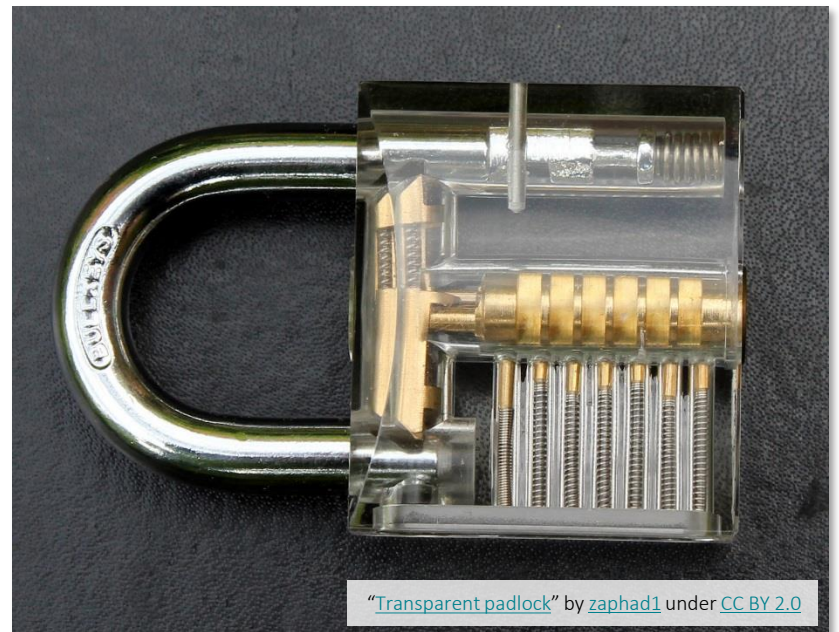
lots of copies keep stuff safe



"Terracotta Warriors" by [Pedro Szekely](#) under [CC BY-NC-SA 2.0](#)

a more interoperable LOCKSS

- **maximize impact** by enabling integration + interconnection
- **improve sustainability** by leveraging standardized community solutions



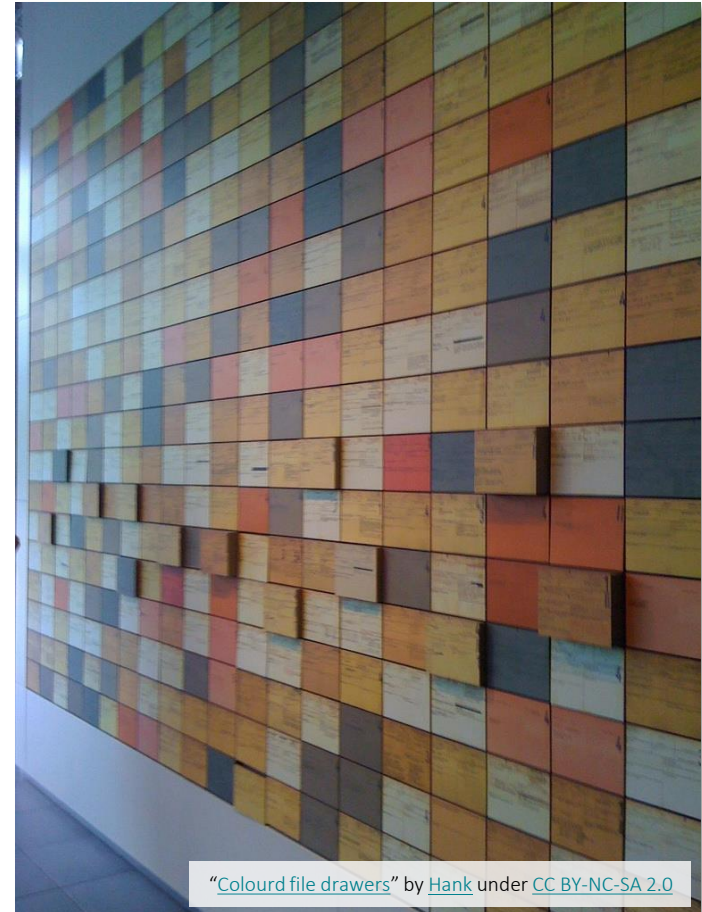
polling + repair

- ensuring bit integrity
 - protocol to **validate content** stored in peers
 - **doesn't depend** on stored checksums
 - based on articulated [threat model](#)
- lots of LOCKSS
 - repository replication
 - other distributed digital preservation systems



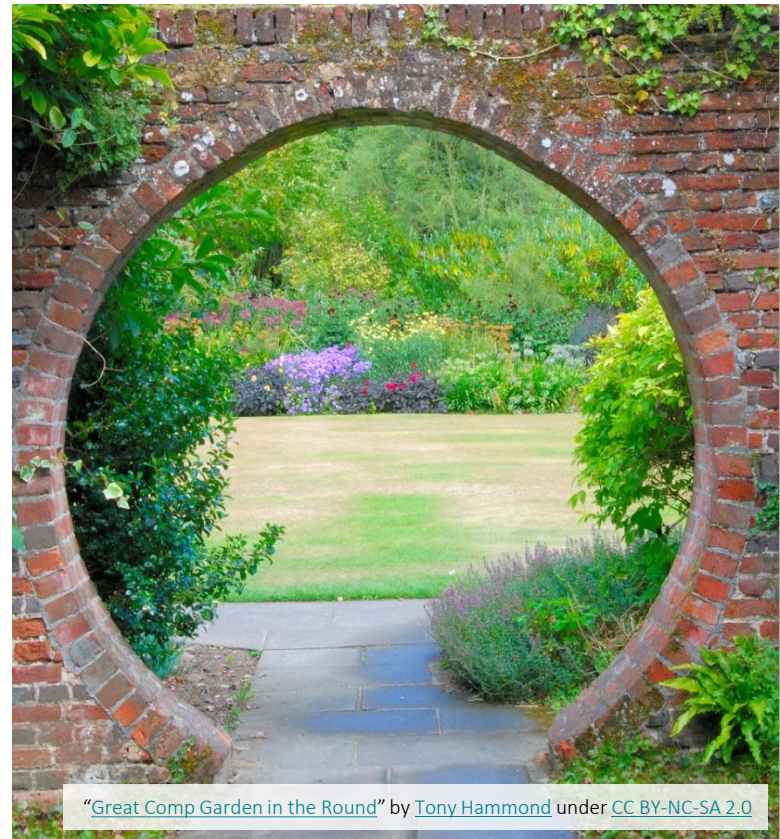
metadata extraction

- parses fields typically used for OpenURL queries
- adapted to DOM for **specific** web content publishing platforms
- handles **specific XML DTDs**
- supports
 - **e-resource** harvesting workflows
 - **web archive** descriptive metadata creation



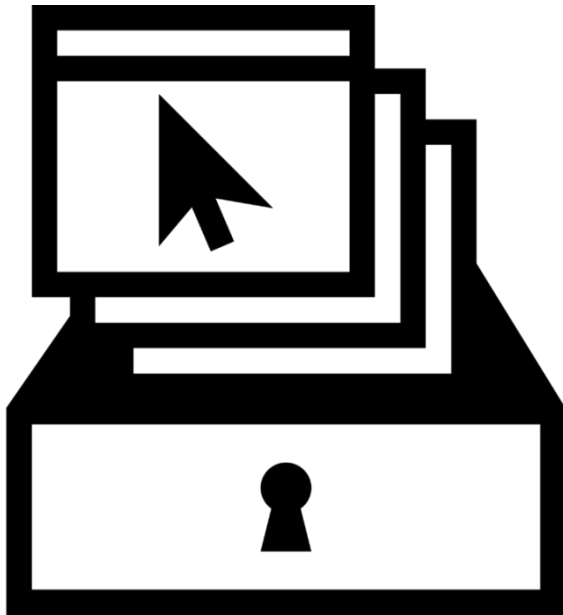
access interfaces

- access to WARC-stored content via:
 - DOI
 - OpenURL
 - Memento
- also, **on-access format migration** in OpenWayback
- integrate archived **open access materials** in scholarly infrastructure



aligning with web archiving

Web ARChive (WARC) format



compatible technologies

- Heritrix
- OpenWayback
- WarcBase
- Web Archiving Proxy

web archiving system APIs (WASAPI)

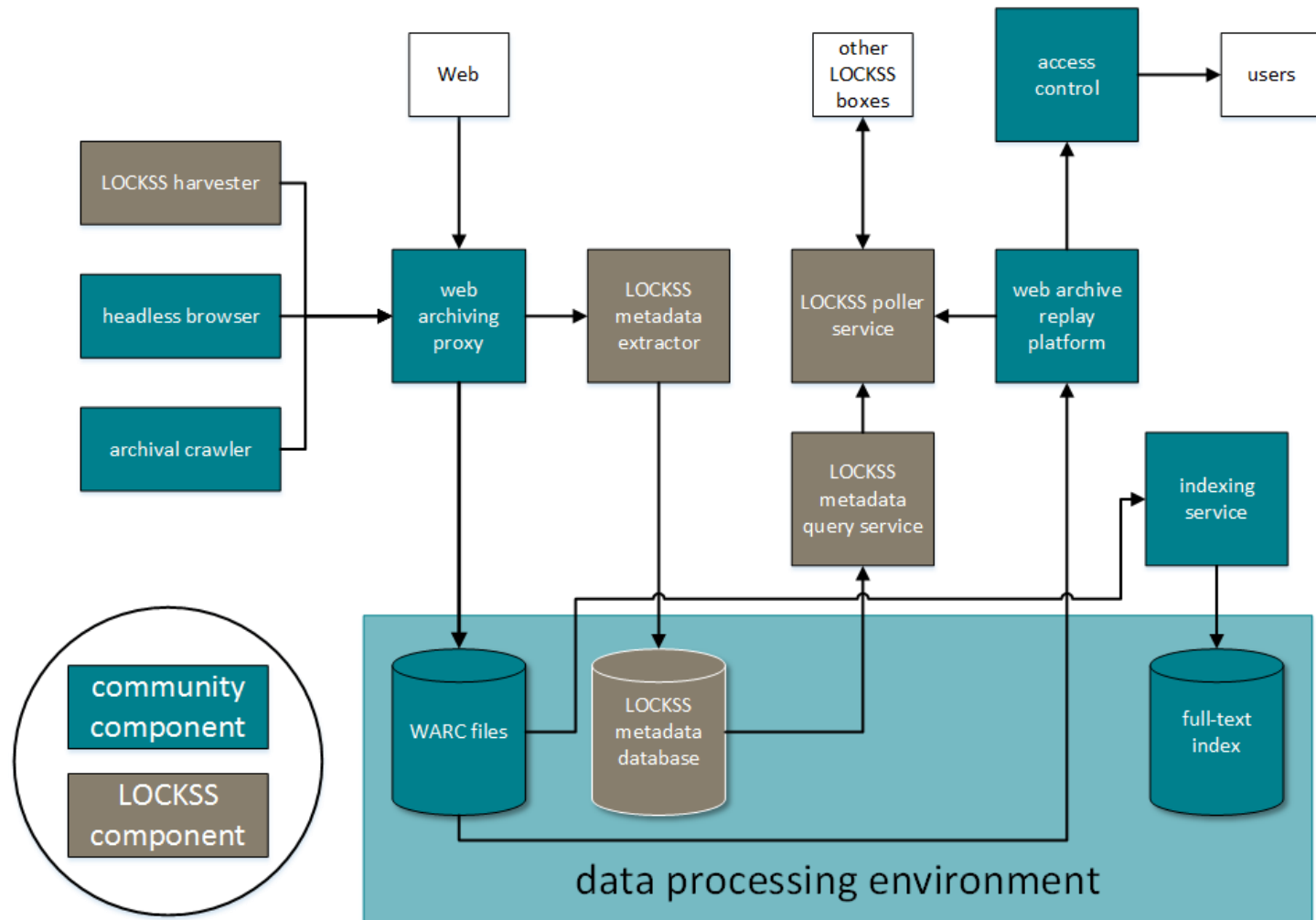
National Digital Platform Projects funded in August 2015

Systems Interoperability and Collaborative Development for Web Archiving

(LG-71-15-0174-15): The Internet Archive, working with partner organizations University of North Texas, Rutgers University, and Stanford University Library will undertake a two-year research project to explore techniques that can expand national web archiving capacity in several areas.



leveraging community components



THANK
YOU