Add your copy into this first column.

Add your copy into this second column.

Add your copy into this third column.

Building a Digital Preservation Strategy Across a Broad University System

CNI Fall 2019
INTRODUCTION TO THE PROCESS

• Overview of the University of California’s Libraries Advisory Structure (UCLAS)

• CoUL states plans and priorities annually.

• Maximize long-term access to digital content has been a priority for a decade. Lots of work led to this point.

• DOC recognized the importance and marrying IT and library preservation activity.
Digital Preservation Working Group

Investigate Internally
• Look at the current and planned UC activity

Develop a Baseline
• Use OAIS as a baseline to develop a high-level overview

Investigate Externally
• Perform a review of external preservation service providers, e.g., CLOCKSS, DPN, HathiTrust and Portico
Phase One Keys

• Articulate gaps between existing UC Libraries digital preservation capabilities and practices compared to current best practices and building blocks.

• The phase one team is not charged to write a prescriptive document that outlines a single approach to digital preservation for the UC Libraries. Rather, the group will develop both a high-level snapshot of current UC Libraries practices and capabilities, as well as an overview of the building blocks an academic research institution the size of the University of California should consider related to digital preservation practices, policy, capabilities, expended resources and potential service providers.

• From these parallel investigations, the phase one working group will build the foundation for subsequent phases as outlined below.
Phase One Focus

- Look Outward
- Look Inward
- Look Across
- Phase Two Charge
Timeline
Scope
Methodology
Exemplars (12)

- CLOCKSS
- DPN
- HathiTrust
- Portico
- LOCKSS (Stanford)
- Preservica
- Rosetta (Ex Libris)
- Chronopolis (UCSD)
- Merritt
- University of Michigan
- University of Illinois
- Internet Archive

UCs (11)

- UC Berkeley
- UC Davis
- UC Irvine
- UC Los Angeles
- UC Merced
- UC Riverside
- UC San Diego
- UC San Francisco
- UC Santa Barbara
- UC Santa Cruz
- CDL
Topics Covered

• Organization
• Mission
• Business Model
• Succession
• Rights Management/IP
• Integrity/Fixity
• Sustainability

• Property
• Architecture
• Ingest
• Metadata
• Access
• Roles/Responsibilities
• Storage/Replication
Data from interviews was compiled in sheets by interviewee and color-coded by topic. These were used to develop the matrices.
## Exemplar Matrix

<table>
<thead>
<tr>
<th>Business model</th>
<th>Architecture</th>
<th>Ingest Req's</th>
<th>Metadata Req's</th>
<th>Storage and Replication</th>
<th>Access</th>
<th>Succession</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chronopolis</td>
<td>University</td>
<td>Multiple Nodes/OAIS</td>
<td>Limited</td>
<td>Agnostic</td>
<td>Distributed/Multiple Copies</td>
<td>Dark</td>
</tr>
<tr>
<td>CLOCKSS</td>
<td>Consortium</td>
<td>Multiple Nodes/OAIS</td>
<td>Strict</td>
<td>Agnostic</td>
<td>Distributed/Many Copies</td>
<td>Dark</td>
</tr>
<tr>
<td>DPN</td>
<td>Consortium</td>
<td>Multiple Nodes</td>
<td>None</td>
<td>Minimal (DC)</td>
<td>Distributed/Amazon</td>
<td>Dark</td>
</tr>
<tr>
<td>Hathi Trust</td>
<td>Consortium</td>
<td>Multiple Nodes/OAIS</td>
<td>Strict</td>
<td>Strict</td>
<td>Semi-Local/Exploring Amazon</td>
<td>Dark/Light</td>
</tr>
<tr>
<td>Internet Archive</td>
<td>Independent</td>
<td>Multiple Nodes</td>
<td>Limited</td>
<td>Agnostic</td>
<td>Distributed/Multiple Copies</td>
<td>Light/Dark</td>
</tr>
<tr>
<td>Illinois</td>
<td>University</td>
<td>Local/OAIS</td>
<td>Limited</td>
<td>Minimal (MODS)</td>
<td>Local/Exploring Amazon</td>
<td>Dark/Light</td>
</tr>
<tr>
<td>LOCKSS (Stanford Libraries)</td>
<td>Vendor</td>
<td>Multiple Nodes/OAIS</td>
<td>Limited</td>
<td>Agnostic</td>
<td>Distributed/Four Copies</td>
<td>Dark</td>
</tr>
<tr>
<td>Merritt (CDL)</td>
<td>Consortium/University</td>
<td>Multiple Nodes</td>
<td>Limited</td>
<td>Agnostic</td>
<td>Distributed/Multiple Copies</td>
<td>Light/Dark</td>
</tr>
<tr>
<td>Michigan</td>
<td>University</td>
<td>Local</td>
<td>None</td>
<td>Minimal</td>
<td>Local/Amazon</td>
<td>Dark/Light</td>
</tr>
<tr>
<td>Portico</td>
<td>Consortium</td>
<td>Multiple Nodes/OAIS</td>
<td>Limited</td>
<td>Agnostic</td>
<td>Distributed/Multiple Copies</td>
<td>Dark</td>
</tr>
<tr>
<td>Preservica</td>
<td>Vendor</td>
<td>Local/OAIS</td>
<td>None</td>
<td>Minimal/XIP</td>
<td>Distributed/Amazon</td>
<td>Dark/Light</td>
</tr>
<tr>
<td>Rosetta (Ex Libris)</td>
<td>Vendor</td>
<td>Local/OAIS</td>
<td>None</td>
<td>Minimal (DC)</td>
<td>Configured by implementer</td>
<td>Dark/Light</td>
</tr>
</tbody>
</table>
### UC Matrix

<table>
<thead>
<tr>
<th>University</th>
<th>Business model</th>
<th>Architecture</th>
<th>Ingest Req's</th>
<th>Metadata Req's</th>
<th>Storage and Replication</th>
<th>Access</th>
</tr>
</thead>
<tbody>
<tr>
<td>Berkeley</td>
<td>University</td>
<td>Tind/Merritt</td>
<td>Limited</td>
<td>Basic</td>
<td>Distributed/Amazon</td>
<td>DAMS, CDL</td>
</tr>
<tr>
<td>Davis</td>
<td>University</td>
<td>Fedora/Archivematica</td>
<td>Limited</td>
<td>Minimal</td>
<td>Distributed/Amazon</td>
<td>DAMS</td>
</tr>
<tr>
<td>Irvine</td>
<td>University</td>
<td>Nuxeo/Merritt</td>
<td>Limited</td>
<td>Basic</td>
<td>Distributed</td>
<td>CDL, Dark/Light</td>
</tr>
<tr>
<td>Los Angeles</td>
<td>University</td>
<td>Various</td>
<td>Limited</td>
<td>Basic</td>
<td>Distributed/Amazon</td>
<td>DAMS, Dark/Light</td>
</tr>
<tr>
<td>Merced</td>
<td>University</td>
<td>Nuxeo/Merritt</td>
<td>Limited</td>
<td>Basic</td>
<td>Distributed</td>
<td>CDL, Dark/Light</td>
</tr>
<tr>
<td>Riverside</td>
<td>University</td>
<td>Nuxeo/Merritt</td>
<td>Limited</td>
<td>Basic (DC)</td>
<td>Distributed</td>
<td>Light</td>
</tr>
<tr>
<td>San Diego</td>
<td>University</td>
<td>Various</td>
<td>Limited</td>
<td>Minimal</td>
<td>Distributed/Data Center</td>
<td>DAMS, Dark/Light</td>
</tr>
<tr>
<td>San Francisco</td>
<td>University</td>
<td>Nuxeo/Merritt</td>
<td>Limited</td>
<td>Basic</td>
<td>Distributed</td>
<td>CDL, Dark/Light</td>
</tr>
<tr>
<td>Santa Barbara</td>
<td>University</td>
<td>Samvera/Local/Merritt</td>
<td>Limited</td>
<td>Basic</td>
<td>Distributed/Data Center</td>
<td>DAMS</td>
</tr>
<tr>
<td>Santa Cruz</td>
<td>University</td>
<td>Samvera/Merritt</td>
<td>Limited</td>
<td>Minimal</td>
<td>Distributed</td>
<td>DAMS</td>
</tr>
</tbody>
</table>
Working Group Findings
Three Tiers of UC Digital Preservation

The Top Tier
- California Digital Library (CDL)
- Chronopolis at UC San Diego

The Bottom Tier
- CDL clients

The Independent Middle Tier
- Digital preservation is largely aspirational
- No one is working together
- Legacy systems continue to be in use
- Staff and economic resources are limited
- Focus is primarily on DAMS
Technology Findings

There’s nothing magical about a preservation repository
• All the issues are well understood
• We have the tools and reference models to build on
• We have standards in place, and they’re keeping up with technology
• Data integrity is no longer a significant problem

Storage
• Almost all storage in the middle tier is local
• Storage costs are a universal concern
Other Findings

Non-technology findings
• The OAIS model is still sound
• There is general agreement about Best Practices
• Certification is essential
• Staff limitations
• Prioritization
Working Group conclusions are supposed to be out-of-scope, but…

- Technology should not be our focus
- System-wide resources are limited and activities not coordinated
- Each campus does not need its own repository
- Our challenges are to **build collaboration** and to develop a well-articulated governance model.

“While research universities and cultural heritage institutions are innately long-running, they operate on that implicitly rather than by making explicit long-term plans.”

---

DPN sunset announcement, Dec 2018
Phase 2

Lots of preservation activity in the system
• Nearly all of the UC libraries using a TRAC Certified repository
• That preservation activity is disjointed and not complete
• Limited or no digital preservation unit or staff in most UC Libraries

Charge specifics
• Develop a high-level taxonomy of content types and associated preservation requirements
• Identify those content types that are common to the majority of UC libraries
• For the identified content type(s), develop an estimate of the amount of content held
• Throughout the process, provide opportunities for campus library staff to participate in activities that build expertise in techniques for assessment of local practices, workflow development, and approaches to content appraisal and selection for preservation decision-making.

Closing Thoughts
Digital Preservation Strategy Working Group, Phase One

John Chodacki, California Digital Library
Mary Elings, UC Berkeley
Todd Grappone (DOC), UCLA
Greg Janée, UC Santa Barbara
Charlie Macquarie, UC San Francisco
Rice Majors (DOC Liaison), UC Davis
Kevin Miller, UC Davis
Erik Mitchell (CoUL Liaison), UC San Diego
Adrian Petrisor, UC Irvine
Chrissy Rissmeyer, UC Santa Barbara
Edson Smith (Chair), UCLA
Roger Smith, UC San Diego
Perry Willett, California Digital Library

Phase One Report