WestVault

https://coppul.ca/westvault

COPPUL’s next generation Community LOCKSS Network

Corey Davis, COPPUL Digital Preservation Coordinator

Slides at https://goo.gl/sHM53f
Council of Prairie and Pacific University Libraries

A consortium of 22 university libraries in Western Canada. [https://coppul.ca/](https://coppul.ca/)
A brief history of COPPUL and LOCKSS...

The COPPUL PLN was established in December 2007, when the Directors of COPPUL libraries agreed to support a two year pilot project.

8 founding members, 11 current members

Mostly OJS content, plus 2 IRs and 1 digital collection
Issues with the original COPPUL PLN

• (Relatively) hard to get stuff in
• No technical or governance mechanisms in place to enable effective scaling
• Only institutions that host a LOCKSS box benefit
Solutions

- Make it easy to get stuff in (and out!)
- Figure out how to expand storage across all nodes to meet emerging demand
- Enable all COPPUL members to use it (even if they won’t/can’t run a LOCKSS box)
Simplifying deposit and withdrawal

Drag and drop Files Here to Upload

Or Select Files to Upload
Scaling storage

WestVault “supernodes” at member institutions with large data centres that can *meet emerging demand*
Preservation storage as a service
Costs

- $2000/yr per Tb for 5 independent copies across Western Canada, all hosted on non-commercial infrastructure
- Support from LOCKSS and COPPUL DP Coordinator
- Annual retainer for technical support provider to SFU (?) + PKP-COPPUL MOU to support future developments
<table>
<thead>
<tr>
<th>Fixed server costs</th>
<th>Fixed cost (regardless of network storage requirements)</th>
<th>5 Tb initial deployment</th>
<th>10 Tb initial deployment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Educloud</td>
<td>$650</td>
<td>$650</td>
<td>$650</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supernode storage costs</td>
<td>Tb/year</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UBC</td>
<td>$150</td>
<td>$750</td>
<td>$1,500</td>
</tr>
<tr>
<td>Alberta</td>
<td>$250</td>
<td>$1,250</td>
<td>$2,500</td>
</tr>
<tr>
<td>Regina</td>
<td>$240</td>
<td>$1,200</td>
<td>$2,400</td>
</tr>
<tr>
<td>Saskatchewan</td>
<td>$176</td>
<td>$880</td>
<td>$1,760</td>
</tr>
<tr>
<td>Manitoba</td>
<td>$250</td>
<td>$1,250</td>
<td>$2,500</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total infrastructure costs</td>
<td>$5,980</td>
<td>$11,310</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Network development and contingency costs</td>
<td>Total maintenance costs</td>
<td>$3,875</td>
<td>$3,875</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total network costs</td>
<td>$9,855</td>
<td>$15,185</td>
</tr>
<tr>
<td></td>
<td>Total Tb/year cost</td>
<td>$1,971</td>
<td>$1,519</td>
</tr>
</tbody>
</table>

https://goo.gl/i51FDA
Use cases

• Specific high-value corpora at some (perhaps larger) member institutions
• Institutions with smaller digital collections might want to ingest everything
• For institutions in the middle, identify several content groups
Supporting documents

- Hosting partner service level agreement
- Depositor agreement
- Succession plan
- Transition plan for existing PLN
- [All available at https://coppul.ca/westvault]
Governance

Subcommittee to be established under the COPPUL Digital Stewardship Network
Building capacity

• Developing curriculum based on DPN’s Digital Preservation Workflow Curriculum available at https://dpn.org/members

• Offering the Digital Preservation Management Workshop in Calgary this June https://coppul.ca/dpm-workshop-2018
On the horizon...

COPPUL is working with OCUL on a research project to test the feasibility of expanding the Ontario Libraries Research Cloud storage service nationally.
WestVault basics
Deposit into a PLN via the OwnCloud app

- User depositing content
- OwnCloud staging server
- OwnCloud notifies OwnCloud to delete files
- LOM notifies LOM of new deposits
- PLN notifies LOM of agreement
- LOM registers deposits with PLN

OC = OwnCloud
LOM = LOCKSS-O-Matic
COPPUL PLN
Withdrawl from a PLN via the OwnCloud app

OC = OwnCloud
LOM = LOCKSS-O-Matic

User withdrawing content

Selects desired files in OwnCloud

OwnCloud staging server

Sends notification email when files are ready

Up to 24 hours later?

OC sends request for file to LOM

OC fetches file from LOM

LOM retrieves files from PLN

COPPUL PLN
ownCloud
Wrong password. Reset it?

☐ Stay logged in
ownCloud

The last cloud collaboration platform you'll ever need.

Explore the features »
Creating bags
What are ‘bags’?

BagIt is a hierarchical file packaging format designed to support disk-based storage and network transfer of arbitrary digital content.
Why use bags?

- Wide adoption in digital libraries
- Received content is ready-to-go in a familiar filesystem tree
- Future indexing services for WestVault will only work for standard BagIt structured packages
- These services will enable more effective discovery and content management
Support for bag creation

- COPPUL’s Archivematica-as-a-service
- AVP’s Exactly desktop tool
Archivematica

Archivematica-as-a-Service

Archivematica is an open-source digital preservation system that supports standards-based, long-term access to collections of digital objects. Archivematica-as-a-Service is designed for libraries wishing to preserve digital collections without the need for installing and locally managing Archivematica instances.

Shared service model

The service is a joint effort of COPPUL, Artefactual Systems (developers) and University of British Columbia (UBC) through CPAUL. COPPUL promotes the service, signs up new institutions at no one-time set-up costs; Artefactual Systems provides installation and user training; and UBC provides a fee-based server hosting and backup service.
<table>
<thead>
<tr>
<th>Transfer</th>
<th>UUID</th>
<th>Transfer start time</th>
</tr>
</thead>
<tbody>
<tr>
<td>elds_2008-01-01.zip</td>
<td>e97fcbec-e9d6-4ceeb-a73b-6534edba8d7</td>
<td>2017-12-27 16:11</td>
</tr>
<tr>
<td></td>
<td>Micro-service: Approve transfer</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Job: Approve standard transfer</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>elds_2008-01-01</td>
<td>713d59c7-7acd-44a3-8733-b04787e6f2a7</td>
<td>2016-11-28 11:23</td>
</tr>
<tr>
<td></td>
<td>Micro-service: Create SIP from Transfer</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Micro-service: Complete transfer</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Micro-service: Identify DSpace files</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Micro-service: Examine contents</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Micro-service: Characterize and extract metadata</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Micro-service: Validation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Micro-service: Update METS.xml document</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Micro-service: Extract packages</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Micro-service: Identify file format</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Micro-service: Clean up names</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Micro-service: Generate transfer structure report</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Micro-service: Scan for viruses</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Micro-service: Quarantine</td>
<td></td>
</tr>
</tbody>
</table>
# All Packages

Packages are Transfers, SIPs, DIPs and AIPs uploaded to a Location managed by the storage service.

Show [10] entries

<table>
<thead>
<tr>
<th>UUID</th>
<th>Description</th>
<th>Originating Pipeline</th>
<th>Current Location</th>
<th>Size</th>
<th>Type</th>
<th>Pointer File</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>414b64ff-f5e2-477e-938d-bcd1-ad953aaf</td>
<td>None</td>
<td>Archivematica on uvic-gold1 (af23374a-44bd-4570-9db9-5ff78cfab95f8)</td>
<td>/storage/aips/414b/64ff/f5e2/477e/938d/bcd1/ad95/3aaf/DSpaceExport325b64ff-f5e2-477e-938d-bcd1-ad953aaf.7z</td>
<td>5.4 MB</td>
<td>AIP</td>
<td>Pointer File</td>
<td>Delete</td>
</tr>
<tr>
<td>44bf252b-c83a-4a73-87e3-e4b0f6ebc12</td>
<td>None</td>
<td>Archivematica on uvic-gold1 (af23374a-44bd-4570-9db9-5ff78cfab95f8)</td>
<td>/storage/aips/44bf252b-c834a73-87e3-c24b0f6ebc12/ETDs-1828-333-44bf252b-c83a-4a73-87e3-e4b0f6ebc12.7z</td>
<td>3.4 GB</td>
<td>AIP</td>
<td>Pointer File</td>
<td>Delete</td>
</tr>
<tr>
<td>551f9e25-7e1d-490c-8e32-9f46d54890e</td>
<td>None</td>
<td>Archivematica on uvic-gold1 (af23374a-44bd-4570-9db9-5ff78cfab95f8)</td>
<td>/storage/aips/551f/9e25/7e1d/490c/8e32/9f46/d548/890e/WARC-4867-551f9e25-7e1d-490c-8e32-9f46d54890e.7z</td>
<td>8.5 GB</td>
<td>AIP</td>
<td>Pointer File</td>
<td>Delete</td>
</tr>
<tr>
<td>5db01b74-8a94-455b-849b-abb6e56e4a36</td>
<td>None</td>
<td>Archivematica on uvic-gold1 (af23374a-44bd-4570-9db9-5ff78cfab95f8)</td>
<td>/storage/aips/5db0/1b74/8a94/455b/849b/abb6e56e4a36/ETD-NoTopAIP-Test5db01b74-8a94-455b-849b-abb6e56e4a36.7z</td>
<td>29.0 MB</td>
<td>AIP</td>
<td>Pointer File</td>
<td>Delete</td>
</tr>
</tbody>
</table>
Create Spaces

Access protocol:

---
Arkivum
Dataverse
DuraCloud
DSpace via SWORD2 API
FEDORA via SWORD2
Local Filesystem
LOCKSS-o-matic
NFS
Pipeline Local Filesystem
Swift

How the space can be accessed.

Size in bytes (optional)

Absolute path to the space on the server:
About

Exactly

Exactly is a free and open source application that is easy-to-use for

<table>
<thead>
<tr>
<th>Label</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source Organization</td>
<td></td>
</tr>
<tr>
<td>Organization Address</td>
<td></td>
</tr>
<tr>
<td>Contact Name</td>
<td></td>
</tr>
<tr>
<td>Contact Phone</td>
<td></td>
</tr>
<tr>
<td>Contact Email</td>
<td></td>
</tr>
</tbody>
</table>

To copy and paste, use Windows shortcuts: "control+c" and "control+v"
LOCKSS-O-Matic

...where the magic happens
WestVault
A Private LOCKSS Network by COPPUL.

This is the administrative front-end for WestVault. End-users wanting more information, or wanting to join the PLN should check out a different page.
# Deposit List

Displaying 19 records of 19 total.

<table>
<thead>
<tr>
<th>Uuid</th>
<th>Received</th>
<th>Action</th>
<th>Provider</th>
<th>Size</th>
<th>State</th>
<th>Pinstate</th>
</tr>
</thead>
<tbody>
<tr>
<td>92523694-2F6B-44B4-A4CB-1ABDC8CF94E</td>
<td>2017-12-14</td>
<td>add</td>
<td>michael</td>
<td>1476332</td>
<td>complete</td>
<td>agreement</td>
</tr>
<tr>
<td>E60C3F1A-8BA9-4AFB-6A4F-6A17FC6D2F6B</td>
<td>2017-12-15</td>
<td>add</td>
<td>mark</td>
<td>10737418240</td>
<td>harvest-error</td>
<td></td>
</tr>
<tr>
<td>4FFBBA5B-9D2B-44B2-8F9C-009386A40ED8</td>
<td>2017-12-15</td>
<td>add</td>
<td>michael</td>
<td>15360000000</td>
<td>harvest-error</td>
<td></td>
</tr>
<tr>
<td>1F732EB3-EE38-4957-BC9E-0BC4663B415A</td>
<td>2017-12-16</td>
<td>add</td>
<td>michael</td>
<td>136494</td>
<td>complete</td>
<td>agreement</td>
</tr>
<tr>
<td>017658B2-BDCB-4B0D-8443-CC9F75ED8F4E</td>
<td>2017-12-17</td>
<td>add</td>
<td>mark</td>
<td>10737418240</td>
<td>harvest-error</td>
<td></td>
</tr>
<tr>
<td>E17D4E5A-2D08-4626-BD9F-BB462C0223FB</td>
<td>2017-12-18</td>
<td>add</td>
<td>mark</td>
<td>20331441</td>
<td>complete</td>
<td>agreement</td>
</tr>
<tr>
<td>29B1A93-6206-4DEA-83E4-3C0F6C14BD26</td>
<td>2018-01-03</td>
<td>add</td>
<td>corey</td>
<td>817</td>
<td>complete</td>
<td>agreement</td>
</tr>
<tr>
<td>8E40E73E-C2B5-4798-81DC-B8660B89C3</td>
<td>2018-01-03</td>
<td>add</td>
<td>corey</td>
<td>1532</td>
<td>complete</td>
<td>agreement</td>
</tr>
<tr>
<td>D64B0C61-1100-48A6-A0EE-6B309AB8E5D3</td>
<td>2018-01-03</td>
<td>add</td>
<td>corey</td>
<td>614</td>
<td>complete</td>
<td>agreement</td>
</tr>
<tr>
<td>6A404823-3B77-41BF-851C-47C01411EB6A</td>
<td>2018-01-03</td>
<td>add</td>
<td>corey</td>
<td>258</td>
<td>complete</td>
<td>agreement</td>
</tr>
</tbody>
</table>
Deposit received.
Deposit harvest succeeded.
Payload checksum validation succeeded.
Virus check passed. No infections found.
Organize succeeded.
Deposit to Lockssomatic succeeded.
Deposit status succeeded.
Getting your stuff back...
## WestVault Deposits

Showing deposits for corey.

<table>
<thead>
<tr>
<th>Path</th>
<th>Status</th>
<th>LOCKSS Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>bag-info.xml</td>
<td>deposited</td>
<td>agreement</td>
</tr>
<tr>
<td>FileSystemData.txt</td>
<td>deposited</td>
<td>agreement</td>
</tr>
<tr>
<td>bag-info.csv</td>
<td>deposited</td>
<td>agreement</td>
</tr>
<tr>
<td>TransferComplete.txt</td>
<td>deposited</td>
<td>agreement</td>
</tr>
<tr>
<td>bag-info.txt</td>
<td>deposited</td>
<td>agreement</td>
</tr>
<tr>
<td>bit.txt</td>
<td>deposited</td>
<td>agreement</td>
</tr>
<tr>
<td>manifest-md5.txt</td>
<td>deposited</td>
<td>agreement</td>
</tr>
<tr>
<td>tagmanifest-md5.txt</td>
<td>deposited</td>
<td>agreement</td>
</tr>
<tr>
<td>davis-cover-2015.docx</td>
<td>deposited</td>
<td>agreement</td>
</tr>
<tr>
<td>davis-cv-2015.docx</td>
<td>deposited</td>
<td>agreement</td>
</tr>
<tr>
<td>davis-cover-2015.pdf</td>
<td>deposited</td>
<td>agreement</td>
</tr>
<tr>
<td>davis-cv-2015.pdf</td>
<td>deposited</td>
<td>agreement</td>
</tr>
<tr>
<td>Corey_CV_EL_AL.zip</td>
<td>restore</td>
<td>agreement</td>
</tr>
<tr>
<td>TransferComplete.txt</td>
<td>deposited</td>
<td>agreement</td>
</tr>
</tbody>
</table>
In summary...
Questions? Comments?

- Costs
- Sustainability
- Scalability
- Enhancements
- Alternatives
Thanks!

- corey@coppul.ca
- @coreyleedavis
- (778) 677-5746