Choose Your Own Adventure:

Digital Preservation Storage Stories

link to Google Slides https://goo.gl/33Mrnv
The Texas Digital Library is a consortium of Texas higher education institutions that builds capacity for preserving, managing, and providing access to unique digital collections of enduring value.

tdl.org
The Infrastructure that Preserves Texas
The Voyage to Preservation Storage

TDL members’ local digital preservation workflows and procedures.

- Content is ingested into TDL DuraCloud and passes through Amazon S3.
- Copies may remain in S3 only and/or be stored in any combination of S3, Amazon Glacier, Chronopolis, and the Digital Preservation Network.
4 DP Storage Stories
Disclaimer: Stories Condensed

Institution names* are changed and use cases have been combined in order to reflect the variety of issues we face in Texas

*(We’ll still tell you who is using a particular workflow when you ask, because we love to share)
The Institution That Made Perfect Its Enemy
“We need to do a full TRAC audit before we start”

“Our digital preservation strategy is non-existent”

“We’re still stuck in the 90s technologically”

“We only have 3 FTE staff”

“IT doesn’t think we need anything else”
Texas Levels of Preservation

Findings based on Spring 2017 interviews conducted with 15 Texas Digital Library member institutions, 3 of which were active users of TDL Digital Preservation Services.

*proposed functional areas

published content in **peril** that only exists in Digital Commons and on local shared drives

needs ‘backup’ function

...chooses Glacier
Searching for a Place to Put the Treasure
content in DSpace and replicated on local storage

+ special collections in ContentDM

...chooses Chronopolis &

...partners with TDL to adapt DSpace >> DuraCloud workflow implemented by Duraspace for DSpace Direct
The Institution With All The Magic Beans
content on shared drives, in a homegrown system and in DSpace
+ Archivematica
+ digitized content workflow and a homegrown metadata engine

...chooses DPN

...partners with TDL for consultation to align local workflows with DPN deposit

tdl.org
<table>
<thead>
<tr>
<th>Content Description</th>
<th>Funding Requirements</th>
<th>How long must this content persist? (Retention=for ever, 7 years, etc.)</th>
<th>Copyright/IP status?</th>
<th>Risk to originals - scale of 1-10, 1 being all originals in good condition and preserved</th>
<th>Media At-Risk?</th>
<th>Scale of 1-10, 1 being not at all, 10 being one copy on a floppy disk from 1989</th>
<th>Complexity (size and diversity of content) - Please indicate on a scale of 1-10, 1 being the least complex (ie 1 format, small files, under 1GB total)</th>
<th>Monetary value, scale of 1-10, 1 being least valuable</th>
<th>Intellectual value, scale of 1-10, 1 being least valuable</th>
<th>Popularity/demand, scale of 1-10, 1 being the least demand</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content Type</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Location</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Related Sy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Choose Your DP Storage Adventure

[Link: https://tdl.org/digital-preservation/]

<table>
<thead>
<tr>
<th>Digital Preservation Storage options via DuraCloud@TDL</th>
<th><strong>Chronopolis</strong></th>
<th><strong>DPN</strong></th>
<th><strong>Amazon S3</strong></th>
<th><strong>Glacier</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-commercial, rooted in cultural heritage community</td>
<td>Non-commercial, rooted in cultural heritage community</td>
<td>Commercial</td>
<td>Commercial</td>
<td>Commercial</td>
</tr>
</tbody>
</table>
Choose Your DP Storage Adventure

https://tdl.org/digital-preservation/

<table>
<thead>
<tr>
<th>Digital Preservation Storage options via DuraCloud@TDL</th>
<th>Chronopolis</th>
<th>DPN</th>
<th>Amazon S3</th>
<th>Glacier</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Geographic Distribution</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Any three from the following technologically diverse partner nodes: UCSD, TDL, UMIACS, and NCAR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Any three from the following technologically diverse partner nodes: HathiTrust, APTTrust, Chronopolis, Texas Preservation Node (TPN)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amazon East</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amazon East</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
# Choose Your DP Storage Adventure

<table>
<thead>
<tr>
<th>File Fixity and Data Integrity</th>
<th>Chronopolis</th>
<th>DPN</th>
<th>Amazon S3</th>
<th>Glacier</th>
</tr>
</thead>
<tbody>
<tr>
<td>DuraCloud@TDL MD5 hashes on upload.</td>
<td>Fixity check type varies per node.</td>
<td>DuraCloud@TDL MD5 hashes on upload.</td>
<td>Amazon does some combination of MD5 hashes and cyclic redundancy checks (CRCs) with unspecified regularity, on upload, and when content is moved.</td>
<td>DuraCloud@TDL MD5 hashes on upload.</td>
</tr>
<tr>
<td>Chronopolis checks SHA-256 hashes every 30 days or as specified by member, not to exceed 6 months.</td>
<td>DuraCloud@TDL MD5 hashes on upload.</td>
<td>Amazon does some combination of MD5 hashes and cyclic redundancy checks (CRCs) with unspecified regularity, on upload, and when content is moved.</td>
<td>DuraCloud@TDL MD5 hashes on upload.</td>
<td></td>
</tr>
</tbody>
</table>
# Choose Your DP Storage Adventure

<table>
<thead>
<tr>
<th>File Fixity and Data Integrity</th>
<th>Chronopolis</th>
<th>DPN</th>
<th>Amazon S3</th>
<th>Glacier</th>
</tr>
</thead>
<tbody>
<tr>
<td>DuraCloud@TDL MD5 hashes on upload.</td>
<td>Fixity check type varies per node.</td>
<td>DuraCloud@TDL MD5 hashes on upload.</td>
<td>DuraCloud@TDL MD5 hashes on upload.</td>
<td>DuraCloud@TDL MD5 hashes on upload.</td>
</tr>
<tr>
<td>Chronopolis checks SHA-256 hashes every 30 days or as specified by member, not to exceed 6 months.</td>
<td>DuraCloud@TDL MD5 hashes on upload.</td>
<td>Amazon does some combination of MD5 hashes and cyclic redundancy checks (CRCs) with unspecified regularity, on upload, and when content is moved.</td>
<td>Amazon does some combination of MD5 hashes and cyclic redundancy checks (CRCs) with unspecified regularity, on upload, and when content is moved.</td>
<td></td>
</tr>
</tbody>
</table>
DuraCloud and the Four Paths

tdl.org
Your Very Own Robot (Sync Tool 1)

What (data do you want to store)
Where (you want to store your data)
How (you want to store your data)
When (you want to sync your data into Duracloud)
Your Very Own Robot (Sync Tool 2)
Space Patrol (content into spaces)
Space and Beyond (storage providers)

S3
Glacier

DPN
THE DIGITAL PRESERVATION NETWORK

CHRONOPOLIS

tdl.org
A Light On Burro Mountain

Content challenges:
Private, Personal and Confidential
Temporary
Embargoed
The Authors & Thanks

c.mumma@austin.utexas.edu @lookouthoney

njw@austin.utexas.edu @nj_woodward

kristi.park@austin.utexas.edu

Thanks to our DPS Members:

Angelo State, Baylor University, UT Austin, UT Rio Grande Valley
Texas A&M, Stephen F. Austin State University, Trinity University

tdl.org