Digital Preservation Practices across the Alliance

Devin Becker, Chair, Digital Preservation Working Group
Digital Preservation Working Group

Sara Amato, Willamette University
Devin Becker, University of Idaho
Laura Buchholz, Reed College
Moriah Caruso, University of Washington
Brian Davis, Oregon State University
Danielle Mericle, University of Oregon
Jodi Allison-Bunnell, Alliance staff
Survey Summary/Methodology

- Based on work out of the National Digital Stewardship Alliance (NDSA)
  - NDSA Table: “Levels of Digital Preservation”
  - “Staffing for Effective Digital Preservation” NDSA Report, December 2013
  - “USGS Guidelines for the Preservation of Digital Scientific Data.” April 2014
- Delivered to Orbis Cascade Alliance Members in May 2016
- 32 of 39 Members responded
- 2 Parts
  - Current practices, size of archive, staffing, and policies/education
  - NDSA Level Evaluation
General Results - Digital Preservation Practices

- Nearly all members report that they create and/or manage digitized photographs (97%), books and maps (84%), and audio or audiovisual materials (87%). Electronic theses and dissertations are also widely managed (84%).
- 25 of 32 indicate they do engage in digital preservation practices
- Of those 25:
  - 8 have a digital preservation policy in place; 2 are developing one
  - 8 offer digital preservation education
- Staffing:
  - Majority have less than 1 FTE assigned to digital preservation
  - All indicate an increase in staffing would be helpful
Digital Archives Size/Growth

- Libraries in the Orbis Cascade Alliance manage about 149TB of digital content for preservation purposes
  - 9 institutions manage 5 or more TB,
  - 5 manage between 1 and 4.9 TB
  - 10 manage less than 1TB
- Digital archives have grown 25% in last year, 86% in past 5 years.
- Estimated growth for next year: 62%
**NDSA**

“Levels of Digital Preservation”

<table>
<thead>
<tr>
<th>Storage and Geographic Location</th>
<th>Level 1 (Protect your data)</th>
<th>Level 2 (Know your data)</th>
<th>Level 3 (Monitor your data)</th>
<th>Level 4 (Repair your data)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Two complete copies that are not collocated</td>
<td>At least three complete copies</td>
<td>At least one copy in a geographic location with different disaster threats</td>
<td>At least three copies in geographic locations with different disaster threats</td>
</tr>
<tr>
<td></td>
<td>- For data on heterogeneous media (optical discs, hard drives, etc.) get the content off the medium and into your storage system</td>
<td>- Document your storage system(s) and storage media and what you need to use them</td>
<td>- Obsolescence monitoring process for your storage system(s) and media</td>
<td>- Have a comprehensive plan in place that will keep files and metadata on currently accessible media or systems</td>
</tr>
<tr>
<td></td>
<td>File Fixity and Data Integrity</td>
<td>Level 2 (Know your data)</td>
<td>Level 3 (Monitor your data)</td>
<td>Level 4 (Repair your data)</td>
</tr>
<tr>
<td></td>
<td>Check fixity on all ingests</td>
<td>- Check fixity of content at fixed intervals</td>
<td>- Check fixity of all content in response to specific events or activities</td>
<td>- Ability to replace/repair corrupted data</td>
</tr>
<tr>
<td></td>
<td>- Use write-blockers when working with original media</td>
<td>- Maintain logs of fixity information; supply audit on demand</td>
<td>- Ensure no one has write access to all copies</td>
<td>- Ensure no one has write access to all copies</td>
</tr>
<tr>
<td></td>
<td>- Virus-check high risk content</td>
<td>- Ability to detect corrupt data</td>
<td>- Ability to detect corrupt data</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Information Security</td>
<td>Level 3 (Monitor your data)</td>
<td>Level 4 (Repair your data)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Identify who has read, write, move and delete authorization to individual files</td>
<td>- Document access restrictions for content</td>
<td>- Perform audit of logs</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Restrict who has those authorizations to individual files</td>
<td>- Maintain logs of who performed what actions on files, including deletions and preservation actions</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Metadata</td>
<td>Level 4 (Repair your data)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Inventory of content and its storage location</td>
<td>- Store administrative metadata</td>
<td>- Store standard technical and descriptive metadata</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Ensure backup and non-collocation of inventory</td>
<td>- Store transformative metadata and log events</td>
<td>- Store standard preservation metadata</td>
<td></td>
</tr>
<tr>
<td></td>
<td>File Formats</td>
<td>- Inventory of file formats in use</td>
<td>- Monitor file format obsolescence issues</td>
<td></td>
</tr>
<tr>
<td></td>
<td>When you can give input into the creation of digital files encourage use of a limited set of known open formats and codecs</td>
<td>- Perform format migrations, emulation and similar activities as needed</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## NDSA Levels

<table>
<thead>
<tr>
<th>NDSA category</th>
<th>Storage &amp; Geographic Location</th>
<th>File Fixity &amp; Data Integrity</th>
<th>Information Security</th>
<th>Metadata</th>
<th>File Formats</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alliance average level</td>
<td>1.17</td>
<td>0.54</td>
<td>1.29</td>
<td>1.79</td>
<td>2.04</td>
</tr>
<tr>
<td>Alliance median level</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

- Average NDSA level for Alliance institutions was 1.37 (out of 4), with the mean being 1
  - Strengths: Metadata and File Formats
  - Weaknesses: File Fixity & Data Integrity
Archive Size vs. NDSA Level
Recommendations

1. Investigate a shared digital archive for digital archival content from Alliance members.
   ○ Why: Making available storage in locations with different disaster threats for Alliance members would greatly improve the levels of digital preservation for many members.
Recommendations cont.

2. Provide in-person or web-delivered education and assistance on file-fixity and data integrity, as well as how to record and store transformation metadata.
   ○ Why: Address a lack of knowledge across alliance regarding file fixity, checksums, etc.
3. Create Alliance portal for digital preservation training kits and checklists/quick-start guides for Alliance institutions looking to improve their digital preservation practices.

   ○ Why: Allows institutions to address/educate themselves on-demand regarding those areas in which they need the most assistance.

   ○ A few specific requests:

   i. Create an online resource that documents which institutions have readers of obsolete file formats.

   ii. Provide lists of tools for various tasks and contacts at institutions using them on Alliance documentation site.

   iii. Provide a template for a digital preservation plan.

   iv. Quick / High Level Digital Preservation FAQ - what it is/isn’t and why it is important.
Thank you for your time!

Devin Becker
dbecker@uidaho.edu

On behalf of the Digital Preservation Working Group