A Snapshot of the U.S. Web Archiving Landscape through the 2013 NDSA Survey Report
NDSA Web Archiving survey background

2011
• 78 respondents
• program info
• tools and services
• access
• policies

2013
• 92 respondents
• program info
  • staff time, metrics, skills, content concerns
• tools and services
• access and discovery
  • new discovery options
• policies
  • embargo, social media, robots.txt, resources
Respondent Characteristics
less than half repeat respondents

- 2011: 40
- 2013: 56
  - 33 repeat respondents
universities still make up most programs

Pie charts showing the distribution of programs in 2011 and 2013:
- **2011**
  - College or University: 47%
  - Archive: 13%
  - State Gov: 13%
  - Other: 12%
  - Fed Gov: 8%
  - Public Library: 2%
  - Museum: 3%
- **2013**
  - College or University: 52%
  - Archive: 15%
  - State Gov: 13%
  - Other: 8%
  - Fed Gov: 5%
  - Public Library: 2%
  - Museum: 1%
## Archive-It and SAA top group affiliations

<table>
<thead>
<tr>
<th>group</th>
<th>2011</th>
<th>2013</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="https://example.com/iipc.png" alt="IIPC Logo" /></td>
<td>8%</td>
<td>7%</td>
</tr>
<tr>
<td><img src="https://example.com/nda.png" alt="NDSA Logo" /></td>
<td>31%</td>
<td>33%</td>
</tr>
<tr>
<td><img src="https://example.com/saa.png" alt="SAA Logo" /></td>
<td></td>
<td>45%</td>
</tr>
<tr>
<td><img src="https://example.com/ai.png" alt="Archive-It Logo" /></td>
<td>72%</td>
<td>71%</td>
</tr>
</tbody>
</table>
most programs are fractionally staffed
web/archiving tech savviness are key skills
data volume and archive use are key metrics
Maturity and Progress
programs have matured slightly since 2011
strong perceptions of progress since 2011
many new programs since 2011
two-thirds of them now use Archive-It
Archiving Focus

“Ant Farm Media Van v.08 (Time Capsule) in Bellewether at Southern Exposure” by Steve Rhodes under CC BY-NC-SA 2.0
more programs are only self-archiving
Content Working Group

concern about social media, databases, video

Number of organizations

<table>
<thead>
<tr>
<th>Media Type</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social Media</td>
<td>69</td>
</tr>
<tr>
<td>Databases</td>
<td>65</td>
</tr>
<tr>
<td>Video</td>
<td>64</td>
</tr>
<tr>
<td>Interactive Media</td>
<td>49</td>
</tr>
<tr>
<td>Audio</td>
<td>40</td>
</tr>
<tr>
<td>Blogs</td>
<td>32</td>
</tr>
<tr>
<td>Art</td>
<td>16</td>
</tr>
</tbody>
</table>
untapped interest in collaboration

- Yes: 21% (2011), 17% (2013)
- No: 72% (2011), 47% (2013)
- Not yet, but interested: 33%
- Don't know: 7% (2011), 2% (2013)
Tools and Services
web archiving as a service still most popular
data not transferred from service provider

- Transferred: 19% in 2011, 20% in 2013
- Haven't transferred: 81% in 2011, 80% in 2013
increased use of tools supporting W/ARC
less granular descriptive metadata
Archiving Policies

“Handle With Care” by ServInt under CC BY-NC-ND 2.0
most don’t notify or seek permission

<table>
<thead>
<tr>
<th>Action</th>
<th>No action</th>
<th>Notify</th>
<th>Request permission</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capture</td>
<td>42</td>
<td>17</td>
<td>14</td>
</tr>
<tr>
<td>Provide restricted access</td>
<td>42</td>
<td>7</td>
<td>13</td>
</tr>
<tr>
<td>Provide public access</td>
<td>45</td>
<td>11</td>
<td>15</td>
</tr>
</tbody>
</table>
more conditional handling of robots.txt

- Always respect robots.txt: 38% in 2011, 22% in 2013
- Sometimes/conditionally respect robots.txt: 33% in 2011, 55% in 2013
- Never respect robots.txt: 8% in 2011, 8% in 2013
- Don't know: 21% in 2011, 16% in 2013
social media archiving policies are uncommon

- Has social media archiving policy: 24%
- Lacks social media archiving policy: 76%
policies based on community practices
Landscape Summary

“Mt Baldy from Box Springs Mountain wi Theodolite” by signal mirror under CC BY 2.0
profile of the average survey respondent

- university archive
- started in last three years
- Archive-It user
- \( \frac{1}{4} \) FTE web-savvy archivist
- concerned w/ content capture, cost, and use
- broad level of description
- ambivalent about collaboration
maturity and convergence

• maturity
  • 75% cite *some or significant progress* since 2011
  • 38% *started programs* since 2011
  • 8% more *programs in active status* since 2011

• convergence
  • 79% using *external service providers*
  • 81% devoting $\frac{1}{2}$ FTE or less to web archiving
  • 67% rely on *community practices* for policy-making
  • 13% more *using Wayback* since 2011
challenges and opportunities

• challenges
  • 53% concerned about data volume growth
  • 47% concerned about fostering access
  • more than 73% concerned about content capture

• opportunities
  • 33% interested but not yet involved in collaborations
  • 76% lack social media archiving policies
  • less than 23% of archived materials are described
implications and questions

• implications
  • web archiving not (yet) a top institutional priority
  • demand for ongoing Archive-It technical investment
  • U.S. web archiving landscape is changing quickly

• questions
  • how to build institutional support?
  • collaboration with whom and on what?
  • what’s not being archived?
  • how well are we curating what we do archive?
Nicholas Taylor
@nullhandle

“Thank You” by vistamommy under CC BY 2.0