2013 NDSA Web Archiving Survey Report Highlights

Nicholas Taylor (@nullhandle)
Web Archiving Service Manager
Stanford University Libraries

SAA Annual Meeting: Web Archiving Roundtable
August 13, 2014
NDSA Web Archiving Survey Working Group

Jefferson Bailey
Internet Archive / Archive-It

Kristine Hanna
Internet Archive / Archive-It

Cathy Hartman
University of North Texas

Abbie Grotke
Library of Congress

Edward McCain
University of Missouri

Christie Moffatt
National Library of Medicine

Nicholas Taylor
Stanford University
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

“Lego People” by Scoobay under CC BY-NC-SA 2.0
universities still make up most programs
SAA WebArch RT tops group affiliations

<table>
<thead>
<tr>
<th>group</th>
<th>2011</th>
<th>2013</th>
</tr>
</thead>
<tbody>
<tr>
<td>IIPC netpreserve.org</td>
<td>8%</td>
<td>7%</td>
</tr>
<tr>
<td>NDSA%</td>
<td>31%</td>
<td>33%</td>
</tr>
<tr>
<td>Web Archiving SAA Roundtable</td>
<td></td>
<td>45%</td>
</tr>
</tbody>
</table>
most programs are fractionally staffed
Maturity and Progress
programs have matured slightly since 2011
strong perceptions of progress since 2011

- Significant progress: 40%
- Some progress: 36%
- About the same: 20%
- Slightly worse off: 2%
- Much worse off: 2%
many new programs since 2011
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
concern about social media, databases, video
untapped interest in collaboration
Tools and Services
web archiving as a service still most popular
data not transferred from service provider
increased use of tools supporting W/ARC
less granular descriptive metadata
most don’t notify or seek permission
more conditional handling of robots.txt

- Always respect robots.txt: 38% (2011), 22% (2013)
- Never respect robots.txt: 8% (2011), 8% (2013)
- Don't know: 21% (2011), 16% (2013)
social media archiving policies are uncommon

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

- ARL Code of Best Practices: 27%
- Section 108 Study Group: 17%
- Oakland Archive Policy: 4%
- Counsel or service provider: 7%
- Statute: 4%
- Don't know: 5%
- Other organizations: 36%
takeaways and questions for SAA WebArch RT

• for individual organizations:
  • if you’re only self-archiving, what’s on your roadmap?
  • how are you preserving your web archive data?
  • how do you describe and enable discovery of web archives?
  • how do you handle robots.txt?
  • what are your plans for social media archiving policy?

• for the group:
  • what is this group (vs. IIPC, NDSA) best equipped to do?
  • what kind of collaboration are you interested in?
Thank You by vistamommy under CC BY 2.0