

NDSA Web Archiving Survey

The National Digital Stewardship Alliance (NDSA) Content Working Group [http://www.digitalpreservation.gov/ndsa/working_groups/content.html] is sponsoring this survey of organizations in the United States who are actively involved in or planning to archive content from the web. The goal of the survey is to better understand the landscape of web archiving activities in the United States, including what organizations or individuals are archiving, what types of web content are being preserved, the tools and services being used, and what type of access is being provided for researchers. More than one response per institution is acceptable if there are separate, distinct archiving programs within a given organization.

The survey will close October 31, 2011.

The information gathered as a part of this survey will be reported to NDSA members and summary results (which will not disclose individually identifiable responses) will be shared publicly, with an initial announcement of the results appearing on the Library of Congress's Digital Preservation blog, The Signal [<http://www.loc.gov/blogs/digitalpreservation>].

If you have any questions about this survey, contact Abbie Grotke, NDSA Content Working Group Co-Chair and Library of Congress Web Archiving Team Lead, at abgr@loc.gov.

Thank you for participating!

About Your Organization

* 1. First, tell us about yourself and your organization.

Name:	<input type="text"/>
Organization:	<input type="text"/>
City/Town:	<input type="text"/>
State:	<input type="text"/>
ZIP:	<input type="text"/>
Email Address:	<input type="text"/>

2. What is the access URL (or URLs, if more than one access point) for your web archives?

* 3. Organization Type:

<input type="text"/>
<input type="text"/>

4. Does your organization belong to either of these two groups? Select as many as apply.

- International Internet Preservation Consortium (IIPC) - netpreserve.org
- National Digital Stewardship Alliance (NDSA) - digitalpreservation.gov/ndsa

Archiving Program Information

NDSA Web Archiving Survey

5. What is the status of your web archiving activities?

- Planning/Considering archiving but haven't started yet
- Pilot/Testing
- Production/Actively crawling
- Have crawled content in the past, but we aren't currently crawling

Note: If you're not yet archiving but have already made some policy decisions, please feel free to continue with the survey with your plans in mind.

6. What are the goals of your web archiving activity? Select as many as apply.

- Archive your own web site as a type of institutional record
- Archive content from other organizations or individuals for future research

Other (please specify), or comments:

7. What year did your organization begin archiving web content?

Collection Areas

8. Does your organization have collection or selection policies that specifically address web archiving?

Comments

9. If yes, and the policies are publicly available and on the web, please provide a URL:

10. If your selection policies are not publicly accessible, would you consider sharing them with NDSA members? If yes, we will follow up with you at a later date.

NDSA Web Archiving Survey

11. Please briefly describe the scope of your web archive collections: what type of events, topics, themes, or approaches you take in archiving content from the web.

12. What types of content are you including in your archives?

- websites
- blogs
- social media

Other (please specify)

13. What subjects are represented in your web archives? Check all that apply.

	Have Archived or Currently Archiving	Planned
Arts and Culture	<input type="checkbox"/>	<input type="checkbox"/>
Current Events	<input type="checkbox"/>	<input type="checkbox"/>
Government, Politics, and Law	<input type="checkbox"/>	<input type="checkbox"/>
Maps and Geography	<input type="checkbox"/>	<input type="checkbox"/>
News, Media and Journalism	<input type="checkbox"/>	<input type="checkbox"/>
Religion and Philosophy	<input type="checkbox"/>	<input type="checkbox"/>
Science, Mathematics, and Technology	<input type="checkbox"/>	<input type="checkbox"/>
Social Sciences	<input type="checkbox"/>	<input type="checkbox"/>
World history and Culture	<input type="checkbox"/>	<input type="checkbox"/>

Other (please specify)

14. If you selected "News, Media, and Journalism" in Question 14, tell us a bit more.

	Have archived or currently archiving	Planned
Newspapers	<input type="checkbox"/>	<input type="checkbox"/>
Broadcast/television	<input type="checkbox"/>	<input type="checkbox"/>
Citizen Journalism/Community News	<input type="checkbox"/>	<input type="checkbox"/>

Other (please specify)

NDSA Web Archiving Survey

15. If you selected "Government, Politics, and Law" in Question 14, tell us a bit more.

	Have archived or currently archiving	Planned
Federal Government	<input type="checkbox"/>	<input type="checkbox"/>
State Government	<input type="checkbox"/>	<input type="checkbox"/>
Local Government	<input type="checkbox"/>	<input type="checkbox"/>
City Government	<input type="checkbox"/>	<input type="checkbox"/>
Local Elections	<input type="checkbox"/>	<input type="checkbox"/>
State Elections	<input type="checkbox"/>	<input type="checkbox"/>
Federal Elections	<input type="checkbox"/>	<input type="checkbox"/>
Government documents (PDFs, etc.) but not entire websites	<input type="checkbox"/>	<input type="checkbox"/>
Other (please specify)	<input type="text"/>	

Collaborative Archiving

16. Often web archivists come together to collaboratively preserve web content around specific events, themes, or domains. Has your organization ever participated in a collaborative web archive?

- Yes (if so, please describe in the comments below)
- No
- Don't know

Comments

NDSA Web Archiving Survey

17. As events occur where information unfolds rapidly on the web (such as natural disasters or terrorist attacks, or recent events in the Middle East) or when the content is too great for one archive to manage alone (such as .gov archiving), web archivists often reach out to as many interested organizations as are able to help. We are hoping to expand our network of collaborators on future projects. Would your organization be interested in future collaborative web archives (if they fit within your collecting scope and interests)?

- Yes
- No
- Maybe

Comments

Crawling/Tools

18. Are you using an external service or organization to archive, or crawling in-house?

- External service or company
- In-house
- Both

Comments

19. If an external service or organization is used, which one?

- Archive-It
- California Digital Library's Web Archiving Service (WAS)
- Hanzo Archives
- Internet Archive's Contract Crawling services
- Iterasi
- Reed Technology's Web Archiving Service

Other (please specify)

NDSA Web Archiving Survey

20. If you are using an external service, have you transferred any of your archived data from that service to your organization?

- Yes
- No

21. If you have not yet transferred any of your data, why not?

- Building our in-house infrastructure but hope to transfer soon
- No place to store/maintain it
- Not sure what we'd do with it once we got it

Other (please specify)

22. If crawling in-house, what tools or software do you use?

- Adobe Web Capture
- Grab-a-Site
- Heritrix
- HTTrack
- Web Curator Tool

Other (please specify)

23. What viewer or software are you using to provide access to your web archive data?

- Wayback Machine
- WERA
- Custom viewer (please describe below)

Other/Comments (please specify)

Researchers and Access

NDSA Web Archiving Survey

24. What kind of access do you provide to researchers? Select as many as apply.

- URL search
- Full-text search
- Browse list by URL
- Browse list by Title
- Catalog records: Collection-level description
- Catalog records: Item-level description

Other (please specify)

25. How are researchers using your archives?

Permissions and Robots

26. Do you ask site owners permission to crawl their websites or content?

- Always
- Sometimes/It depends
- Never
- Don't know

27. Do you ask site owners permission to allow you to provide access to archived content publicly (that is, permission to provide access outside of your organization's physical location)?

- Always
- Sometimes/It depends
- Never
- Don't know

Other (please specify)

NDSA Web Archiving Survey

28. Do you respect robots.txt when crawling?

- Always
- Never
- Custom (please explain in comments)
- Don't know

Other/Comments

Thank you!

This ends our survey. Thank you for time!

Subscribe to the NDIIPP's Digital Preservation blog, The Signal [<http://blogs.loc.gov/digitalpreservation/>], for announcements about the results of this survey.