Advocating for Web Archivability

Nicholas Taylor
Web Archiving Service Manager
Digital Library Systems and Services

NDSA Content Working Group
May 7, 2014
archivability?
Why ADVOCATE?
ignorance not indifference
good return on investment
opens up new opportunities
Making a

COMPELLING CASE

“Sassy yet Sophisticated” by Bailey under CC BY-NC-ND 2.0
future users are users, too

“a connection between past and future” by Gioia De Antoniis under CC BY-NC-ND 2.0
maintain web usability
improve temporal web usability

http://www.house.gov/wayback

Saved 425 times between September 21, 2006 and January 2, 2013.

PLEASE DONATE TODAY. Your generosity preserves knowledge for future generations. Thank you.
improve temporal web usability
recover your lost website
refer to earlier website versions
archivability is the new accessibility
institutional history
websites are cultural artifacts

The World Wide Web project

The WorldWideWeb (W3) is a wide-area hypermedia[1] information retrieval initiative aiming to give universal access to a large universe of documents.

Everything there is online about W3 is linked directly or indirectly to this document, including an executive summary[2] of the project, Mailing lists[3], Policy[4], November’s W3 news[5], Frequently Asked Questions[6].

What’s out there?[7] Pointers to the world’s online information, subjects[8], W3 servers[9], etc.

Help[10] on the browser you are using

Software Products[11] A list of W3 project components and their current state. (e.g. Line Mode[12], X11 Viola[13], NeXTStep[14], Servers[15], Tools[16], Mail robot[17], Library[18]).

Technical[19] Details of protocols, formats, program internals etc

<ref.number>, Back, <RETURN> for more, or Help:  

“The World Wide Web project”
facilitate compliance
optimize for other crawlers
How to IMPROVE ARCHIVABILITY
follow web standards and accessibility guidelines
use a site map, transparent links, and contiguous navigation
maintain stable URLs and redirect when necessary
use semantically-meaningful URLs
be careful w/ robot exclusion rules

```
Disallow: /scripts/
Disallow: /themes/
```

“drupal/robots.txt at 7.x”
minimize reliance on external assets necessary for presentation
minimize reliance on external assets necessary for presentation
serve reusable assets from a single, common location
specify HTTP response headers for caching and content encoding
embed metadata, especially character encoding
use durable data formats

“Lascaux cave painting” by Christine McIntosh under CC BY-ND 2.0
prefer responsive design over user-agent personalization
examine your site in the Internet Archive Wayback Machine

Greetings & Welcome to
A Multidimensional Perception ~/\= & PCGuru

RealAudio 3.0 format

Free Speech Online
Blue Ribbon Campaign

Last Updated December 5, 1996
Best viewed with either

Netscape Navigator or Internet Explorer

We are constantly growing in this cyberdimensional reality.
The following list of skills & talents are currently available.

- Audio Ambient/Soundboard Mastering.
- Video Producing/Directing/Editing.
- 35mm Still Photography.
- PC Consulting/Training/Upgrade/Repair/Installation.
- HTML3 Home Page Programming & Designing.
- Internet Consulting/Training.
- Technical Advising.
assess archivability w/ Archive Ready

Is your website Archive Ready?

What is ArchiveReady?
An online tool which evaluates if your website will be archived correctly by web archives, such as the Internet Archive.

Why bother?
Because not all websites are archive ready and this may result in invalid web archives and ultimately in information loss.

Who is it for?
Web professionals who need to check if their websites are archive ready.
Web archive engineers who need to evaluate target websites before harvesting and ingestion.

How does it work?
ArchiveReady analyses your website (i.e. HTML, Images, CSS, JS, Sitemaps) and calculates a set of Archivability Facets: Accessibility, Cohesion, Metadata, Performance & Standards Compliance.

How much does it cost?
ArchiveReady.com is completely FREE for personal use. Large scale use of the ArchiveReady.com API as well as technical and scientific support is available on a fee. Please contact the author to learn more.

Do you need more information?

Service started in 2012/10/01, websites checked: 37018.
Opportunities for COLLABORATION

“collaboration” by Jennifer Leonard under CC BY-NC-ND 2.0
discover content to be archived

“Pecio en las Brother I” by Felix Esteban under CC BY-NC-ND 2.0
custom 404 page

Internet Archive: “Free ’404: File Not Found’ Handler for Webmasters to Improve User Experience”
deprecate without deletion
archive and redesign
thank you!

Nicholas Taylor
ntay@stanford.edu