

Web Archiving for University Records

Elliot Williams

School of Information, The University
of Texas at Austin

Roadmap

Defining web archiving

Why university archives should care

Web archiving methods

Appraisal challenges

“Web archiving is the process of collecting websites and the information that they contain from the World Wide Web, and preserving these in an archive.” (UK National Archives, 2011)

Also, providing access to those materials!

Why should university archives care?

“At present the Internet is an essential record of human life and continues to grow in size daily, which has instigated a need and an opportunity for archivists to be actively documenting this base of human knowledge. No single organization will be able to fully archive the entire Internet; therefore it will become imperative that as many archives as can take part in web archiving.”

- SAA Web Archiving RT, “Goals and Objectives”

Why should university archives care?

An increasing amount of material only exists online.

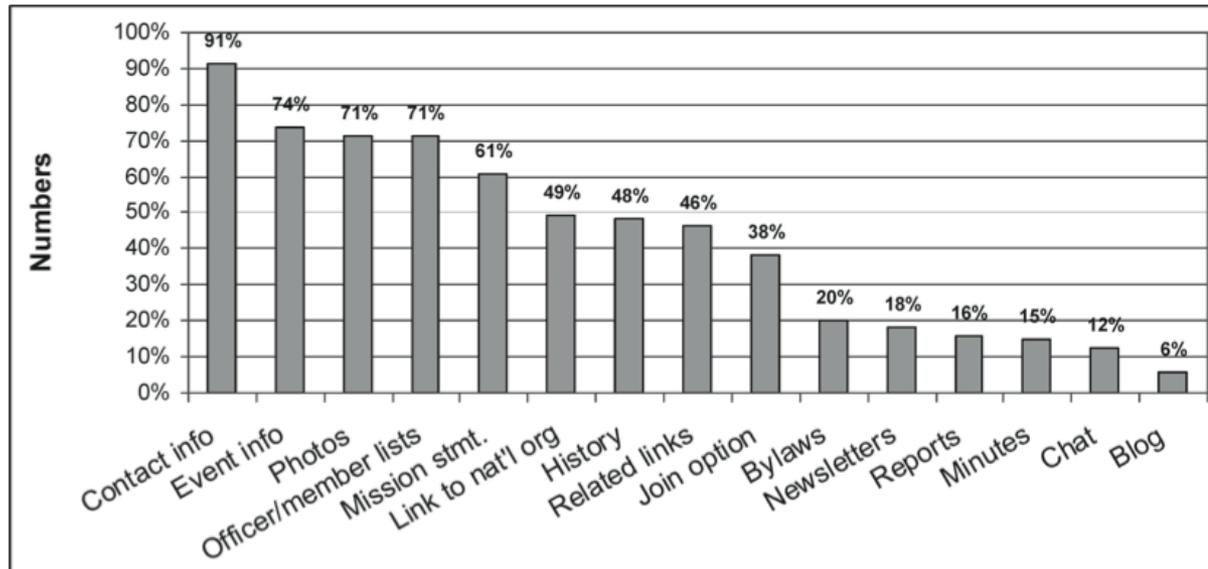


FIGURE 7. Website Features/Content (N=294)

Content stored on student organization websites, from Prom and Swain, "From the College Democrats to the Falling Illini," (2007).

Web archiving methods

- Web crawlers - e.g. Heritrix
- Offline browser software – e.g. HTTrack
- Archiving constituent files obtained from site creator or administrator

Challenges: Technological limitations

Depending on the method used, certain kinds of content cannot be captured or preserved.

- Social media sites (especially Facebook)
- Flash and Javascript
- Streaming video or audio
- Database-driven content

Challenges: Technological limitations

Database-driven content:



The screenshot shows the top navigation of the School of Information website. The logo features a stylized 'i' and 's' in a circle, with 'SCHOOL OF' above and 'information' to the right. Navigation links include 'ABOUT+', 'PROGRAMS+', and 'ADMISSIONS+'. A breadcrumb trail shows 'Home » Courses'. Below a decorative hatched line, the word 'COURSES' is displayed in large orange letters. Underneath, the year '2013:' is followed by links for 'Spring', 'Summer', and 'Fall'. Further down are links for 'Other Semesters' and 'School of Information Course Descriptions'.

SCHOOL OF
information

ABOUT+ PROGRAMS+ ADMISSIONS+

Home » Courses

COURSES

2013: [Spring](#) | [Summer](#) | [Fall](#)

[Other Semesters](#)

[School of Information Course Descriptions](#)



The screenshot shows the 'CLASSES' page. It features a note about data transfer errors for classes before Spring 2005. Below the note are two dropdown menus for 'Semester' (set to 'Fall') and 'Year' (set to '2013'). A 'Display Class List' button is located to the right of the year dropdown.

CLASSES

Note: Class listings before Spring 2005 may have data transfer errors and are provided for historical interest only.

Semester: Year:

Challenges: Privacy and IP

Is all material that is posted online fair game to be captured by an archives?

Is it feasible for archival staff to thoroughly review everything that is captured for privacy and intellectual property issues?

Challenges: Engaging with donors

Web records can be captured and preserved without disrupting the ongoing existence of the site, and even without notifying the record creator.

Saves time for archivists and expands universe of possible materials for inclusion

But, also isolates archivists, eliminates the possibility of incorporating creators' knowledge, and diminishes opportunities for education and outreach

Conclusion

Web archiving can and should be a part of any attempt to documenting the history, culture, and functioning on colleges and universities.

Traditional archival tools and perspectives will continue to be important in this new realm.

Works Cited

The National Archives (UK). “Web Archiving Guidance” (2011)
<http://www.nationalarchives.gov.uk/documents/information-management/web-archiving-guidance.pdf>.

Prom, Christopher J., and Ellen D. Swain. “From the College Democrats to the Falling Illini: Identifying, Capturing, and Appraising Student Organization Websites.” *The American Archivist* 70 (2007): 344-363.

Society of American Archivists Web Archiving Roundtable. “Goals and Objectives” (2013) http://saa.archivists.org/4DCGI/committees/SAATBL-WEBRT.html?Action=Show_Comm_Detail&CommCode=SAA**TBL-WEBRT&.

Wagner, Jessica L., and Debbi A. Smith. “Students as Donors to University Archives: A Study of Student Perceptions with Recommendations.” *The American Archivist* 75 (2012): 538-566.

Links to these sources, a copy of this presentation, examples of university web archives, and general web archiving resources can be found at: www.elliotdwilliams.com/ssa-web-archiving