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

Appraisalal 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.

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:
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


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