Web Archiving Comics at the Library of Congress

Megan Halsband, Reference Specialist mhals@loc.gov

Ingest/Acquisition

The Library of Congress uses open source and custom-developed software to manage different stages of the overall workflow[1]:

Selection and permissions: The Library of Congress has developed and implemented the DigiBoard, a tool that allows curatorial staff to select websites for archiving and facilitates management of legal permissions for website capture and/or offsite access as well as web archive quality review processes. Development of the platform is ongoing. This tool performs a variety of collections management functions including managing site owner contact information, allowing review of multiple collections through the "Reviews" page, and assigning various permission levels to allow multiple users to work on the same collection.[2]

Acquisition: Web archives are created using the Heritrix archival web crawler.

Storage: Web archives are packaged in BagIt-conformant packages using the BagIt Library.

Access: Web archive replay is enabled by a local installation of the Wayback Machine (Internet Archive).

See FAQ for more information:

http://www.loc.gov/webarchiving/faq.html

[1]Grotke, A., & Jones, G. (2010). DigiBoard: A tool to streamline complex web archiving activities at the Library of Congress. Proceedings from IWAW '10: 10th International Web Archiving Workshop, Vienna. http://iwaw.europarchive.org/10/IWAW2010.pdf
[2]Grotke, A. (2014) Digiboard Curator Tool. Open Wayback Developers Group. IIPC Consortium General Assembly, Paris.
http://netpreserve.org/sites/default/files/attachments/IIPC_CuratorTools_DigiBoard2.ppt
[3]Garrity, S. (2014) Award! Award! Award! The Comics Journal.
http://www.tcj.com/award-award-award/
[4]Comics and Cartoons. Library of Congress Collection Policy Statement.
http://www.loc.gov/acq/devpol/comics.pdf
*These collections are still in process and are not yet available online.

Web Archiving Comics at the Library of Congress

Megan Halsband, Reference Specialist mhals@loc.gov

Small Press Expo Collection

Collection established in 2011
24 sites currently crawled*

Started in 2011 as a pilot collection for acquiring online comics content; initially included only the winners of the Ignatz Award for Outstanding Online Comic as well as the SPX website and social media pages. There are 13 new active crawls for 2014.

Collection: Many comic materials now start online and then move to print; webcomics are the new 'ephemeral' material, such as sketches or notes, that libraries have collected for years. For example, American Elf, the 2003 & 2004 Ignatz Award Winner for Outstanding Online Comic, was recently 'completed' by the artist, and was set to be pulled offline; this comic is the original place of publication for a set of characters/stories that would later be published as print comics/graphic novels.

Challenges: The Ignatz Award for Outstanding Online Comic dates back to 2001; some of these sites are now defunct or dead, and many have little or no contact information.

Webcomics Collection

Collection established in 2014
25 sites currently crawled*

Complements sites included in the Small Press Expo Collection; selections for this collection include award winning webcomics[3] (excluding the Ignatz Awards, which are collected in the Small Press Expo Collection), as well as webcomics that have significance in the field due to longevity, reputation, etc.

Collection: This collection supplements the Library of Congress' extensive holdings in both comic books, graphic novels, and original comic art, and fulfills criteria in the Library's Collection Policy Statement.[4] Webcomics are an increasingly popular format utilized by contemporary creators in the field. This particular material often has very limited or no accompanying print materials that could be collected by the Library, though some creators work in both formats. Capturing these websites provides the Library with a method to acquire content not represented in other collection formats, as well as attempting to preserve the experience of the original intended method of accessing this content.