

September 19, 2011

Background

In November 2010, the Library of Congress convened a Citizen Journalism workshop that brought together researchers, bloggers, journalists, academics and archivists to address selection and preservation issues for hyper-local community news on the web. One of the themes that came out of that meeting were

For more information about this meeting, visit

http://www.digitalpreservation.gov/news/2010/20101202news_article_citizen_journalism_mtg.html. Minutes have been posted here:

http://www.digitalpreservation.gov/news/events/other_meetings/CitizenJournalism10/index.html.

During that workshop, participants identified a need for an easy way for content creators to “opt-in” to preservation, by way of a plugin or other mechanism built into blogging tools.

In July 2011, the Content Working Group of the National Digital Stewardship Alliance held a workshop to discuss categories of at-risk digital content. Once again, blogs were on that list, and the idea of a blog preservation plugin, integrated within blogging software, was again brought up. With renewed interest and a now a working group to sponsor a project, Abbie Grotke and Kristine Hanna reached out to Andrew Nacin, who had represented WordPress at the November 2010 meeting, to explore moving forward with a blog preservation plugin. Nacin is now full-time at WordPress and has offered to write code and help implement whatever is developed.

Potential Approaches

In September Grotke, Hanna and Nacin met to brainstorm potential approaches. Nacin believes that building any of these should not be too time consuming or difficult, we just need to determine what the specifications are. He also felt that any plugin written for WordPress, once it was set, it could likely translate to other blogging platforms.

Here are a few ideas that emerged during the call:

1) Export to Archive Format for Delivery to Preservationists

A plugin could be developed to export a blog in a structured data format that would meet the needs of archivists and preservationists.

Questions to be determined:

- a) What format? (is WARC possible?) Can it export to multiple formats, so there is a choice? Technical specifications would need to be conveyed to Nacin.
- b) How often? Is this determined by the blog owners or by the preservationists? Is this each time the content has been updated? Or on a pre-determined schedule?
- c) What would this export include (comments, imbedded images/files)?
- d) Where would the exported file go, if exported for preservation? A few ideas were brainstormed:
 - a. File sent directly to one organization for preservation (*What organization?*)

- b. File downloaded locally then handed off to preservationists (for example, an author donating papers to a university might want to donate a copy of their blog as well).

2) Export to Archive Format for Personal Archiving

A similar export function as described #1 could be useful for those interested in personal archiving. It would be useful to have a way to allow blog owners to export their blogs for backup and preservation. Similar questions about what format(s) to export to remain, and what is included in this export (comments? Imbedded files?). The assumption would be this could occur at any time; the blog owner would trigger the export.

3) Flag for Opt-In to Preservation/Harvesting

A third approach might allow a blog owner to opt-in to preservation by filling out a “permission form” within a WordPress plugin. The plugin could be set up to automatically notify someone (see b below) that the blog owner has opted in and that it may be crawled, OR, a some sort of flag or banner could appear on the site alerting anyone that the blog owner has opted in for preservation (like LOCKKS, or Creative Commons).

There are a number of policy decisions that would need to be determined:

- a) If a blog owner opts-in, would that guarantee preservation? (there is the question of volume and value – if some of the content is not something that libraries or archives are willing to preserve, what then?)
- b) Where would notice of opt-in go? One idea is the NDSA Content Working Group’s clearinghouse (perhaps a list is generated announcing blogs willing to be preserved, and members or others can “sign up” to preserve certain ones). Another idea is that all notices could go to one organization such as the Internet Archive to be added to the global archive at web.archive.org (if they agree). Some combination of these approaches could also occur.
- c) How often/frequently would the notifying occur?

Recommendations

Grotke, Hanna and Nacin agreed that a pilot might be useful to test volume and to explore some of the policy issues. After further discussion with Martha Anderson, LC proposes the following next steps:

- 1) LC to convene a meeting by end of October (via Webex or in person at LC – or some combination) of stakeholders, including interested parties in NDSA and in the broader community, including archivists, librarians, WordPress staff, others, to discuss policy questions and further determine requirements for the plugin(s), and to set forth a plan for a WordPress pilot.
- 2) Pilot could then lead to branching out to other blog platforms, further refinement of the process and tools.

Additional recommendations are for the CWG to review the blog preservation work conducted in the UK: <http://ukwebfocus.wordpress.com/2011/07/18/blog-preservation-and-plugins/> contains links to key reports.