

Columbia/Mellon Web Archiving Incentives Award – Awardee Final Report

Project: Perma.cc: Mitigating the Pervasive Problem of Link Rot In Scholarly Works and Preserving Online Content
Project ID: 6678163---01
Subrecipient: Harvard Library Innovation Lab
P.I.: Jonathan Zittrain
Date: November 3, 2015

This is the Harvard Library Innovation Lab's Final Report on the status of work supported by the Columbia/Mellon Web Archiving Incentives Award. We received \$24,000 on October 3, 2014 to develop new features for our Perma.cc online link preservation service.

We spent a total of \$20,625 to complete the work on the grant.

The new features developed include an API to support batch processing of Perma.cc links, a WordPress plug in for Perma.cc links, implementation of support for the Memento framework, implementation and optimization of a caching layer for the Perma.cc website, and the creation of tools for the preservation of the Perma.cc archive in case of an unanticipated shutdown.

Work Accomplished

The largest portion of work completed under the award went to developing a stable and extensible API, which supported the other subgoals of the award. We developed an API that allows programmers to perform any and all actions related to creation and editing of Perma Links. To ensure that the API would remain stable and up-to-date, we invested substantial effort updating the frontend perma.cc site to itself use the Perma API for all user actions.

Once we were satisfied with the API, we used it to implement the other subgoals of the award, including: a WordPress plugin based on the Amber project developed by the Berkman Center for Internet and Society; Memento support; and an import/export script for users to create their own Perma instances.

We also used a portion of the award to support configuration of a caching layer to allow for Perma's continued growth. We ultimately settled on caching via the CloudFlare CDN service, which is now configured to cache Perma pages for logged-out users. Implementing caching required code changes as well as configuration changes.

When it became clear that we would achieve our originally proposed work without spending down the entire award, we obtained a no-cost extension to perform additional work on Perma's resiliency and scalability. We then completed additional work to modify the Perma front-end to comply with the "12 Factor App" principles that will allow it to run equally well on platform-as-a-service offerings like Heroku, local servers, or any mix of the two.

Outcomes and Deliverables

The Perma API is now live and in production. Documentation is at <https://perma.cc/docs/developer>.

The API is already supporting third-party users such as <https://canarywatch.org/>, a project that tracks changes to "warrant canaries" on websites. The API is also in use by our experimental Word plugin: <https://github.com/harvard-lil/perma-word-plugin>

Our demonstrated integration with the Amber WordPress plugin persuaded the Amber team to offer Perma as a first-class archival option to their users, alongside their existing web archival options. Perma support will be included in the next version of Amber.

Memento support is live and in production. For example: https://perma-archives.org/warc/pywb/*/http://example.com

Front-end caching is live and in production.

The import/export script is publicly available at https://github.com/leppert/perma_export

Regarding scalability, our internal staging server is now running across a mix of Heroku cloud-based instances and local instances. We anticipate deploying the same change to production over the coming weeks.

Description of any issues or obstacles encountered that might affect the ability to complete work within the revised project timeline

None.

Changes to the Original Product Plan and Proposed Deliverables

As described above, an additional scalability-oriented deliverable was added and completed during the course of the project.

Presentations about the Grant-funded Work.

We spoke about Perma and acknowledged the grant award at a number of conferences. They include:

- Link rot symposium hosted by Georgetown Law Library entitled, 404/File Not Found: Link Rot, Legal Citation and Projects to Preserve Precedent, <http://www.law.georgetown.edu/library/404/>. Jonathan Zittrain, Principal Investigator, was the Keynote Speaker and Kim Dulin, Director of the Harvard Library Innovation Lab, was a panelist. Jonathan's keynote was a broad based

discussion of the necessity of libraries taking on the responsibility for preserving our digital trail. Kim's presentation was focused on Perma.cc as a solution. Streams of the presentations can be found at <http://www.law.georgetown.edu/library/404/streaming.cfm>.

- LIPA (Legal Information Preservation Alliance) program at the 2015 American Association of Law Libraries Annual Meeting. Kim Dulin spoke about Perma.cc's growth in the legal community and described the Perma.cc service.
- MIT Libraries Program on Information Science Brown Bag Series. Kim Dulin spoke about how libraries can mitigate link rot in scholarship. <http://informatics.mit.edu/blog/brown-bag-report-kim-dulin-permacc>
- New England Library Association Annual Conference. Jack Cushman and Sebastian Diaz spoke about link-rot and Perma.cc, highlighting work completed under this award. <http://nelib.org/nela2015/>