

Archive-It Quarterly Open Call, 11/8/2017 (Q3)

All you ever wanted to know about Archive-It, but were not afraid to ask or at least hear about

Presented by Lori

("AIT" = our acronymization of Archive-It and "email us" = ait@archive.org or via support ticket)

Program Updates

- Q3 statsathon: 35 new partners (including CommWebs cohort), 2.1 billion URLs & ~61TB archived in Q3. (continuing a ~5-10% increase per quarter)
 - 30,500 seeds added in Q3 (1.17 million total)
 - 407 collections added in Q2 (9,087 total)
 - 425,223 pageviews for www.archive-it.org
- Partner support stats:
 - 346 new tickets filed by partners
 - 314 tickets solved by Web Archivists
 - 1048 WA responses to tickets (That's a lot of back and forth!!)
 - 95 Support Jiras solved (many things WAs can solve on their own but the trickier technical issues with capture/replay need to be escalated to engineers)
- Advanced Trainings:
 - [Quality Assurance](#) (8/8)
 - Advanced Scoping (9/19)
 - (upcoming - 11/14) Social Media
 - (upcoming - 12/5) Advanced Scoping
- Blog posts:
 - 7 posts, including ones from [NYARC](#), [GPO](#) and [Northwestern](#)
 - Volunteers for future posts welcome/encouraged
- Community Building -- Events attended -- find us and say hello at the next event you attend! Who knows, we might buy you dinner!
 - ARA -- Manchester
 - Archive-It Mid Atlantic Users Group -- NYC
 - IIPC GA -- Ottawa
 - iPRES -- Kyoto
 - Talk @ MIT Libraries -- Cambridge
 - DLF + NDSA -- Pittsburgh
 - MARAC -- Buffalo
 - (upcoming) CNI -- DC
 - (upcoming) DocNow -- St. Louis
- [NDSA Web Archiving Survey](#): If you haven't filled it out yet, please do! Closes November 17.

Engineering Activities

- WARC uploader

- A service integrated into the web application for uploading WARCs created outside Archive-It to be added to your collection. Currently in beta release. <https://partner.archive-it.org/1036/collections/7013>. Additional user testers wanted!
- Realtime reports rebuild
 - Total rebuild of the many systems supporting crawl reports and real-time stats. Real-time analytics are a pain! Released in early September.
- WASAPI
 - API functionality complete, including derivative job submission and results! <https://github.com/WASAPI-Community/data-transfer-apis/tree/master/ait-specific-ation>. Also wrapping up the white paper attendee reports from our National Symposium on Web Archiving Interoperability in February. User testers wanted!
- Auto-add Social Media rules
 - For Social Media sites, including Facebook, Instagram, YouTube, etc we'd create rule templates that would apply to new seeds that are added, and will automatically update/add new rules when our recommendations change.

Grant & Special Project Updates

- We have kicked off our IMLS-funded project, *Community Webs: Empowering Public Libraries to Create Community History Web Archives*, to provide training, resources, and support for fostering a network of public libraries building local history web archive collections.
 - Two year IMLS Grant
 - 27 Public Library partners
 - Project page: <https://archive-it.org/blog/projects/community-webs/>
 - Grant details: <https://www.imls.gov/grants/awarded/re-85-17-0060-17>
 - Kickoff meeting at IA last Friday! Look for related OERs soon

Future Planning

- Brozzler
 - Brozzler development ongoing. Focusing this Q on scaling out the hardware infrastructure dedicated to Brozzler instances, which should allow for expanded availability and more usage and crawling!
- Partner Data API
 - Preliminary investigation of existing data models around the API that powers "partner data". Let us know if you have any use cases for AIT data you'd like to use externally to the system.
- Moving Seeds from one collection to another
 - Still on the roadmap, though no timeline yet

In Other News

The department that Archive-It is in at IA also includes IA's Web Group (domain-scale web archiving, research, and data preservation services) and some staff contribute to both areas. Some quick hits of what's up on that side of the digi-shop:

- We are archiving datasets now! And open access journals! We have crawled well over 15million PDFs and are working with multiple machine learning firms on journal article detection and metadata extraction. Expect more news soon.
- So far this year we have archived the web domains of .cu, .au, .nz, .gov, .lu, .mil, .id -- mucho ccTLDs! More forthcoming!
- We have End of Term web data from Library of Congress & UNT, have generated the full-text search index, have made datasets, and have some researcher partnerships. We expect CDL to update the portal by the end of the year, <http://eotarchive.cdlib.org/> and all our data will be publicly available.

QUESTIONS AND ANSWERS!!!