Electronic Currents

Assistant Editor: Sarah Dorpinghaus, University of Kentucky. Contact Sarah at sarah.dorpinghaus@uky.edu if you would like to guest author a column or have a good idea to share.

Increasing Access to Web Archives: Archive-It and the Discovery Layer

By Christina Beis, Kayla Harris, and Stephanie Shreffler, University of Dayton

Background

Web archiving involves the process of selecting, capturing, and making accessible content from the web. The Internet Archive has been capturing websites since 1996 as part of its mission to provide universal access to all knowledge. A lofty goal, but its web archiving tools, including the subscription-based Archive-It, are some of the best known in the field. Many Internet users may be familiar with the Internet Archive's Wayback Machine, a portal to view the sites that have been archived.

According to a report in *Scientific American* in 1997, the average lifespan of a website was just 44 days; in 2003, the lifespan was 100 days.[1] Estimating the lifespan of a website depends on a variety of factors and definitions, but regardless, the numbers demonstrate how ephemeral web content can be, and why web archiving is so important.

Project Planning

The University of Dayton purchased a subscription to Archive-It in 2014 to begin capturing web content that fit within the collecting scope of each of the three special collections areas, including the Marian Library, the US Catholic Special Collection, and the University Archives and Special Collections. Archive-It allows users to harvest, organize, and catalog their collection materials and then access them through the Archive-It website, where each institution has a homepage with a search box and facets.
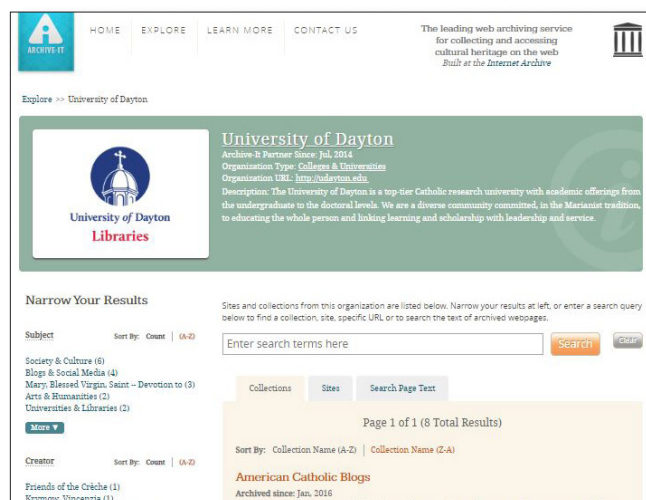
The University of Dayton Libraries' Archive-It Collection homepage at https://archive-it.org/home/udayton

Although the public could access the archived web content through the collection page, the material was not actively being promoted and the page was not linked from the University Libraries website. This was partly due to the inconsistent metadata for each of the sites, as consistent standards or guidelines did not yet exist. A cross-departmental team was formed in early 2018 consisting of archivists from each of the three special collection units, the Discovery Services librarian, the coordinator of cataloging, and the digital projects manager. The team's goal was to explore ways to increase access and use of the Archive-It collections.

One potential solution was to integrate the Archive-It records into UDiscover, the University of Dayton's locally branded name of EBSCO Discovery Service. The team did some initial research and sent messages to Archive-It and to our EBSCO's senior software as a service catalog specialist to inquire if others had already completed this process, but no examples of this specific implementation existed. We did learn that the New York Art Resources Consortium (NYARC) used an API to pull results from Archive-It into its Primo discovery layer, so there was some precedence for adding archived websites to a discovery layer.

Process

The first step in adding the Archive-It collections to the discovery layer was to submit an Institutional Repository Database Questionnaire to EBSCO. The questionnaire allows you to set a harvesting schedule, metadata format, and specifications for the data and collections. On the questionnaire, we selected a name, Archive-It @ UD, and for the database to be set up as a Digital Archive. Dublin Core was selected as the metadata standard to be harvested via OAI-PMH. We had originally set a monthly harvesting schedule, but later changed it to weekly due to the volume of edits the archivists were making to the metadata in Archive-It. We submitted the questionnaire to EBSCO in March 2018, and our Archive-It @ UD collection was live in UDiscover within 10 days. Typically, EBSCO estimates four to six weeks for collection building, but because our collection was only 30 records when we started, the process went much faster.

(Continued from page 13)

Once the Archive-It @ UD collection was live, we noticed a few issues that needed to be fixed. The largest clean-up needed was adding metadata fields for each of the seeds in Archive-It. With three different collection owners, as well as a staffing change that occurred during the process, it was important to standardize each seed's metadata. Some records were missing titles and authors at the seed level, while others needed subject headings or descriptions. In February 2018, the OCLC Research Web Archiving Metadata Working Group released recommendations on descriptive metadata.[2] Based on these recommendations, the coordinator of cataloging created local guidelines that dictate which elements we want to require at the seed and collection levels to maintain consistency.[3] The archivists used this data dictionary for metadata clean-up.

We also noticed that the source type icon and text in UDiscover did not match up to the true source type of the records as all records defaulted to "academic journal." Using a defined set of vocabulary for dc:type, the archivists went into their records on Archive-It and updated the metadata. We then submitted a Publication Type Lookup Table to EBSCO that contained appropriate mapping for the type source and icon.

Publication Type Lookup Table

The <TypeSource> mapped and Icon are fixed fields that are used to reflect your publication and document types present in your collections. This data is commonly located in <dc:type>, <dc:format> or <setName> tags within your data. Please review the list of viable publication types that EBSCO supports within its database and create a key in the table below to identify how your publication types should be represented. Add or delete rows as needed. If a document type in your data is not present in this lookup table, the value will be displayed as is in the database.

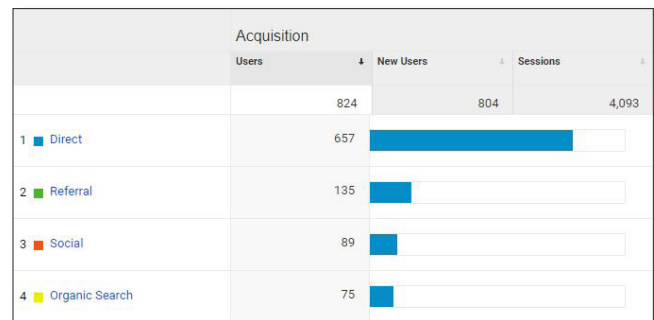| <dc:type> Type Source Field Content	<TypeSource> mapped	Icon (fixed per pub type)
Social Media	Electronic Resource	Electronic Resource
Facebook page	Electronic Resource	Electronic Resource
Instagram page	Electronic Resource	Electronic Resource
Twitter page	Electronic Resource	Electronic Resource
Blog	Website	Electronic Resource
Website	Website	Electronic Resource
Video	Video	Video

Publication Type Lookup table submitted to EBSCO details the type source field content used in the Archive-It seeds and how they are mapped to the records in UDiscover.

Promotion and Usage

In addition to making the Archive-It collections searchable in UDiscover, we also added an Archive-It catalog record and link in our listing of databases and created a LibGuide. To increase awareness of these collections, we published a blog post on our library website, which was shared through the university electronic newsletter, and reached over 19,000 subscribers. The blog was also promoted on our library Facebook page. We worked with the libraries' coordinator of marketing and engagement to reach out to other campus partners. She shared information about Archive-It with all of the university social media managers. In addition, all university website managers were e-mailed information about our web archiving program and were encouraged to utilize the "Save Page Now" function on the Wayback Machine when updating pages on the University of Dayton website. In honor of Archives Month in October 2018, we hosted an internal professional development event for librarians and staff to learn more about web archiving, our Archive-It collections, and how to access them.

We are tracking the use of the Archive-It collections in a few different ways. On the University of Dayton Archive-It homepage, we have added our Google Analytics Tracking ID. This allows insight into our audience, including referral channels. There are also options in EBSCOAdmin to run targeted usage reports of custom collections and repositories, including COUNTER 5 reports.

Acquisition			
	Users ↓	New Users ↓	Sessions ↓
	824	804	4,093
1 ▪ Direct	657		
2 ▪ Referral	135		
3 ▪ Social	89		
4 ▪ Organic Search	75		

Google Analytics data show the referral traffic to the University of Dayton Libraries' Archive-It Collection from October 1, 2018, to January 1, 2020.

Future Plans

Now that we are capturing statistics from both Google Analytics and EBSCOAdmin, we hope to investigate different ways to increase our traffic. For example, when we wrote the blog post on web archiving in October 2018, we noticed an increase in social media referrals to our collections. Our average number of users has remained fairly steady, but as the archivists continue to add to the Archive-It collections, we will also promote their use.

After the integration of our Archive-It records into the discovery layer, Archive-It published an article on different access integrations.[4] Several examples, like NYARC, include using APIs, while other Archive-It partners have built custom portals to integrate their web content into

their websites. Given our institution's size and available resources, using an API or creating a custom portal would not have been a feasible solution. We also believe that adding the archived web content to our discovery layer avoids creating another silo of special collections materials, as we are already using a digital commons institutional repository and ArchivesSpace for collection management.

This project was a rewarding collaboration between archivists and technical services librarians. We all gained insight into the workflows of each area and hope to find ways to collaborate again in the future.

Notes

This article was adapted from a presentation (bit .ly/37C8MMd) given at the 2019 Best Practices Exchange in Columbus, Ohio. A full-length article (bit .ly/3aO0RgQ) is available in the *Journal of Web Librarianship,* vol. 13, issue 3, or as a postprint (bit.ly/2RUxdhi) in the University of Dayton's institutional repository, eCommons.

1. Mike Ashenfelder, "The Average Lifespan of Webpage," *The Signal* (blog post), Library of Congress, November 8, 2011, https://blogs.loc.gov/thesignal/2011/11/the-average-lifespan-of-a-webpage.

2. Jackie Dooley and Kate Bowers, "Descriptive Metadata for Web Archiving: Recommendations of the OCLC Research Library Partnership Web Archiving Metadata Working Group," https://www.oclc.org/research/publications/2018/oclcresearch-descriptive-metadata.html.

3. Ione Damasco, "Archive-It UD Data Dictionary," https://docs.google.com/document/d/1BqY3OLPE KD3npyOhwokMkAhFTkAZS6w7yL9vcTXVYG8/ edit?usp=sharing.

4. Karl-Rainer Blumenthal, "Archive-It APIs and Integrations," https://support.archive-it.org/hc/en-us/articles/360001231286-Archive-It-Access-Integrations.