Allie Whalen, Blake McDowell, Genevieve Havemeyer, Karl McCool
October 28, 2014
Archive-It Web Crawl

INTERFERENCE ARCHIVE

<u>**Introduction: (background about Interference)**</u>
        This project seeks to archive the website and associated social media platforms for Interference Archive, a volunteer-run library and archive collecting five decades of activism history, based in Brooklyn, New York. Fairly new, Interference Archive opened in December of 2011. Their archive maintains a collection of media, objects, and ephemera from social movements both national and international, including posters, flyers, zines, books, comics, photographs, moving images, audio recordings, T-shirts, and buttons. Run by a core collective of ten members, Interference Archive is anything but hierarchical or bureaucratic.
        Our discussions with our contacts at Interference Archive and the decisions both they and our group have made during the course of this project have been shaped by the collectively run nature of the organization. Although Bonnie Gordon is the named curator for this project, our discussions with her and fellow core collective member Molly Fair have been shared with the Interference born-digital working group and through that we have also received input from Anna Perricci. Furthermore, the collectively-run nature of the organization has meant, not only that we had more than a single curator to communicate with, but also that no one single person at the organization had all the answers for us. As we would discover, different collective members had different knowledge to offer about what Interference-associated web content needed archiving and   what   did   not.

<u>**Getting Started:**</u>
        In our early discussions, our group had considered a few different potential curators and organizations, all dealing with politically engaged or challenging work, much like Interference Archive, including both Fales and Tamiment Libraries at NYU. As it would turn out, although we did not choose to contact anyone at Tamiment Library to be a curator for our group, Tamiment archivist Nicole Greenhouse would prove instrumental in helping us with this project. After noticing that the Interference Archive website did not seem to be archived online, our group decided to approach core collective member Molly Fair, initially through email and shortly after that in a face-to-face conversation at the AMIA conference. Because she has recently taken a less active role in Interference Archive, Molly recommended contacting Bonnie Gordon, who became our official curator for this project.
        In our initial conversation with Molly, her immediate response was to suggest we use Archive-It to capture the website for Justseeds Artists' Cooperative, a separate

collective, nevertheless associated with Interference Archive. To her knowledge, the Justseeds website had not been archived. There are currently plans to move publishing platforms for the Justseeds website from Movabletype to Wordpress and Molly is concerned that there will inevitably be some loss of data during this migration of web content. We would later learn, through Nicole Greenhouse, that Tamiment Library was in fact archiving the Justseeds website, which Molly had not known. Our group let Nicole know about the plans to move web platforms for Justseeds and let Molly and Bonnie that they could contact her to alert her before the move took place, so she could get one last capture of the website. Although Nicole, Bonnie, and Molly all know each other, they did not share all the same knowledge about the web archiving needs of Interference Archive and associated websites like Justseeds. Tamiment Library does not as a matter of policy inform organizations that their websites are being archived by the library. Furthermore, as stated above, the nonhierarchical, collectivist nature of Interference Archive meant that our group needed to communicate with more than just one member of the collective.

Planning & Scoping:

After our initial proposal to Bonnie and Molly, the parameters for our crawl were outlined in email correspondence with members of Interference Archive's Born-Digital Working Group. We discussed what aspects of Interference Archive's web presence were high priority, and what, if anything, was already being archived. To get more information about the status of Tamiment's crawl of the Justseeds website, our group met with Nicole Greenhouse, who recommended focusing on aspects like social media and video content, like Vimeo, which they aren't able to capture. We decided to focus just on Interference Archive, rather than including their partner organizations, in order to limit the amount of content captured for this project. We figured it would serve as a good example to provide IA with if they want to further pursue a more extensive archiving project for born digital works made in collaboration with other organizations. We have included single crawls of the Justseeds blog and Twitter feed to provide IA and Justseeds with a reference for any differences in how Archive-It captures content, since Tamiment does not capture social media, and there may be link to social media sites via their blog.

The Interference Archive website is well organized and contains digital representations of several publications, as well as serving as the home for the IA Blog, where unique posts are made weekly about ongoing activities and projects. A lot of new content and announcements are shared on Interference Archive's social media websites, including links to events and content from partner programs (like Justseeds). Since social media has been such a major platform for political and social activism in the past few years, we felt that it would be an interesting way to document IA's participation in this online social movement, since they are an archive of social movement production. The inclusion of the social media crawls also seemed

appropriate to do in addition to capturing the entire Interference Archive website because IA's Blog posts are less frequent than their Twitter feed, which seems fairly active.

Technical Implementation and Problems:

The changes needed to make our original crawls were the frequency. The original scope of each crawl - Interference Archive's website, www.interferencearchive.org, and their Twitter, Vimeo and Facebook pages – proved to be satisfactory. The issue with the frequency of the crawls was that we were running most of them far too frequently. Interference Archive does not have a fast changing website. It notes events and programs in-line with its political leanings and has a blog, but updates do not appear to happen more than a few times a month. There is no listing of weekly events or a constant flow of posts on the blog. As such, we set our frequency to two times a month which feels adequate for capturing the rate of changes to the website. Our seed URL for this crawl is http://interferencearchive.org/

The Interference Archive Vimeo website has but one video, uploaded 10 months ago. We set this to monthly with the idea that if Interference starts using it more we would capture the channel page at a reasonable pace. Archive-It does not currently support capture of videos from a user's channel page. Videos can be captured from their specific URL. The URL for IA's one uploaded video has been captured from http://vimeo.com/81646906. Archive-It does not support playback of videos in archived pages viewed on Wayback Machine, but they can be viewed in the Video Reports for each crawl. Unfortunately, this does not provide the public with access to archived videos, but it's better than nothing. Our seed URL for this crawl is http://vimeo.com/interferencearc.

The Interference Archive Twitter crawl we left as the most most frequently crawled seed. In keeping with the nature of Twitter, this is the most frequently updated area of Interference Archive's online presence. Our Twitter crawl is pulling in about 63,000 URLs per crawl vs. about 2,000 URLs for the crawl of the main Interference Archive website. 63,000 seemed like an excessive number for a single Twitter account, but on further inspection we realized the number is largely due to our crawl archiving the Interference Archive Twitter account in multiple different languages. It is also archiving all of the retweets by different users of any Interference Archive tweet. This aspect of the capture was unanticipated, but seems to align with IA's interest in documenting multicultural materials, and considering much of their work spreads into international causes. The archiving of retweets is also something we felt IA, and most any Twitter user, would want to access to as the retweets are an integral part of the Twitter process.

A technical issue that arises with the Twitter is that when clicking on some of the links on the archived pages the Twitter sign in box pops up. Clicking on "sign in" brings up a page saying that the requested URL is out of scope. This makes sense because it is highly doubtful that Twitter would allow their sign-in process to be archived. But, it is

unclear why certain links bring up the sign in box and others don't is not clear. Other than this one strange issue, Twitter appears to be a very friendly URL to crawl. Our seed for this crawl is https://twitter.com/InterferenceArc/

Our initial test and subsequent crawls revealed that the scope of the Facebook crawl needed to be adjusted in order for it to be captured. We had to add host constraints to block robots.txt files, for facebook.com to enable capture of content on facebook.com, and to fbcdn.net, and akamaihd.net - to ensure that content would display properly. Archive-It clearly states that crawls of Facebook do not always display well when archived, and that many features are limited, such as the ability to expand comments and capture every 'like'.

Conclusion:

Our project to archive the Interference Archive website and its associated social media content developed in response to the members of the Interference collective, as well as the realities of their web presence. As our report has made clear, due to the collectivist nature of our chosen organization, their web archiving needs had to emerge from a series of conversations with different collective members. Later, we would have to adjust our crawls to suit the nature of the website, its Twitter account, Facebook page, and Vimeo account. The Interference Twitter account, for instance, required more frequent crawls than the organization website. Crawls had to be calibrated to the needs of particular web platforms. As noted, Facebook also offered an extra challenge, as did attempting to preserve the video content on the Interference Vimeo page. This project has underscored the benefit of having good communication with a client or clients, as well as the necessity of running test crawls and being prepared to revise methods.

Our Interference Archive Archive-it collection can be viewed at: https://www.archive-it.org/collections/4932.