

# INTERFERENCE ARCHIVE

## Web Crawl Project

Digital Preservation 2014, MIAP

Allie Whalen, Blake McDowell, Genevieve Havemeyer, Karl McCool

# The Library

## Mission – Structure – Collections



- Volunteer-run library and archive collecting five decades of activism history
- Based in Brooklyn
- Opened December of 2011
- Hosts exhibitions, lectures, workshops, screenings, and performances
- The archive maintains a collection of media, objects, and ephemera from social movements both national and international
- Includes posters, flyers, zines, books, comics, photographs, moving images, audio recordings, T-shirts, and buttons.
- Run by a core collective of ten members

*“We consider the use of our collection to be a way of preserving and honoring histories and material culture that is often marginalized in mainstream institutions. As an archive from below, we are a collectively run space that is people powered, with open stacks and accessibility for all.”*

# The Web Crawl

## Finding a curator

- Project was shaped by the collectively run nature of the organization.
- Received input from:
  - Our official curator, Bonnie Gordon
  - Fellow core collective member, Molly Fair
  - Interference born-digital working group member, Anna Perricci
- Different collective members had different knowledge to offer about what Interference-associated web content needed archiving and what did not.
- Gathered answers and information from several people associated with Interference Archive

# The Web Crawl

## Finding a project

Exploring communications between libraries and ongoing web crawls



Initial plan:

use Archive-It to capture the website for Justseeds Artists' Cooperative, a separate collective associated with Interference Archive

*However...*

- Tamiment Library is already archiving the Justseeds website, which the Justseeds group was unaware of
- Tamiment Library policy doesn't inform organizations that their websites are being archived by the library
- Although Nicole, Bonnie, and Molly all know each other, they didn't share all the same knowledge about ongoing web archiving and the web archiving needs of Interference Archive and associated websites like Justseeds
- The nonhierarchical, collectivist nature of Interference Archive meant that our group needed to communicate with more than just one member of the collective

# The Web Crawl

## Our Project:

Archive the Interference Archive website and their social media sites (Blog, Twitter, Facebook, and Vimeo)



PROPAGANDA A WORKER! **INTERFERENCE ARCHIVE** EXPLORING THE RELATIONSHIP BETWEEN CULTURAL PRODUCTION AND SOCIAL MOVEMENTS. —LEARN MORE 131 8TH STREET NO. 4 BROOKLYN, NY 11215 E-MAIL US FACEBOOK / TWITTER

HOME ABOUT EXHIBITIONS EVENTS BLOG PUBLICATIONS DONATE VISIT / ACCESSIBILITY EDUCATION

Education Working Group Fall 2014 classes and workshops

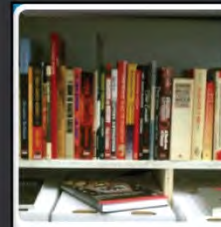


The Education Working Group at Interference Archive hopes to create educational opportunities that activate the collection at the archive, whether by drawing on similar themes or using the collection for inspiration and research.

In fall 2014, we've put together a schedule of classes and workshops that members of the Interference Archive community as well as newcomers have proposed. Be sure to visit the [Education section](#) of our website for more info!

October 26, 2014

Today's processing- Autonomía newspapers



**Interference Archive**

@InterferenceArc

The Interference Archive explores the

TWEETS 1,052 FOLLOWING 612 FOLLOWERS 1,080 FAVORITES 112 LISTS 3

Tweets Tweets & replies Photos & videos

**Interference Archive** @InterferenceArc · 2h

Coming up in November, Collective members Jen, Josh and Greg will be presenting "Activating Social Movement..." [fb.me/3px2i44zf](https://fb.me/3px2i44zf)

**Interference Archive** @InterferenceArc · 5h

Education Working Group Fall 2014 classes and workshops: The Education Working Group at Interference Archive h... [bit.ly/1spOlox](https://bit.ly/1spOlox)

Interference Archive retweeted

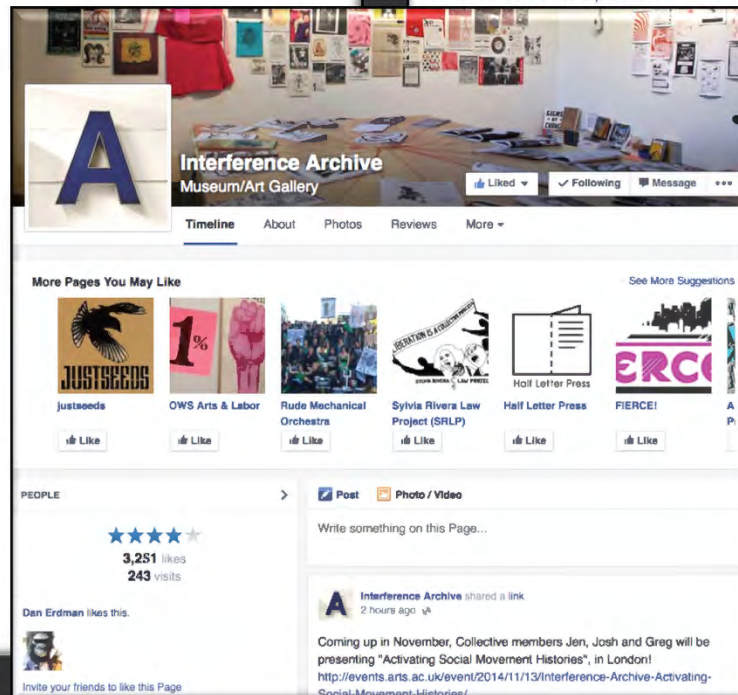
**laura whitehorn** @laurawhitehorn · 16h

with @snidelines after @InterferenceArc political prisoner event 2day [facebook.com/photo.php?fbid...](https://facebook.com/photo.php?fbid...)

Interference Archive retweeted

**Louise Barry** @louisebarry · Oct 24

co-teaching this upcoming workshop @InterferenceArc : [interferencearchive.org/workshop-explo...](http://interferencearchive.org/workshop-explo...)



**Interference Archive** Museum/Art Gallery

Timeline About Photos Reviews More

More Pages You May Like

- Justseeds
- OWS Arts & Labor
- Rude Mechanical Orchestra
- Sylvia Rivera Law Project (SRLP)
- Half Letter Press
- PIERCE!

3,251 likes 243 visits

Dan Erdman likes this.

Interference Archive shared a link 2 hours ago

Coming up in November, Collective members Jen, Josh and Greg will be presenting "Activating Social Movement Histories", in London!  
<http://events.arts.ac.uk/event/2014/11/13/Interference-Archive-Activating-Social-Movement-Histories/>

# The Web Crawl

## Our Project:

Archive the Interference Archive website and their social media sites (Blog, Twitter, Facebook, and Vimeo)

## Planning and Scoping the Crawl:

- Parameters for crawl were outlined in email correspondence with members of Interference Archive's Born-Digital Working Group
- Evaluated high priority aspects of website, and what, if anything, was already being archived
- Met with Nicole Greenhouse for more information on status of Tamiment's crawl of the Justseeds website
- Evaluated Tamiment recommendations to focus on social media and video content, like Vimeo, which they aren't able to capture.
- Focus just on IA, not including their partner organizations, in order to limit the amount of content captured for this project.
- Good example for IA for future archiving project for born digital works made in collaboration with other organizations.

# The Web Crawl

## Our Project:

Archive the Interference Archive website and their social media sites (Blog, Twitter, Facebook, and Vimeo)

### The Interference Archive Website:

- Well organized
- Contains digital representations of several publications
- 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 IA's social media websites, including links to events and content from partner programs (like Justseeds).

### Focus on Social Media:

- Document IA's participation in the online social movement of digital platforms for political and social activism
- Include social media crawls, in addition to capturing the entire Interference Archive website, because IA's Blog posts are less frequent than their active Twitter feed.
- Include 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 Web Crawl

## Technological Implementation

### Scope

Original scope of each crawl:

- Interference Archive's website
- Twitter, Vimeo and Facebook pages

### Seed URLs

<http://interferencearchive.org/>

<https://twitter.com/InterferenceArc/>

<http://vimeo.com/81646906>

<https://www.facebook.com/InterferenceArc/>

### Frequency



Daily

- Most frequent
- Most updated
- Twitter: 63,000 URLs per crawl vs. IA Website: 2,000 URLs per crawl

Daily

Once a month

- Only 1 video
- Reasonable frequency to capture channel page in case usage increases

Twice a month

**Additional Crawls:** Justseeds

- Twitter (One-Time) <https://twitter.com/Justseeds/>
- Blog (One-Time) <http://justseeds.org/blog/>



# The Web Crawl

## Technological Problems and Crawl Analysis

### Frequency

- Running most of them far too frequently.
- IA does not have a fast changing website.
- Updates only a few times a month.
- No listing of weekly events or a constant flow of posts on the blog.

# The Web Crawl

## Technological Problems and Crawl Analysis

### Vimeo

- Archive-It's lack of video capture
- Archive-It does not support capture of videos from a user's channel page.
- Videos can be captured from their specific URL.
- No supported playback of videos in archived pages viewed on Wayback Machine, but they can be viewed in the Video Reports for each crawl.
- No public with access to archived videos

# The Web Crawl

## Technological Problems and Crawl Analysis

### Twitter

- Twitter appears to be a very friendly URL to crawl.
- Large amount of crawls (63,000) as a result of crawling the IA Twitter account in multiple different languages.
- Archiving all of retweets by different users of any IA tweet.
- “Sign in” Box: When clicking on some of the links on the archived pages the Twitter sign in box pops up.


# The Web Crawl

## Technological Problems and Crawl Analysis

### Facebook


- 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 added host constraints to block robots.txt files on:
  - Facebook
  - fbcdn.net
  - akamaihd.net
- To ensure that content would display properly. Archive-It 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'.

# Our Crawls




HOME EXPLORE LEARN MORE CONTACT US

The leading web archiving service for collecting and accessing cultural heritage on the web  
*Built at the Internet Archive*



Explore >> New York University - Moving Image Archiving and Preservation Program >> Interference Archive



## Interference Archive

Collected by: [New York University - Moving Image Archiving and Preservation Program](#)

Archived since: Oct, 2014

Description: A web crawl for Interference Archive's website and social media pages: "Interference Archive explores the relationship between cultural production and social movements. This work manifests in public exhibitions, a study and social center, talks, screenings, publications, workshops, and an online presence. The archive consists of many kinds of objects that are created as part of social movements by the participants themselves: posters, flyers, publications, photographs, books, T-shirts and buttons, moving images, audio recordings, and other materials." ([www.interferencearchive.org/our-mission](http://www.interferencearchive.org/our-mission))

Subject: [Arts & Humanities](#), [Society & Culture](#), [Politics & Elections](#)

### Narrow Your Results

Sites for this collection are listed below. Narrow your results at left, or enter a search query below to find a site, specific URL or to search the text of archived webpages.

Subject Sort By: Count (A-Z)

Social Movements (1)

Publisher Sort By: Count (A-Z)

Interference Archive (1)

Contributor Sort By: Count (A-Z)

Alycia Sellie (1)  
Anika Paris (1)  
Blithe Riley (1)  
Bobby Moeller (1)  
Bonnie Gordon (1)

[More](#)

Enter search terms here

Sites  Search Page Text

Page 1 of 1 (7 Total Results)

Sort By: [Title \(A-Z\)](#) | [Title \(Z-A\)](#) | [URL \(A-Z\)](#) | [URL \(Z-A\)](#)

Title: Interference Archive  
URL: <http://interferencearchive.org/>

Description: The official website for Interference Archive, a volunteer-run archive and collective: "Interference Archive explores the relationship between cultural production and social movements. This work manifests in public exhibitions, a study and social center, talks, screenings, publications, workshops, and an online presence. The archive consists of many kinds of objects that are created as

<https://www.archive-it.org/collections/4932>