Collection & community building through web archiving: engaging with faculty and students in a collaborative web archiving project

Andrea Schuler
Tufts University, andrea.schuler@tufts.edu

Follow this and additional works at: https://digital.sandiego.edu/symposium

Part of the Library and Information Science Commons

Schuler, Andrea, "Collection & community building through web archiving: engaging with faculty and students in a collaborative web archiving project" (2017). Digital Initiatives Symposium. 2.
https://digital.sandiego.edu/symposium/2017/2017/2

This 45-minute concurrent session is brought to you for free and open access by Digital USD. It has been accepted for inclusion in Digital Initiatives Symposium by an authorized administrator of Digital USD. For more information, please contact digital@sandiego.edu.
Collection & community building through web archiving: engaging with faculty and students in a collaborative web archiving project

Presenter 1 Title
Librarian for Digital Collections

Session Type
45-minute concurrent session

Abstract
Tisch Library at Tufts University has recently begun a pilot web archiving project, aiming to deepen Tufts’ collections in areas of strategic importance and support more “traditional” library collection development activities, while collecting material that is not known to be comprehensively collected by other institutions. Additionally, the project offers an opportunity for collaborative collection building with faculty and students that serves as a unique way to deepen our community’s engagement with the library.

The initial pilot collection focuses on environmental justice, selected due to its relevance to the Tufts community and curriculum and to build on existing Tisch Library collection strengths. Two undergraduate courses related to environmental justice were identified and invited to partner in the pilot project. This partnership would leverage student research to expand the initial collection while introducing students to concepts of web archiving and information literacy around websites and providing them with the opportunity to contribute to shaping the scholarly record. Both courses added a brief assignment to their syllabus: while doing research on their chosen topics, students would identify 3-7 web sites they felt would benefit from preservation and submit the sites to the library, to be evaluated and added to the web archive as appropriate.

This presentation discusses the process of beginning a subject-based web archiving project, focusing on the collaborative project with two undergraduate classes. It addresses decisions made when starting and scoping the project; collection development issues; the logistics, benefits, and outcomes of the student and faculty collaboration; and future directions.

Location
KIPJ Room EF

Keywords
web archiving, outreach, collaboration, collection development, digital collections, undergraduates

This 45-minute concurrent session is available at Digital USD: https://digital.sandiego.edu/symposium/2017/2017/2
Collection & community building through web archiving: engaging with faculty and students in a collaborative web archiving project

ANDREA SCHULER, LIBRARIAN FOR DIGITAL COLLECTIONS
TISCH LIBRARY, TUFTS UNIVERSITY
MAY 2, 2017
Agenda for today

Intro to a subject-based web archiving project
Scoping & collection development strategies
Logistics, benefits & outcomes of a student and faculty collaboration project
Concerns & evaluation
Future directions (???)
What, and why?

What: A pilot subject-based web archiving project focused on environmental justice, and built in collaboration with two undergraduate classes.

Why: goals of the project:
- To explore a new method of collecting born-digital content.
- To deepen our collections in an area of strategic importance.
- To contribute to the overall scholarly record by collecting material that is not known to be comprehensively collected by other institutions.
- To pursue collaboration opportunities with faculty and students in order to increase engagement with our communities, contribute to student learning, and increase the project’s output.
Starting the project: scoping

Options: topically, geographically, by type of resource

“This collection contains websites, blogs, and social media feeds of citizen groups and non-profit organizations who deal with environmental justice issues. Sites were selected based on their topical and geographic focus: the topical scope includes all environmental justice issues broadly, and currently, the geographic scope is limited to the City of Boston and the cities and towns that comprise Greater Boston. Sites hosted at other educational institutions or governmental sites that otherwise meet the selection criteria are not included.”
Choosing search terms

High level terms

*Environmental justice*  *Climate justice*

Combined with geographic terms

*Boston  Greater Boston  Cambridge  Somerville*

Plus reading on the subject

Led to more specific terms

*asthma  toxic waste  flooding  pipeline  tree cover*

*air pollution  Water pollution  mercury  water safety*

*green space  food deserts  lead poisoning  climate change*

*healthy food  inadequate transportation  sea level*
Gathering seeds

Methods:
- Aggressive Googling
- Follow links from already-identified sites
- Search blog domains – e.g. site:wordpress.com or site: blogspot.com
- Look at Facebook pages of identified groups – “Pages liked by this page”
Considerations – ask yourself:

Who is responsible for the content?

What is the content – is it useful/important, or is there a lot of irrelevant content?

Is there unique information – only available on that site or only available online?

Is the content at risk?

Consider context – does the website stand alone, or do additional sites/documents need to be archived along with it?
<table>
<thead>
<tr>
<th>Organization</th>
<th>Website</th>
<th>Crawl 1 results</th>
<th>Crawl 2 results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alternatives for Community &amp; Environment</td>
<td><a href="http://acce-omg.org/">http://acce-omg.org/</a></td>
<td>good</td>
<td>good</td>
</tr>
<tr>
<td>Chelsea Collaborative – Environmental Chelsea</td>
<td><a href="http://www.chelsea.org">http://www.chelsea.org</a></td>
<td>site formatting &amp; images still lost</td>
<td>added domain for links in, site formatting &amp; images still lost</td>
</tr>
<tr>
<td>Green Justice Coalition</td>
<td><a href="http://massclu.org">http://massclu.org</a></td>
<td>good</td>
<td>good</td>
</tr>
<tr>
<td>Somerville Climate Action</td>
<td><a href="http://www.somenv.org">http://www.somenv.org</a></td>
<td>looks mostly okay</td>
<td>good</td>
</tr>
<tr>
<td>Boston Climate Action Network</td>
<td><a href="http://bostoncan.org">http://bostoncan.org</a></td>
<td>needs to be recrawled</td>
<td>good</td>
</tr>
<tr>
<td>Watertown Citizens for Peace, Justice, and</td>
<td><a href="http://watertownca.org">http://watertownca.org</a></td>
<td>images not loading</td>
<td>good</td>
</tr>
<tr>
<td>NOAH Chelsea Creek Action Group</td>
<td><a href="http://noahcdc.org">http://noahcdc.org</a></td>
<td>4 “read more” links same as crawl 1</td>
<td>4 “read more” links same as crawl 1</td>
</tr>
<tr>
<td>Friends of the Maiden River</td>
<td><a href="https://maidenriver.org">https://maidenriver.org</a></td>
<td>need to add link</td>
<td>good</td>
</tr>
<tr>
<td>Creation Coalition</td>
<td><a href="http://creationcoal.org">http://creationcoal.org</a></td>
<td>good except for ir links same as crawl 1</td>
<td>good except for ir links same as crawl 1</td>
</tr>
<tr>
<td>Roxbury Safety Net</td>
<td><a href="http://fairfood.org/stopthebiolab/stc.org">http://fairfood.org/stopthebiolab/stc.org</a></td>
<td>good, though less useful without external links?</td>
<td>good, though less useful without external links?</td>
</tr>
<tr>
<td>Fairmount Indigo Collaborative</td>
<td><a href="http://fairmountindy.org">http://fairmountindy.org</a></td>
<td>looks good if you same as crawl 1</td>
<td>same as crawl 1</td>
</tr>
<tr>
<td>City Growers</td>
<td><a href="https://citygrowers.wordpress.com/">https://citygrowers.wordpress.com/</a></td>
<td>good</td>
<td>good</td>
</tr>
<tr>
<td>Stop the West Roxbury Lateral Pipeline</td>
<td><a href="https://swrl.info/">https://swrl.info/</a></td>
<td>good</td>
<td>good</td>
</tr>
<tr>
<td>Resist the Pipeline</td>
<td><a href="http://www.resistthepipeline.org/">http://www.resistthepipeline.org/</a></td>
<td>good</td>
<td>good</td>
</tr>
<tr>
<td>Environmental Justice at Tufts University blog</td>
<td><a href="http://tuftsaction.blogspot.com/">http://tuftsaction.blogspot.com/</a></td>
<td>good, pull in Google+ profile</td>
<td>good, pull in Google+ profile</td>
</tr>
<tr>
<td>Tufts Climate Action</td>
<td><a href="https://tuftsdivest.org">https://tuftsdivest.org</a></td>
<td>not added yet, pending DCA decision</td>
<td>not added yet, pending DCA decision</td>
</tr>
<tr>
<td>Groundwork Somerville</td>
<td><a href="http://www.groundworksomerville.org">http://www.groundworksomerville.org</a></td>
<td>good</td>
<td>good</td>
</tr>
</tbody>
</table>
Metadata! (a work in progress)

At the collection level:
- Title
- Creator
- Description
- Publisher
- Dates (range of crawls)
- Identifier/Source (associated collection page in Archive-IT, e.g. https://archive-it.org/collections/7635)
- Rights
- Collector
- Language

At the seed level:
- Title
- Creator
- Subject (LCSH)
- Description
- Dates (of crawls, a single date or range)
- Identifier/Source (associated page in Wayback, e.g. https://wayback.archive-it.org/7635/*/http://ace-ej.org/)

"Some cheerful data", by Dirkuys, via Flickr, CC BY SA 2.0
Course collaboration

Environmental Justice, Security, and Sustainability

Women and Water: Fighting for Environmental Justice
Reaching out to faculty

Highlight benefits to the students

- experience with information literacy around websites
- exposure to web archives (a resource they’re probably not familiar with)
- the opportunity to directly contribute to a scholarly project – especially important for undergrads who don’t always have the opportunity to work on a “real-life” project
5-page Literature Review Choose an issue in environmental justice that you are interested in and find a set of articles (or book) to review. Write a 5-page (2500-3500 words) paper reviewing this issue, providing an overview of the problem, data and evidence framing the issue, strategies and solutions pursued, and your own critical analysis of the methods and conclusions. Use the readings and themes from the course in your analysis. Try to find books and articles published in the last ten years. If you are reviewing articles, try to find at least 5 sources, with at least 3 from scholarly peer-reviewed sources. Grey literature (by government, think tanks, advocacy groups, and others) are acceptable for the remaining articles.
Collaboration proposal – how will this actually work?

**Proposal:** Collaborate with students in “Environmental Justice, Security, and Sustainability” to build and expand Tisch Library’s pilot web archiving project focused on environmental justice.

**Objectives:** Leverage student research on a particular aspect of environmental justice to identify websites to add to the web archive, ensuring the sites’ preservation for future researchers and practitioners; introduce students to concepts of web archiving and information literacy around websites; and provide students with the opportunity to directly contribute to a scholarly project.

**Possible workflow:**

1. Students are given a brief introduction to the concept of web archiving, reasons why web archiving is important, and the Tisch EJ project.
2. In conjunction with the Literature Review assignment, students are instructed to identify three to five websites that they believe are important for context or research around their topic and that should be preserved for the future. Students will be given basic information about types of sites that are potentially appropriate for the project.
3. By December 15 (due date of the Literature Review assignment), students will submit their list of websites via a Google Form. The form will include a small amount of metadata for each site, including a brief description and a space for students to identify the importance of the site.
4. The Librarian for Digital Collections will evaluate the list of submitted sites for appropriateness, ability to be crawled for the archive, etc., and add to the web archive as applicable. Students will be able to visit the EJ collection web archive homepage to see their work live.
Link submission

Google form:

- Your name
- Website URL to archive
- Please provide a brief description of the site that could be used to describe it on the Archive-It web archive public interface (see https://archive-it.org/collections/7635 for description examples)
- Why do you think this site is important for the EJ movement?
- Is there anything else we should know about this site?

Environmental Justice, Security, and Sustainability

Use this form to submit websites for Tisch Library’s Environmental Justice web archiving project. Please submit 3-5 sites that you have identified as being important for context or research around your topic and that should be preserved for the future.

Note that you must complete this form one time for each link you would like to submit (i.e., if you have 3 links, you will submit the form 3 times)

Any questions? Email Andrea Schuler, Digital Collections Librarian, at andrea.schuler@tufts.edu.

* Required

Your name *

Your answer

Website URL to archive *

Your answer

Please provide a brief description of the site that could be used to describe it on the Archive-It web archive public interface (see https://archive-it.org/collections/7635 for description examples) *

Your answer
Presentation to students

What

Why
- Increasing amounts of content are available only online. It can disappear or change in an instant
- It’s important for future researchers, practitioners, historians, the general public
- Stats on link rot

Wayback Machine vs. Archive-It

Considerations
- There are no right or wrong links
- Types of sites to consider
- Collection development considerations

How
Links were submitted! Now what?

Evaluate each link for relevance to the collection, quality of the site, any IP concerns, and ability to be crawled

Group sites into themes

Re-write scope to reflect expansion of the collection & to guide further scoping

Gather further seeds to flesh out each theme
Evaluation: benefits

The opportunity to engage with undergrads and faculty

Students learn about web archiving & information literary re: web sites

Faculty subject expertise

Students gathered links that might not come up in “environmental justice” searches

Submitted links helped illuminate important themes

Student-written descriptions saved time in metadata creation

Some excitement from the faculty & students

“Stars”, by Quinn Dombrowski, via Flickr, CC BY SA 2.0
Sept. 19, 2016 issue of The ExPress: Weekly ExCollege News & Ideas

Course Highlight

Each week, The ExPress gives readers a glimpse into subjects being taught at the ExCollege this semester. Contact us if you would like to see your class featured, or if you would like to read more about a topic being offered!

Women & Water: Fighting for Environmental Justice

This week we’re excited to have Andrea Schuler, Librarian for Digital Collections at the Tisch Library joining us in class to talk about her work. Curious what Digital Collections and Environmental Justice have in common? Digital Collections are one way we can record what’s transpiring with these issues so that researchers in the future will also have access to the same websites, blogs, and social media sources we have today. Collecting sources for the Tisch Digital Archives is part of the semester-long projects each student in our class will be working on. In addition to the digital archive work, each student will be working on a paper and presentation about the course topic. Students will choose their own story to research – whether it’s a public figure like Rachel Carson (author of Silent Spring) or a more hidden story like the fight by four women to stop an oil refinery from being built in New Hampshire’s Great Bay in the late 1970’s.
Concerns

Is this random?
How to document collecting decisions
Scope creep – is it strictly environmental justice?
Privacy – should I ask for permission to crawl?
Rights
What are people going to use this stuff for?

“questions”, by airpix, via Flickr, CC BY 2.0
Evaluating the project overall

Did the collaboration project actually save me any time? Probably not

Did it find things I wouldn’t have found otherwise? Absolutely

Was it successful overall? Yes

What would I change?
What next?

Provide better access to the collection. Some options:

- MARC records

Collection-level MARC record in Columbia’s Clio catalog
What next?

Provide better access to the collection. Some options:

- MARC records
- Webpage on the library site contextualizing the collection & linking to Archive-It, or a custom-built portal

NC State page explaining web archiving, their collection activities, and links to collections

https://www.lib.ncsu.edu/web-archiving/collections

Our Web Archive Collections

Below you will find links to, and a brief description of, each of our unique web archive collections that are housed in the Internet Archive. These collections are intended to compliment our existing primary collecting areas. As our web archiving initiative grows we will continue to add additional collections to this page in the future.

NC State University Websites Collection

The bulk of the NCSU Web Archive is dedicated to preserving the unique cultural heritage of our institution as it is represented on university websites. To meet these goals we have partnered with the Internet Archive. The Internet Archive provides us with the tools and technology required to crawl, capture, store, and provide access to the websites in our collection. We have identified many of the primary websites used by the university and are regularly capturing those sites for preservation.
What next?

Provide better access to the collection. Some options:

◦ MARC records
◦ Webpage on the library site contextualizing the collection & linking to Archive-It, or a custom-built portal

Columbia Human Rights Web Archive portal

https://hrwa.cul.columbia.edu/
What next?

Provide better access to the collection. Some options:
- MARC records
- Webpage on the library site contextualizing the collection & linking to Archive-It, or a custom-built portal
- Finding aid

UNC Chapel Hill finding aid

http://finding-aids.lib.unc.edu/40417/#
What next?

Provide better access to the collection. Some options:

◦ MARC records
◦ Webpage on the library site contextualizing the collection & linking to Archive-It, or a custom-built portal
◦ Finding aid
◦ Online exhibit?

*Online exhibits at Tisch create narratives using library collections. Exhibits highlight our unique resources and facilitate user access, education, and research for the Tufts community and a wider public audience.*
What next?

How to engage users???
- Is the collection there mostly for the future?
- Data mining?
- Another collaboration project?

“Big question mark”, by Benjamin Reay, via Flickr, CC BY NC 2.0
Questions?

Or stories to share about engaging users with your web archives?

Thank you!
andrea.schuler@tufts.edu

“Conversation”, by Valerie Kensky, via Flickr, CC BY 2.0