Preserving News on the Margins

KATHERINE SKINNER, Executive Director, Educopia Institute
EMILY GORE, Director for Content, Digital Public Library of America
with JAMES DANKY (University of Wisconsin), ANA KRAHMER (UNT), and KOPANA TERRY (UKY)

Dodging the Memory Hole, San Francisco, November 16, 2017
Team

**STAFF**
Emily Gore, PI
Katherine Skinner, research lead
Donald Brown, consultant
Sarah Lippincott, consultant
Nathan Brown, TrueBearing

**ADVISORS**
Franky Abbott (DPLA)
Cheryl Ball (West Virginia University)
Randall Burkett (Emory University)
James Danky (University of Wisconsin)
Nick Graham (UNC Chapel Hill)
Reinette Jones (UKY)
Ana Krahmer (UNT)
Nancy McGovern (MIT)
Sam Meister (Educopia)
Jessica Meyerson (Educopia)
Cal Shepard (State Library of NC)
Kopana Terry (UKY)
Tyler Walters (Virginia Tech)
Frederick Zarndt (Consultant)

Photo: https://www.barcelonasegwaytour.com/los-castellers-discover-one-particular-attractions-barcelona/
Why is this particular project needed?

- Historic preservation efforts (microfilm projects, etc.) have usually focused on mainstream voices - even if the efforts have been focused on smaller communities.

- Many of you may participate in a statewide newspaper digitization project and/or NDNP. We all LOVE and APPRECIATE those projects, but we realize that ALL projects have limitations.

- Papers that are born-digital or potentially have copyright challenges are usually avoided.

- We have often brokered deals with vendors who have digitized our materials to place behind a paywall. While there are cases where this makes sense, it is still restricted/privileged access and might include preservation? (trusted source?)
“Titles should reflect the political, economic, and cultural history of the state or territory. Such newspapers usually have **statewide or regional influence**. Preference should be given to titles that are **recognized as “papers of record” at the state or county level** and that contain published **legal notices, news of state and regional governmental affairs, and announcements of community news and events**. In addition, **selected titles should provide state, or at least multi-county, coverage of the majority of the population areas**. Newspapers with a **broad chronological span** are preferred to those with short runs and those that published sporadically.”
From the NDNP Technical Guidelines

- Digital reproductions should be made from a preservation copy of microfilm, a clean second-generation duplicate silver negative.

- Provide the master page images, delivered to LC, as uncompressed images in TIFF 6.0 format.

- Provide access to Public Domain content.

- For each newspaper page, structural metadata must be provided to relate pages to title, date, and edition, to sequence pages within issue or section; and to identify associated image and OCR files.

- Include technical metadata to support the functions of a trusted repository.
Realize there is a need!

- The Digital Public Library of America was awarded a planning grant by the John S. and James L. Knight Foundation to research a role and develop a plan for DPLA regarding newspaper content.
- DPLA had an original goal for the project to look at statewide projects and build an aggregation of non-NDNP newspapers.
- After spending months doing initial research and surveying statewide projects about their newspaper content, we realized that most of the digitized non-NDNP newspapers were similar voices.
- We decided to change directions and had to make a hard call with our Program Officer to ask for an extension and a significant shift in direction.
- We wanted to research news content of voices that have rarely been aggregated and/or digitized and potentially develop a project to make those news sources available via DPLA.
Sometimes you have to do a...
What happened next

- DPLA staffers Emily Gore and Franky Abbott started trying to tackle this research ourselves and began to focus on researching newspapers in the African American and LGBT communities.

- We knew we wanted to do research, bring together colleagues and experts to advise us on developing a large scale project.

- We didn’t have the time we needed to do this work and we needed to engage with a knowledgeable consultant who could help us figure out this landscape.

- Emily called Katherine, and we worked to develop a scope of work, and we are here today to learn about those findings and discuss next steps.
News on the Margins

Researchers rely on news sources by and for marginalized communities to reveal the density of perspectives and experiences embedded in U.S. local and national cultures.
FREEDOM’S JOURNAL.

RIGHTEOUSNESS EXALTETH A NATION.

NEW-YORK, FRIDAY, MARCH 26, 1825.

Vol.

Our Patrons.

In viewing our first number to our Patrons, we feel the advantage that is attached to the publication of a Journal, the expediency of which we have in view by the publication of this Journal; the expediency of works of trivial importance, we shall consider it a part of our duty to recommend to our young readers, such authors as will not only enlarge their stock of useful knowledge, but such as will also serve to stimulate them to higher attainments in science.

We trust also, that through the columns of the FREEDOM’S JOURNAL, many practical narratives which they have published; the establishment of the republic of Haiti after years of sanatory warfare; its subsequent progress in all the arts of civilization; and the advancement of liberal ideas in South America, where despoticism has given place to free governments, and where many of our brethren now hold important civil and military stations.

We wish to plead our own cause. Too long have others spoken for us. Too long have our friends been deceived by misrepresentations, in things which concern us dearly, though in the estimation of some mere trifles; for though there are many in society who exercise towards us benevolent feelings, and who actuated by the cordial approbation of every friend with humanity.

The peculiarities of this Journal, render it important that we should advertise to the world the motives by which we are actuated, and the objects which we contemplate.

We wish to plead our own cause. Too long have others spoken for us. Too long has the public been deceived by misrepresentations, in things which concern us dearly, though in the estimation of some mere trifles; for though there are many in society who exercise towards us benevolent feelings, and who actuated by the cordial approbation of every friend with humanity.

The peculiarities of this Journal, render it important that we should advertise to the world the motives by which we are actuated, and the objects which we contemplate.

We wish to plead our own cause. Too long have others spoken for us. Too long have our friends been deceived by misrepresentations, in things which concern us dearly, though in the estimation of some mere trifles; for though there are many in society who exercise towards us benevolent feelings, and who actuated by the cordial approbation of every friend with humanity.

The peculiarities of this Journal, render it important that we should advertise to the world the motives by which we are actuated, and the objects which we contemplate.

We wish to plead our own cause. Too long have others spoken for us. Too long have our friends been deceived by misrepresentations, in things which concern us dearly, though in the estimation of some mere trifles; for though there are many in society who exercise towards us benevolent feelings, and who actuated by the cordial approbation of every friend with humanity.

The peculiarities of this Journal, render it important that we should advertise to the world the motives by which we are actuated, and the objects which we contemplate.
Obstacles

- Where on earth *is* the content
- Most content is still print/microfiche only
- What has been digitized is hard to find
- Paywalls
- Born-digital content isn’t collected yet (!!!)
- Rights. Rights. Rights. (right?)
So what can we do?
**Goal:** to make more accessible the news outputs of marginalized communities, past and present.

**Project Steps:**

1. **Research** existing work in this area, seek out partnerships and avenues

2. **Create an open data collection framework and methodology** for recording info about print, microfiche, and digital news of marginalized communities as comprehensively as possible

3. **Pilot** the use of the framework and **collect data** for two communities: LGBT and African American

4. **Learn** from the process

5. **Scope** what DPLA and other partners might undertake in the future to increase access to these crucial historical sources.
So what have we accomplished?
1. Research existing work

Bibliographies
- Tracy Baim, *Gay Press, Gay Power* list (125 pubs)

Lists
- *The Black Press Held by the LOC* (Pluge, 1991)
- *Wisconsin Historical Society African American Newspapers*
- *Afro Newspaper Morgue Collections*
- *Internet Public Library*

Microfilm Collections
- Underground Newspaper Collection (LC)
- Atlanta Lesbian Feminist Alliance archives
- African American Newspapers and Microfilms (LC)
- Lesbian Herstory Archives
- Gay, Lesbian, Bisexual and Transgender Historical Society

Digitized content
- *NDNP and Chronicling America*
- *Muncie Times* Collection (Ball State)
- *Qnotes* (UNC Charlotte)
- *Southern Voice* (Kennesaw State)
- *Freedom’s Journal - Wisconsin Historical Society*
- *Central Florida Future - UCF*
- *Newsbank-Readex - African American newspapers*
- *Gale - Archives of Sexuality and Gender*
- *Accessible Archives - African-American Newspapers*
- *Ebsco - LGBT Life*
- *ProQuest - Ethnic Newswatch, Historical Black Newspapers*

Aggregators
- *HathiTrust*
- *CRL*
- *OCLC*
2. Create a Data Framework

1. Building on existing works (especially James Danky’s *Bibliography of African American Newspapers and Periodicals*)

2. Data framework and model vetted by 10+ leaders in news aggregation and cataloging

3. Ambition: balance between enough data to be actionable but not so much that it deters participation

3. Collect Data

- Crowdsourcing plea
  - Listservs (45+ association leaders across LAMs assisted)
  - Direct calls and emails to well-known archives in these two collecting areas

- Calls to major aggregators (e.g., OCLC, CRL, Readex, ProQuest, Gale, etc)

- Direct calls and emails to current news publishers in these two areas
## 4. Analysis (preliminary)

<table>
<thead>
<tr>
<th>Framework</th>
<th>African American</th>
<th>LGBT</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Crowdsourcing</strong></td>
<td>1,796 records</td>
<td>3,129 records</td>
<td>4,925 records</td>
</tr>
<tr>
<td></td>
<td>1,794 titles</td>
<td>3,126 titles</td>
<td>4,119 titles</td>
</tr>
<tr>
<td></td>
<td>96 institutions</td>
<td>66 institutions</td>
<td>132 institutions</td>
</tr>
<tr>
<td><strong>Born-digital calls</strong></td>
<td>62 contacted</td>
<td>65 contacted</td>
<td>127 calls/emails</td>
</tr>
<tr>
<td></td>
<td>11+ conversations</td>
<td>2 conversations</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3 responses</td>
<td>6 responses</td>
<td></td>
</tr>
<tr>
<td><strong>Aggregators</strong></td>
<td>51,635 records</td>
<td>5,543 records</td>
<td>57,178 records</td>
</tr>
<tr>
<td></td>
<td>2,118 titles</td>
<td>251 titles</td>
<td>2,362 titles</td>
</tr>
<tr>
<td></td>
<td>2,389 institutions</td>
<td>849 institutions</td>
<td>2,453 institutions</td>
</tr>
</tbody>
</table>
Data visualizations (beta)

https://public.tableau.com/profile/truebearing#!/vizhome/NOTMCombinedSurvey/NOTMCompositeData

https://public.tableau.com/profile/truebearing#!/vizhome/OCLCNOTMData/OCLCData
Next steps

- *Bibliography of African American Newspapers and Periodicals* (Danky, Harvard, 1998)- digitize and bring up live (!!!)
- Directory of African American Newspapers and Periodicals - beta version launch
- White paper about the News on the Margins findings and recommendations
- Project proposals to start the work ahead