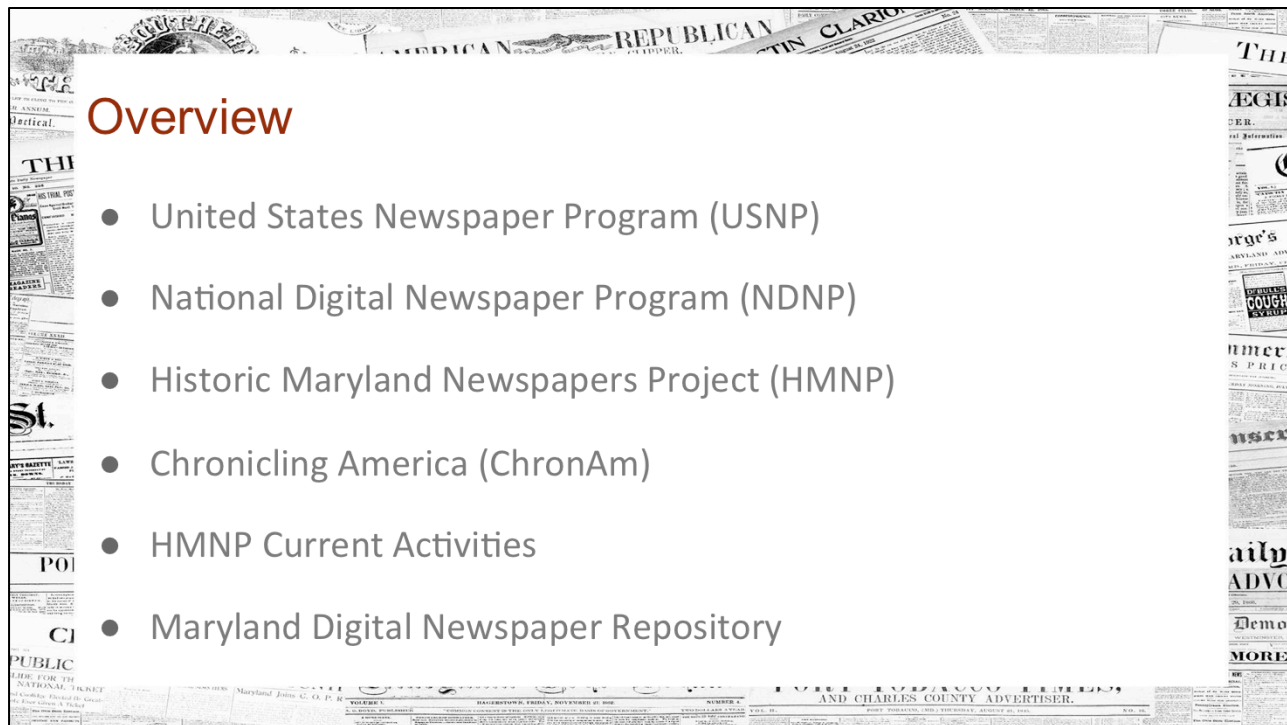# Maryland Digital Newspapers

Robin C. Pike, Rebecca E. Wack
University of Maryland Libraries

Overview

- United States Newspaper Program (USNP)

- National Digital Newspaper Program (NDNP)

- Historic Maryland Newspapers Project (HMNP)

- Chronicling America (ChronAm)

- HMNP Current Activities

- Maryland Digital Newspaper Repository

Robin:
Today, Rebecca and I are going to discuss the United States Newspaper Program, and how that led to the National Digital Newspaper Program. We will also inform you on the progress and activities of the Historic Maryland Newspapers Project, and demonstrate how to use Chronicling America. We will conclude by telling you about the progress made in hosting newspapers at the University of Maryland, College Park Libraries, and how we hope to turn this into the Maryland Digital Newspaper Repository.

US Newspaper Program (USNP), 1982-2001

- Enhance access to historic American newspapers
- Locate, catalog, and preserve on microfilm newspapers published in the United States
  - 150,000 titles cataloged, 600,000 holdings, more than 75 million pages filmed
- Established the USNP Preservation Microfilming Standards
- All 50 states participated

NATIONAL ENDOWMENT FOR THE
Humanities

Robin:
The US Newspaper Program (USNP), a National Endowment for the Humanities grant-funded initiative that occurred between 1982-2011, facilitated newspaper preservation and access across all 50 states through the location of newspapers, catalog record creation, and the creation and circulation of microfilm copies of newspapers. Each state in the program designated a repository for the microfilm, with additional use copies sent to many state public libraries and historical societies. In total, this program enabled the cataloging of 150,000 titles, 600,000 holdings, and more than 75 million pages filmed. This project established preservation microfilming standards for newspapers that are still employed today. This project greatly facilitated patrons' free access to newspapers, although it still presented accessibility issues. Despite the increase in access across more organizations, many patrons still faced technical and physical limitations posed by the media and preferred published newspaper abstracts. Indexes to births, marriages, and death announcements were essential, particularly when identifying ancestors before 1850 when the census and vital records are incomplete, however, most indexing projects focused on large dailies and county newspapers of record and neglected many titles from smaller towns. Indexers also did not create indexes containing more than names and places, making research into news events and daily life featured in the news articles and advertisements just as inaccessible.

National Digital Newspaper Program (NDNP), 2005-present

- Project publication date range: 1836-1922; 1690-1963
- One award per state/territory
  - Multiple content partners
- ~100,000 pages per award
  - Multiple award opportunities
- Title selection criteria
- Microfilm selection criteria
- Other deliverables: title essays, survey data

LIBRARY OF CONGRESS
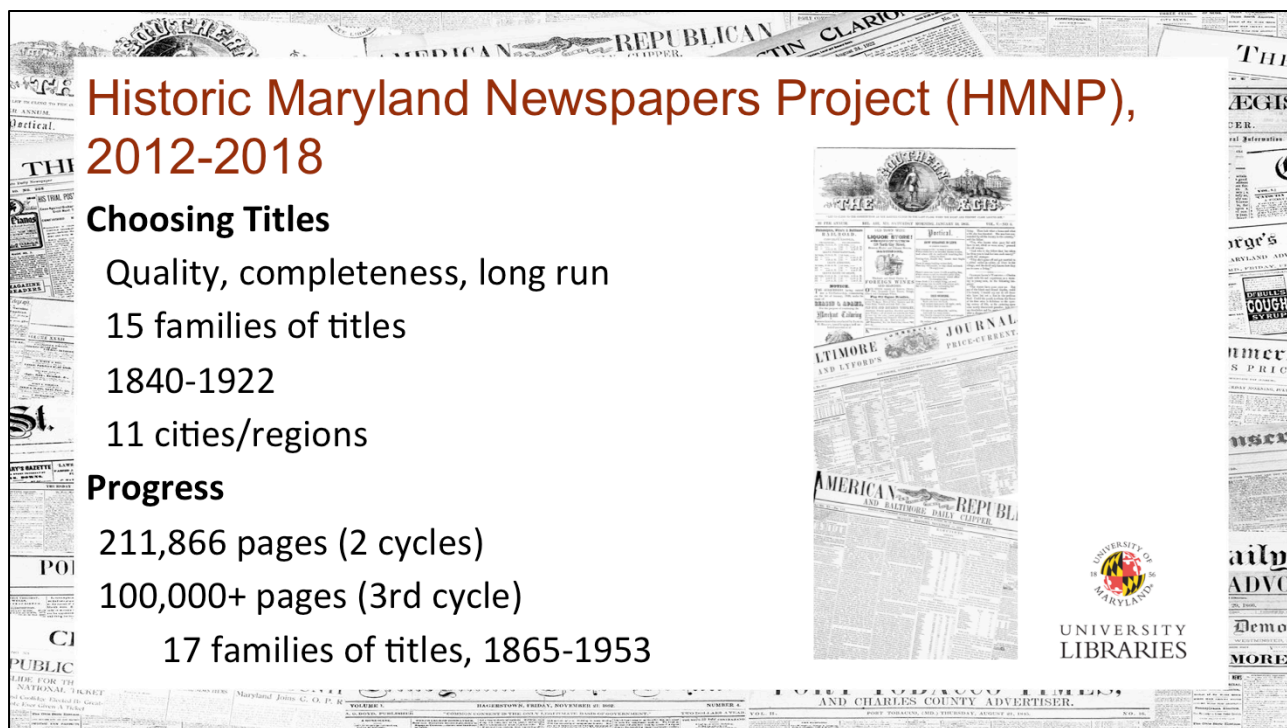
NATIONAL ENDOWMENT FOR THE Humanities

Robin:
Modeled after the access provided by licensed newspaper databases but with an intent for preservation and access, free newspaper access in the 21st century has evolved from microfilm to a freely accessible database of digitized newspapers. Established in 2005 and growing out of the USNP, "The National Digital Newspaper Program (also known as NDNP), a partnership between the National Endowment for the Humanities and the Library of Congress, is a long-term effort to develop an Internet-based, searchable database of U.S. newspapers with descriptive information and select digitization of historic pages." "In order to plan for phased development, the annual award program began with targeting digitized material for the decade 1900-1910. In subsequent award years, the time period was gradually extended decade by decade, to cover the historic period 1836-1922." The project now focuses on newspapers from 1690-1963, and requires the title be in the public domain. NEH makes one award to each state partner, who in turn collaborates with relevant state partners to include additional content. The awardee digitizes 100,000 pages in each two-year award. State partners are encouraged to seek second and third awards to produce a total of approximately 300,000 digitized pages, though some states have now been granted fourth awards.
The NDNP award program funds digitization with the goal of funding the content contribution from all US states and territories. Partners are asked to select titles based on high research value, geographic and temporal coverage, and reflecting a variety of ethnic, racial, political, economic, religious, or other special audiences and interest groups. The titles should also not be available on another database. Partners are also encouraged to select higher quality microfilm and with the newspapers on the film being of sufficient completeness to enable the most accurate Optical Character Recognition (or OCR). In addition to digitized microfilm, awardees are expected to provide 500-word essays about the history of a title, or a family of titles, created through title changes or mergers. Awardees also need to perform a survey of digitized newspapers, and provide this data including links for the titles that are freely available elsewhere, such as via the Google mass digitization project.
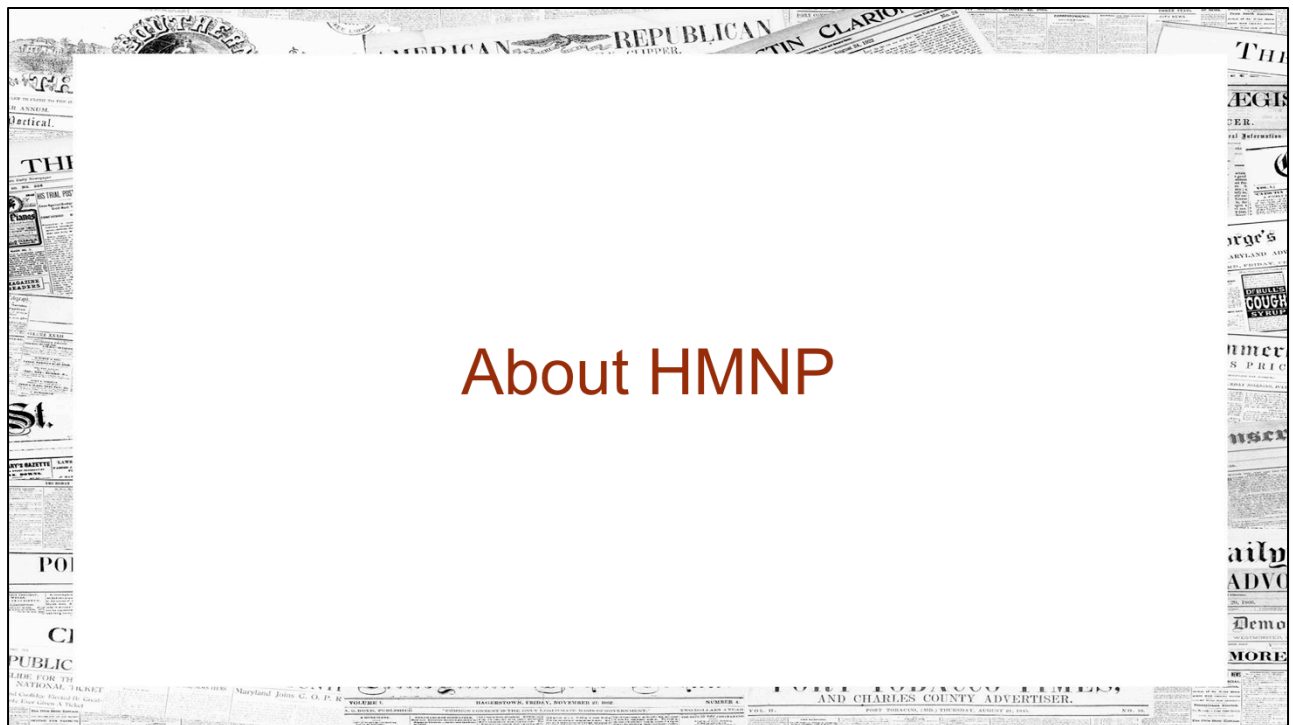
About HMNP

Historic Maryland Newspapers Project (HMNP), 2012-2018

Choosing Titles

Quality, completeness, long run

15 families of titles

1840-1922

11 cities/regions

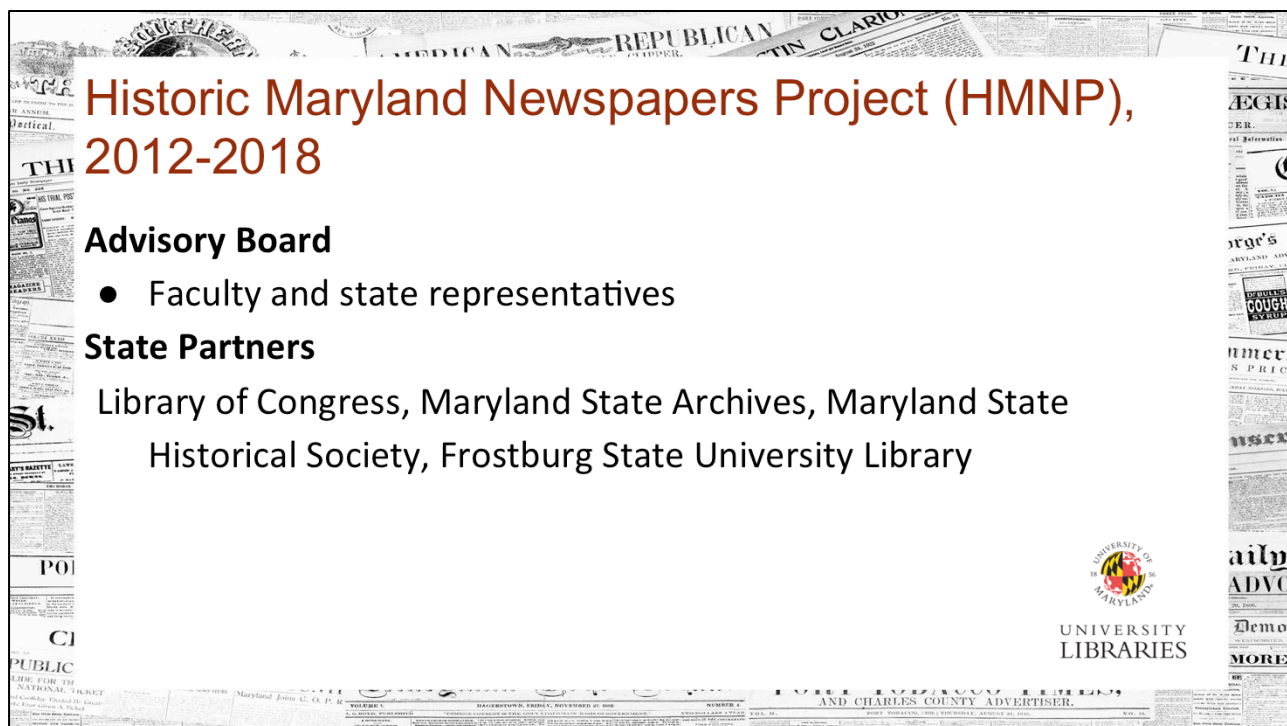**Progress**

211,866 pages (2 cycles)

100,000+ pages (3rd cycle)

    17 families of titles, 1865-1953

Robin:
The Historic Maryland Newspapers Project (HMNP) began in 2012 as the Maryland State newspaper digitization project. From 2012-2016 during the first two two-year grant cycles, the project digitized 211,866 pages across 15 families of newspaper titles from 11 cities or regions dating between 1840-1922. In the third grant, we will add an additional 100,000 or more pages from 17 families of titles dating from 1865-1953.

Historic Maryland Newspapers Project (HMNP), 2012-2018

Advisory Board
- Faculty and state representatives

State Partners

Library of Congress, Maryland State Archives, Maryland State
Historical Society, Frostburg State University Library

UNIVERSITY
LIBRARIES

Robin:
The project team and Advisory Board, a group comprised of archivists, librarians, educators, and historians, have proposed Maryland newspaper titles for all three grant awards based on patron interest or knowledge of the importance of the title, and prioritized these titles based on NEH content selection criteria and potential use by researchers, building a body of newspapers representing the people, businesses, and culture of Maryland. Because the project is not permitted to digitize newspapers previously digitized and available through other resources, such as commercial databases, the HMNP focused on digitizing community-based and small-town newspapers, which feature a wealth of information about local life. As the Maryland State Partner, UMD does not hold any of the newspapers we are digitizing, and are in a unique position among other NDNP awardees. We have partnered with the Library of Congress, Maryland State Archives, Maryland State Historical Society, and the Frostburg State University Library to gain access to content for this project.

HMNP Titles (Grants 1-3)
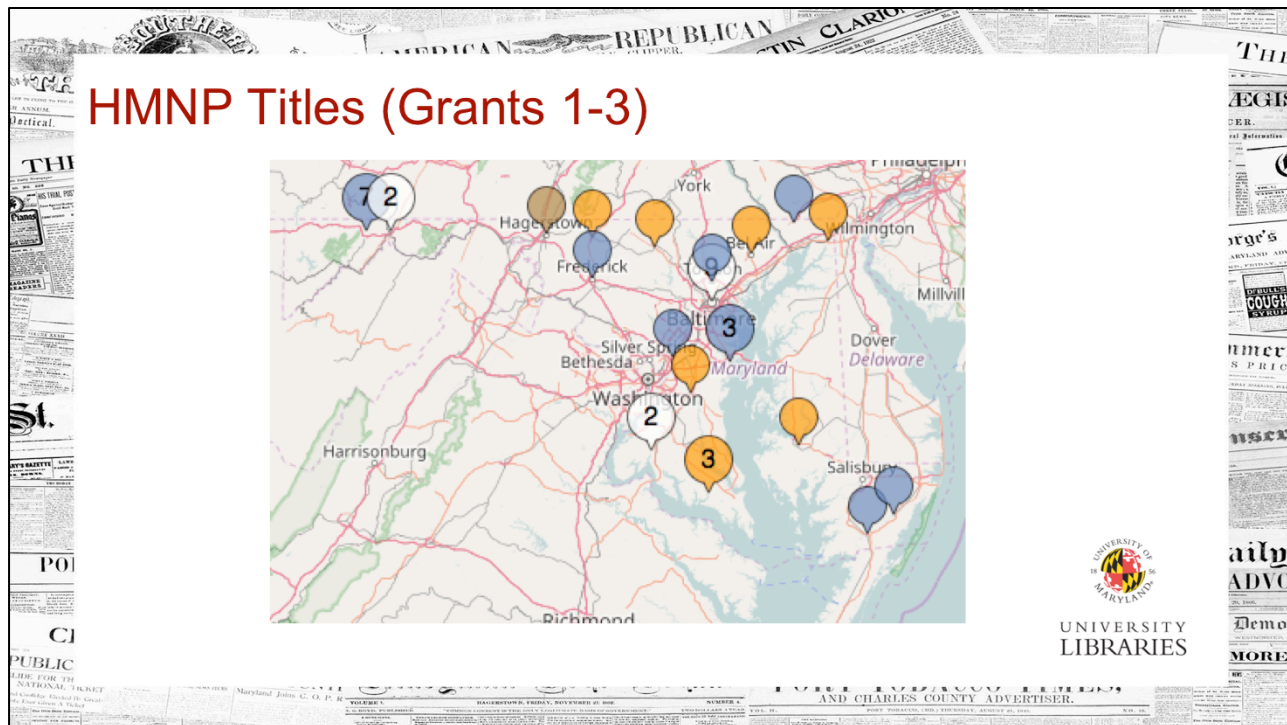
Rebecca:
Geographic representation across the state and growing with each grant cycle.
Western Md
Central Md/Baltimore
Eastern Shore
Southern Md
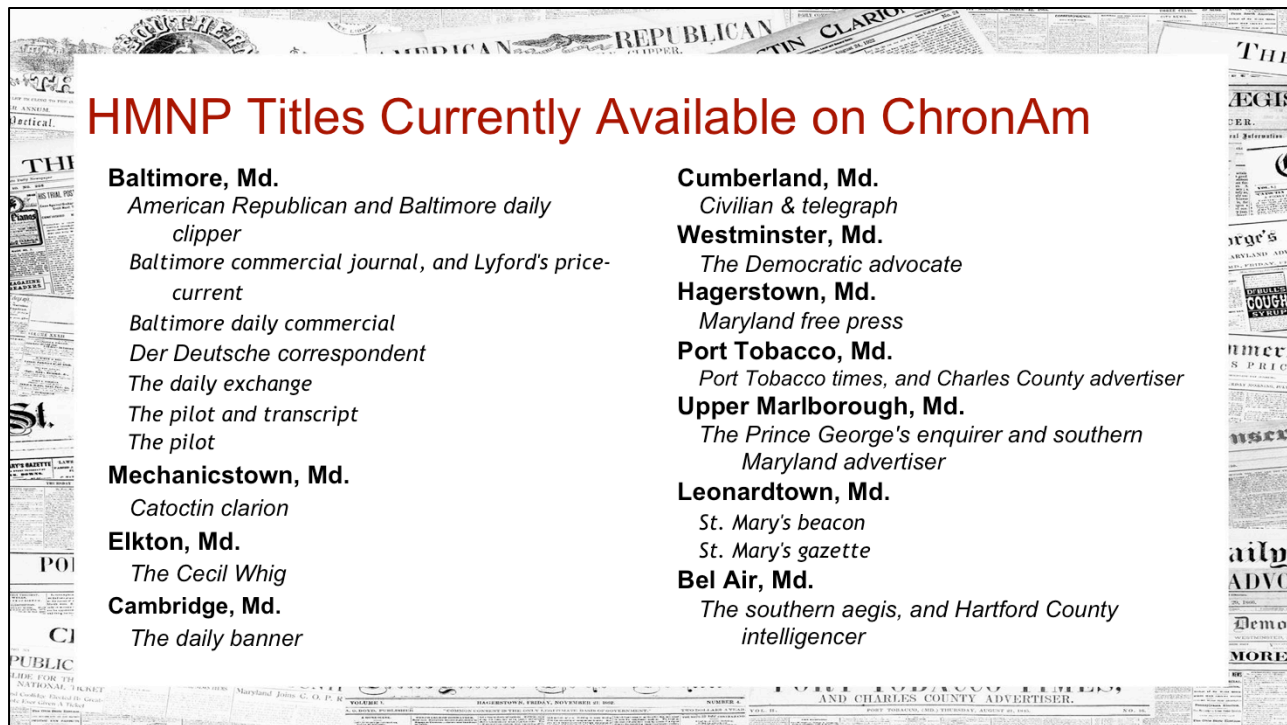
HMNP Titles Currently Available on ChronAm

Baltimore, Md.
 *American Republican and Baltimore daily
  clipper*
 *Baltimore commercial journal, and Lyford's price-
  current*
 Baltimore daily commercial
 Der Deutsche correspondent
 The daily exchange
 The pilot and transcript
 The pilot

Mechanicstown, Md.
 Catoctin clarion

Elkton, Md.
 The Cecil Whig

Cambridge, Md.
 The daily banner

Cumberland, Md.
 Civilian & telegraph

Westminster, Md.
 The Democratic advocate

Hagerstown, Md.
 Maryland free press

Port Tobacco, Md.
 Port Tobacco times, and Charles County advertiser

Upper Marlborough, Md.
 *The Prince George's enquirer and southern
  Maryland advertiser*

Leonardtown, Md.
 St. Mary's beacon
 St. Mary's gazette

Bel Air, Md.
 *The southern aegis, and Hartford County
  intelligencer*

Rebecca:
These titles cover a time span of approximately 82 years, with the earliest title, *The Pilot/The Pilot and Transcript* covering Baltimore-area news in 1840.
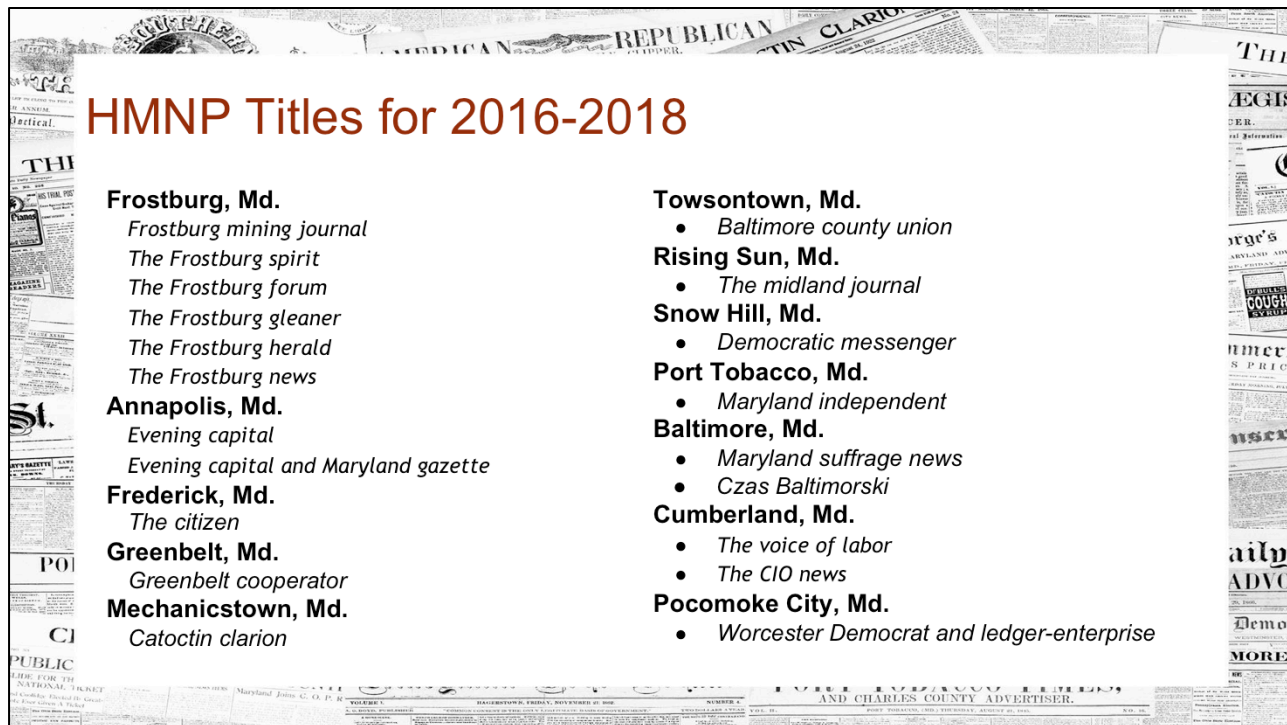
Regional Breakdown:
Western Maryland (4 titles)
Central Maryland/Baltimore (10 titles)
Eastern Shore (2 titles)
Southern Maryland (4 titles)

HMNP Titles for 2016-2018

Frostburg, Md.
  Frostburg mining journal
  The Frostburg spirit
  The Frostburg forum
  The Frostburg gleaner
  The Frostburg herald
  The Frostburg news
Annapolis, Md.
  Evening capital
  Evening capital and Maryland gazette
Frederick, Md.
  The citizen
Greenbelt, Md.
  Greenbelt cooperator
Mechanicstown, Md.
  Catoctin clarion

Towsontown, Md.
- *Baltimore county union*
Rising Sun, Md.
- *The midland journal*
Snow Hill, Md.
- *Democratic messenger*
Port Tobacco, Md.
- *Maryland independent*
Baltimore, Md.
- *Maryland suffrage news*
- *Czas Baltimorski*
Cumberland, Md.
- *The voice of labor*
- *The CIO news*
Pocomoke City, Md.
- *Worcester Democrat and ledger-enterprise*

Rebecca:
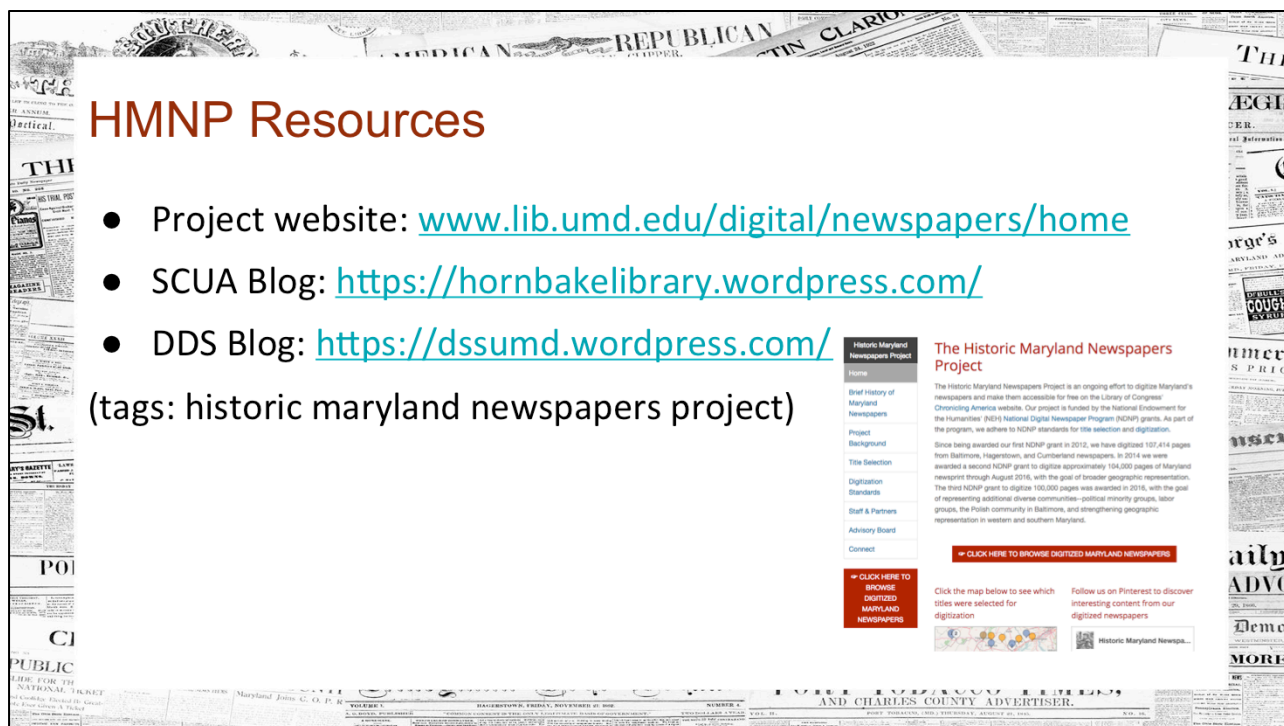
Titles we're particularly excited about in this grant cycle:

Maryland suffrage news published out of Baltimore from 1912 to approximately 1922, the paper was designed to help unite various suffrage organizations across the state and to serve as a source of information about suffrage to the women of the state of Maryland. It contains information about the movement, discussions of the needs or working class women, and news about education.

The voice of labor/CIO news published in Cumberland beginning in 1938, the paper covered the activities of textile workers, rubber workers, coal miners, and laundry and railroad workers in Western Maryland.

The Evening capital is the paper of record for Annapolis, publishing daily issues (except Sundays) beginning in 1884.

HMNP Resources

- Project website: www.lib.umd.edu/digital/newspapers/home
- SCUA Blog: https://hornbakelibrary.wordpress.com/
- DDS Blog: https://dssumd.wordpress.com/

(tags: historic maryland newspapers project)

Robin:

In 2012, UMD first started promoting the HMNP via the project website. We then had students do small research projects in their spare time and write short entries about Maryland history for the Special Collections and University Archives blog. The former project manager, Liz Caringola, discovered that these stories did not generate much web traffic based on the effort that was involved in creating the essays, so she discontinued this approach. We continued to blog about the project and its process on the Digital Systems and Stewardship division blog, which focused more about the technological process rather than the content.

HMNP Resources

- Wikipedia: https://en.wikipedia.org/wiki/Wikipedia:Meetup/DC/Chronicling_America
- Gateway: www.lib.umd.edu/newspapergateway
- Brochure

Gateway to Digitized Maryland Newspapers

The Gateway to Digitized Maryland Newspapers is a resource that compiles information about Maryland newspapers that have been digitized and are accessible online. Our goal is to provide a central location for users to find digitized Maryland newspapers and provide links to the websites that host them.

MD Newspapers
Search
Historic Maryland Newspapers Project
Maryland Newspapers at Maryland State Archives

UNIVERSITY OF MARYLAND

UNIVERSITY LIBRARIES

Robin:
At the end of the first grant cycle, Liz Caringola turned to Wikipedia, a well-indexed place where people frequently turn to find information, and had students elaborate on the title essays, creating more links to the digitized resource. She also hosted several edit-a-thons where she taught community members how to edit Wikipedia and they spent a few hours creating pages that related to Maryland history using the digitized content in Chronicling America. At the end of the first grant, Liz also expanded the survey data to include titles that were available via paid resources as well, and with the assistance of developers at UMD, created the Gateway to Digitized Newspapers, a database of all these titles with links back to the free or paid sources. Rebecca Wack, new project manager, has maintained this information.

Prior to the Rebecca coming to the project, I created a brochure intended to be an easy guide for genealogists, a target audience, but one that UMD, as a university, has a harder time performing outreach to. The brochure is available to be picked up [will it?], and is intended to be used as a handy guide for the Advisory Board members and other Chronicling America advocates to perform outreach at their institutions. This brochure is intended to be the first in a series of brochures or other resources to reach out to different demographic groups, such as K-12 teachers or higher education professors. In the next year, we intend to create a much larger Resources section of our website.

NDNP Resources and Outreach Initiatives

- LOC NDNP Extras (http://www.loc.gov/ndnp/extras/)
  - Tips for searching Chronicling America
  - Teaching resources
  - Projects that utilize ChronAm for research and scholarship
  - NDNP state awardee blogs and original video content
- meta|morphosis (http://www.uky.edu/Libraries/NDNP/metamorphosis/index.html)
  - Partnership between University of Kentucky, LoC, and NEH to prep non-NDNP states for inclusion in future NDNP projects
  - Offers video lectures on: historic newspapers, microfilm evaluation, proposal writing, calculating storage needs, and metadata collection and dissemination

LIBRARY OF CONGRESS

NATIONAL ENDOWMENT FOR THE
Humanities

Rebecca:
NDNP Extras
Multiple search tutorials provided by Ohio, Vermont, Florida, and South Carolina
newspaper projects
Teaching resources
        Lesson plans from LOC, NEH, other state newspaper projects
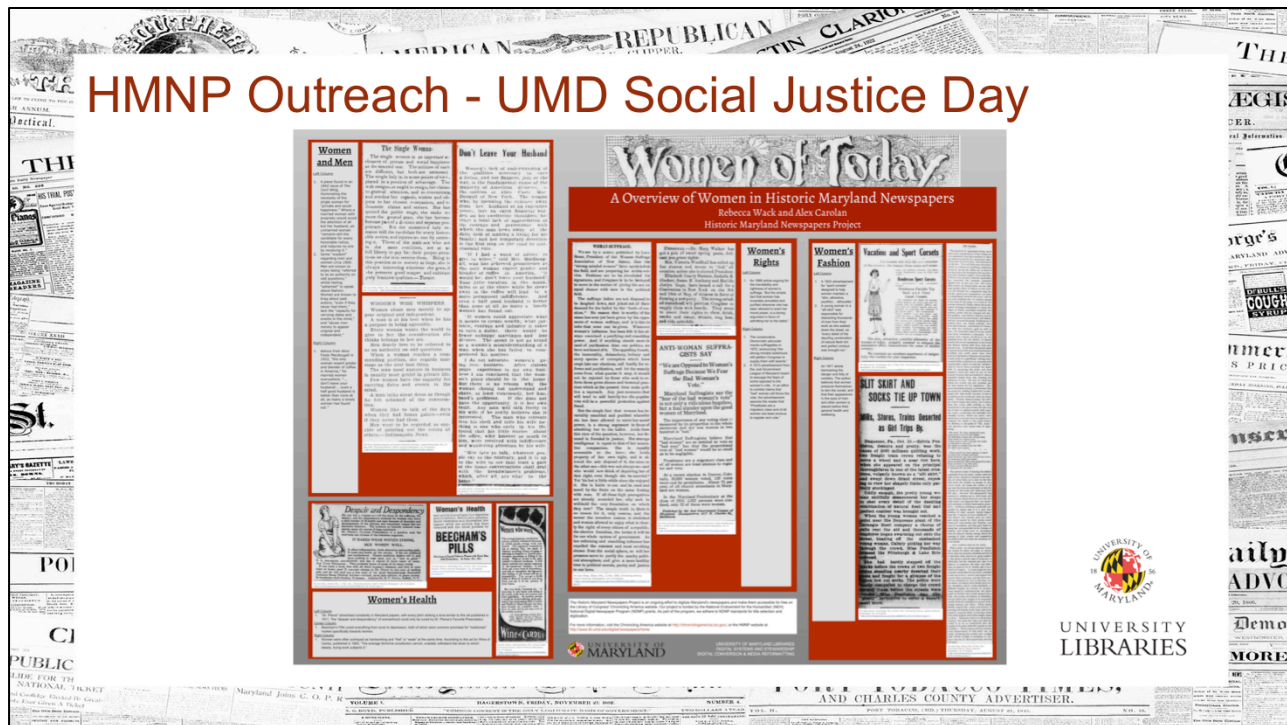   Teaching with historical resources
        Topic search guides
Using ChronAm (this will be covered later by Robin)
State newspaper blogs
Video content
Meta|morphosis
Amazing video tutorial series from U of Kentucky offering guides on microfilm
evaluation, storage needs, metadata collation, etc.

HMNP Outreach - UMD Social Justice Day

Rebecca:
Social Justice Day 2017 (Tuesday April 18, 2017)
Poster entitled "Women of Today - An Overview of Women in Historic Maryland Newspapers" covered issues of women and men, women's health, women's rights, and women's fashion. We selected a variety of articles from Maryland papers on ChronAm and discussed them with students and conference attendees. The female students in particular seemed rather shocked that marriage and health issues do not seem to have changed much, at least as far as media handling is concerned.

HMNP Outreach - Social Media

Rebecca:
Facebook allows us to go more in-depth on particular subjects and generally yields more click-throughs to Chronicling America.
Twitter allows us to stay topical and reach younger users with greater volume. We tend to interact with other state newspaper projects on Twitter as well.
Instagram allows us to have fun. We post absurd ads and articles and have seen respectable engagement as far as "likes" and "comments" are concerned.

#ChronAm - Other NDNP Social Media Accounts

Paper Bot
@paperbot

Just a bot announcing 100 year old news from OCR'ed newspapers using the Chronicling America API.

Somewhere on the Web
github.com/edsu/paperbot
Joined September 2010

NV Newspapers
@nvdnp FOLLOWS YOU

Nevada Digital Newspaper Program (NVDNP) is an extension of the National Digital Newspaper Program (NDNP) based at UNLV - University Libraries

Las Vegas, Nevada
nvdnp.wordpress.com
Joined March 2015

Library of Congress ✔
@librarycongress

World's largest library. Explore collections, services & plan a visit. Follow @LibnOfCongress to meet Carla Hayden, 14th Librarian of Congress.

Washington, DC
loc.gov
Joined June 2007

DigMichNewspapers
@DigMichNews FOLLOWS YOU

Digitizing Michigan's past, one historic newspaper at a time at: DigMichNews.cmich.edu

Central Michigan University
DigMichNews.cmich.edu
Joined September 2013

FL & PR Newspapers
@UFNDNP FOLLOWS YOU

Historic #Florida and #PuertoRico newspapers provided by the @UFlorida & @uprrp digitized and made available for free online

Gainesville, FL
ufdc.ufl.edu/ufndnp
Joined November 2013

Rebecca:
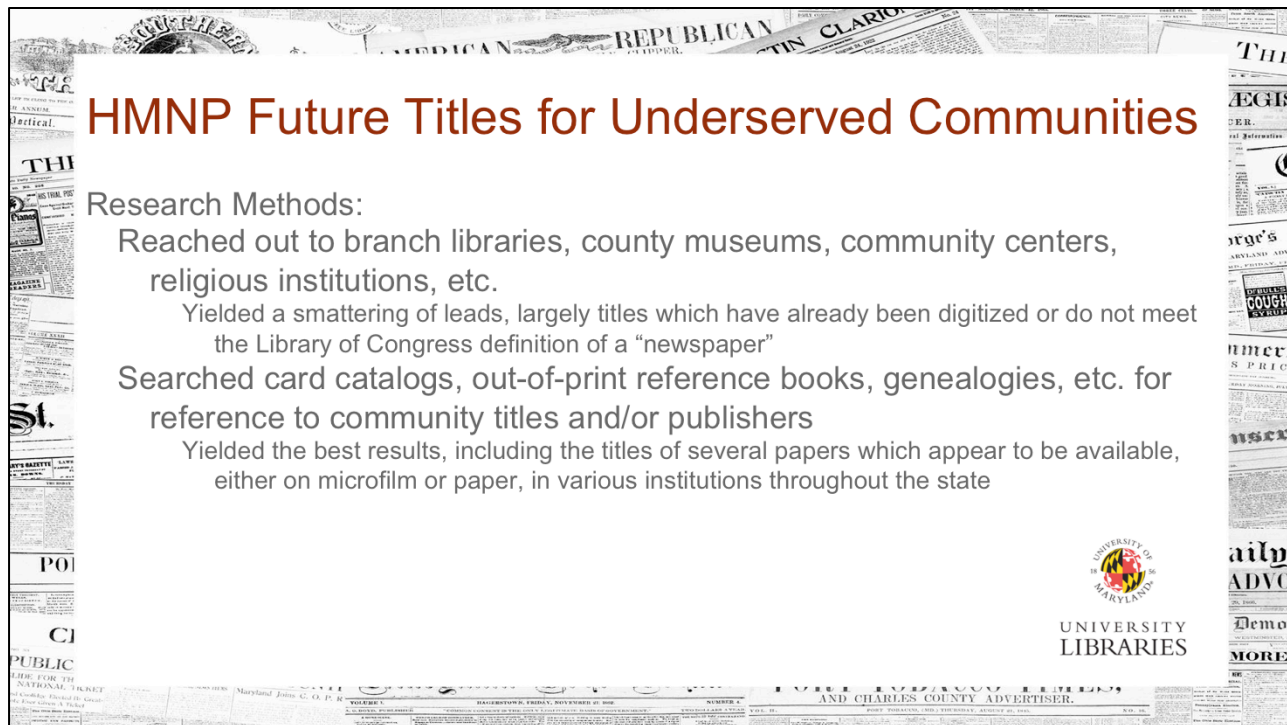@PaperBot retweets #ChronAm posts on Twitter
@LoC is a general account which occasionally posts ChronAm content
@DigMich and @UFNDNP are as active as we are

Live Tweet Demo

Rebecca

HMNP Future Titles for Underserved Communities

Research Methods:
  Reached out to branch libraries, county museums, community centers,
    religious institutions, etc.
      Yielded a smattering of leads, largely titles which have already been digitized or do not meet
        the Library of Congress definition of a "newspaper"
  Searched card catalogs, out-of-print reference books, genealogies, etc. for
    reference to community titles and/or publishers
      Yielded the best results, including the titles of several papers which appear to be available,
        either on microfilm or paper, in various institutions throughout the state

UNIVERSITY
LIBRARIES

Rebecca:
Librarians are busy (go figure). Plenty of assistance when I've requested information on specific titles, but minimal responses concerning general collection contents. Reference books were incredibly helpful, especially earlier publications. Memory is short, especially for niche titles, so finding books published in the early 1900s concerning specific communities yielded the most fruitful results.

HMNP Future Titles for Underserved Communities

Research Findings (or: Have you seen these papers?)

Der Neue Correspondent	1869-187?	German
Telegraf	1909-1951	Czech
The Negro Appeal	1899-19??	English
The Afro-American	1892-1899	English
The ledger	1898-1899	English
Afro-American ledger	1900-1915	English
The American citizen	1879-1???	English
The race standard	18??-1???	English
The Jewish chronicle	1875-1880	English, German
Katholische Volkszeitung	1860-1914	German
Baltimore Wecker	1851-1867	German
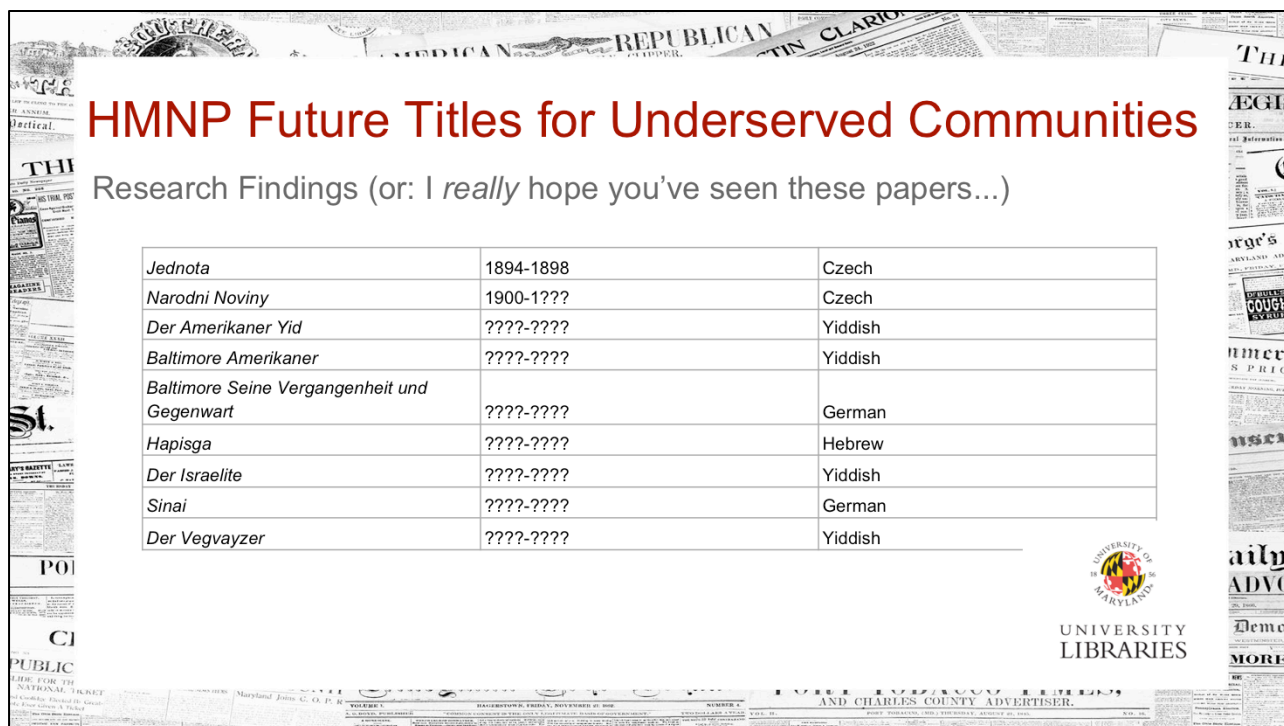
UNIVERSITY
LIBRARIES

Rebecca:
All titles published out of Baltimore; all titles covered in ChronAm database

Some of these titles included article references in books, but most were simply cited as having existed. Many of the African American titles are deceptive in that a holdings search will yield lots of microfilm across the country… but it's generally of the same issue(s), as a one-off of the paper would've been included in a grab bag black newspapers reel.

HMNP Future Titles for Underserved Communities

Research Findings (or: I *really* hope you've seen these papers...)

Jednota	1894-1898	Czech
Narodni Noviny	1900-1???	Czech
Der Amerikaner Yid	????-????	Yiddish
Baltimore Amerikaner	????-????	Yiddish
Baltimore Seine Vergangenheit und Gegenwart	????-????	German
Hapisga	????-????	Hebrew
Der Israelite	????-????	Yiddish
Sinai	????-????	German
Der Vegvayzer	????-????	Yiddish
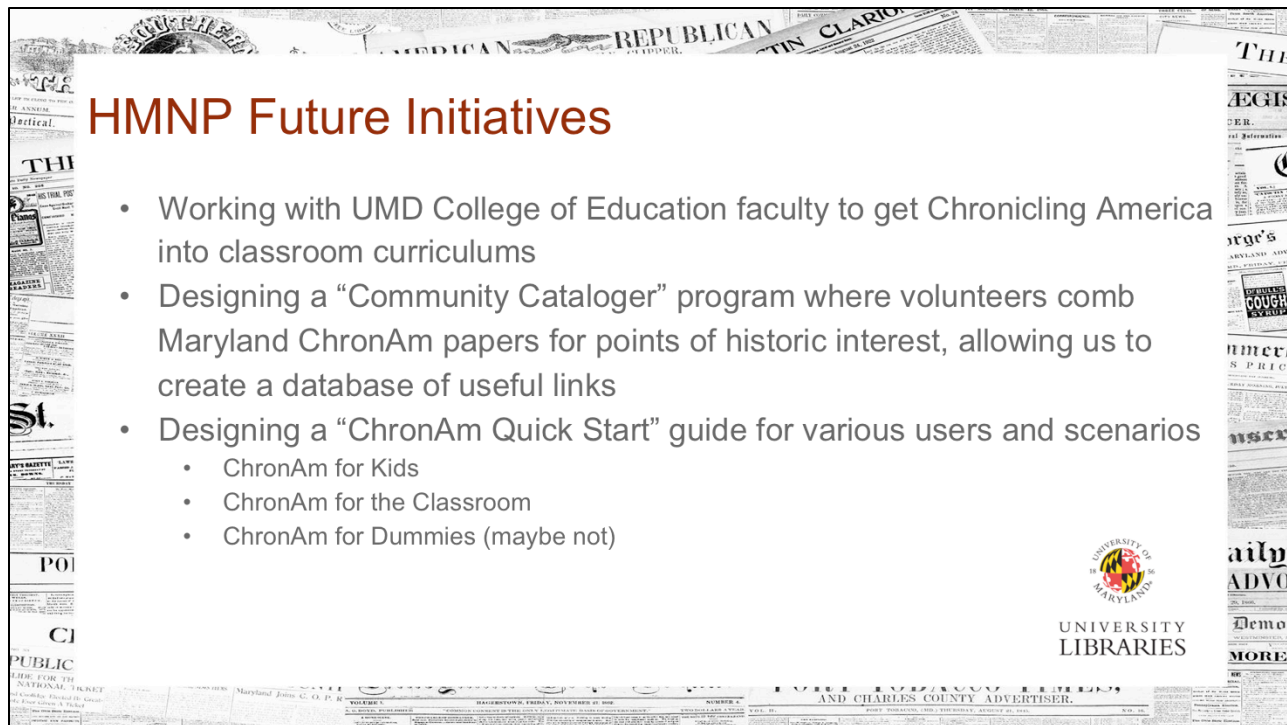
UNIVERSITY LIBRARIES

Rebecca:
All titles published out of Baltimore; none of these titles have ChronAm newspaper database listings

These titles are ghosts. All were found cited in bibliographies with little direct reference in the text itself. The Yiddish titles are particularly tricky, as the community seemed to recycle popular newspaper titles across the country.

HMNP Future Initiatives

- Working with UMD College of Education faculty to get Chronicling America into classroom curriculums
- Designing a "Community Cataloger" program where volunteers comb Maryland ChronAm papers for points of historic interest, allowing us to create a database of useful links
- Designing a "ChronAm Quick Start" guide for various users and scenarios
  - ChronAm for Kids
  - ChronAm for the Classroom
  - ChronAm for Dummies (maybe not)

UNIVERSITY OF MARYLAND
UNIVERSITY LIBRARIES

Rebecca:
Get feedback from room to see what resources they're interested in

We need your input concerning who would be best served by a ChronAm quick starter… who are your library patrons and what would they be using the resource for? We want to design printables for multiple scenarios to ensure as much usage as possible.

Focusing on reaching out to social studies teachers, both nationally and within Maryland, and working with education students to get ChronAm front and center in their minds before they enter their own classrooms.

Questions About HMNP

Chronicling America

Chronicling America

- http://chroniclingamerica.loc.gov/
- Database that searches NDNP digital newspapers
- Freely-accessible, no login required

Robin:

Chronicling America is the freely accessible database of all the newspapers digitized in the NDNP, including all the Maryland newspapers, and is developed and permanently maintained at the Library of Congress. The corpus is word-searchable and browsable in multiple ways. Chronicling America also facilitates a second point of access to the Newspaper Title Directory, the catalog records created during the USNP; the 500-word historical essays about the titles digitized and available on Chronicling America; bulk data downloads via an Application Program Interface (API); and NDNP Extras, projects and resources expanding building on the newspaper data.

ChronAm Facts

Representation:

- 43 states
- 1 territory (Puerto Rico)
- Washington, District of Columbia
- 2,228 titles
- 11,972,816 pages

(Included as of 5/31/2017)

Robin:
Chronicling America includes nearly 12 million pages and over 2,000 titles from 43 states, Puerto Rico, and Washington, DC.

ChronAm Usability

Search for:

- News articles
- Life events
    - Obituaries
    - Marriage
    - Birth
- Advertisements
- Cartoons
- Poetry
- Literature
- Sheet music
- And much, much more!

Robin:

These historic newspapers include news articles, announcement of life events, advertisements, Cartoons, poetry, literature, sheet music, and much more.

ChronAm Searching Options

Simple Search
★ Most inclusive search

Advanced Search
★ Narrow down many results

Search Pages | Advanced Search | All Digitized Newspapers 1789-1924

US Newspaper Directory, 1690-Present

All states | + from 1789 to 1924 + | enter one or more search words | GO

All Digitized Newspapers 1789-1924
★ Search within a specific newspaper title

US Newspaper Directory 1690-present
★ Part of the original US Newspaper Project
★ Search all newspapers (digitized or not)

Robin:
The search interface offers multiple options for simple and advanced searches, and links to the directory of all cataloged newspaper titles. Within the simple search, you can limit by state, a handy feature if you want to search in Maryland only, and the date range.
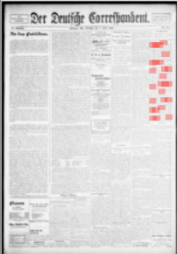
Robin:

After you perform a search, the keywords will highlight, leading you to successful results. This simple search shows the results of a Maryland search for "John Wilkes Booth," with no other limits. As you see, the these results contains many red highlights, showing that this article might be highly useful in your research, whereas if the page doesn't show many red highlights, it might not provide as many details.

Using Primary Sources in Research

Oyster Wars

From Wikipedia, the free encyclopedia

The **Oyster Wars** were a series of sometimes violent disputes between oyster pirates and authorities and legal watermen from Maryland and Virginia in the waters of the Chesapeake Bay and the Potomac River from 1865[1] until about 1959.

Contents [hide]
1 Background
2 Clashes
3 Government responses
4 See also
5 References
6 Further reading
7 External links

Background [edit]

In 1830, the Maryland General Assembly passed legislation which authorized only state residents to harvest oysters in its waters.[2] Maryland outlawed dredging, while Virginia continued to allow it until 1879.[3] In 1865, the Maryland General Assembly passed a law that required annual permits for oyster harvesting[2] and this has been described as the start of the Oyster Wars.[1]

Oyster Wars

Oyster War engagement in 1886.

Date	1865–1959
Location	Chesapeake Bay Potomac River

Parties to the civil conflict	
Oyster pirates	Legal watermen
New England fishermen	State of Maryland
	Commonwealth of Virginia

Robin:

For those of you who don't know, the oyster wars occurred between Maryland and Virginia in the Chesapeake Bay and Potomac River, from about 1865-1959, due to unequal harvesting laws between the two states, resulting in illegal activities and clashes on the water. This Wikipedia page relies on secondary sources and modern sources for citations, and is an example of resources that could be elaborated on by using historical newspapers as primary source material. Because this search would require looking at two states' newspapers from a discrete era, it is an ideal example of when to use Advanced Search.

ChronAm Advanced Searching Options

Robin:
The Advanced search offers many more limitations. In addition to limiting by state and date, you can limit by newspaper title, front page only, language, and have more options for searching words. You can narrow/expand using Chronicling America Advanced Search: "with any of the words," "all of the words," "with the phrase," and "with the words within x number of spaces." In this example, I chose to search through Maryland and Virginia papers for the words "oyster" and "war" within five words of each other.

Advanced Search in ChronAm

Robin:
The results provide over 5,000 pages of articles documenting oyster wars from Maryland and Virginia.

ChronAm Search Tips

Family Names:
- Search full name, first and middle initial and last name (e.g., W.D. Barrows), maiden name, or last name
- Spelling variants
- Quotation marks around exact names
- First name abbreviations (e.g., Chas. for Charles, Wm. for William)
- Women may be identified by husband's name (e.g., Mrs. John Smith)

Vocabulary and Spelling:
- Vocabulary and spellings have changed. Browse newspapers for potential search terms contemporary to the period.
- Synonyms: "wedding," "marriage," "nuptial," etc.
- Typos occur often in historic newspapers and text recognition is imperfect (e.g., "e" for "o")

Robin:
Beyond searching for famous people and popular topics, for genealogical research or other name-based research, it's important to consider that names may be worded in several ways, such as first and last, or just initials and a last name. Spelling variants are also common with many names, including people, but also the names of places as towns and cities went through incorporation process. Using quotation marks around names can be helpful, particularly if the last name is common or the name for something else (such as "Pike"), but it can also rule out too many results. Finally, married women were often referred to by their husband's name, so family history may require more research.

The vocabulary or spellings for words have changed through time, so if search results are not yielding results, browse the newspapers around the period in which you want to find information, and then use synonyms for words.

Finally, typos in the newspaper or unclear text can result in imperfect OCR. If your search results are still not yielding results, try substituting letters--to OCR, "e" and "o" are often interchangeable.

Robin:
After you click on a results page, you can zoom in closer to read the text. The top inset shows what area of the page you are viewing. If you are interested in downloading your results, you have multiple file options, as OCR text, PDF, and JPEG 2000.

Robin:

The viewing box can be used in conjunction with the clipping tool, the tool on the far right of the toolbar. Creating a clipping creates an image file with metadata so you know where the clipping came from, which is helpful to continue your research.

ChronAm Database/OCR Limitations

- Poor-quality originals
- Inconsistent/mixed fonts
- Faded text
- Missing characters
- Page bleed-through
- Tight gutters/missing text
- Column confusion

The Goldsboro star. volume (Goldsboro, N.C.), 25 June 1881. *Chronicling America: Historic American Newspapers.* Lib. of Congress. <http://chroniclingamerica.loc.gov/lccn/sn91068338/1881-06-25/ed-1/seq-2/>

Robin:

Chron Am has some limitations, caused by newspaper incompleteness or less than ideal quality issues, such as inconsistent or mixed fonts, faded text, missing characters, bleed through from the reverse side of the page, or tight gutters if the newspaper is bound, or column confusion caused by an error in the OCR script reading the newspaper.

ChronAm v. Other Newspaper Databases

- Only undigitized titles
- Public domain restriction
- More data capabilities
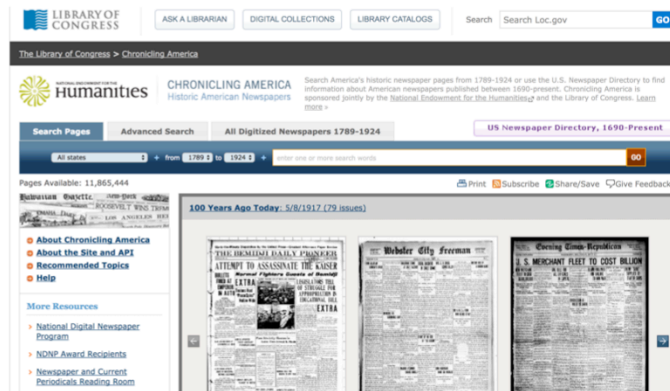
Robin:
Though Chronicling America has nearly 12 million pages of newsprint from across the country and a substantial subset from Maryland, the collection is not exhaustive. It excludes specific papers because the NDNP will not digitize newspapers available elsewhere, including commercial database products like Proquest, Readex, or Newspapers.com. For example, the archives of the *Afro-American*, the premiere African American title in Maryland, and is available via Proquest. Readex holds limited runs of four additional African American newspaper titles. These titles form the known, extant copies of African American newspapers in Maryland, and are not included in the Maryland corpus on Chronicling America, creating a knowledge gap for those who lack access to these databases.
ChronAm does provide much greater capabilities for repurposing the newspapers and OCR, expanding the usage of the data beyond that of straight research.

ChronAm Demo

http://chroniclingamerica.loc.gov/

Using ChronAm in Projects

- About the Site and API:
  http://chroniclingamerica.loc.gov/about/api/
    - Bulk data download
- Using ChronAm in projects
    - http://www.loc.gov/ndnp/extras/

Using Chronicling America to Advance Research and Scholarship
The following examples re-use digital objects from Chronicling America in different ways. Some are displays of selected digital items, some use machine-harvested datasets from the site, some use standard APIs (application programming interface) to connect data in Chronicling America to other systems. There are lots of ways to discover more about historic newspapers - perhaps even make your own!

America's Public Bible: Biblical Quotations in U.S. Newspapers
Lincoln Mullen
George Mason University

American Lynching: Uncovering a Cultural Narrative
Andrew Bales
University of Cincinnati

An Epidemiology of Information
Virginia Tech
University of Toronto

Animated Visualization of Arizona Sentinel Weekly, 1872-1911
Visualizing newspapers history: The Hawaiian Star, 5930 front pages, 1893-1912
Software Studies Initiative
University of California, San Diego
Cyrus Kiani

Bookworm
Culturomics

Chronicling Hoosier
Kristi Palmer
Indiana University-Purdue University

Create Modern Microfiche with the Chronicling America API
Tim Kulp - MSDN Magazine

Data Visualization: Journalism's Journey West
Rural West Initiative, Bill Lane Center for the American West
Stanford University

Digital APUSH: Revealing History with Chronicling America
A.P. U.S. History Students
Sunapee High School

Hawai'i newspaper content in Chronicling America
Flickr set of images from Hawai'i newspapers cropped from Chronicling America
University of Hawai'i Manoa Library

Historic Newspapers (Library of Congress Flickr pilot project)
Library of Congress

Robin:
In addition to performing research-based searches and browsing the papers, Chronicling America also provides access to information about historic newspapers and select digitized newspaper pages. The Library of Congress provides information on how to access the data in different ways via the API on the page "About the Site and API." To encourage a wide range of potential uses, the Library of Congress designed several different views of the data, all of which are publicly visible. Each uses common Web protocols, and access is not restricted in any way. Users do not need to apply for a special key to use them. Together they make up an extensive application programming interface (API) which researchers can use to explore all of the data in many ways.

One of the more revolutionary aspects of this database is that the Library of Congress allows a bulk download of all the batch and OCR data, or subsets of that data, which can be facilitated by their staff so they ensure the researcher is not a hacker. For example, researchers can try out new indexing techniques on the millions of pages of OCR data in Chronicling America, or a service provider could support a high volume of full-text searches across the corpus, outside of the Chronicling America API. LC also provides a website called NDNP Extras, which features examples of how other institutions and individuals have used ChronAm data or research topics to create education resources to advance research and scholarship. While people can use the projects to facilitate their research, they can also use these projects as inspiration and create their own notable projects, which can be included on this NDNP Extras site.
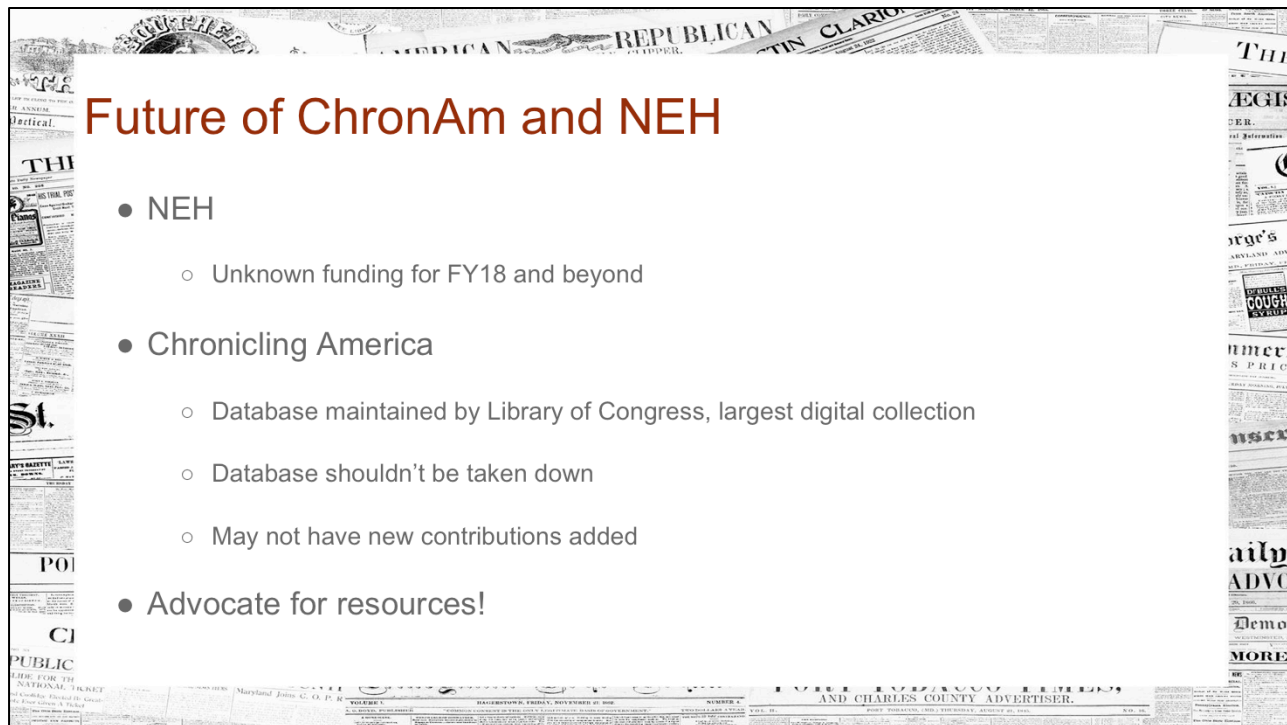
NEH Data Challenge

- **Criteria:** Create web-based project using Chronicling America data
- Nationwide competition
- Cash prizes awarded to top four projects ($1K-$5K)
- Winners announced in July 2016

- "*America's Public Bible* uncovers the presence of biblical quotations in the nearly 11 million newspaper pages in the Library of Congress's Chronicling America collection."
- http://americaspublicbible.org/

Robin:

The NEH Data Challenge was a national competition to create web-based project using Chronicling America data. Cash prizes were awarded to top four projects in July 2016. America's Public Bible was one of the winning projects and "uncovered the presence of biblical quotations in the nearly 11 million newspaper pages in the Library of Congress's Chronicling America collection."

Future of ChronAm and NEH

- NEH
  - Unknown funding for FY18 and beyond
- Chronicling America
  - Database maintained by Library of Congress, largest digital collection
  - Database shouldn't be taken down
  - May not have new contributions added
- Advocate for resources!

Robin:
We haven't heard more from NEH beyond that they are waiting for Congress's final budget proposal. The White House's current budget proposal includes enough funding to shut down the agency in FY18, but this budget is far from a proposal.
The Library of Congress separately maintains Chronicling America as its largest digital collection. Though we haven't had confirmation, we can more accurately predict that the database will not be taken down, and the content will continue to be freely available and searchable to all. If NDNP grant projects wind down due to the closure of that agency, new contributions may cease, but we suspect that this will be either in the fall of 2018, or more likely, in the fall of 2019, as the 2017-2019 grants come to a close.
If you start using Chronicling America and find it useful, we highly encourage you to contact your representatives to let them know you value this resource and don't want NEH shut down.

Questions About ChronAm
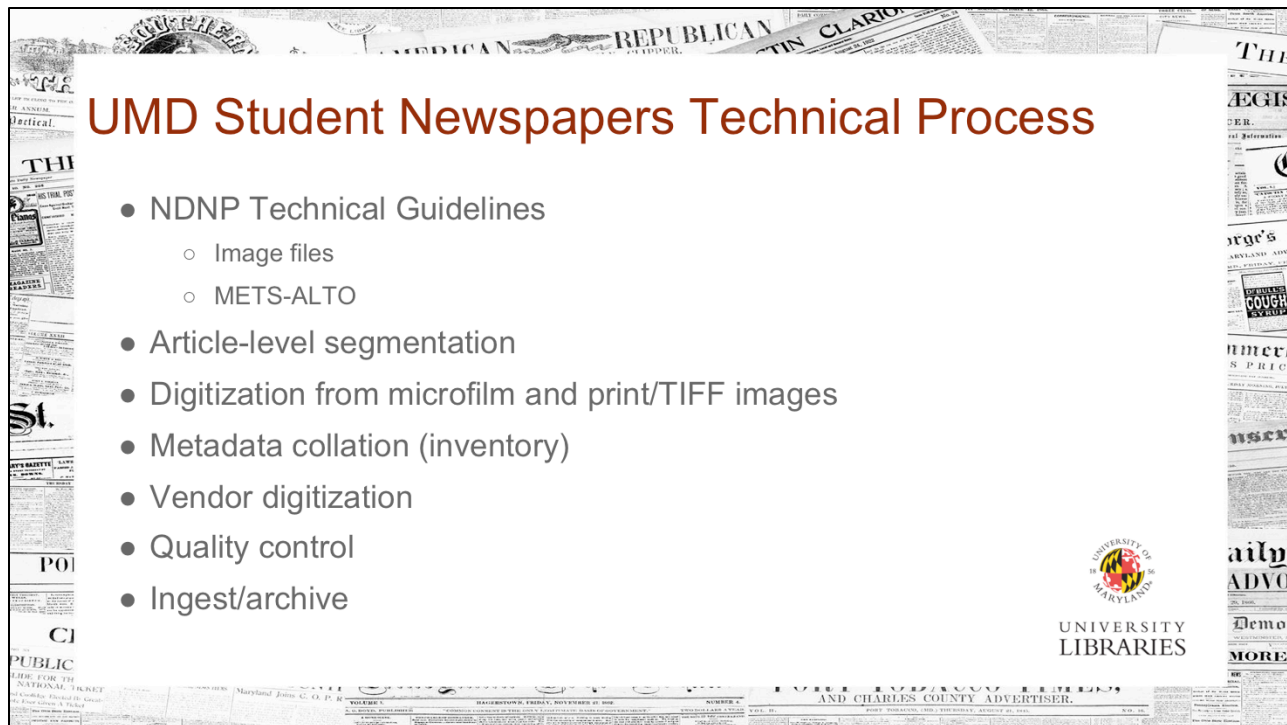
Maryland Newspaper Repository

UMD Student Newspapers Digitization Timeline

- 2013: *The Diamondback* project conception
- 2015: UMD Launch
- 2016: *The Diamondback* Digitization Project, Phase 1
- 2017: *The Diamondback* Digitization Project, Phase 2
- 2017: UMD Launch
- 2017: UMD student newspapers, Phase 1
- 2018: *The Diamondback* Digitization Project, Phase 3
- 2019: UMD student newspapers, Phase 2

Robin:
With the growing expertise in newspaper digitization and the METS-ALTO metadata standard promoted by NDNP, in 2013, the Manager, DCMR was approached by the University Archivist, Anne Turkos, to digitize The Diamondback, the student newspaper of record, and its earlier preceding titles, from 1910-present. Because there was no likely funding for this project, the University Archivist created a crowd-funded campaign in 2015 to raise money for an initial project phase, which resulted in raising $30,550 from the crowd-funding campaign, an additional $1,000 from the Student Government Association, and an additional $2,000 from an Anacostia Trails Heritage Area Inc. grant. The DIC approved the project to begin for the fiscal year 2016 project cycle. The first phase of the project was limited to approximately 28,500 pages, from 1910-1971 due to funding and also because publication and copyright was transferred from the university to Maryland Media, Inc., a private company, beginning in October 1971, and the University Archivist was still in discussions to allow permission to digitize content after 1971.
The second digitization phase of approximately 48,500 pages, from 1971-1998, began in January 2017, and is nearly completed, using the remaining fundraised money and additional Libraries gift money. Anne Turkos will initiate a second UMD Launch campaign this fall to try and fund the remaining approximately 48,000 pages from 1999-2016 in early 2018. University Archives gift funds will allow *Black Explosion*, another student newspaper, to be digitized this fall. Additional student newspaper titles will be digitized next year.

UMD Student Newspapers Technical Process

- NDNP Technical Guidelines
  - Image files
  - METS-ALTO
- Article-level segmentation
- Digitization from microfilm and print/TIFF images
- Metadata collation (inventory)
- Vendor digitization
- Quality control
- Ingest/archive

UNIVERSITY
LIBRARIES

Robin:
Project stakeholders, including the myself, Liz Caringola (former HMNP project manager), and University Archivist Anne Turkos, agreed that the project would primarily follow the NDNP standards including the image files and METS-ALTO metadata. METS-ALTO is a combination of two XML metadata schemas. METS provides descriptive information about the page, issue, microfilm reel, etc. ALTO encodes OCR and the page style and layout, essentially creating coordinates for each word. Together, METS-ALTO creates a complete metadata picture of the newspaper page. It's what also enables the word-searching and the word-highlighting of the words you searched for. This is why creating simple PDF or TIFF files won't work. Anne determined that in addition to the NDNP deliverables and UMD-unique metadata fields, she wanted patrons to have access to article-level segmentation, which would provide additional METS-ALTO files of text-corrected OCRed titles to assist patrons with searching and content indexing.
While the project primarily relies on microfilm for digitization because it is more cost effective, early issues would also be digitized from print after a conservation vendor repaired the aged and damaged issues; these issues were included in the second project phase. UMD students created an issue-level inventory, with added metadata for the reel and batch, through a process called metadata collation. A student also performed automated checks on the batch using the Digital Viewer and Validator tool developed by the Library of Congress for NDNP batches, and proceeded to manually inspect 20% of the issues digitized.
We have ingested and archived the batches from the first project phase, and will soon be doing so with the second phase.

Robin:

At the same time as project digitization, development in 2016 was progressing on the UMD Student Newspapers Repository.

The Software Systems Development and Research department was provided technical requirements for the administrative and public interfaces as guidelines to build the user interfaces of the repository. They first began work on the backend of the repository, including the image server and the image viewer. A Systems Librarian ingesting digitized content into repo beginning in February 2017 and completed in early March.

Developers are currently working on the public interface and have a "soft launch" product out to facilitate internal searching at the beginning of June, two weeks ago. Developers are continuing to work on the repository, adding the OCR display and search capabilities to the administrative and public interfaces by the end of summer. The interface will be officially rolled out in the fall semester.

Through the rest of 2017, into 2018 and 2019, developers will continue to work on adding additional features and allowing editing in the administrative interface.

UMD Student Newspapers Browse

Robin:
This view shows the ability to browse by publication date, decade, newspaper title, and newspaper component.

UMD Student Newspapers Viewer
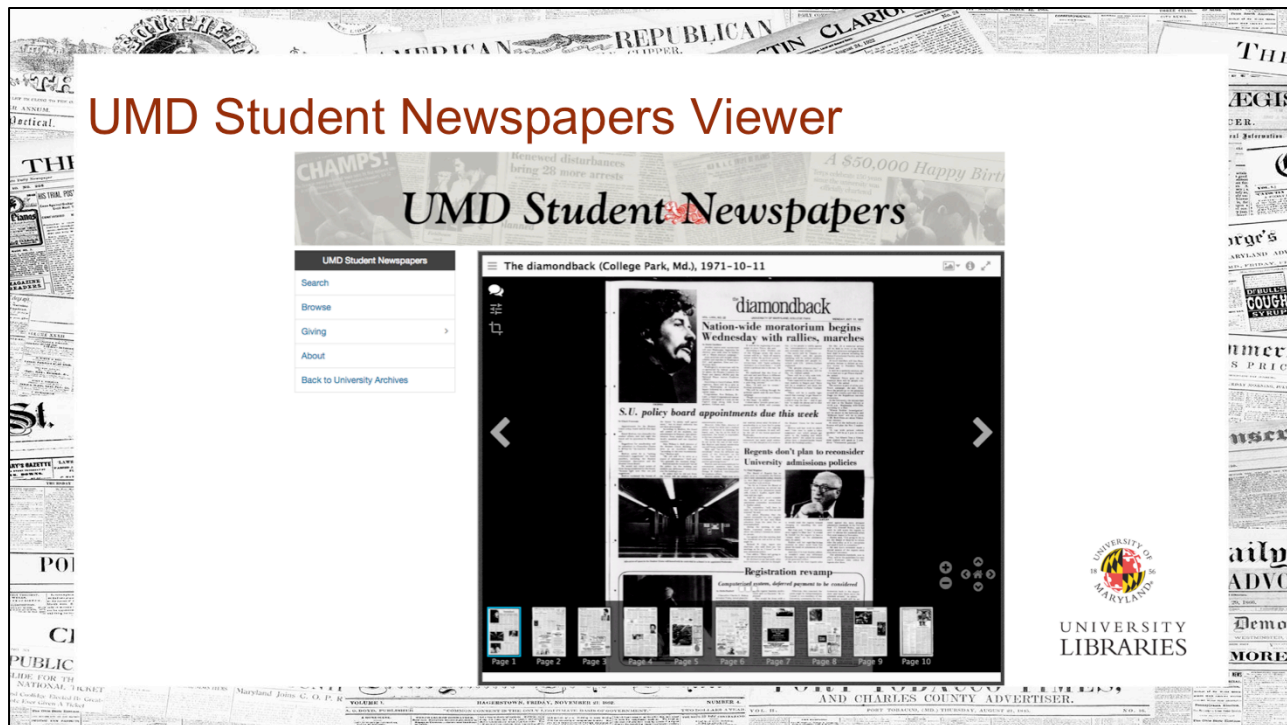
Robin:
This is the default viewer for each issue, which shows a main page viewer, options for scrolling through each page, and tools to manipulate the image in the viewer.

Maryland Digital Newspaper Repository

- Newspapers published in MD, freely available
- Repurpose developed infrastructure
- Timeline
  - Beginning 2017
    - Gauging interest
    - Partnerships
    - Metadata
    - Funding
  - 2018
    - Digitization
    - Infrastructure
  - Late 2018-2019
    - Ingest

Robin:
Because we are building the technical expertise to create a separate newspaper repository for our student newspapers, and we acknowledge many newspapers exist outside the NDNP criteria, we aim to create a Maryland Digital Newspaper Repository to create a publicly, freely available digital collection of newspapers published in the State of Maryland.
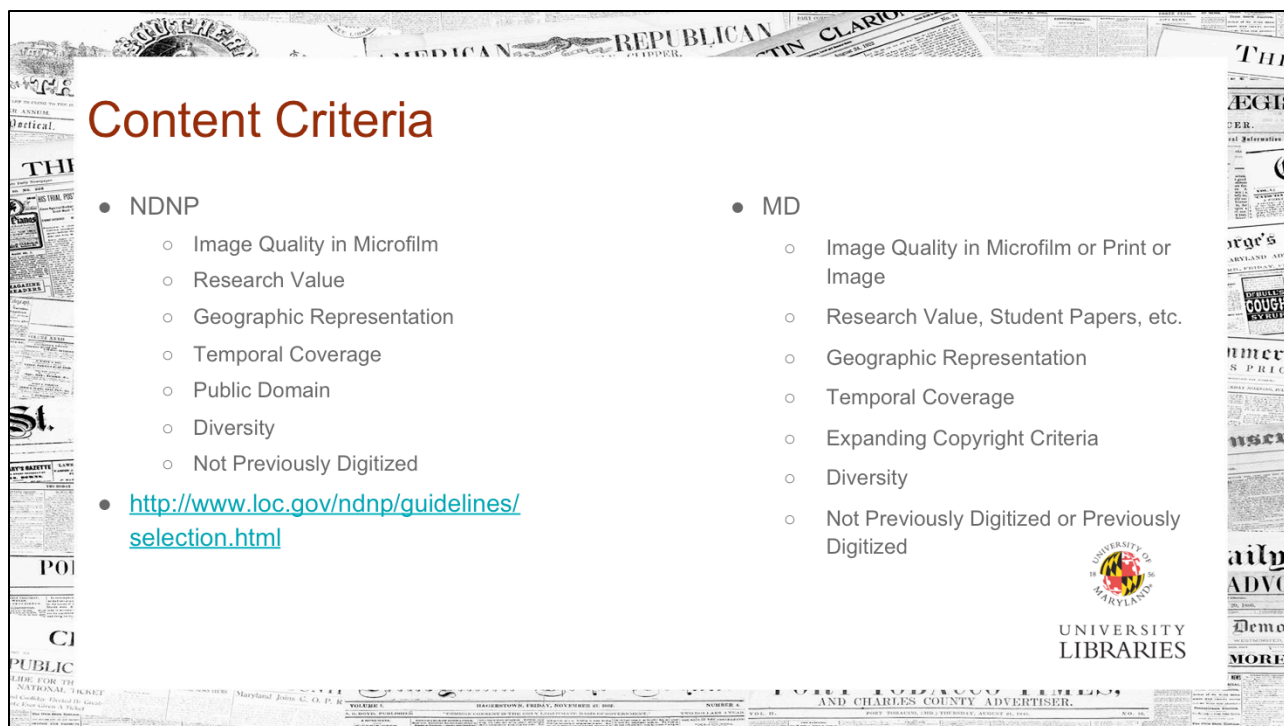
Technically, we plan to repurpose much of the infrastructure and public interface we have developed for UMD Student Newspapers with separate branding. We would allow file download, as we would for the UMD Student Newspapers, and apply the same filters and browse capabilities. UMD's developers would make future application development and migration decisions, with the feedback of key stakeholders in the Libraries.

If you have technical questions, Ben Wallberg, the Manager of SSDR will be available at the end of the session to answer questions.

In an ideal world, we will gauge interest throughout the rest of this year and begin partnerships for content. We will advise on metadata creation and could assist with this process in 2017. If other aspects require research, such as hunting down issues or determining copyright, those should happen in 2017 as well. If the holding institution doesn't have allocated funds for this project, we would recommend fundraising or applying to grants during this period.

Digitization could begin as early as 2018, with the necessary funding. All of this preparation work would afford UMD time to continue development on the repository infrastructure and user interface.

Ingest would most likely not happen until later in 2018 or 2019.

Content Criteria

- NDNP
  - Image Quality in Microfilm
  - Research Value
  - Geographic Representation
  - Temporal Coverage
  - Public Domain
  - Diversity
  - Not Previously Digitized
- http://www.loc.gov/ndnp/guidelines/selection.html

- MD
  - Image Quality in Microfilm or Print or Image
  - Research Value, Student Papers, etc.
  - Geographic Representation
  - Temporal Coverage
  - Expanding Copyright Criteria
  - Diversity
  - Not Previously Digitized or Previously Digitized
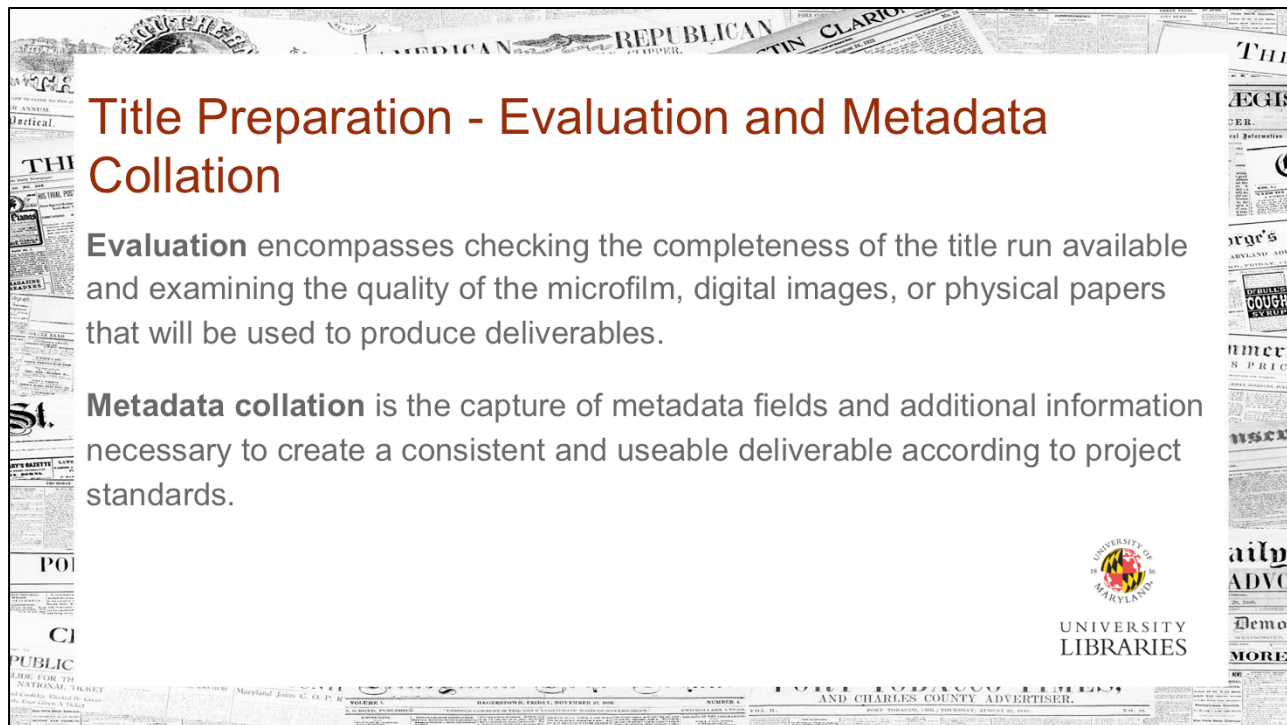
UNIVERSITY LIBRARIES

Robin:
The differences between the two content selection criteria include:
The inclusion of newspapers from print where no microfilm exists, which is not prioritized in the NDNP project since digitization from microfilm is the most cost-effective. We would also accept high-quality digital images which we could send to a vendor to create the metadata deliverables and file structure necessary for ingest.
As a university, we have slightly different research needs and want to include student newspapers or other special interest community papers that may not have qualified under "research value" for the NDNP. We are still developing what these guidelines would exclude, though some examples of non-newspaper publications are church bulletins and magazines.
If an entity provides funding for papers previously digitized and available via another database, such as a commercial database, we could consider re-digitizing and including that newspaper in the repository.

Title Preparation - Evaluation and Metadata Collation

Evaluation encompasses checking the completeness of the title run available and examining the quality of the microfilm, digital images, or physical papers that will be used to produce deliverables.

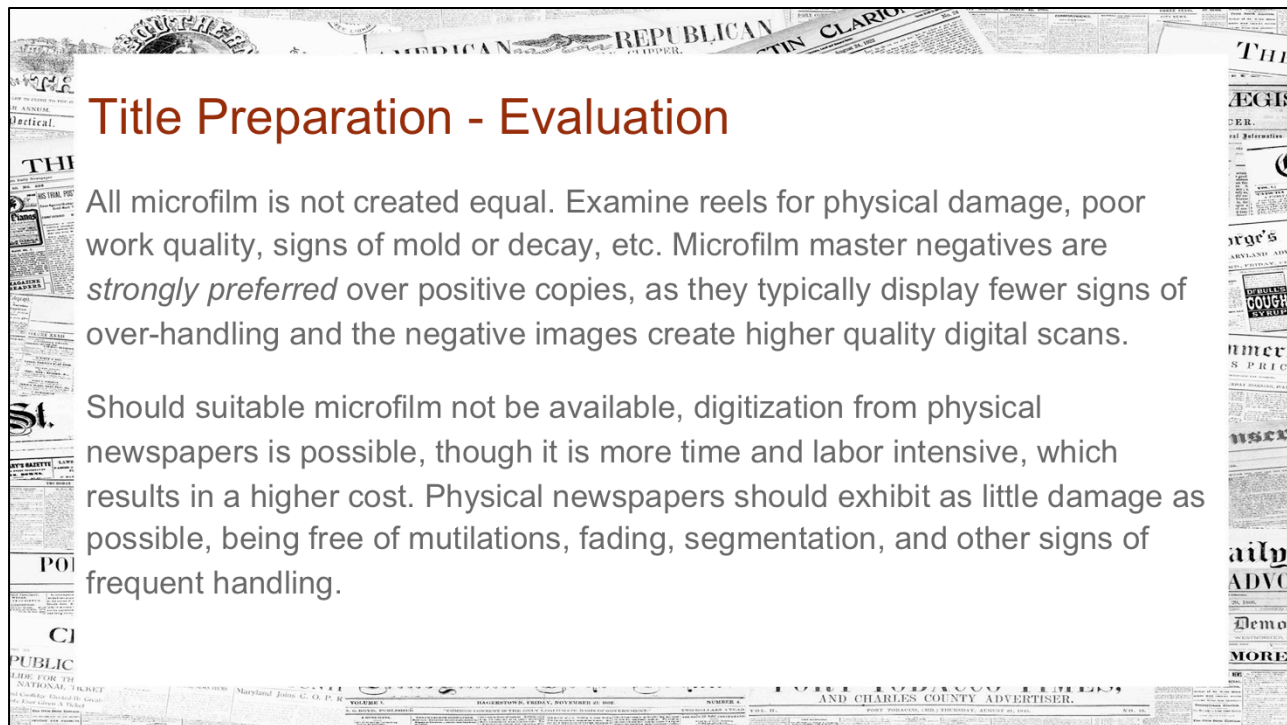**Metadata collation** is the capture of metadata fields and additional information necessary to create a consistent and useable deliverable according to project standards.

UNIVERSITY
LIBRARIES

Rebecca:
Good evaluation is how you preempt a wasted effort. Take your time to do accurate inventories and condition assessments before proceeding forward.

Metadata collation is brutal and boring but necessary to yield a useable product.

Title Preparation - Evaluation

All microfilm is not created equal. Examine reels for physical damage, poor work quality, signs of mold or decay, etc. Microfilm master negatives are *strongly preferred* over positive copies, as they typically display fewer signs of over-handling and the negative images create higher quality digital scans.

Should suitable microfilm not be available, digitization from physical newspapers is possible, though it is more time and labor intensive, which results in a higher cost. Physical newspapers should exhibit as little damage as possible, being free of mutilations, fading, segmentation, and other signs of frequent handling.

Rebecca: Evaluation for excessive damage or quality concerns.

Title Preparation - Metadata Collation

Metadata collation generates the object-level data necessary for a usable deliverable. The process is straightforward but time-consuming, resulting in an issue-by-issue entry for each title run including publication date, page count, supplement information, misprints and missing pages, etc. Issues missing from the run must also be accounted for.

The HMNP's standard metadata collation spreadsheet looks like this:

Reel Info		Title Information										Page Information					UMD Notes			
Reel number	MSA file name	LCCN	Title	Publication location	Vol	Edition order	Edition label	Issue no	Issue date	Issue date as labeled	Issue present indicator	Page count published	Dup pages	Missing pages	Section label		Comments	QR	QR notes	Page mutilations / microfilming issues

Rebecca:

Copyright Criteria

- Public Domain
    - 1690-1922
    - 1923-1962, not registered with the Copyright Office, registered but not renewed after 28 years

- Orphan Works
    - Inability to locate the copyright holder after due diligence research

- Copyrighted with Permission
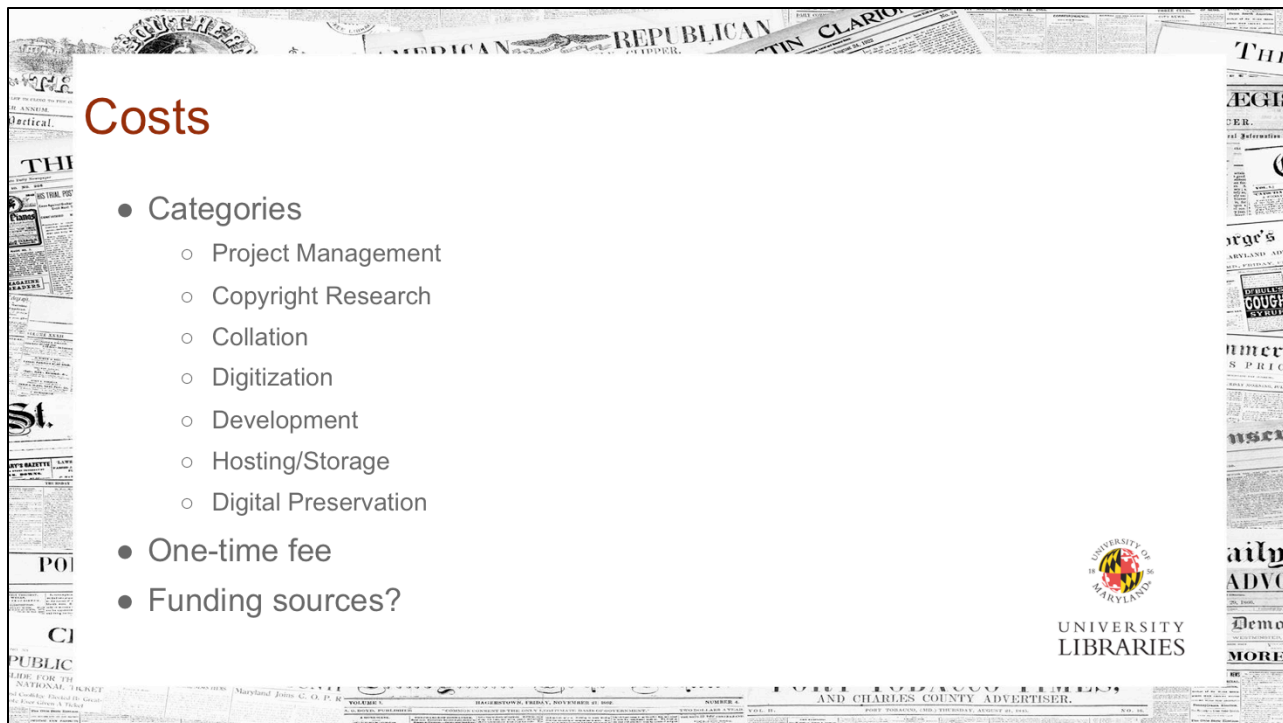   - Signed agreement between the copyright holder and UMD

UNIVERSITY LIBRARIES

Robin:
We want to expand beyond public domain papers, which would be published from 1690-1922 or published between 1923-1963 where the copyright was not registered with the Copyright Office, or was not renewed after 28 years. This research is easy enough to perform online, assisted by the digitized copies of the Catalog of Copyright Entries, held by the Library of Congress.
We also want to include orphan works newspapers, often small town papers whose publication were not bought by a larger entity when it ceased publication, and for which you cannot locate the copyright holder. Digitizing orphan works would require that due diligence research be performed and is documented. We would also allow papers under copyright with signed permission from the rights holder.
Finally, we are willing to partner with current newspapers and sign an agreement to act as their archive, host their copyrighted content, and make their newspapers freely available online.

Robin:
As we are still developing the UMD Student Newspaper Repository and are refining the ingest and archiving of files, we are still developing the future cost model for projects. We will break down the project into:
Project management costs, which will account for staff time to manage the project, liaise with the content holder and digitization vendor, and advise on project tasks such as fundraising or outreach.
Collation, if the holding institution does not wish to receive training and complete this project themselves. The metadata must conform to guidelines set by UMD.
Copyright research, to account for staff time if the holding institution does not wish to complete this themselves. The level of research must conform to guidelines set by UMD.
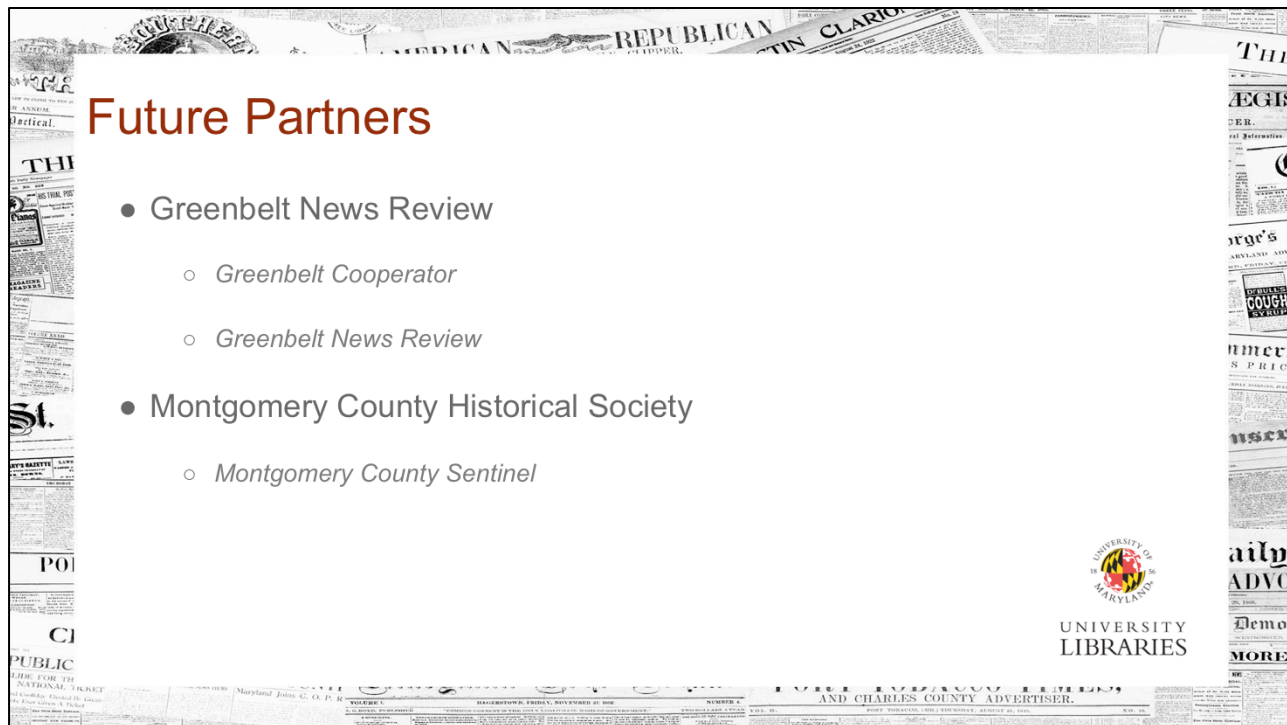We would outsource digitization and post-production to a vendor, but would manage the creation of the files necessary for inclusion in the repository.
Any custom development or scripting work that needs to be performed to ingest the files would be additional charges, which is why creating standard deliverables and putting them into the single repository is essential.
Finally, hosting and storage costs, and digital preservation costs, storing the files for the long-term must be integrated into the costs.
Our goal is to be able to create a process for collection and create a standard, one-time fee around that process, so that you could fundraise or receive grant funding for that amount.
While NEH funding may be drying up soon, we do want to consider state funding sources, such as local grants, local or private foundations, and state money, such as bonds. If you have any relationships with these organizations or agencies, please let us know.

Future Partners

- Greenbelt News Review
  - *Greenbelt Cooperator*
  - *Greenbelt News Review*
- Montgomery County Historical Society
  - *Montgomery County Sentinel*

UNIVERSITY LIBRARIES

Robin:
We want to close by acknowledging two future partners that will be key for getting the Maryland Digital Newspaper Repository off the ground--the Greenbelt News Review and the Montgomery County Historical Society--who approached us with an interest in digitizing newspapers. The Montgomery County Historical Society is interested in digitizing the Sentinel either through NDNP or separately through 1963. The Greenbelt News Review will be a particularly important partner for the beginning of the MD Digital Newspaper Repository because they are the publishers for their newspaper, and are interested in digitizing the entire run from 1937 to the present, and would like UMD to be the archive for their paper, making the issues freely available online. With their upcoming 80th anniversary, they plan to start fundraising to support this project, and are also interested in seeking grants.

Questions

Contact Us!

Robin C. Pike
Rebecca Wack
Historic Maryland Newspapers Project
University of Maryland, College Park

Email: lib-newspapers@umd.edu
Website: http://www.lib.umd.edu/digital/newspapers/home