

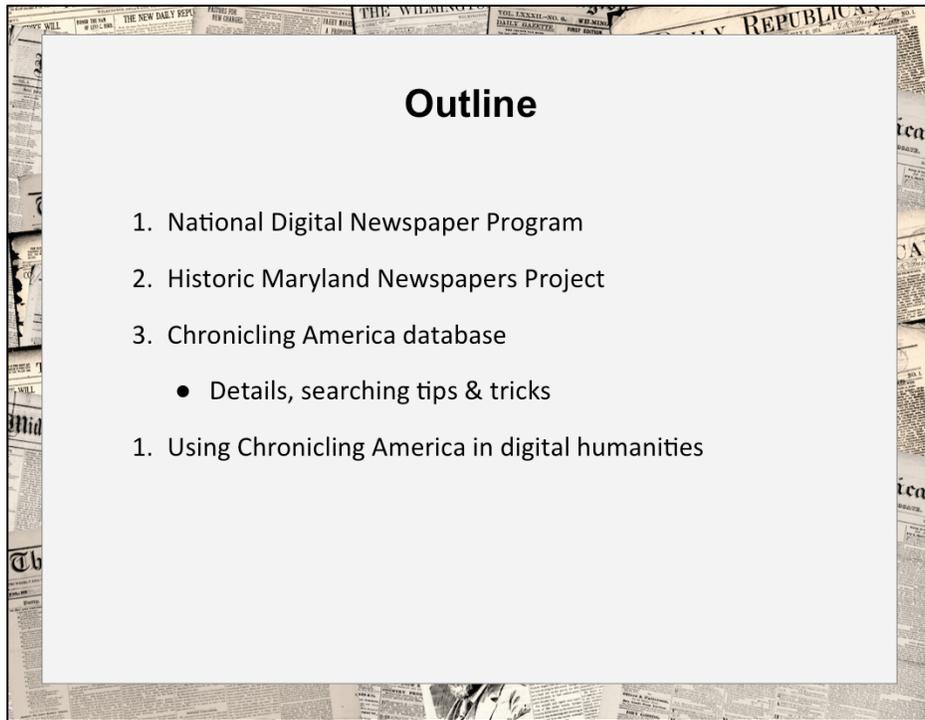
Maryland Newspapers In Chronicling America

The National Digital Newspaper Program &
Chronicling America

Robin C. Pike
Historic Maryland Newspapers Project
University of Maryland, College Park
rpik@umd.edu



UNIVERSITY
LIBRARIES



Today I am going to speak about the National Digital Newspaper Program or NDNP, the Historic Maryland Newspapers Project or HMNP--the Maryland contribution to the NDNP, the Chronicling America Database and how to use it, and some examples of projects with the goal of introducing you to the resources available on Chronicling America so that you can teach users how to use this resource in their research.

National Digital Newspaper Program (NDNP)

- Project timeline: 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

NATIONAL ENDOWMENT FOR THE

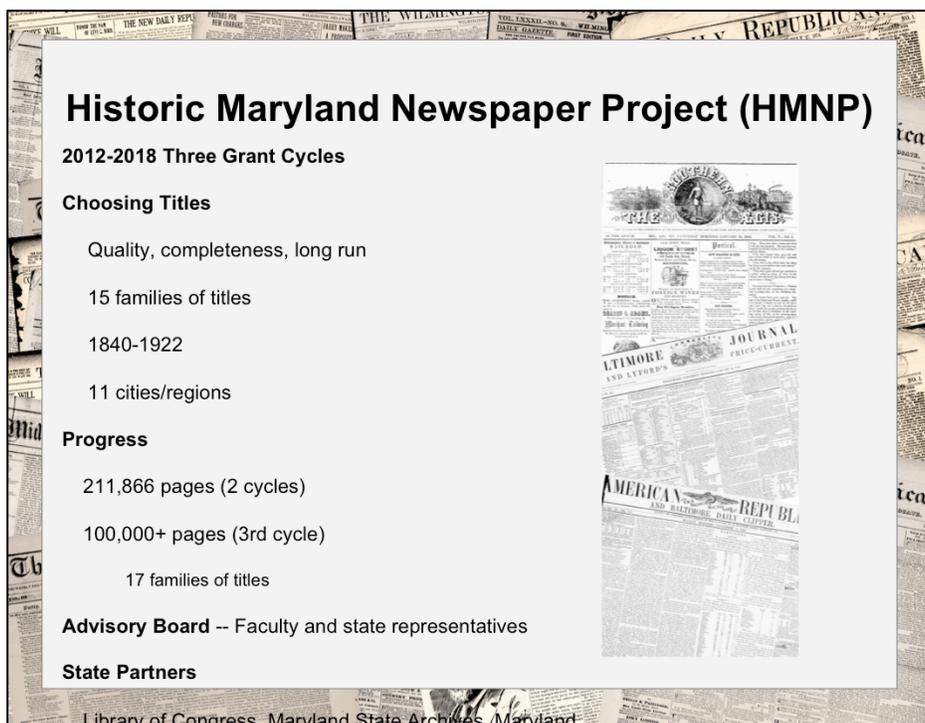
Humanities



LIBRARY OF CONGRESS

Established in 2005, “The 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.” 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 content contribution from all US states and territories. Partners 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. In addition to digitized microfilm, awardees are expected to provide 500-word essays about the history of a title, or a family of titles, and to provide a list of digitized newspapers that are freely available elsewhere.



Historic Maryland Newspaper Project (HMNP)

2012-2018 Three Grant Cycles

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

Advisory Board -- Faculty and state representatives

State Partners

Library of Congress, Maryland State Archives, Maryland

The Historic Maryland Newspapers Project (HMNP) began in 2012. 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. The project team and Advisory Board proposed Maryland newspaper titles for all three grant awards, 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. 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 because UMD does not hold any of the newspapers we are digitizing.

HMNP Resources/Promotion (1)

- Project website:
www.lib.umd.edu/digital/newspapers/home
- DDS Blog: <https://dssumd.wordpress.com/>
 (tags: historic maryland newspapers project)
- Wikipedia:
https://en.wikipedia.org/wiki/Wikipedia:Meetup/DC/Chronicling_America
- Gateway:
www.lib.umd.edu/newspapergateway

To receive more information on the HMNP, you can visit the project website or the Digital Systems and Stewardship division blog.

Under the direction of Liz Caringola, former project manager, content about Maryland newspapers has been added to Wikipedia. Liz also expanded the survey data to include titles that were available via paid resources as well, and with the assistance of SDDR and Josh Westgard, 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.

HMNP Resources/Promotion (2)

- Facebook: @HistoricMDNews
- Twitter: @HistoricMDNews
- Instagram: @HistoricMDNews
- Brochure

Rebecca Wack has created Facebook, Twitter, and Instagram accounts, using the additional text allotment, timely events, and trending hashtags to engage with new users, with a focus on engaging with students.

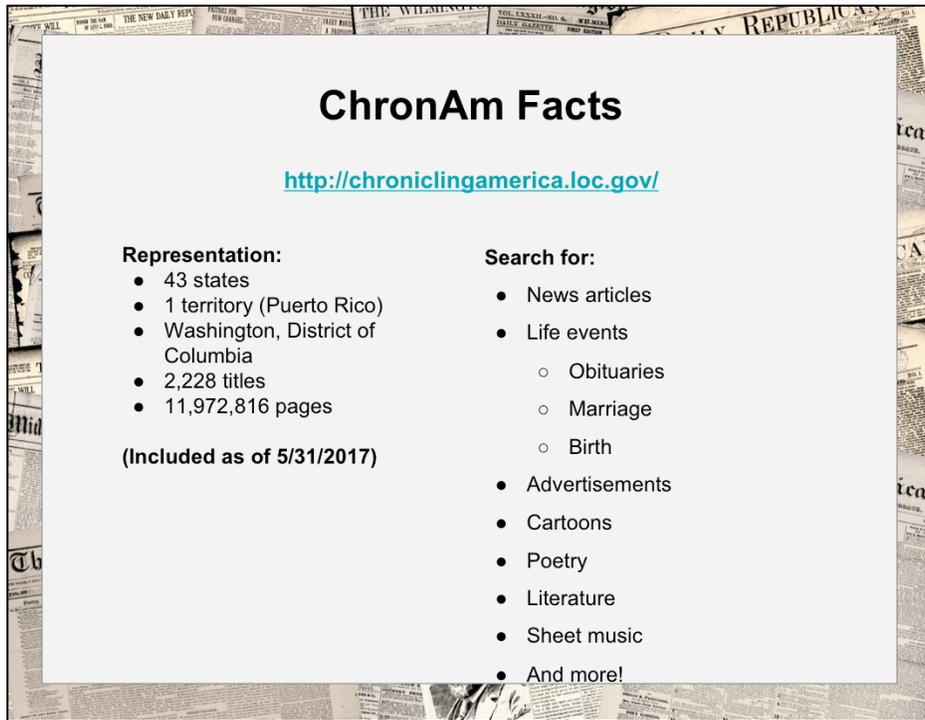
I also created a brochure intended to be an easy guide for genealogists, a target audience. The brochure is intended to be used as a handy guide for the Advisory Board members and other Chronicling America advocates to perform outreach at their institutions. This brochure is intended to be the first in a series of 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.

What is Chronicling America?

- Database that searches NDNP digital newspapers
- Freely-accessible, no login required

The screenshot shows the Chronicling America website interface. At the top, it features the Humanities logo and the title "CHRONICLING AMERICA: Historic American Newspapers". A search bar is present with filters for "All states" and "All Digitized Newspapers 1789-1924". Below the search bar, it indicates "Pages Available: 11,865,444". The main content area displays a "100 Years Ago Today, 5/2/1917 (171 ISSUES)" section with three newspaper thumbnails: "The Northwest Worker" from Everett, Wash.; "Perrysburg Journal" from Perrysburg, Wood Co., O. (Ohio); and "Bisbee daily review" from Bisbee, Ariz. A sidebar on the left contains navigation links such as "About Chronicling America", "About the Site and API", "Recommended Topics", and "Help".

The database, Chronicling America is developed by and permanently maintained at the Library of Congress. The corpus is word-searchable and browsable in multiple ways. An accompanying national newspaper directory of catalog and holdings information on the website directs users to newspaper titles available in all types of formats. Historical essays about the titles digitized are linked to each title page. ChronAm also links to NDNP Extras, which are projects and resources building on the newspaper data.



ChronAm Facts

<http://chroniclingamerica.loc.gov/>

Representation:

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

(Included as of 5/31/2017)

Search for:

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

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

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

The screenshot shows the ChronAm search interface. At the top, there are two callout boxes: 'Simple Search' (Most inclusive search) and 'Advanced Search' (Narrow down many results). Below these is the search interface with tabs for 'Search Pages', 'Advanced Search', 'All Digitized Newspapers 1789-1924', and 'US Newspaper Directory, 1690-Present'. The search bar contains 'All states' for location, 'from 1789 to 1924' for dates, and a text input field for search words. Below the search bar are two more callout boxes: 'All Digitized Newspapers 1789-1924' (Search within a specific newspaper title) and 'US Newspaper Directory 1690-present' (Part of the original US Newspaper Project; Search all newspapers (digitized or not)).

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)

The search interface offers multiple options for simple and advanced searches, and links to the directory of all cataloged newspaper titles.

Searching in ChronAm

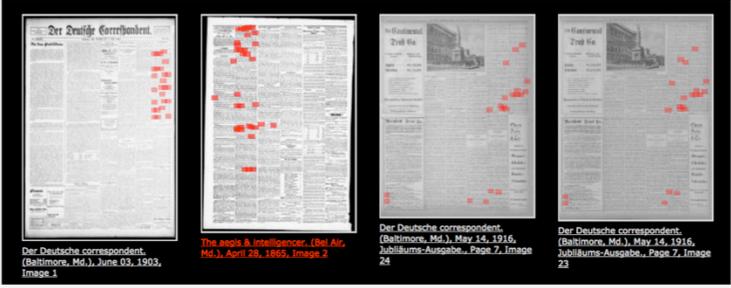
Search terms: "John Wilkes Booth"

- Highlighted in search results

Results 1 - 20 of 111 | 1 | 2 | 3 | 4 | 5 | 6 | Jump to page: GO View: Gallery List

111 results containing "John Wilkes Booth"

Show only front pages Sort by: Relevance Results per page: 20 5



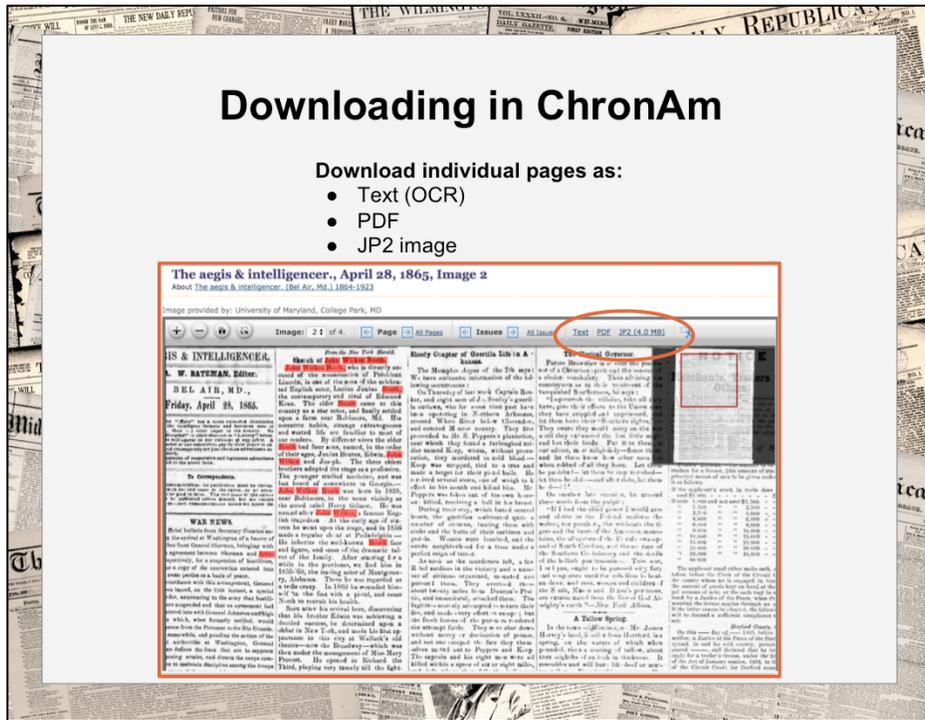
The screenshot displays four newspaper images with red highlights indicating search results. The first image is from 'Der Deutsche Correspondent' (Baltimore, Md.), June 03, 1903, Image 1. The second image is from 'The seas & intelligence. (Bei Air, Md.), April 26, 1865, Image 2. The third image is from 'Der Deutsche correspondent. (Baltimore, Md.), May 14, 1916, Jubilums-Ausgabe, Page 7, Image 24. The fourth image is from 'Der Deutsche correspondent. (Baltimore, Md.), May 14, 1916, Jubilums-Ausgabe, Page 7, Image 23.

After you perform a search, the keywords will highlight, leading you to successful results.

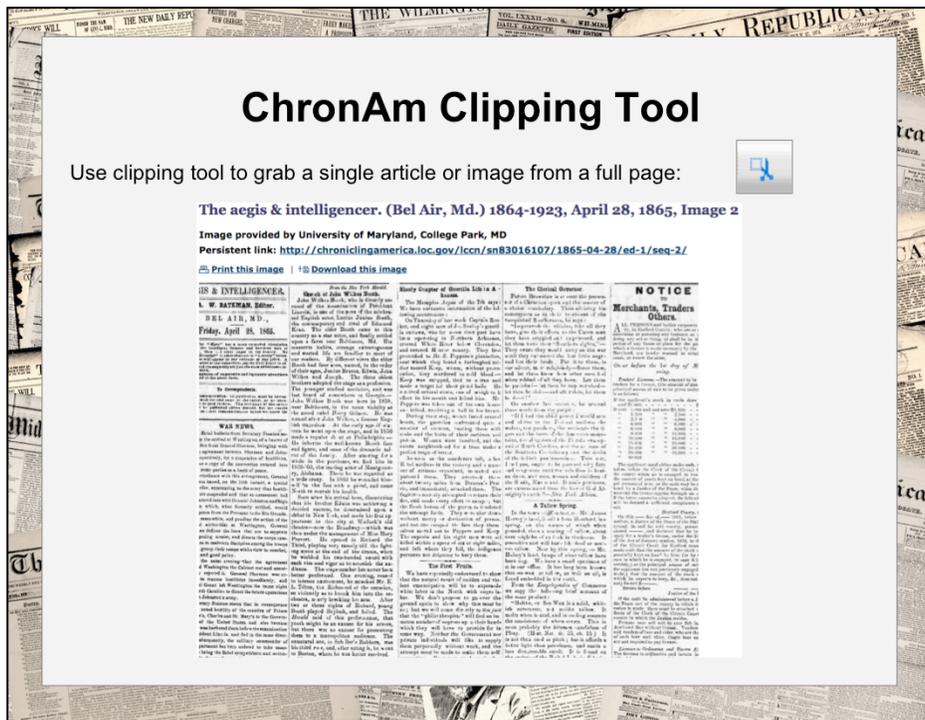
Downloading in ChronAm

Download individual pages as:

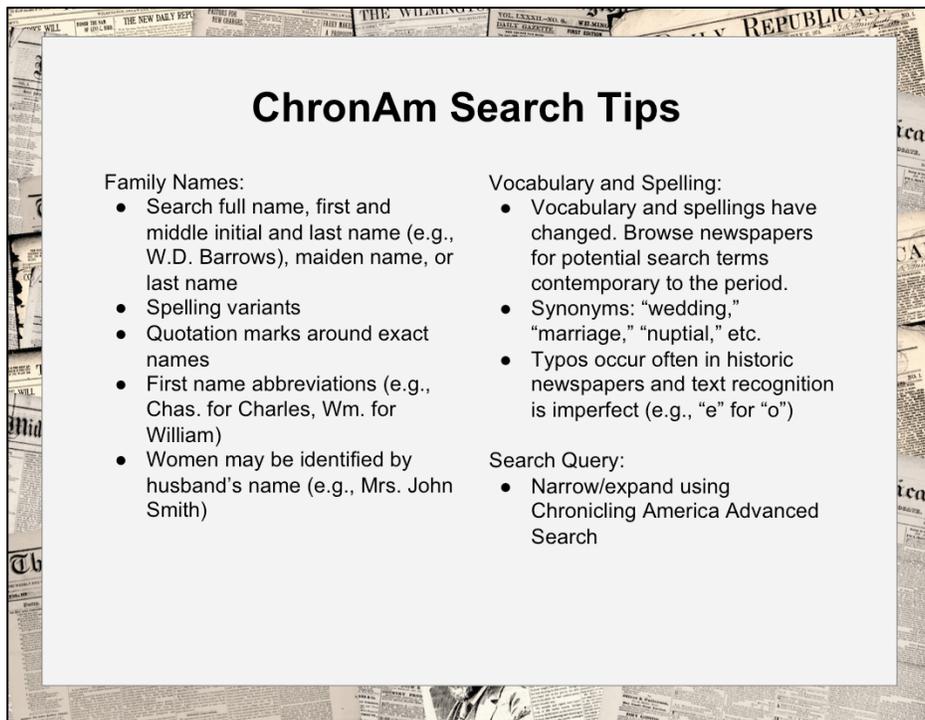
- Text (OCR)
- PDF
- JP2 image



After you click on a 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.



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 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")

Search Query:

- Narrow/expand using Chronicling America Advanced Search

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.

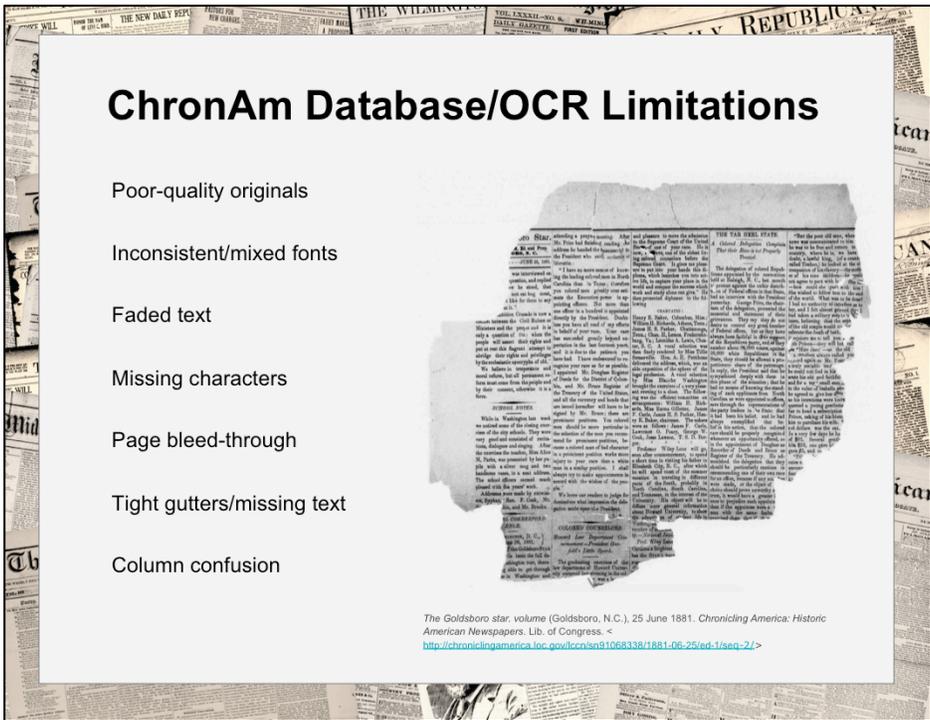
The vocabulary or spellings for words have changed through time, so if search results are not yielding results, browse the newspapers around the period in which you want to find information, and then use synonyms for words.

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.

Finally, you can narrow or expand results using Chronicling America Advanced Search, including dates, newspaper title, "with any of the words," "all of the words," "with the phrase," and "with the words within x number of spaces."

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, *Chronicle America: Historic American Newspapers*. Lib. of Congress. <
<http://chroniclingamerica.loc.gov/lccn/sn91068338/1881-06-25/ed-1/seq-2/>>

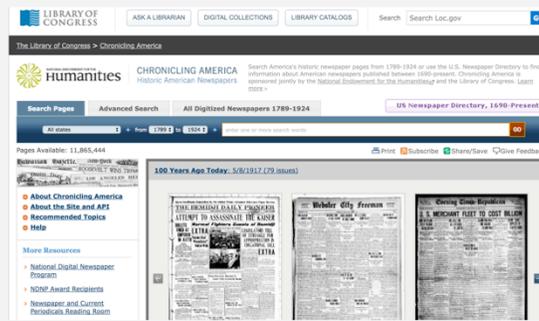
Chron Am has some limitations, caused by newspaper incompleteness or less than ideal quality issues, such as inconsistent or mixed fonts, faded text, missing characters, bleed through from the reverse side of the page, or tight gutters if the newspaper is bound, or column confusion caused by an error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

ChronAm vs. Other Databases

Only undigitized titles

Public domain restriction

More data capabilities



Though ChronAm 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 ChronAm, 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.

Using ChronAm in Projects

- About the Site and API:
 - <http://chroniclingamerica.loc.gov/about/api/>
 - Bulk data download
- NEH Data Challenge
 - America's Public Bible
 - <http://americaspublicbible.org/>
- UMD Campus
 - Social Justice Day
 - Classroom



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.

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."

In addition to promoting the project via social media, and performing outreach to genealogical and research communities, the UMD project team is finding ways to demonstrate the resource as a research tool on campus. On April 18, 2017, Rebecca Wack and her student assistant participated in the on-campus Social Justice Day, focusing on the Media and Integrity poster and discussion roundtable. This roundtable led to Rebecca working with an English faculty member to integrate Chronicling America into her classroom. We hope this will serve as a