Overview of the Session

- Do you have the only run?
- What titles are others working on so I don’t duplicate work?
- Funds for the Project?
- What about copyright? What can I do without a legal challenge?
- How does paper differ from microfilm if I want to digitize the paper?
- Should I microfilm then digitize or digitize directly from the paper copy?
- What’s involved with requesting a quote for digitization?
- What digitization standards should I follow?
- Can’t I simply take camera shots of the pages or scan the pages on a copier myself?
- What do I do with the file once the company completes the digitization?
- Questions

Do you have the only copy of the newspaper title in question?

- Title search in *Chronicling America* or *OCLC WorldCat*
- Check with your local historical society and the State Library
What titles are others working on so I don’t duplicate work?

- Check Chronicling America for digital files (https://chroniclingamerica.loc.gov/newspapers/)
- State Library's List of Digitized PA Newspapers (https://www.statelibrary.pa.gov/GeneralPublic/Learn/Pages/Newspapers.aspx)
- Pennsylvania Newspaper Archive (hosted at the Penn State University Libraries) (http://panewsarchive.psu.edu/)
- Email to Bill Fee at the State Library (wfee@pa.gov)

How do I find Funds for the Projects?

- Local Historical Society or Donor
- Local Charity Grant
- State Library’s funds from LSTA
  - Must be willing to work with the State Library
  - Complete the documents needed for a quote
  - Upload the file to a public repository such as Pennsylvania Newspaper Archive hosted at the Penn State University Libraries

What about copyright? What can I do without a legal challenge?

Public Domain Newspapers (U.S.)

Paper
- Pre-1923 (soon to be 1924!)
- 1923 - 1977 without a notice
- 1923 - 1963 with a notice but not renewed

Microfilm
- Who created the microfilm? Check the film target.
- Who holds the copyright and master negative?

Questions can be addressed to https://libraries.psu.edu/services/scholarly-publishing-services/contact-copyright-publishing-and-open-access
How does paper differ from microfilm if I want to digitize the paper?

Scanning directly from the paper will retain the “real” look and feel of the original, however it may be more expensive depending on the condition of the paper and its binding, if bound.

Scanning from microfilm, while a more cost efficient process, will produce only black and white images from the original film. Poorly filmed papers will produce poor scanned images and can affect the accuracy in searching.

Should I microfilm then digitize or digitize directly from the paper copy?

Both are acceptable practices, however, ask yourself:

• What is the “scannability” of the original paper? Are the pages intact and can be easily hand-turned?

• If bound, is there a good interior margin for scanning the entire page without distortion? If not, can the volume be disbound to achieve single pages for scanning?

• Scanning from microfilm? Can you locate the second-generation negative for scanning?

What’s involved with requesting a quote for digitization?

**Paper or Microfilm?**

**Paper**

• Count pages per issue
• Count pages per volume; note missing pages/issues
• Measure size of pages

**Microfilm**

• Location of the master or print negative
• Paper filmed 1-image per frame (1Up) or 2-images per frame (2 Up)?
  • Generally 400 frames per reel (2 Up) or 1600 images on a full reel
• Pages per issue?
• Paper have sections?

Title level metadata (title, publication location, volume and numbering of start and end date, LCCN)

Companies working in the area
What digitization standards should I follow?

The Library of Congress’ National Digital Newspaper Program technical standard is the de facto standard for all newspaper digitization. Why?

- Proven track record for accessibility over the web.
- Page images produced with this standard support effective OCR representation.
- Produces usable digital formats with high probability of sustainability into the future.

What is the NDNP Technical Standard?

Summary of the standard:

- Produce grayscale images (scanned for maximum resolution possible between 300-400 dpi, relative to the original material) primarily from microfilm.
- Create OCR with word-bounding boxes, with recognition of columns, but without segmentation of pages into articles.
- Capture structural metadata for pages, issues, editions, and titles to support a chronologically based browsing interface.

Can’t I simply take camera shots of the pages or scan the pages on a copier myself?

- Images versus OCR text
- Size of pages versus scanning equipment
- Segmenting the files into dates and/or articles
- Where would you house this file and who could access it?
What do I do with the digital images once the digitization is complete?

- Library of Congress -- Chronicling America
  (https://chroniclingamerica.loc.gov/newspapers/)

- PA Photos and Documents
  (https://www.powerlibrary.org)

- Pennsylvania Newspaper Archive
  (http://panewsarchive.psu.edu)

- Your Institutional Repository

- Other? Vendor hosted Digital Reader

---

**Pennsylvania Newspaper Archive**

http://panewsarchive.psu.edu/

- Launched in March 2017; hosted at the Penn State University Libraries
- Utilizes the Open Online Newspaper Initiative, OPEN ONI, a Library of Congress open source software
- This is a freely accessible web resource with the goal to provide access to the Commonwealth's historical newspapers as a single portal

---

**Pennsylvania Newspaper Archive**

http://panewsarchive.psu.edu

A look under the hood of this unique Pennsylvania Resource

More than just Civil War titles!

All PA titles from Chronicling America and many more - 180 titles and growing!

The planned future home of all digitized Pennsylvania titles!
Pennsylvania Newspaper Archive

Features

- Look and feel as LC's Chronicling America Newspaper database
- Browse by title
- Browse by date (single title)
- Browse all issues by date
- Page-level segmentation
- Keyword searching and advanced searching
- Highlighted search results
- Different file format displays available including text

Planned Features:

- Website upgrade with new content and contact information
- Clickable map
- How to digitize and add newspapers; digitization specs; how to implement your project; preservation and storage of originals, how to deliver files to the Penn State University Libraries
- Use and copyright of content; how to cite
- Other PA newspaper websites with online content
- Acknowledgment of partners contributing content or funding

How do I add a paper to the Pennsylvania Newspaper Archive?

- Contact the University Libraries
- Confirm title and holdings have not already been loaded
- Create or convert TIFF files to the “Penn State specs”
- Scans from microfilm or direct scans from the original paper are acceptable
- Can add non-English titles
- Owner assumes responsibility for the digital preservation of the files