LIBRARY OF MICHIGAN

Volume 1, Issue 2 August 2025

Guide to Digitizing Newspaper

INTRODUCTION TO NEWSPAPER DIGITIZATION

Unknown author, CC BY-SA 3.0, via Wikimedia Commons

Paper Problems

Newspapers can be found in almost every public library. While library patrons use newspapers to learn about current events, both local and abroad, historic newspapers also have value to researchers, whether they are exploring something specific, like their genealogy, or broader, like trends over time.

Unfortunately, newspapers are fragile, printed on paper meant to last only as long as the news was relevant. This short lifespan causes an issue for researchers who are unable to access deteriorating newsprint. To help preserve content over time newspapers were often bound into volumes and, starting in the 1930s, were put on microfilm, which also had the benefit of condensed storage.

A New Barrier

Currently, libraries are faced with a new barrier to providing access to newspapers: microfilm machines are becoming harder to repair and replace, and patrons are more comfortable navigating the digital world than handling a microfilm reader. This change has led to an increased interest in and need for newspaper digitization at public libraries.

There are many factors to consider before starting a newspaper digitization project, from copyright to digital preservation. This guide will walk librarians through the process of deciding whether and how to digitize their newspapers utilizing many Michigan-specific resources.

INDEX

LIBRARY OF MICHIGAN

Volume 1, Issue 2 August 2025

Guide to Digitizing Newspaper

PREPARATION

Inventory & Selection

The first step in any digitization project is to know what you have. If you plan to digitize your newspapers, not only will you need to provide information to a vendor in order to get an accurate cost estimate for the project, but also for your own assessment of a newspaper's copyright status and therefore your ability to digitize and provide access. You can start with a simple spreadsheet listing the title and other relevant information, such as format, page count, condition, and publication date. You will also want to include other important information you may have, such as whether this was a daily or weekly publication, or if there are any misprints or missing issues. You can use this **Newspaper Inventory Template** to identify what newspapers you have and record relevant information. Once you know what you have, you can move on to determining whether what you have should be digitized.

As with all digitization projects, not everything should, or even could, be digitized. Before moving ahead, you can assess your newspaper collection for digitization suitability. The **Library of Congress** recommends determining whether a newspaper title should be selected for microfilming and digitization by assessing its research value, physical condition, and completeness.

Research Value – Does the paper have historical significance? Is the content consistently requested by researchers for access? Is this content available elsewhere in alternate editions? Digitizing newspapers is time and cost intensive, so make every effort to be sure that there is sufficient demand for the final product.

Physical Condition – Is the paper stable enough to be digitized without damaging the original? Is the quality of the microfilm sufficient to create a readable digital surrogate? Digitization cannot improve a bad microfilming job or replace torn or weak newsprint. Digital images should be taken from the best quality version available.

Completeness – Is the content complete? Are there several missing issues? Due to the nature of newspapers, anything less than reasonably complete makes for unsatisfactory use. Are you able to collaborate with other organizations to combine material to complete a run, if necessary?

If a newspaper title has a significant research value, has sufficient quality, and is relatively complete, it is a good candidate for digitizing! It's time to move onto the next step!

Guide to Digitizing Newspaper

PREPARATION

Check Availability

Digitization requires many resources, such as time, money, and expertise, so before you spend your limited resources on digitizing newspapers, confirm that the newspaper is not already digitized and available online. The Clarke Historical Library at Central Michigan University developed the **Digital Michigan Newspaper Portal**, a comprehensive list of digitized versions of Michigan newspapers by county, to help with research. Once you have selected a newspaper title for digitization, check to see if the newspaper is already available digitally by searching for it via the portal, the Library of Congress' **Chronicling America**, and a Google search. If it has already been digitized by another institution, there is no need to digitize the newspaper in your collection, simply point researchers to the digital version. If the newspaper is behind a paywall, you may be able to access it through one of the many online databases that are freely available to **Library of Michigan cardholders**.

If the newspaper has been partially digitized, this may be an opportunity to collaborate with another organization to complete the full run of or fill gaps in the title. Digitization is a naturally collaborative process, so this is a great opportunity to start the conversation and identify ways to work together.

If the newspaper does not appear to be available digitally already, check the **Newspaper Family Histories** page to confirm that the paper was not published under a different title at another point in time. If the paper has changed names, perform the same search for the alternate name(s). If you confirm that the newspaper has not been digitized at all, proceed to assessing whether you have the right to digitize the paper.

"Counties in Michigan" map from Digital Michigan Newspaper Portal

LIBRARY OF MICHIGAN

Volume 1, Issue 2 August 2025

Guide to Digitizing Newspaper

PREPARATION

Copyright Assessment

Before you digitize an item, you should assess its copyright status. **What is copyright?** "Copyright is a type of intellectual property that protects original works of authorship as soon as an author fixes the work in a tangible form of expression." While copyright does provide rights to the author or copyright holder regarding how their works can be used, their rights are not indefinite. The length of copyright protection will depend on several factors, including when it was created and whether it was published, registered with the copyright office, or if that registration was renewed. Once copyright has expired or a work has been placed in the public domain, works can be used freely. Currently, anything published before 1930 is in the public domain, but for items published in 1930 going forward copyright status is a bit harder to determine.

Works published in the United States are covered for 95 years after the original publication date. Works enter the public domain on January 1 of every year.

It is prudent to perform a copyright assessment of any material that you are planning to digitize, especially if you are not the copyright owner.

To perform a copyright assessment on newspapers published in the United States, use a tool, like the **Rights Review flowchart** or the **PA Digital Rights Statement Selection Tool**, to determine a work's copyright status. Be sure to document your findings on the **Newspaper Inventory Template** as you determine the copyright status of serial issues.

Works Published Pre-1978

Starting with either the year the paper was published or the last year the paper could have been in the public domain, search the **Catalog of Copyright Entries (1906-1977)** to identify whether the title was either registered with the copyright office and/or the copyright registration was renewed.

For pre-1978 works, if the work was published:

- **before 1930** the work is in the public domain
- **between 1930-1977** without a copyright notice the work is in the public domain
- **between 1930-1963** with a copyright notice but copyright was not renewed the work is in the public domain
- **between 1930-1963** with a copyright notice and the copyright was renewed the work is covered for 95 years after the original publication date
- **between 1964-1977** with a copyright notice the work is covered for 95 years after publication date

Physical property rights and intellectual property rights are not the same thing, and you must ensure that you are legally authorized to digitize a work.

Guide to Digitizing Newspaper

PREPARATION

Works Published in 1978 and beyond

Starting with the year the paper was published, search the **Public Catalog of Copyright Entries (1978-present)** to identify whether the title was published with a copyright notice.

For works published in 1978 and beyond, if the work was published:

- **between 1978-March 1, 1989** and created after 1977 and published with notice the work is covered for 70 years after the death of author, or 95 years from date of publication for works with corporate authorship, or 120 years from creation, whichever expires first
- **between 1978-March 1, 1989** and created before 1978 and first published with notice in the specified period the work is covered for 70 years after the death of author, or 95 years from date of publication for works with corporate authorship, or 120 years from creation, whichever expires last or December 31, 2047
- **between March 1, 1989-2002** and created after 1977 the work is covered for 70 years after the death of author, or 95 years from date of publication for works with corporate authorship, or 120 years from creation, whichever expires first
- **between March 1, 1989-2002** and created before 1978 and first published in this period the work is covered for 70 years after the death of author, or 95 years from date of publication for works with corporate authorship, or 120 years from creation, whichever expires last or December 31, 2047
- **after 2002** the work is covered for 70 years after the death of author. If a work of corporate authorship or 95 years from date of publication for works with corporate authorship, or 120 years from creation, whichever expires first

As you perform the copyright assessment, be sure to document the publication date, copyright status, copyright expiration year, copyright owner, and owner permission or attempted contacts. If there is ever litigation surrounding copyright infringement of the work this documentation can be used to show the intent to be a “good actor” in respect to intellectual property rights.

Many papers contain syndicated or separately copyrighted content, such as advertisements, serialized fiction, and comic strips, which may have different copyright protections than the rest of the paper. Be sure to assess the paper and that no more than 50% of the content is separately copyrighted content.

Guide to Digitizing Newspaper

PREPARATION

Copyright Assessment

If a newspaper title is in the public domain, you are free to digitize and provide access to it. If a newspaper title is still protected under copyright, you have multiple paths forward. First, you can attempt to contact the publisher of the newspaper or copyright holder and request permission to digitize and provide access to the newspapers in your collection. If you do receive permission from the copyright holder, create and maintain documentation of their permission to digitize and provide access to your collection with your administrative records.

If the newspaper is still being published and you either have or can develop a relationship with the publisher, you may be able to request digital versions of the newspaper directly from the publisher. Most newspapers are now born-digital, meaning that they were created on a computer and sent directly to print or hosting online. By working directly with the publisher, you can circumvent the costly process of digitization and request a PDF version to be shared directly via email, File Transfer Protocol (FTP), or external storage media (DVD, hard drive, flash drive, etc.).

Three Categories of Rights Statements from RightsStatements.org:
In Copyright, No Copyright, Other

Identifying and contacting the publisher may be difficult as several publishers of local newspapers are now defunct or have been acquired by large conglomerates. If you are unable to identify or contact the copyright holder to request permission to digitize and provide access to the newspaper, you may still decide to digitize and make them available under **Fair Use**. See the resource list at the end of this document for tools to determine whether a Fair Use exemption is appropriate for the work being reviewed.

Once you have determined an item's copyright status, consider using Standardized Rights Statements from RightsStatements.org to clearly communicate the copyright and re-use status of digital objects to the public.

Once you have performed your copyright assessment and determined whether and by whom copyright is held for the work, you can determine whether to move forward. Consult stakeholders and legal counsel if the work is not clearly in the public domain and/or you are not the rights holder of the work to be digitized to determine your level of risk. There will always be some risk involved in digitization, it is up to you and your organization to determine the level of risk you are comfortable with.

Guide to Digitizing Newspaper

PRODUCTION

Determining Format

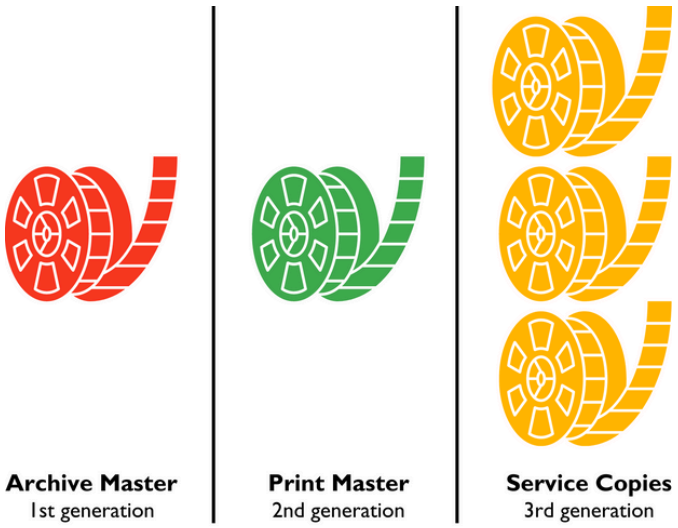
Now that you have determined that you are able to digitize the newspaper, either due to its copyright status, receiving permission from the copyright holder, or have determined that you will use a Fair Use argument, it is time to move forward with planning. Depending on what formats you have, either microfilm or print, you can decide how to move forward.

If you are working with microfilm, assess its condition. The condition of the microfilm can have a huge impact on the digitized version. If possible, scan from the second generation of microfilm, often called the print master or print negative. Negative film generally has less “noise” than positive prints, such as the third generation (service copy), which are used for patron access. Due to the additional noise and frequent use, which can cause scratches, third generation film would produce poor scans and have a lower success rate in optical character recognition (OCR).

Poorly microfilmed newspaper may not be able to be digitized successfully. If film is out of focus or has other issues you may not be able to capture content through digitization. Remember: garbage in, garbage out! A vendor should be able to assess the microfilm to determine whether it is suitable for digitization.

If you only have paper or your microfilm is a service copy in poor condition, explore whether another organization or the original microfilming company has a master copy of the microfilm that you could borrow or duplicate to use in your digitization project.

Microfilm Generations



Adapted from “Microfilm Selection for Digitization,” OCLC Online Computer Center Library

If you only have access to the paper, decide whether you will microfilm and digitize or digitize only. Some vendors can only work from microfilmed versions and would need to microfilm the print version first and then digitize from microfilmed versions, while others can digitize directly from the newspaper. Most vendors would prefer to digitize directly from microfilm, but it is a discussion to have with your vendor about their capabilities.

If the vendor can digitize directly from newspaper, you will need to decide wither to digitize only, or to microfilm and digitize. Microfilm is still considered the preservation format for newspaper content, so it is still beneficial to microfilm paper.

LIBRARY OF MICHIGAN
Volume 1, Issue 2 August 2025

Guide to Digitizing Newspaper

PRODUCTION

Finding a Vendor

Newspaper microfilming and digitization is a challenging process that is most often contracted out to a vendor. It requires specialized equipment and expertise to not only digitize, but also create appropriate metadata for management and discovery. If you plan to digitize in-house (using your own staff and equipment), consider whether you have the resources (staff time and expertise, equipment, etc.) to digitize to standards and/or best practices.

When looking for a **vendor** there are several resources available to help you identify potential digitization partners. There are vendors located in Michigan that microfilm and digitize newspapers, as well as many vendors nationally. Consider asking colleagues for vendor recommendations or explore some of the digitization vendor resources listed in the appendix of this guide.

Once you have identified potential digitization vendors, be sure to check references and, if possible, do a test batch to confirm specifications, quality, and communication. Most vendors will provide a sample scan at little or no cost. By doing a sample scan both you and the vendor can have a more accurate idea of what the project will entail, including the quality and orientation of the original microfilming. Some newspapers have multiple pages microfilmed in the same frame, which may have an impact on the overall workflow and cost of a digitization project.

Considerations:

- Will you microfilm and digitize paper? Or digitize paper only? Or digitize microfilm? What are the cost differences and implications?
- Is the cost per frame or per page? Is there a cost to split images into individual pages?
- What type(s) of files will you be receiving? Do they meet preservation standards?
- What does the cost include? Metadata? OCR? Deskewing or other adjustments?
- What level of quality control will the vendor perform?
- What is the procedure if files fail to meet quality standards?
- Will your vendor host the files or is something you will need to manage?
- Will the newspaper be recycled upon completion of the project or returned?
- Will the vendor maintain copies of the files?
- Are file management costs one-time or ongoing?

Many of the answers to these questions will depend on your individual circumstances, resources, and needs, as well as whatever your vendor's capabilities. Whatever you and the vendor decide, be sure to document it in a contract or agreement to protect both your interests.

Make sure you get a copy of the preservation master files, even if the vendor will be hosting your files!

Guide to Digitizing Newspaper

PRODUCTION

File Specifications

One of the benefits of using a vendor is that you are able to hold them to higher standards than if you were digitizing material in-house, so choose file formats wisely to ensure long-term access and ability to digitally preserve the content. There is no sense in spending valuable time and financial resources to digitize something that you will not be able to access in the future. Below are the recommended technical specifications for digitizing newspapers based on the **NDNP Technical Guidelines & Specifications**.

File Format Specifications

- TIFF 6.0, uncompressed, file per page
- If newspapers were microfilmed two pages per frame, split into two separate page files
- 8-bit grayscale (if digitizing from microfilm or black-and-white originals)
- 24-bit color, if digitizing from color originals
- Avoid 1-bit bitonal, which can reduce the accuracy of OCR
- Maximum resolution possible, between 300 to 400 dpi (relative to the size of original page and considering the size of lettering)
- Crop to the page edge, not to the edge of the text, with up to ¼" beyond the page
- Deskew, if skew is greater than 3°
- Optional: Multi-page searchable PDF per issue

File Naming Advice

- Make sure that your filenames are **consistent** and **identifiable** across your digitization project.
- Use a standard identifier, such as an **OCLC number** or **LCCN**, to identify the newspaper.
- Record the issue date in **YYYYMMDD format**
- Include the page sequence number as “001”, “002”, “003” and so on.

Example:
Issue folder for the January 2, 1941, issue of the Thompsonville Press (OCLC number 27354139):
o27354139_19410102

Page 1 of the January 2, 1941, issue of the Thompsonville Press:
o27354139_19410102_001.tif

Guide to Digitizing Newspaper

PRODUCTION

Optical Character Recognition

Optical character recognition (OCR) is the process of converting a static image into machine-readable text. It is one of the biggest advantages of digitizing your newspapers and makes them more accessible to researchers. Because OCR is a technology with some limitations, there are a number of factors that will influence OCR quality:

- Stains, tears, faded text, or bleed-through on the original newspapers
- Damaged or deteriorating microfilm
- Lighting issues during microfilming
- Complex newspaper layouts
- Fonts of different styles and sizes
- Page curvature
- Technical specifications used for digitization

To ensure that you are getting the best OCR results possible, make sure that you follow the recommended technical specifications and digitize from the best quality microfilm or originals that you have access to.

Funding Digitization

Newspaper digitization can be very expensive, with costs between $0.10-$0.80 per page, depending on desired outputs. When considering years of daily newspaper content, the costs can add up quickly!

Luckily, you don't have to fund a digitization project all on your own. The CMU **Digital Michigan Newspaper Grant program** funds digitization of Michigan newspapers every year, whether in print format or on microfilm.

Another alternative is working with a content subscription provider. In this model, content subscription vendors digitize newspapers at no cost to a contributing library, but access to digitized content is available only to the library within the library building and through the content provider's website behind a paywall. This embargo period (usually lasting 3-5 years) allows the content subscription vendor to recuperate their costs before the library is able to provide free and open access to the content.

Although newspapers should be digitized in complete runs ideally, it is not something that needs to be completed all at once. Many libraries have ongoing contracts with digitization vendors to digitize a certain amount of material annually in order to stay within their budget. If a newspaper was relatively recently microfilmed and kept in **suitable storage conditions**, microfilm can last well into the future, giving you time to undertake the necessary research and preparation for digitizing, providing access to, and digitally preserving your historic newspapers.

Guide to Digitizing Newspaper

PROVIDING ACCESS

All About Access

The goal of any digitization project is to enable and improve access to material, so be sure to include a plan for providing access to your digitized newspapers.

If you use a vendor to digitize your newspaper you can often have them provide access to the digitized content. If your vendor hosts your files, ensure that you receive a copy of the files and can access them outside of the vendor's platform (that they are not in a proprietary format or require a vendor's proprietary software to access).

If your vendor does not provide hosting services, you may explore hosting on a variety of platforms. If you choose a vendor to host your files, talk to them ahead of time to confirm the type of files they require and any associated metadata.

Digital historical newspapers and magazines from the Grand Rapids History Center

Takedown Policies

Takedown requests are inevitable with any form of digital content, so having a **takedown policy** in place will help prepare you to handle one if and when it arises. Takedown policies should include:

- What the request relates to (preferably with a content identifier or URL)
- Who to contact and how (provide an email address or other method of contact)
- What the next steps are (estimated response time and what the results could be)

D-CRAFT (Digital Content Reuse Assessment Framework Toolkit) provides a takedown policy template and useful examples of policies related to digital collections topics, including Takedown Policies for both Copyright and Sensitive Information, and can be used as a reference for creating your own policy. As always, also have an internal policy for handling any takedown requests received.

Guide to Digitizing Newspaper

PRESERVATION

More Than a Backup

Whether you host files locally or use a vendor, it is essential to digitally preserve your files. Always request a copy of your files and save them in at least three places. A good practice is following the **3-2-1 Rule**, meaning you have three copies of your files on two different storage media, one of which is offsite.

Have at least three copies of your files

stored on two different media types

with one copy stored offsite

Although it is easy to think of digital files as being effortlessly preserved, digital bits can be even more fragile than their analog counterparts. This is why it is important to take steps to protect your investment. You will want to generate **checksums** for your files and validate them at regular intervals. Checksums are like digital fingerprints that let you know if there have been any changes to your files over time. If a checksum fails, replace the corrupt file with a valid file from one of your other copies.

The **Michigan Digital Preservation Network** (MDPN) is a great resource for Michigan cultural memory organizations to digitally preserve their newspapers over time.

Conclusion

Digitizing newspapers requires a significant investment, not only financially but in time and energy. Pick titles that will benefit from digital access and don't duplicate efforts. If something is already digitized, don't re-digitize so you can have a copy at your library. Assess a paper's condition and digitize from the best, most complete version. Be sure to back up your files and ensure their validity over time.

Guide to Digitizing Newspaper

RESOURCES

Copyright Resources

- U.S. Copyright Office
- How to Investigate the Copyright Status of a Work
- Cornell University Library: Copyright Term and the Public Domain
- Determining copyright status of serial issues CC BY 4.0 John Mark Ockerbloom
- The Online Books Page FAQ, especially "How Can I Tell Whether a Book Can Go Online?" and "How Can I Tell Whether a Copyright Was Renewed?"
- Peter Hirtle's Copyright and Cultural Institutions: Guidelines for Digitization for U.S. Libraries, Archives, and Museums, especially the Safe Digitization Workflow (page 203).
- Fair Use and Copyright Research For Newspaper Digitization: What You Need To Know
- The Online Books Page: Copyright Registration and Renewal Records
- U.S. Copyright Office Public Records System
- U.S. Copyright Office Fair Use Index
- Cornell University Library, Checklist For Conducting A Fair Use Analysis Before Using Copyrighted Materials
- Standardized Rights Statements for Cultural Heritage Institutions: RightsStatements.org
- Minnesota Digital Library: Rights Review
- PA Digital Rights Statement Selection Tool

Michigan-Specific Resources

- Digital Michigan Newspaper Grant Program
- Digital Michigan Newspaper Portal
- Library of Michigan Digitization Page
- Michigan Newspaper Family Histories
- Michigan Digital Preservation Network
- Newspaper Inventory Template
- Newspapers on Microfilm at LM
- Online Resources for Library of Michigan Cardholders

Guide to Digitizing Newspaper

RESOURCES

Digital Preservation Resources

- Digitization Vendor Resources
- FADGI Guidelines: Technical Guidelines for Digitizing Cultural Heritage Materials
- Michigan Digital Preservation Network
- National Digital Newspaper Program
- NEDCC Preservation Leaflet Microfilm & Microfiche
- NDNP Microfilm Selection Guidelines
- NDNP Technical Guidelines for Applicants
- Preservation Measures for Newspapers
- United States Newspaper Program

Notices

The research and resources above are for informational purposes only and not for the purpose of providing legal advice. You should contact your attorney to obtain advice with respect to any particular issue or problem. The Library of Michigan cannot endorse specific vendors or brands. Libraries should contact vendors and other experts for advice and to ensure that their choice reflects the unique needs of that library.

This is made possible in part due to funding from Institute of Museum and Library Services.