Digital Scanning Recommended Practice

Patrick Walston[1], Madalyn Massey[1]
[1]UNAVCO

Sep 19, 2021

2 Works for me Share

dx.doi.org/10.17504/protocols.io.bwr3pd8n

Madalyn Massey

ABSTRACT

A protocol regarding digitization of documents stored in paper format. Topics covered: minimum standards, best practices, digital scanning hardware, software, file storage, things to consider, and resources available.

DOI

dx.doi.org/10.17504/protocols.io.bwr3pd8n

PROTOCOL CITATION

Patrick Walston, Madalyn Massey 2021. Digital Scanning Recommended Practice. **protocols.io** https://dx.doi.org/10.17504/protocols.io.bwr3pd8n

LICENSE

This is an open access protocol distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

CREATED

Jul 20, 2021

LAST MODIFIED

Sep 19, 2021

PROTOCOL INTEGER ID

51739

Introduction

1 *What is it?*

Digitization is the process of turning physical information into digital data. The USGS and National Geological and Geophysical Data Preservation Program (NGGDPP), aim to preserve and curate historical geological information using Digitization. By following standard processes, scientists can create quality products for research and maximize resources to preserve valuable information. This protocol will focus on providing workflows for "*Documents and Related Materials", "Large-Format Sheets and Maps",* and, *"Aerial Photography".*

1.1 *History*

Digital Scanning has become a staple process in repositories. By optimizing digitization, repositories are able to store physical materials for long term preservation. These digital scanning guidelines should be updated frequently enough for the use of modern technology. "Practices are developed to assist participants in the National Geological and Geophysical Data Preservation Program (NGGDPP) and others actively preserving geoscientific materials, data, and artifacts." (*National geological and geophysical data preservation program*), sourced via National Geological and Geophysical Data Preservation Program- Digital Scanning.

Citation: Patrick Walston, Madalyn Massey (09/19/2021). Digital Scanning Recommended Practice. https://dx.doi.org/10.17504/protocols.io.bwr3pd8n

This is an open access protocol distributed under the terms of the **Creative Commons Attribution License** (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

1.2 *Why preservation and curation is important*

"Since it is unlikely that materials will be rescanned in the future due to their fragile and deteriorating nature, limited funds and time, and access limitations, it is critical to produce high-quality digital scans that closely resemble the original materials. Adopting an appropriate workflow will reduce the cost, time involved, and excessive handling of materials" (NGGDPP).

Minimum Standards

2 Source: Bowman, S. D. (2016). GUIDELINE AND BEST PRACTICES FOR GEOLOGIC DATA PRESERVATION (NGGDPP) DIGITAL SCANNING. (Draft)

A few precautions should always be exercised in areas used for document access.
- Tabletop areas should be clear before accessing and spreading documents out.
- Static electricity can build when scanning, and a static dispersal mat should be used beneath the scanner.

- *Tips and Suggestions:*
- When sorting materials for scanning, sort in groups that allow use for their appropriate workflows.
- "assembly line" procedure (Bowman deems efficient practice)

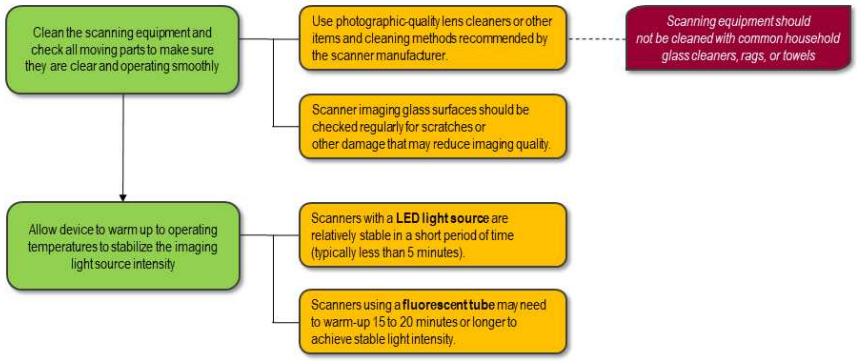
Best Practices

3 ***Typical Digital Scanning Workflow:***

1. Sort materials
2. Clean materials
3. Scan
4. Process and clean up
5. Create distribution files
6. Archive original files

Sourced: USGS Digital Scanning Workflow

4 **1. Prepare the scanning device**



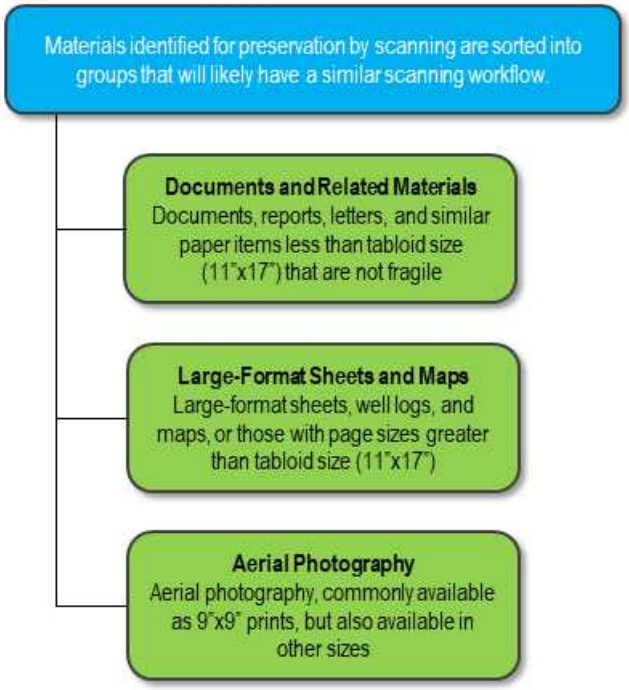
- The scanner shouldn't be placed near a source of direct sunlight/lighting. Sunlight and strong ambient lighting can lead to the degradation of materials and poor quality scans.
- The scanner shouldn't be placed near a significant source of dust (eliminate clutter, avoid placing it in an abandoned room/space, clean environment often).
- Equipment should be placed in a secure environment (do not allow other machines/objects/furniture to cause movement of the scanner).
- See scanner manual for notes about warm up time.
- A power conditioner can minimize light source fluctuations (Bowman 2016). (What is a power conditioner?)

Citation: Patrick Walston, Madalyn Massey (09/19/2021). Digital Scanning Recommended Practice. https://dx.doi.org/10.17504/protocols.io.bwr3pd8n

This is an open access protocol distributed under the terms of the **Creative Commons Attribution License** (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

5

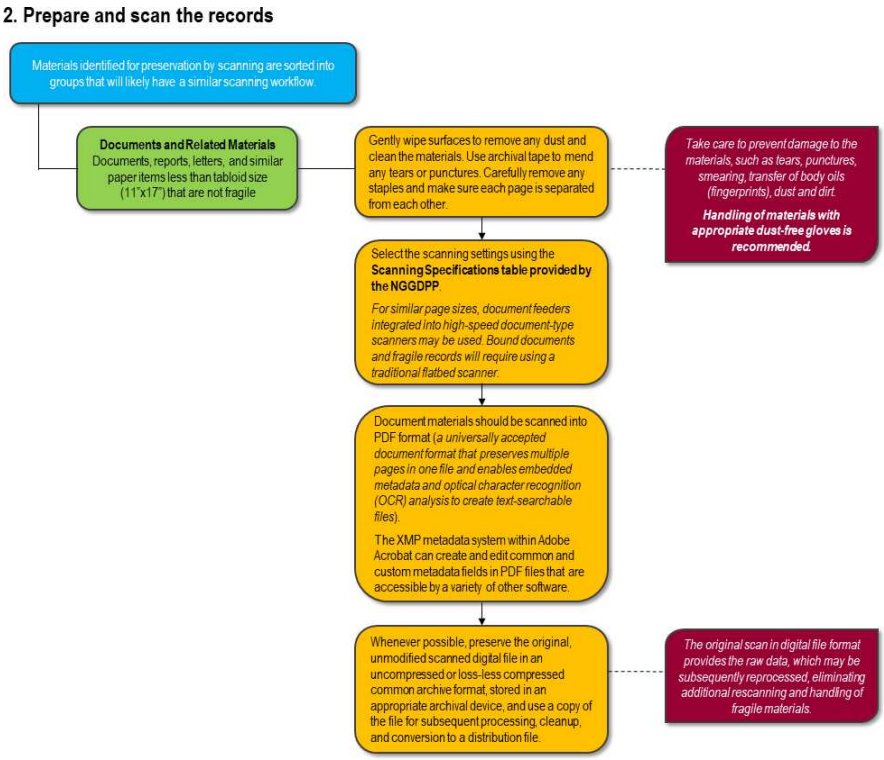
2. Prepare and scan the records



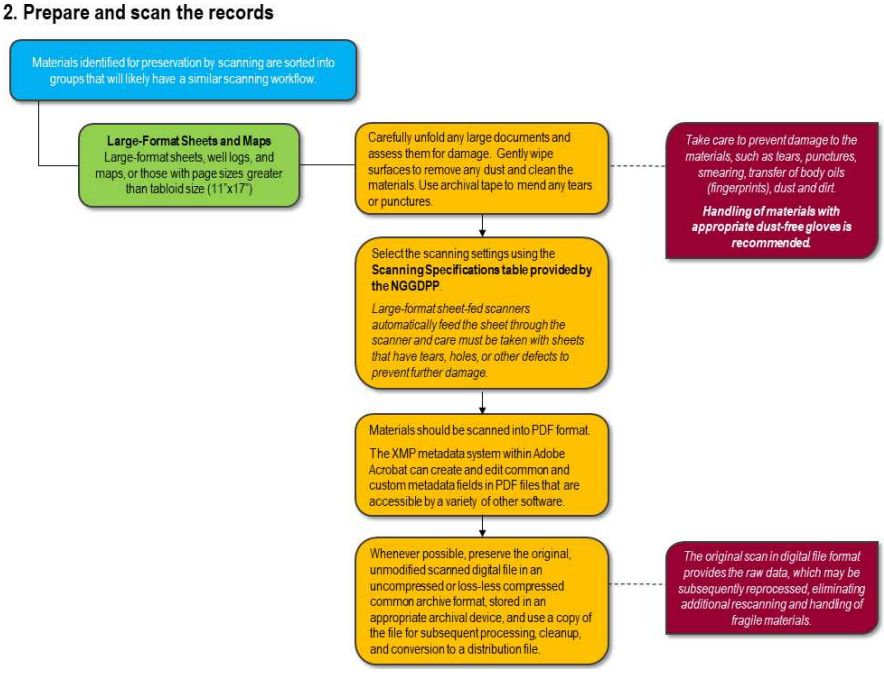
5.1 Workflow for *Documents and Related Materials*

Citation: Patrick Walston, Madalyn Massey (09/19/2021). Digital Scanning Recommended Practice. https://dx.doi.org/10.17504/protocols.io.bwr3pd8n

This is an open access protocol distributed under the terms of the **Creative Commons Attribution License** (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited



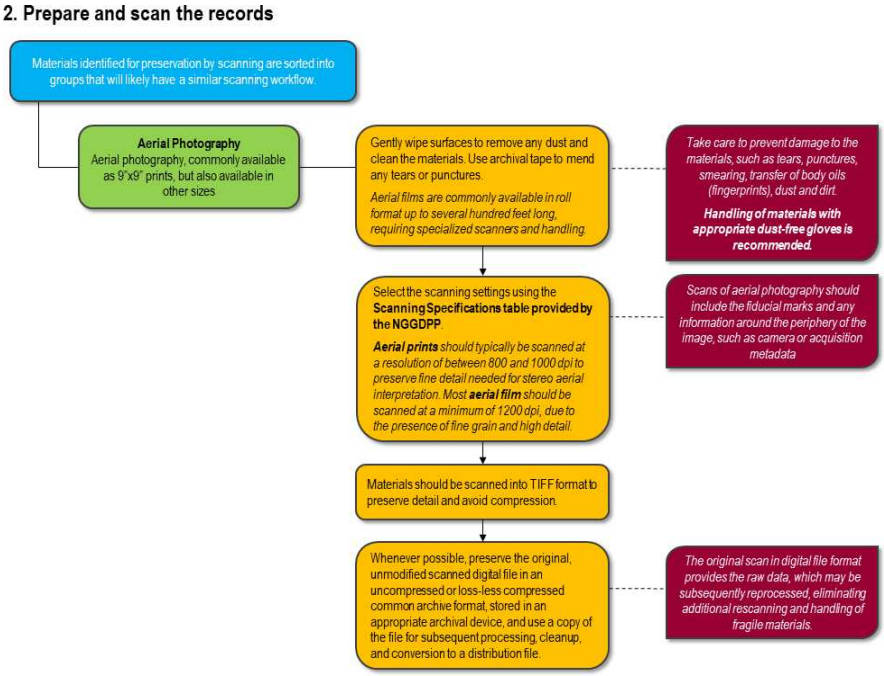
5.2 Workflow for *Large-Format Sheets and Maps*



5.3 Workflow for *Aerial Photography*

Citation: Patrick Walston, Madalyn Massey (09/19/2021). Digital Scanning Recommended Practice. https://dx.doi.org/10.17504/protocols.io.bwr3pd8n

This is an open access protocol distributed under the terms of the **Creative Commons Attribution License** (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited



Digital Scanning Hardware

6 *Types of Digital Scanners*

- Flatbed
- Sheet fed
- Integrated
- Drum
- Portable

6.1 *Document Scanners*

Flatbed and sheet fed scanners are the most common in homes and offices. A flatbed scanning machine moves its scanner over a stationary document. Sheet fed scanners move documents over the image scanner. Flatbed scanners work well in scanning notebooks and other bound materials, but sheet fed scanners are recommended for loose materials because they can scan images much faster and have greater hardware reliability.

6.2 *Large-Format Scanners*

- Sheet fed scanners are available in large sizes ranging from standard 8.5 inches width to 44 inches width.
- Larger scanners should also include a higher DPI and 600 is recommended.
- Sheet fed scanners are the most useful for large documents but care should be taken to start the scan slowly and make sure the oversized document is fed into the scanner correctly to avoid damage.

- Drum scanners are useful for scanning very large documents like maps and including a very high DPI, however they are not widely manufactured and are very expensive.
- Documents are placed on a large horizontal drum which rolls to feed documents into the scanner.
- Useful for digitizing very large maps with intricate features, drum scanners should be used only if documents are too large to fit into a large sheet fed scanner.

6.3 *Film Scanners*

Citation: Patrick Walston, Madalyn Massey (09/19/2021). Digital Scanning Recommended Practice. https://dx.doi.org/10.17504/protocols.io.bwr3pd8n

This is an open access protocol distributed under the terms of the **Creative Commons Attribution License** (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

- Flatbed and sheet fed image scanners are generally unable to scan film.
- Film needs to be backlit in order to view the image, which most scanners do not achieve.
- Scanners should be able to back light and digitize color negatives, color positives and black and white negatives.
- Film scanners are available from companies such as Kodak, Pacific View, and Epson.

7 *Digital Scanner Specifications*

- Repositories should have access to 4 types of digital scanners depending on collections materials, a flatbed to scan bound paper materials such as field notebooks, a sheet fed scanner to scan large amounts of loose paper materials or well logs, an over-sized sheet fed scanner for maps and pictures larger than 8.5 inches wide, and a scanner for creating digital images of film from color negatives, color positives, and black and white negatives.

- Flatbed scanners should include at least 24-bit color scale and 8-bit grayscale with at least 600DPI scanning capability. Flatbed scanners can either be part of a printer or a stand-alone device. Inexpensive models exist that meet and exceed these requirements.

- Sheet fed scanners for loose document digital scans should have at least 24-bit color capability and 300 DPI. Scan speeds should be variable from 1 to 10 inches per second.

- Over-size sheet fed scanners have the same specifications as regular size sheet fed scanners, and their width should accommodate any large documents in the repository. A width of at least 34 inches can be recommended but 44 inchers can ensure scanning capabilities of any newer over-sized documents.

- Film scanners should be selected for specific film width, however some devices allow for variable film width.

7.1 *Resolution*

- Scanner resolution is measured in Dots Per Inch (DPI) and some scanners will include variable DPI settings.
- Scanners have a range of 50 to around 7000 DPI.
- When scanning text documents 300 DPI is adequate and clear, however with larger or more intricate images, such as map documents, a higher DPI setting of 600 is recommended.

7.2 *Optical Density*

- Optical density refers to a devices ability to produce an image balanced in highlight and contrast detail.
- Scanners are measured on a scale from 0.0 to 5.0, the latter being the highest optical density possible.

- Scanners other than drum scanners will have an optical density of 2.5 - 3.0.
- Drum scanners produce the highest optical densities possible at 3.0 - 4.5 which is one reason that they are so expensive and rarely used outside of industry.

7.3 *Color or Greyscale Depth*

- Scanner color quality is measured in bits.
- 8-bit images include 256 different colors in grayscale, 24-bit images include 17 million different full colors, and 36-bit image scanners can produce images with over 68 billion different colors.
- Most scanners are built with at least 24-bit color scanning capabilities and are great for scanning any kind of colored document.

7.4 *Bit Depth*

Citation: Patrick Walston, Madalyn Massey (09/19/2021). Digital Scanning Recommended Practice. https://dx.doi.org/10.17504/protocols.io.bwr3pd8n

This is an open access protocol distributed under the terms of the **Creative Commons Attribution License** (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

- Bit depth refers to the sensitivity of scanner head and the number of coded bits of optical information.
- A higher bit depth means a higher number of coded bits for reproducing large ranges of color. In general, the higher the bit depth the higher the image and color quality.
- Scanners such as flatbed and sheet fed typically include very high bit depth specifications capable of recording extremely vibrant detailed images, however these digital files will be much larger and require more space to house.
- With increasing scanner size comes decreased bit depth, and film scanners operate differently than normal optical scanners by measuring image quality in megapixels (MP).

7.5 Reliability

- The expected number of cycles or length of time between failures, or Mean Time Between Failures (MTBF) is the tested lifespan of a printer or scanner. Check scanner and printer instructions for specifics, or contact the manufacturer regarding details. The MTBF is also useful in considering which scanner or printer to purchase based on what it will be used for and how long it will last in specific scenarios.

7.6 *Color Space/Profile*

- Color space and profile is an encoded file included with digital image scans that tells printers how to print the included colors.
- Color profile is necessary to include with every image scan because it bridges the gap from one scanner type and brand to any printer type or brand, producing color accurate images every print.
- These files are automatically created with new image scans and included in the file code.
- If issues arise with incorrect print color, color space coding could be the problem.

7.7 *Output File Format*

- .JPG, .TIF, .GIF, .JPEG, .PNG, .PDF which will be compressed to .ZIP for storage and transfer after editing.

Digital Scanner Software

8
- Scanning software is useful for image crop, edit, and annotations.
- Software also allows users to change files from one type to another, access and work with many different file types, and include metadata in image files for easy inter operation and use.
- Excellent software suggestions include Adobe photoshop, or the open source application Archivematica.

Digital File Storage

9
- A Local Area Network (LAN) should be maintained by repositories.
- The LAN allows for local digital archiving and shared storage space among many computers and servers.
- Repository digital storage should be very large to include any form of digital information regarding the records collection, and a minimum capacity of 50 terabytes is recommended.
- Individual scanned image files can be a gigabyte or more in size when including a high DPI setting, but can be compressed to .ZIP format when storing or transferring across the web.
- File compression software greatly reduces file size to use less space in local digital storage. Software to compress images to .ZIP files is free and accessible on the internet.

Things to Consider

10
- Scanners may require intermittent cleaning. Consult manual for scanners specifics. A dry microfiber cloth should be used to clean dust and other contaminants from scanning equipment.

Resources

11 Bowman, S. D. (2016). GUIDELINE AND BEST PRACTICES FOR GEOLOGIC DATA PRESERVATION (NGGDPP) DIGITAL

Citation: Patrick Walston, Madalyn Massey (09/19/2021). Digital Scanning Recommended Practice. https://dx.doi.org/10.17504/protocols.io.bwr3pd8n

This is an open access protocol distributed under the terms of the **Creative Commons Attribution License** (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

SCANNING. (Draft)

Collections care. Preservation Guidelines for Digitizing Library Materials - Collections Care (Preservation, Library of Congress). (n.d.). https://www.loc.gov/preservation/care/scan.html.

National geological and geophysical data preservation program. Digital Scanning Workflow. (n.d.). https://www.usgs.gov/core-science-systems/national-geological-and-geophysical-data-preservation-program/digital-scanning.

National geological and geophysical data preservation program. Digital Scanning Hardware. (n.d.). https://www.usgs.gov/core-science-systems/national-geological-and-geophysical-data-preservation-program/digital-0.

National geological and geophysical data preservation program. Scanning Specifications. (n.d.). https://www.usgs.gov/core-science-systems/national-geological-and-geophysical-data-preservation-program/scanning.

National geological and geophysical data preservation program. Scanning Instrumentation. (n.d.). https://www.usgs.gov/core-science-systems/national-geological-and-geophysical-data-preservation-program/scanning-0.

Archivematica: open-source digital preservation system. (n.d.). Www.archivematica.org. Retrieved September 19, 2021, from https://www.archivematica.org/

Citation: Patrick Walston, Madalyn Massey (09/19/2021). Digital Scanning Recommended Practice. https://dx.doi.org/10.17504/protocols.io.bwr3pd8n

This is an open access protocol distributed under the terms of the **Creative Commons Attribution License** (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited