WISCONSIN HISTORICAL SOCIETY

Collecting, Preserving, and Sharing Stories since 1846

NATIONAL ARCHIVES
NATIONAL HISTORICAL PUBLICATIONS & RECORDS COMMISSION

WISCONSIN HISTORICAL SOCIETY

Digitization Project Guidance

For Local Units of Government

October 2017

CONTACT US
Division of Library, Archives and Museum Collections | govarc@wisconsinhistory.org

wisconsinhistory.org

Digitization Project Guidance

Contents

INTRODUCTION

Wisconsin's local governments are increasingly turning to the digitization of their public records records. There are numerous reasons for this, among them:

- To provide access to staff working in multiple locations
- To save physical space in an office environment
- To cut offsite storage costs
- To preserve fragile records that have degraded
- To provide online access to items for the public

There are, however, many considerations for appropriately managing a digitization project as well as managing and preserving the resulting files. This guidance is intended to provide support to local units of government to help ensure the success of their projects and the long term management of the records created by the projects. The sections will walk government employees through the various decision points in digital project planning, project implementation, and long term maintenance of digitized records. It is hoped that this guidance will help ensure consistent quality across projects, so that the integrity of reformatted records can be efficiently managed across time and rapidly changing technologies.

RESPONSIBILITIES

It is important for local government employees to understand that, according to Wis. Stat. § 19.21(1), they are responsible for the maintenance and care of all public records created as part of their position, including any records created by prior employees. This responsibility pertains to any public record created by the government unit regardless of format.

If a unit chooses to scan paper records and keep them as electronic records, they are required to follow procedures established under Wis. Stat. § 16.61(7)(a). Standards and guidelines for the management of electronic records are outlined in Chapter Adm. 12 to ensure that electronic records remain:

- Accessible
- Accurate
- Authentic
- Reliable
- Legible
- Readable

After scanning is completed, local governments must:

- Ensure the records are stored, maintained and migrated to accessible formats throughout the lifecycle of the record until deletion or transfer according to the records schedule.
- Ensure there is a procedure in place to delete or transfer digitized records at the end of their approved retention time.
- Ensure the scanned records are not subject to accidental or unauthorized destruction.

Particular care must be applied when scanning confidential records. During the scanning process and throughout the record's lifecycle, units must maintain confidentiality and limit access to those persons authorized by law, administrative rule or established local policy.

The digitization of public records does not negate the responsibility that local governments have to notifying the Wisconsin Historical Society on the destruction of public records. Notification is required when:

- The original paper records will be destroyed after they have been scanned and verified.
- Digitized records have met the required retention period.

Please consult with the Wisconsin Historical Society if you have any questions regarding the notification process or the retention of paper records after they have been digitized.

SO YOU THINK YOU WANT TO DIGITIZE….

A successful digitization project requires a commitment of organizational resources in hardware, software, staff time, file organization, storage, and preservation activities to ensure the digital content remains accessible over time. In actuality, digital items can be much more challenging to manage and maintain over time than paper records.

It is recommended that you review "*So You Want Think You Want To Digitize"* prior to starting a digitization project. This document was designed to help local units of government ask critical questions about both the scanning process and the management of the files throughout their lifecycle. It may be most beneficial to review this document with a working group of those most heavily invested in the project (key staff, IT, records officer) since the answers may vary depending on their perspective and responsibilities within the government unit. Having everyone understand the parameters of a project will most likely eliminate surprises and give you the best possible outcome for your project.

WHAT DOES IT MEAN TO DIGITIZE YOUR RECORDS?

A digitization project consists of much more than running documents though a scanner in the backroom and calling it "done". A successful project includes:

Selecting your materials

When faced with boxes of records to digitize, first take the time to assess and organize your originals. There will be many items such as duplicates, drafts or internal / routine documents that you will not want to scan and maintain in a digital format.
Don't scan a mess!

Identify the records schedule associated with the records. This will be important further along in the project in terms of assigning metadata, determining how you want to provide access, and how you will manage the items long-term. If the records schedule specifies that the records will be deleted in the near future, you may want to reconsider taking the time to digitize them at all and instead concentrate resources on records of a longer lasting nature.

Refer to the *"Scanning Decision Tree"*, for guidance regarding content selection for digitization projects. This document walks through the questions you should ask to determine if a collection of documents should or should not be scanned and help you choose the items most likely to give your unit the best return on investment when deciding amongst many potential projects.

In-House Staff vs. Hiring a Vendor

Once your unit has decided to begin a digitization project, one of the first decisions is whether the scanning and related work (quality control and metadata) will be performed onsite by in-house staff or offsite by a vendor.

In some cases, it may make sense to bring a vendor onsite to manage the digitization project using their staff, or to hire a vendor to only digitize large items or odd formats you don't have the technology to capture and have no long-term interest in purchasing.

There are advantages and disadvantages to each approach, and no one-size-fits-all solution. Some of the key advantages and disadvantages of managing the digitization project yourself vs. contracting it offsite with a vendor are outlined in the "*In-House vs. Contractors"* document.

Purchasing Equipment

If you decide to digitize materials on your own, you will be required to evaluate and purchase the scanners and any other image capture devices that will be used for the project. As with most technology, the devices change with some frequency. Equipment decisions should be based on the items you want to digitize. Fortunately, there are many options within every price range. The *"Technology Considerations"* document has been put together to help you evaluate your options and narrow the field.

DISCOVERABILITY

A well-managed scanning project will allow you to efficiently find and provide access to the digitized content. This is particularly important to ensure an efficient response to open records requests and/or litigation. You could complete every other part of the digitization project correctly, but if you can't locate the files when they are needed, your unit has wasted staff time and resources. Failure to manage your files also leads to documents being left on unit servers far past their retention schedules which wastes IT resources such as management costs and storage space. Fortunately, there are several techniques, that when used in conjunction with each other can help you locate and manage your files correctly.

Metadata

Metadata describes the who, what, when, where, and why of your scanned records and is one of the core components of being able to easily locate those records when needed. Examples of common metadata used for discoverability are:

- Basic information about the document's contents
    - title
    - document date
- Technical information about the scanned item itself
    - file size
    - format type

Generally speaking, metadata fields should be standardized, consistent and searchable. Your organization should determine and document what metadata will be collected before starting your project and how it will be written (ie: all dates will be written *yyyymmdd)*. Metadata should be applied to the records at the point of digitization for ease and efficiency. You *can* apply metadata post-scanning, but it will be considerably more difficult to go back after the digitization process is complete.

The accompanying *"Digitization Metadata"* document describes some of the primary metadata fields to consider for your digitization project. There should be some key fields that your unit uses for every digitization project. These can be combined with some specialized options that may change between projects. For instance, you will likely always want some sort of name or date field, but there are others like "location" you would want to capture if you had images or a geolocation field for maps.

File Naming and Organization

Once the records have been scanned and appropriate metadata applied, following file naming and organization best practices are next in your line of defense for managing and locating your records over time. Establishing a file naming convention is a key part of managing e-records throughout their lifecycle. This applies to scanned collections as well. Consistent file naming conventions help:

- Organize the digitized items.
- Better maintain files during active use.
- Provide more efficient access for public records requests and legal needs.
- Ensure proper disposition based upon retention schedules.

Developing standards for the way you organize electronic files helps you identify what you have, how they are arranged, where they are located and related retention information. Rules around file naming and organization should be well documented for your digitization project to enable others to locate and manage the files into the future.

Optical Character Recognition (OCR)

In addition to adding metadata and using file nameing conventions, making the scanned documents full-text searchable provides another tool to help quickly find them at a later date. Optical Character Recognition (OCR) software works with your scanner to create a text file of words in the document. This text is separately indexed by the computer's operating system and allows you to search the body of that document. It is recommended that scanning projects take advantage of this technology whenever possible.

OCR software will not work on all scanned documents. It works well on clear, typeset-type documents and very poorly (if at all) on handwritten cursive writing. For documents that are difficult for the OCR software to read, metadata and file naming / organization will prove to be even more critical for locating your documents. Checking the success of how well the OCR process worked across a selection of documents in your project should be incorporated into the Quality Control (QC) process.

Making Content Available

Once the content has been scanned, processes and technologies should be implemented to provide that content to others in your organization.

MANAGING A DIGITIZATION PROJECT

To ensure your digitization project is successful, it is essential the process is consistent from inception to completion. This is easily accomplished by creating a project plan, following a set digitization workflow, and performing specific quality control steps for each and every project.

Digitization Project Worksheet

Digitization projects benefit greatly from project planning which should include documenting what will be digitized and listing any special instructions surrounding the objects in that project. The "*Digitization Project Worksheet*" is intended as a tool for local governments to help define what will and will not be scanned for each "project" as well as to help document the many facets of a digitization project.

Specifying what will be digitized is especially important if the people performing the scanning are not overly familiar with records schedules and will need to separate the "record" part of a very large folder of material that hasn't been reviewed. *Ideally, the unit records officer is involved in this step and can provide some guidance to those responsible for the scanning process.*

Documenting your decisions in some manner is also helpful in providing consistency between digitization projects for your organization as well as providing a template for future projects.

Digitization Workflow

The digitization workflow should be clearly documented and accessible to all people working on the project. If resources permit, a project manager tracks the project from start to finish to ensure all phases have been properly completed.

Depending on your resources, it is possible that some staff may only work on one specific part of the project. As such, it is critical that each phase of the workflow be clearly documented and accessible to all project staff in order to clarify necessary accomplishments and expectations for each specific task. A sample "*Digitization Workflow*" has been provided to help manage this process and provide consistency with each project. While this covers the key steps of a digitization project, the details of each step can be adjusted to cover your own individual circumstances.

Quality Control Process

After your records are digitized, the images need to go through a review process to ensure all pages have been scanned correctly, the image quality is acceptable and all images are in the correct order and rotation. This is the Quality Control (QC) process which verifies the quality, accuracy, and consistency of digital images. This step is critical for projects which plan to destroy the paper documents after scanning. The "*Quality Control for Digitization Projects*" document guides you through the various steps of the process to ensure your digitized records are a strong representation of the original documents.

STORING AND MAINTAINING DIGITIZED RECORDS

Once the files are digitized, they must be safely stored and maintained over time. This is one of the most important, yet overlooked, components of a digitization project. The storage of the records and any derivatives should be discussed with all relevant parties and

documented as part of your digitization plan. This should be understood by those managing the digital files to help protect your organization in case of natural disasters, cyberattacks, computer hijacking, accidental deletion, or file corruption.

It is important to calculate the storage space you will need for the newly created files. It is recommeded you should have at least two copies of each electronic file and that they be stored in geographically separate locations in case of a catastrophic event. This is particularly true if you have determined that the digital file is now the official record and the original paper version has been destroyed. Ideally, at least one copy of the digitized material should be stored on the agency (or state) controlled server. If your agency does not have the storage space necessary to store the digitized records, refer to the Public Records Board's Guidance on the Use of Contractors For Records Management Services in Cloud Computing Environments.

Maintaining the electronic files over time will require constant vigilance in terms of testing files to ensure they have not degraded, monitoring backups to make sure they are working properly, and periodically replacing the hardware on which they are being stored. The "*Storage and Maintenance*" document provides guidance to help you determine how to evaluate storage options for your organization given its capabilities, staffing and budget.

Digital records that have been stored and maintained for their required retention period, as defined through their proper record schedules, are required to have written notification sent to the Wisconsin Historical Society BEFORE they can be deleted. This final step in the records lifecycle is vital and cannot be overlooked when planning and implementing a digitization project. Failure to plan for records disposition will exponentially increase your unit's storage costs over time as you deal with an ever increasing number of records you are storing, make it more difficult to find records and burden IT staff with migration of digital content over time.

There are occasional business reasons to digitize non-records as well. Planning for the disposition of non-records that an agency decides to scan is equally important, and perhaps more challenging as this content does not have a schedule-driven disposition date. It is critical that the project plan for these items includes a review date at a minimum to ensure the content is periodically reviewed over time and deleted when no longer of use to the agency.

RISK MANAGEMENT

Once records are digitized it is tempting to keep them "forever" since they no longer take up physical space, and are, for the short-term, easy to manage. Keeping records, either digitized or paper, beyond their retention puts your unit at risk. These same risks are present when setting unreasonably long retention times for electronic record series.

Risks include:

- Wasted time - As your organization's servers fill up, searching for specific material becomes more time consuming. This problem is compounded by lack of metadata.
- Discovery - Any record your unit has on its server is subject to discovery in litigation and must be produced. This includes records that have been kept beyond their retention period. Holding these records will also likely increase your costs during the discovery process due to the resources needed to sort through items that should have been disposed of and possible court costs if you can't produce the information in an appropriate amount of time.

- Loss of public favor - Failure to properly manage records can result in negative public perception and can cause the public to view unit activities with suspicion.
- Destruction of records before their time - Records with long retention periods will need to be migrated several times during their lifetime. Each occasion when records are moved from one system to another or one, they are at risk of loss and/or lack of readability.
- Unauthorized deletion - Electronic records must be protected from unauthorized or accidental deletion.

The easiest way to mitigate these risks is to store digitized records in systems that can delete or export electronic records in accordance with their approved retention schedules. Records should be deleted or transferred according to those schedules.

For all digitized records, but particularly those with permanent or long-term retentions, ensure they are stored in a system that can export records to other programs or systems without loss of meaning.

ACHIEVING WELL-MANAGED DIGITIZED RECORDS

In sum, digitized records need to follow the characteristics for any well-managed digital project in your organization:

- You have documented basic information about each project /collection/series that was digitized.
- You have minimal metadata for each project and all digital objects within the project.
- The digitized objects are stored in common, non-proprietary file formats.
- The digitized objects are stored in a well-managed environment with appropriate security.
- You can quickly identify content that may contain Personally Identifiable Information (PII) or confidential information.
- You have multiple copies of the each digitized object stored in at least two locations and have determined which is the official record and which is the backup.

GLOSSARY

The following links provide comprehensive glossaries for the digitization process:

Federal Agencies Digital Guidelines Initiative (FADGI) http://www.digitizationguidelines.gov/glossary.php?alpha=A (viewed 2/23/17)

BCR's Collaborative Digitization Program (CDP) Digital Imaging Best Practices Version 2.0 (viewed 2/23/17)

FOR ADDITIONAL INFORMATION

Public Records Board Guidance on the Use of Contractors for Records Management Services Managing Records in Cloud Computing Environments (viewed 2/23/17) provides considerations if your plans to cloud services or contractors.

The *Primer State of Wisconsin Electronic Records Management: Guidance on Ch. ADM 12* (viewed 2/23/17) provides background information and guidance for State agencies and local units of governmentin Wisconsin to implement Ch. ADM 12.

Wisconsin Administrative Rule Ch. ADM 12 (viewed 2/23/17) outlines requirements for state agencies and local units of government for the management of electronic records including.

Wis. Stat. § 16.61 (viewed 2/23/17) outlines the records management responsibilities of state employees. Of particular note for this guidance are 16.61(4) and 16.61(5).

Wisc. Stat. §§ 19.31-19.39 (viewed 9/25/17) details open records requirements for state agenices and local units of government.

UMass Amherst Libraries Guidelines for Digitization (viewed 2/27/17)

Minimum Standards For Digital Imaging or Scanning of Textual Documents and Minimum Standards for Digital Imaging or Scanning Textual Documents (viewed 2/27/17) Arizona State Library, Archives and Public Records

NISO Framework Working Group. A Framework of Guidance for Building Good Digital Collections, 3rd edition December 2007

So You Think You Want To Digitize

The trend of local governments digitalizing public records has a number of contributing factors. Advances in technology, increased demand for access through electronic means, and a general desire to reduce the volume of paper records stored in offices or at off-site locations all play a role in that decision. But as with all projects, it is important to evaluate a number of factors **before** you begin. Carefully thinking through the components of a scanning project will go a long way towards ensuring a successful and useful outcome. This document is intended to help you consider how the digitization fits into your unit's records management and technology plans; all the while maintaining compliance with retention schedules, statutes, and ordnances. Asking the following questions will help determine if a digitization project is worth your time, effort, and money.

1) What is your purpose?

- How are the digitized materials going to be used and what are the benefits of having the records in digital format?
- Is there a demand for the content of these records in digital format?
- What are the goals of your project/what do you hope to accomplish?
    - To increase access to the records or decrease the handling of fragile originals (preservation)? Or both?
    - To reduce the quantity of your paper files?
    - To provide access to the records online to the public?
    - To provide access to the records to staff in remote locations?
- Will the digital records replace existing paper records?
    - If "Yes" - do you plan on destroying the paper records? Make sure to notify WHS before destruction.
    - If "No" - have you determined whether the paper or digital version will be the official public record?
    - Depending on your answer above, would you need to update a records schedule in some way?
- Will the digitized records be made available online and if so, do you have a plan for that?

2) How do the records retention schedules affect your digitization plans?

- Have the records met their required retention period specified by record schedule or ordinance? Are they ready for destruction or transfer to the State Archives at the Wisconsin Historical Society?
- Are the records slated to be destroyed or transferred within the next 10 years? If so, it may not make sense to take the time or resources to scan them.
- If the records are long-term or permanent, is it more cost effective to store them in their current format?
- Is there an essential business reason/demand to digitize items that are not determined to be unit records?

3) How is the project being funded?

- Have you secured funding for the term of the scanning project?
- What parts of the project will your funding support? (physical resources, hardware, software, networked access, personnel, dedicated space, vendor services, etc.)
- Is there a long-term commitment to cover ongoing, post-scanning costs for preserving and maintaining access to the digitized items?

4) How will you perform the actual digitization?

- Where will the digitization take place--in a central location or off-site? If off-site, does the site meet the security needs for the content you are digitizing and the hardware/software used in the process?
- Will the project be done with in-house staff or will you contract with a vendor?
- What is the scanned image quality you are planning on for this project and do those specifications meet the recommended best practices for access and preservation or potential litigation requirements?
- Who will perform the quality control and what is your procedure for determining an "acceptable image"?
- Which file format will you use for the scanned images and does that format meet the recommended best practices for access and preservation? Keep in mind it must be clear on both a computer screen and printed out on paper.
- Will you create multiple variations of the image, such as thumbnails or user copies, of the master file for viewing and downloading?
- What are the limitations of your hardware and software (file size, file format standards, proprietary file formats, interoperability, scanner limitations, etc.)?

5) How will you locate and access the scanned records?

- What type of description already exists for the records, and at what level (item level or collection level)?
- What metadata do you plan to use to find and locate the records?
- Will the metadata be able to track any derivative files (thumbnails, user copies) created from the master record?
- If the records will eventually be transferred, are you applying minimum metadata standards required by the Wisconsin Historical Society?
- Have you determined a consistent naming convention / file organizational structure for the files that meets best practice standards?

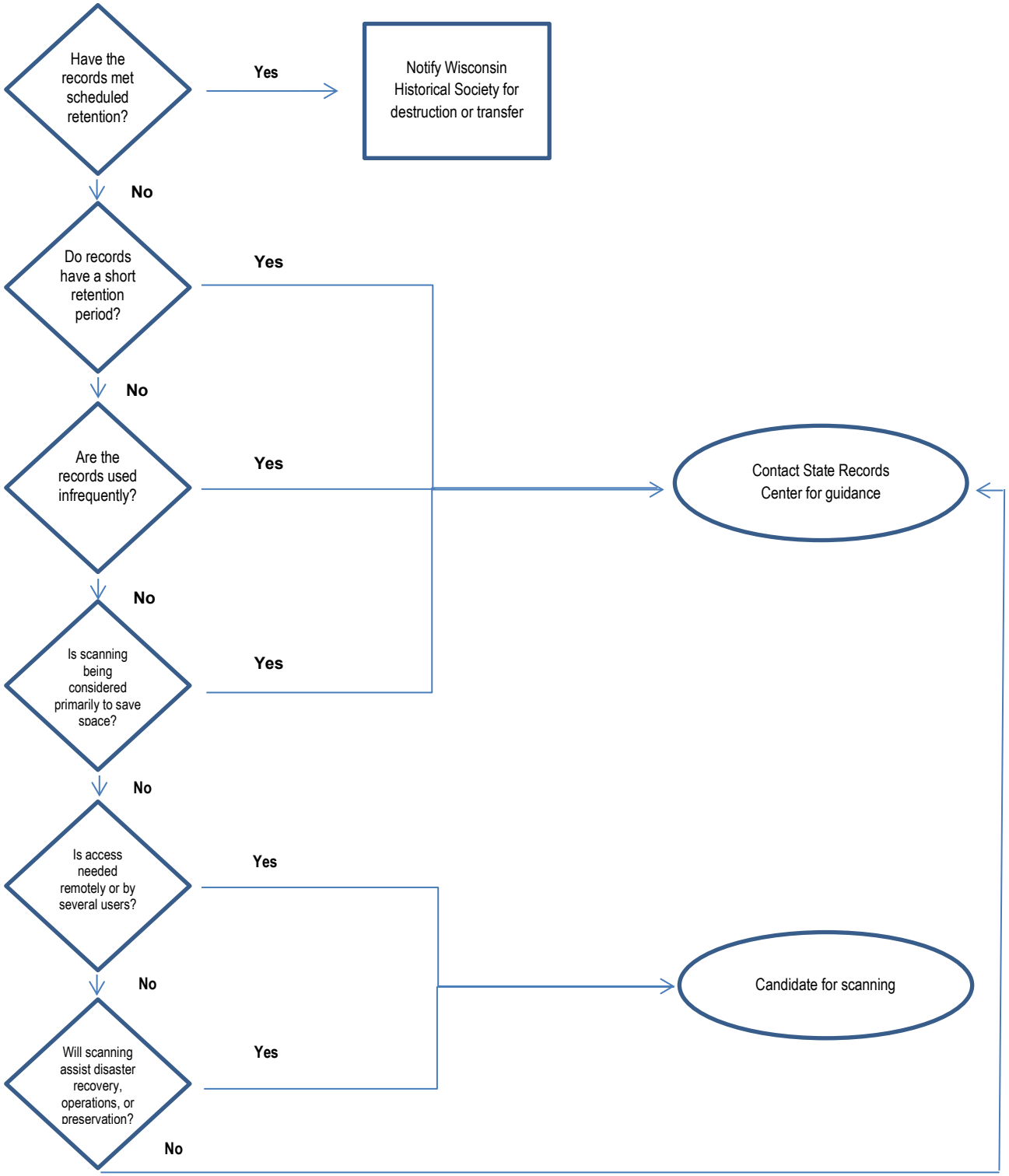
6) How are you planning to maintain the records into the future?

- How do you plan to store the master files and where?
  - Will you be creating copies of the files for access and preservation?
  - How many copies of each of these will you be storing and where? Ideally you need at least 2 copies of your files and preferably in geographically disperse locations.
- Have you consulted with your IT department or your office considered current and future storage needs?
  - What is your current digital storage capability?
  - Do you have enough storage space available to begin your project?
- What kind of a backup procedure do you have in place in case of hardware/software failure?
- Does your IT department or your office have a documented plan to monitor formats and ensure they are migrated appropriately over time as technology changes?
- Do you have funding in place for future storage migration and refreshment?

Resources referenced for this document

NY, South Central Regional Library Council. *Questions to Ask Before Starting a Digitization Project.* n.d. https://scrlc.org/Questions-to-Ask-Before-Starting-a-Digitization-Project (accessed May 2017)

Scanning Decision Tree



Have the records met their scheduled retention time?

Before digitizing any records, evaluate them to see if they are beyond their required retention period and can be destroyed. If they are out of retention, identify them for destruction and do not scan. Notify the Wisconsin Historical Society as required by state statues before destroying any public records. The Wisconsin Historical Society will reply with a decision on whether the records can be destroyed or will be transferred. Consult your corresponding local government records schedules for information on which particular records have been granted waivers for notification.

Do the records have a short retention period?

Generally, it is not cost effective to scan records with short retention periods if the intention is to replace the paper record with a digital copy. The costs of equipment, scanning, metadata creation, quality control and storage may actually be higher than the cost of storing paper records for ten (or less) years. There may be some cost savings if you are creating lower-resolution digital user-copies for frequently requested items your unit wishes to post on a website.

Are the records used infrequently?

If records are accessed only a few times a year, it is usually not cost effective to scan them. Proper records management of the paper will likely be a more effective strategy. If space is a concern, contact the State Records Center for guidance and to verify if storing records at an offsite storage facility is a more practical alternative.

Are records being scanned to save space?

While the business decision to scan records and dispose of the paper to "save space" is often used, the costs associated with scanning and the long term storage and maintenance of electronic records may very well be higher than the cost of properly managing and storing records in paper format. Contact the State Records Center for guidance and to verify if offsite storage is a more practical alternative once the paper records have been reviewed.

Is remote access, or access by multiple people simultaneously, needed?

Electronic access can be the best method for providing access to several people simultaneously and/or to remote users. It can also allow for increased collaboration among staff. The costs associated with increased productivity may outweigh the costs to scan, add metadata, and create a search interface.

Will scanning assist with operations, disaster recovery, and/or preservation?

Before beginning a scanning project, make sure it will benefit the organization in a tangible way. Will the project result in faster or more accurate service to customers? Does the proposed scanning project play a role in disaster recovery planning? Are records being scanned for preservation purposes? If the answers to these questions are generally "yes" the costs connected with scanning and indexing records may be outweighed or justified by these factors.

For additional help and information

State Records Center

The State Records Center (SRC) is a secure, low cost facility that offers state agencies and local units of government (near Madison, WI) off-site control over semi-active and inactive records. More information is available on their website.

Wisconsin Historical Society

govarc@wisconsinhistory.org

Resources referenced for this document

Illinois Secretary of State. *Should I Scan?*. 2013. https://www.cyberdriveillinois.com/departments/archives/records_management/scanningtree.pdf (accessed June 2017).

In-House vs. Contractors

One of the first questions in any digitization project is whether the scanning and related work (quality control and metadata) will be performed by in-house staff or offsite by a vendor. There are advantages and disadvantages to both approaches, and there is no one-size-fits-all solution.

In-House Scanning Projects

Advantages

Control
- All aspects of the project take place within your organization, allowing staff to collaboratively resolve issues as they arise and make decisions.

Material Handling
- Fragile or valuable materials remain on site and are not subject to changing environmental controls, or potential damage through transportation.
- Security and proper handling of records ensured because the local unit can monitor access and ensure proper handling.

Develop Staff Skills
- Digitization may allow staff to learn new skills, including project management.

Project Development
- The project and its requirements can develop incrementally.
- Easier to ensure desired quality requirements are met and upheld.

Disadvantages

Equipment Investment
- The purchase of hardware and software falls on the organization.
- The organization is responsible for replacing obsolete technology.

Costs
- The staff learning curve makes it difficult to set a specific price per digitized item and determine the costs for the entire project.

Staffing
- The organization will either need to reallocate staff to the project, or hire new/temporary staff.
- Staff turnover may create variables that can affect the quality, cost, and ability to meet deadlines.

Space/Resources
- Need to be dedicated to the project including computers, tables/desks, and outlets.

Standards
- The organization is responsible for researching standards and staying abreast of changes.

Project Timeline
- Projects will take more time to get up and running as staff learn and/or develop skills.

Offsite Contractor-Run Scanning Projects

Advantages

Costs
- Project costs are more transparent because many vendors will allow an organization to negotiate a set price per item.
- The vendor is responsible for the upfront costs for technology including scanners, servers, computer terminals, and network connections.
- Problems and costs associated with obsolesce are the responsibility of the vendor.

Project Timeline
- Vendors can typically perform a higher volume of scanning then can be done in-house because trained staff are working in a setting designed specifically for scanning, and using top of the line equipment.

Experience
- Vendors may be able to offer project advice based on past experience.

Staffing
- Expenses related to staff training and turnover are incurred by vendor.
- No need to increase/reallocate unit staff to the project.

Disadvantages

Removed From the Process
- Because the scanning is offsite you may have less control over the image processing and quality control.

Contract Negotiation
- The contract must be negotiated to exact specifications at the outset of the project. Any changes to specifications based on results will need to be re-negotiated.

Material Handling
- All records need to be transported to the vendor. Fragile or large format material may be hard to transport.
- Vendor staff may not be trained to handle confidential records.
- Records may be damaged or lost during transport.

Access to Records
- Records are inaccessible while they are with the vendor.
- Handling of confidential and/or sensitive records will need to be specifically addressed with the vendor.
- Some vendors use proprietary software which limits access or leads to extra costs.

Resources referenced for this document

Northeast Document Conservation Center. *Outsourcing and Vendor Relations.* n.d. https://www.nedcc.org/free-resources/preservation-leaflets/6.-reformatting/6.7-outsourcing-and-vendor-relations (accessed June 2017).

Technology Considerations

Introduction

A successful digitization project must consider the hardware and software needed to capture and manage digital content. The chosen technologies should account for all aspects of capture, preservation, and presentation of the digitized records. The Wisconsin Historical Society does not recommend specific hardware or software but instead suggests you consider the following aspects of hardware and software.

Computers

A computer that balances reliable components with speed and storage is important to the success of any digitization project. Projects that require the purchase of computer hardware should consider the following:

- Purchase a computer that can be dedicated solely to digitization initiatives.
- Purchase as much Random Access Memory (RAM) as your budget allows. Memory allows the computer to more quickly process large amounts of image data.
- Purchase a computer with a processor optimized for image manipulation.
- Purchase as much hard drive space as possible. Portable hard drives that attach via the USB port can be used to supplement the hard drive workspace on the computer(s) being used for digitizing.

Storage

The plan for the storage of digitized files should be considered at the beginning of a digitization project:

- Consult with your organization's IT staff to confirm you will have enough storage space to complete your project.
- Depending on project needs and goals you may need either online or offline storage, or some combination of both.
    - Offline storage: storage of digital data on devices that require human intervention to access (e.g., on backup tapes).
    - Online storage systems: storage of digital data that is accessible and available for immediate use on demand without human intervention.
- Consider how much storage you will need for the current project and possible future projects if necessary. This includes storage of the original document, any copies or thumbnails determined to be necessary, as well as backups.

Monitors/Displays

Monitors give the end-user instant feedback showing text or graphics. For that reason, it is very important to keep your monitor regularly calibrated.

- Liquid crystal display (LCD) monitors tend to be smaller, thinner, and lighter than CRT monitors. They are more energy efficient and take up less space on a desk than the bulkier CRT units. LCD monitor's fixed resolution has been known to lessen adaptability and color rendering; however, they have good image quality.
- Cathode Ray Tube (CRT) monitors render colors of high quality and range. They have more options for handling graphics, so if highly detailed, graphical records are a key part of your project a CRT monitor may be worth the investment for you.

Image Capture Devices

Selection of the image capture device (scanner or camera) is the critical element of a successful digitization project. The correct device for your project depends on numerous factors including:

- Overall project goals
- Format of the object to be scanned

- Size of the object to be scanned
- Condition of materials to be captured
- Available budget
- Litigation requirements for the scan (either current or future)

Several technical factors will also influence your purchase including:

- Cost
- Required expertise needed to operate the device correctly
- Size of scan area
- Speed of the capture device
- Ability to bulk scan many objects at one time
- Connectivity
- Ability to handle different formats and materials

In order to select the correct model with the appropriate features, consider both the objects to be captured and your budget. The nature of the records will impose restrictions and direct your decisions regarding which type of device is used for capture. For example, if the documents are paper contracts, meeting minutes, etc., perhaps a flatbed scanner would be the best option.

- **Flatbed Scanners** are one of the most popular types of scanners. An important consideration when selecting a flatbed scanner is the size of the scan area, a scan area of 11" by 17" is necessary to accommodate a variety of materials. Advantages of flatbed scanners are they:
  - Provide quality scans of two-dimensional objects that fit on the surface of the scanner
  - Require no external studio lighting
  - Generally fit on a table top
  - Can easily be used by someone with only a basic understanding of the digitization process

- **Slide/Film Scanners** are specially designed to digitize slides and film. Although a flatbed scanner with a transparency lid can be used for this purpose, a dedicated film scanner has much higher quality scanning capabilities through higher dynamic tonal ranges and optical resolutions. Optional slide feeders can be purchased which allow batch scans of many slides.

- **Large Format Scanners** are useful for scanning maps, blueprints, architectural drawings, site plans, and posters. They operate like a flatbed scanner but are much larger. Professional grade models come with high optical resolutions; however, they may be cost prohibitive for many projects. Institutions needing to digitize this type of material may want to consider outsourcing to a digital imaging vendor or using a digital camera for image capture (see Digital Cameras below).

- **High-End Book Scanners** allow for overhead copying of bound books and oversized and/or fragile materials that cannot be placed on a flatbed scanner. These scanners are more complex and, consequently, higher priced than flatbed scanners. As with the large format scanners, institutions needing to digitize items from overhead may want to consider using a digital camera for image capture.

- **Wide-format Scanners** were developed to digitize large format materials such as engineering drawings and architectural blueprints. Materials are drawn over the scanning sensor through a pair of drums. Due to the danger of mechanical damage (ripping, tearing), these types of scanners are not recommended for fragile materials.

Image Capture Software

Scanners generally come packaged with the software needed to operate the device. Before purchasing a scanner it is best practice to make sure the software also suits your needs which may include:

- Support for scanning images directly into Photoshop and other image editing applications
- Control over image resolution, bit-depth, and the option to turn on and off automatic adjustment features
- Efficiency and ease of use for the operator
- Production of images in formats appropriate for long-term preservation (not proprietary)
- Support for capture of all unit-determined metadata

OCR Software

Optical Character Recognition (OCR) software works with your scanner to create a text file of words in the document, which allows for full text searching. Using OCR software adds considerable value to your scanning project, because it makes scanned documents easily accessible and searchable once the scanning process has been completed. Be aware that OCR does not work on all types of documents.

Cameras

Just as with scanners, there are many types of cameras suitable for image capturing purposes. For best results combine a camera stand with a 35mm (DSLR) camera or better. Another useful feature is the ability to connect directly to a computer so the image appears onscreen soon after it is captured. This allows for the Quality Control process to be tightly integrated into the workflow and retakes of poor captures to be done almost immediately.

The use of consumer-oriented “point and shoot” cameras and cell phone cameras are not recommended for an image capture project. Lens quality and file size is usually insufficient.

Resources referenced for this document

CDP Digital Imaging Best Practices Working Group. BCR's CDP Digital Imaging Best Practices V.2.0. 2008. http://sustainableheritagenetwork.org/digital-heritage/bcrs-collaborative-digitization-program-digital-imaging-best-practices-version-20 (accessed June 2017).

Digitization Metadata

Appropriate use of metadata is a critical component of a successful digitization project. Metadata provides the structure to identify key information about a collection of records (organization, department, collection name etc.) as well as the individual records within the collection (title, date created, file format, access rules, etc.) Metadata is used for the identification, management, access, use and preservation of individual digital records and provides context for the collection as a whole. Digitization projects that combine the collection of sufficient metadata in conjunction with the use of OCR technologies to enable the ability to search within a document stand the greatest chance of success for management over time.

While no single metadata set will apply to all records series since record sets vary, it is recommended that you select a basic list of metadata fields that will be collected regardless of the records being digitized. Collection level metadata is applied to the overall record series being digitized. This is the "big picture" information that describes what the records are, who manages them, what happens to them at the end of their records lifecycle, and how they are arranged. This data tends to be fairly straightforward and well defined.

File level metadata is more challenging and perhaps more critical as it will help you locate a specific file among potentially thousands within a records series. It is important to evaluate the records prior to scanning and determine what data is needed to separate one item from another and how to find it when searched. A basic set of text documents around a project might be fine with minimal metadata (title, date, author, doctype). A scanned photo collection might require additional metadata such as photo location, event, and subjects. A map collection may require that metadata include a map ID, county code, street address, or GPS coordinates for items to be located in the future.

Part of your evaluation should also review the record series to see if the same data can be found on all records over time. For records that span decades, things like the format of document IDs may have changed over time, or fields in forms where metadata can be extracted from may have been added, deleted, changed or not filled in at all. You need to determine how you will reconcile these anomalies and collect sufficient data to find the digital object when searched. It may make sense to use your QA process to review and manually enter fields where only a minimal amount of data is missing or to clean up misspellings which would hamper future search efforts.

A note about Access Restrictions

Each list below has a place for documenting access restrictions due to legal reasons (PII or Confidentiality) at both the collection level and file level. Documenting access restrictions at the collection level is the least desirable in a records series where restricted and non-restricted documents live side-by-side because it often has the effect of restricting the entire collection or requiring an intermediary each time a document is requested to ensure no restricted items are released. If done correctly, the co-location and identification via metadata of restricted digital objects during a digitization project can actually make this task easier through the creation of security groups on a network drive or in a content management system that allow select individuals to access the items. Depending on the records, it may be most effective to gather restricted documents into a subfolder and apply the restriction to the subfolder which would then propagate down to the items within the folder rather than on individual documents scattered throughout a collection of records. An evaluation and determination of how restricted items will be managed should be part of your project plan prior to starting the digitization process.

The below sections list recommended metadata fields for digital collections and individual records. Your organization should identify and document key items that will be required for every project. This will help ensure consistency across projects and enable them to be searchable and accessible well into the future. Other fields are going to be "nice to have"; items that would be beneficial but not mandatory for the long-term management of the item. These are not meant to be comprehensive lists of all possible metadata fields. If there are other fields that make sense for your unit or your specific records, please use them.

Collection Level Metadata

The following metadata fields can be applied to the record series as a whole. Items in **BOLD** text are fields that, if populated, will provide the greatest opportunity for successful long-term management of the records.

Metadata	Description
Division/Department	Division/Department name that manages the records
Project Title	Common name for the group of records in the digitization project
Event	Event that initiates the start of the retention time period
Retention	Retention time period for the records
Disposition	Specify: Destroy, Destroy Confidential, Transfer to WHS
Access Restrictions (Y/N)	Yes or No – whether there are access restrictions on these records (PII, Confidential)
Description	Description summarizing the specific records being digitized so that others can understand the materials in context.
Arrangement	Description of the materials' organizational scheme. (ex: docket number, alphabetically by name)
Keywords	Name of a person, organizational unit, place, or topic that provides information about the records being digitized or which might be searched
Location	Names of places that are relevant to the collection which can be searched (community, county, state)

File Metadata

The following metadata fields are applied to individual records. Items in **BOLD** text are recommended fields. Given the variety of records, it is highly likely additional metadata fields **not** included on this list will be identified through the evaluation process to support the long term management of your digitized items.

Metadata	Description
Title	File name of the record
Author	Author of record (if known)
Date Created	Creation date of original item
Date Digitized	Date when record was digitized
Disposition Date	Date when WHS should be notified. "Permanent" if it stays with the unit
Format	Format of digitized record (.tif, PDF…)
File Size	File size of digitized object
Pages	Number of pages (helpful in multi-page documents)
Description	Abstract or summary of items contents
Location	Names of places that are relevant to the record which can be searched (community, county, state)
Access Restrictions (Y/N)	Yes or No – whether there are access restrictions on this record (PII, Confidential)
Keyword	Terms that might be searched, but may not appear in the body of the text or other metadata fields.
Unique ID #	Unique ID that was assigned to the paper document at the time of creation. (This may or may not exist depending on the process the paper document was created under)
Document Type	This is assigned during the digitization process and will help in locating the object while searching (ex: report, map, image, presentation)

Digitization Project Worksheet

This form is intended as a resource for local government units to aid them in fulfilling their public records responsibilities. Local units may find it useful as a project planning tool or as a list of pre-project considerations. Additionally, using this form will ensure consistent handling by unit staff or contractors. ***It is recommended that this document be created or reviewed by the project head prior to starting the scanning process and then returned to that person upon completion.***

Each project should begin with a review the records schedule pertaining to the items that will be digitized to provide some basic guidelines about the specific types of documents that fall under a records schedule and should be migrated to digital format. There is a good chance that folders contain a mix of records that should be scanned and non-records that should not be scanned. Next, complete this document and review a sampling of the items to be scanned with the staff doing the scanning so records can be appropriately identified and managed throughout the process.

Section 1: General Project Information

Records to be scanned:

Responsible Department/Division:

Records Schedule or GRS which applies to above records:

Event and disposition:

Personally Identifiable Information (PII): Yes: ☐ No: ☐ Confidential: Yes: ☐ No: ☐

Size of collection to be scanned (boxes or file drawers): Date range of records content:

Describe the materials you plan to digitize. Review the schedule and the physical records and identify items that will be included in the scanning project.

What materials do you NOT plan to digitize? Review the schedule and the physical records and identify items that should not be included in the scanning project.

	Yes	No
Will the scanned items become the official record?	☐	☐
Will the paper records be destroyed after the scans are verified?	☐	☐
Will the records schedule need to be updated at the completion of this project?	☐	☐

Project Team

List key members of the project team below who can provide details and answer ongoing questions about the project:

Project Lead:

Scanning Lead:

Records Officer:

Content Owner:

IT Lead:

Other:

Section 2: Selection

Consider the following for the documents you have identified for scanning, are they:

	Yes	No
Frequently requested by the public or accessed by staff?	☐	☐
Currently active records according to their Records Schedule?	☐	☐
Scheduled to be destroyed or transferred more than 10 years from now?	☐	☐
In stable condition and will not be damaged by scanning or other handling?	☐	☐
Able to be accurately scanned with the equipment available?	☐	☐

If you have answered "No" to any of these questions, you may want to evaluate your project to make sure the benefits outweigh the costs.

Section 3: Access

	Yes	No
Are there any access restrictions associated with these records? If yes, describe the restriction:	☐	☐
Do the records contain a mix of restricted and unrestricted materials? If yes, describe:	☐	☐

Additional notes regarding access to these records:

	Yes	No
Will this content be provided online?	☐	☐

If "Yes", will this content be provided via intranet / internal network or via a public site?

Intranet: ☐ Public: ☐

How will you provide access to the digitized items? (Check all that apply)

Link to original	☐
Create smaller-sized user copy	☐
Thumbnail only	☐
Provide access on a variety of mobile devices	☐

Section 4: Creation

What are the formats of items in the collection that to be scanned? Please refer to the "*Digitization Guidelines*" document to fill in the below chart with the digital format and resolution you will use for each file type.

Current Format	Master Format	PPI	User Copy Format	PPI	OCR (Y/N)

Please provide any additional information below regarding the items that will be scanned or if additional versions (ie: thumbnails, access copies) will be needed.

	Yes	No
Have you acquired OCR software to generate searchable transcripts of printed text?	☐	☐
Do you need to manually generate transcripts of handwritten materials?	☐	☐
Do you have staff available to manually generate transcripts of handwritten material?	☐	☐

Section 5: File naming and organization

How will files be named for this project? (Provide examples of file names)

How will folders be organized for this project? (Provide an example of the folder structure)

Section 6: Metadata

Appropriate use of metadata is a critical component of a successful digitization project. Metadata is used for the identification, management, access, use and preservation of individual digital records and provides context for the collection as a whole. The below lists contain some of the most important metadata fields to collect in order to manage and locate digitized files. Please refer to the "*Digitization Metadata*" document for more complete descriptions of the below fields as well as additional metadata which may be useful to your project.

	Yes	No
Does metadata already exist for individual items?	☐	☐

If "Yes", is it: Detailed ☐ or Minimal ☐

Which metadata elements will be applied at the **collection level**?

- Division/Department
- Project title ☐
- Event date ☐
- Retention ☐
- Disposition ☐
- Access Restrictions ☐
- Description (summary of documents within this collection) ☐
- Keywords ☐
- Other Elements:

Which metadata elements will be collected at the **file level**?

- Title ☐
- Date created ☐
- Disposition date ☐
- Format (Internet Media type, e.g. image/jpeg, application/pdf, etc.) ☐
- Access Restrictions ☐
- Keyword ☐
- Unique ID # ☐
- Document Type ☐
- Other Elements:

Section 7: Maintaining Your Digital Collection

What will you use for short-term storage of your digitized files during the project?

- Local area network/local server ☐
- RAID device ☐
- External hard drives ☐
- Cloud storage ☐
- Other ☐
  - Describe:

How often will new digital items be moved to long-term storage locations?

- N/A - files are being moved directly to long-term storage ☐
- Daily ☐
- Weekly ☐
- Monthly ☐
- When project is completed ☐

What will you use for long-term storage of your digital master files? (choose at least two)

- Local area network/local server ☐
- RAID device ☐
- External hard drives ☐
- Cloud storage ☐
- Other ☐
  - Describe:

Directions for the Project Planning Worksheet

Section 1: General Project Information

This section provides high-level information about the digitization project

Documents to be scanned: Common organizational title of the group of documents that will be scanned.
For example: Executive Board Meeting Minutes; Highway Interchange Reconstruction Project

Responsible Department and Division: The Department/Division within your organization that is responsible for the management of the records.

Records Schedule or GRS which applies to above documents: Record the records schedule or GRS number and name which covers the documents. Also record the Event and Disposition which will trigger the disposition (destruction or transfer) of the records.

PII and Confidential: The answers to these two sections will dictate how you manage both the scanning process and the storage of the resulting scans. If the answer is "Yes" to either of these questions, determine if the scanning staff is qualified to handle these documents throughout the scanning process or if they need additional clearance. This information should also be added as metadata during the scanning process. The scans need to be securely stored in such a way that they are easily identifiable and safe from inappropriate access.

Collection Size: How big is the collection of records you are going to digitize (how many boxes or file drawers) and what is the Date Range or time span of the materials to be scanned? This is important because it helps to determine if any portion of those records has met its disposition date.

What materials do you plan to digitize (or not): The next two sections require you to evaluate both the materials that are covered (or not) under a specific Records Schedule **and** the actual materials that have been identified for scanning. These sections can be used to list both high level descriptions as well as specific documents that would be important to preserve the record or those that are not needed.
For example:
Keep: Final policy documents and reports, Form 685, Final meeting minutes
Don't Keep: any document drafts or copies; Form 223, working copies

You next need to determine what happens once the records have been scanned, will the Scanned items become the official record and replace the paper. What will happen to the paper after scanning? Will the paper records be destroyed after the scans are verified? Finally, Will the records schedule need to be updated at the completion of this project to accommodate changes in the record format and retention.

Identify the key personnel, or Project Team, for the scanning project. This information is important to document not only for questions during the project, but also for reference post-project in case any issues arise about what was scanned or how it was scanned.

Section 2: Selection

Use this section to continue reviewing the materials you've identified for scanning. If the answer is "No" to any of these, it might not make sense to scan these items.

You may want to start your project with collections that are in high demand from the public or require offsite access from staff. Scanning for either one of these reasons tend to make good business sense and lowers costs that can arise from having staff respond to public inquiries or down-time due to staff not being able to access critical documents while working out of the office.

For collections of records covering a long time span, it may make more sense to split it into items that you will hold for at least ten more years and those that would reach their disposition date before that and continue to hold those in paper format. This process also provides you the opportunity to dispose or transfer items that have already met their disposition date.

Section 3: Access

This section specifies not only ***who*** should have access to the scanned materials, but helps you define ***how*** you are going to provide access – two very separate but important questions.

Determining access restrictions for any of the material to be scanned will help you decide who (staff or external audience) will be able to access the records. You must also determine how you will provide access via internet or unit intranet and whether or not the original digitized file will be accessible or if you need to create an additional user copy.

Section 4: Creation

The standards you use to create your digital files will determine the long-term success and usefulness of your project. Refer to the "*Digitization Format Guidelines*" for details necessary to complete this section.

Section 5: File naming and organization

File and folder names are important pieces of metadata. Make sure to determine consistent file naming and organizational practices prior to beginning your project and document those decisions.

Section 6: Metadata

Good metadata makes scanned documents easy to find and use. Metadata should be added to each digital collection and file to ensure it is usable, saved correctly, and findable. Refer to "*Digitization Metadata*" for an explanation of metadata and the fields listed on this document.

Section 7: Maintaining Your Digital Collection

Storage: determining exactly where the digitized files will live and where copies will be kept is often overlooked in digitization projects. Generally, having two copies of digitized content, in geographically dispersed areas is recommended.

Resources referenced for this document

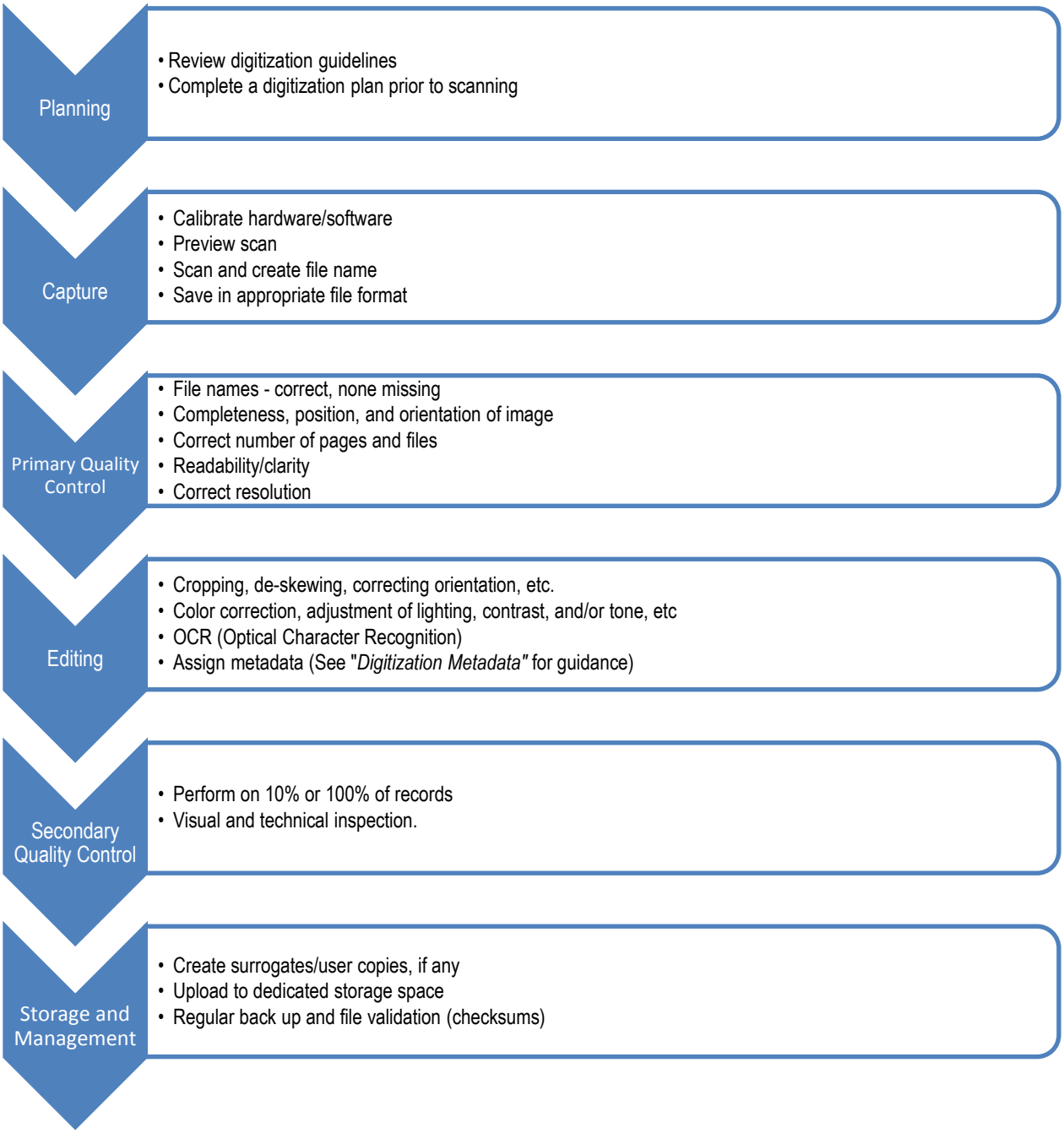
Recollection Wisconsin. *Digital Project Planning Worksheet*. 2013. recollectionwisconsin.org/wp-content/uploads/2013/10/Digital-Project-Planning-Worksheet.docx (accessed June 2017)

Digitization Workflow

Introduction

A documented workflow streamlines the many processes and tasks involved in a digitization project, facilitating consistency and reliable results, especially when dealing with a large volume of objects to be digitized. Documenting the workflow is also important for tracking material through each stage of the process and identifying errors or problems.

Basic Workflow



Planning

It is important to plan all steps of a digitization project at the outset to maximize efficiency and minimize the risk of making mistakes, potentially costing significant time and resources.

- Complete a "*Digitization Project Worksheet*" before any scanning begins.
- Determine a procedure for documentation in order to track records through each stage of the project. For example, a spreadsheet may be created in which a digital file name, the date an item was scanned, and the name of the person who scanned it may be recorded.

Capture/Scanning

The basic steps of scanning records are listed in the chart on page 1. Please see related sections in this document on hardware/software, minimum recommended standards, technical specifications, file naming conventions and file formats for guidance.

Primary Quality Control

The bulk of quality control tasks are performed during the scanning process, including:

- Confirming all file names follow a naming convention and no files are missing
- Verifying images are complete and positioned/oriented correctly
- Confirming the number of pages and files are correct
- Ensuring the images were scanned at an appropriate resolution. They should be readable and clear on both a computer screen and printed paper.

The "*Quality Control for Digitization Projects*" document will provide you with more detail on steps to be taken during this phase.

Editing

Common standards for the images will be determined during the planning phase, and may include cropping, de-skewing, or correcting the orientation of each image. Color correction or adjustments in lighting, contrast, and tone may be necessary, but it is important to keep in mind that the goal is to create an accurate representation of the item, and not correct permanent defects, such as stains. Conversion to Optical Character Recognition (OCR) may also be performed during this stage, which makes text in the document searchable.

Secondary Quality Control

A final inspection of the files should be conducted to catch any defects that may have been missed previously. If you have a small number of digitized records, it is recommended that you perform a high-level secondary quality control check and review all of the files. For large numbers of digitized records, a low-level secondary quality control check is more common. This entails reviewing 100% of the files at the onset of the project until the proof of concept is complete at which point many projects review only 10% of the remaining files. This includes a visual inspection of the images (e.g. as thumbnails), as well as checking that a file's name, size, resolution, and format are correct. Keep in mind that if you are digitizing records with the intention of disposing of the paper, you may want to plan for a much higher secondary quality check since the digital version will now exist as the legal copy.

The "*Quality Control for Digitization Projects*" document will provide you with more detail on steps to be taken during this phase.

Ongoing Storage and Management

If desired, user copies can be created from the masters and the digitized records will be uploaded to dedicated long-term storage. Ongoing management of digital files requires significant time and resources, which should be considered when planning a digitization project. Regular backups and file integrity testing should be performed in an effort to avoid files being lost or corrupt.

Resources referenced for this document

FADGI - Still Image Working Group. *Technical Guidelines for Digitizing Cultural Heritage Materials*. 2016. http://www.digitizationguidelines.gov/guidelines/digitize-technical.html (accessed June 2017).

University of Exeter. *Digitisation Workflow and Guidelines*. 2009. https://projects.exeter.ac.uk/charter/documents/DigitisationWorkflowGuidev5.pdf (accessed June 2017).

Quality Control for Digitization Projects

Quality Control (QC) is a process that verifies the quality, accuracy, and consistency of digital images to ensure the products match the specifications outlined in the project plan. **The Quality Control process is particularly important for scanning projects where paper records will be destroyed after scanning.**

Quality control standards should be developed and widely shared at the beginning of the project and need to remain consistent throughout. Additionally, the specific QC responsibilities of individual staff members must be clearly communicated.

Developing a QC program:

- *Review the project goals*
  The goal of the scanning project will impact the QC criteria. For example, if the goal is to create a replica of the paper record, the digital images should look as close as possible to the original item. However, if the goal of the project is to create the best quality digital images, then ensuring the digital images accurately represent the originals will not be part of the QC process.
- *Identify the products to be evaluated*
  Determine if QC will be performed on the master files and any copies, printouts, metadata, and/or the OCR'ed (Optical Character Recognition) files.
- *Control the QC Environment*
  - A number of factors can impact the QC process including viewing conditions, monitor calibration, and color management.
  - View scanned documents on a monitor and as print outs during the QC process. Documents must be legible both ways.
- *Determine what percentage of files will be reviewed*
  - At the outset, review 100% of files, after the proof of concept many projects review 10% of files.
  - Clearly define baseline characteristics for "acceptable" and "unacceptable" digital images.
  - Determine the number of errors that will render a batch unsuccessful.

Quality Control Checklists

There are two stages of the QC process. The initial is performed by the scanner operator as a regular part of their workflow. The second QC process should be performed by separate staff, after a batch is scanned, at a dedicated QC station.

Scanner Operator QC Responsibilities

- File Specifications
  - Is the file name correct?
  - Is the file format correct?
  - Are the pages in the correct order?
- Image Quality
  - Is the image rotated or backwards?
  - Is the image skewed or off center?
  - Is the content readable?
  - Is the scan "clean"? (no hair, dust, specks of paper, or finger prints visible in the scan)

Secondary Quality Control

- Image Quality
    - Are there moiré patterns? (wavy lines or swirls, usually found in areas where there are repeated patterns)
    - Are digital artifacts present? (such as very regular, straight lines across picture)
    - Is the image too light or too dark?
    - Are details lost in highlight or shadows?
    - Does the image have clean edges, clear contrast, and legible text?
    - Are individual pixels apparent to the naked eye?
- File Properties
    - Is the image the correct size/resolution in long dimension?
    - Is the bit depth correct?
- Is the OCR error free?
- Are all files and pages accounted for?
- Were any items skipped between scanning sessions?
- Is the metadata accurate and complete?

Resources referenced for this document

FADGI - Still Image Working Group. *Technical Guidelines for Digitizing Cultural Heritage Materials*. 2016. http://www.digitizationguidelines.gov/guidelines/digitize-technical.html (accessed June 2017).

Digitization Guidelines

This document provides recommendations for the format and resolution of digitized content based on length of time the item will be kept and its purpose. Items that are destined to come to the State Archives or held long-term or permanently by the local government unit will have specific formats and higher resolutions than content that will only be held for a short period of time before deletion, according to records schedules, or are intended to be user copies of a permanent record. Also included here are brief summaries of digitization terms that will be helpful for you to understand in developing your project and filling out the Digitization Project form. This is not meant to be an in-depth explanation of these terms which can be found in more detail in resources listed in the bibliography of the main "*Digitization Project Guidance*" document.

Resources for Recommendations

This document contains recommendations for scanning the most common types of records which you may encounter during your digitization projects. The below recommendations primarily come from the National Archives and Records Administration (NARA) as they have a comprehensive guide of scanning specifications for a wide range of items that are intended for permanent retention and access. This is supplemented by BCR's CDP Digital Imaging Best Practices and FADGI's Technical Guidelines for Digitizing Cultural Heritage Materials. If you have records that don't meet any of the below categories, you may refer to one of these resources or contact the State Archives staff at the Wisconsin Historical Society for additional support.

Digitization Recommendations

For each of the below categories, the highest recommended resolutions are for items that will be held permanently at a local unit or transferred to the State Archives for permanent retention according to records schedules. For records that will be held at the local unit and eventually deleted according to records schedules, the decision as to whether the resolution is High or Minimal will be determined by the unit through an evaluation of the records to be scanned. The PPI resolutions listed in each category below are the minimal recommendations for that document. A unit can go higher if they so choose.

High Resolution

- Records that will be transferred to the State Archives at the end of their retention period according to record schedules.
- Records that will remain permanently with the local unit based on approved record schedules.
- Unit records that will remain with the unit for some period of time prior to deletion according to records schedules and are of significant value to the unit's mission and reputation. The high resolution option should be used if the act of not having a high-quality scan could have legal ramifications or have a negative impact on the unit's reputation.
- Because of the high capture specifications, this file could confidently take the place of the original record if the original record no longer exists. This option should be used where there is a need to ensure that all significant visual information is carried forward.
- Files must be uncompressed with no application/product dependencies or extensions.

Minimal Resolution

- Records that will remain with the unit for some period of time prior to deletion according to records schedules and are **not** of significant value to the unit's mission and reputation. These are records that are traditionally of low use by staff and the public after they have reached the inactive stage.
- Because the capture/transfer specifications are closer to or at the minimum recommended level, this file would not be an optimal substitute for the original record if the original record is no longer viable.

Access Copies

- Unit records that will remain in paper for its retention period but will be distributed electronically.
- Unit records that have a high or mid resolution master copy, but need a lower resolution, smaller sized copy for distribution.
- This file is retained only as long as needed to satisfy a specific "on demand" distribution need. It is not retained after delivery to the requesting customer and is not actively managed over a long term.

Text Documents

Scanned text is a photograph of a printed page produced by digital camera or scanner.

- Bitonal (1-bit black and white) images must be scanned at 300-600 ppi. Scanning at 600 ppi is recommended. This is appropriate for documents that consist exclusively of clean printed type possessing high inherent contrast (e.g., laser printed or typeset on a white background).

- Gray scale (8-bit) must be scanned at 300-400 ppi. Scanning at 400 ppi is recommended. This is appropriate for textual documents of poor legibility because of low inherent contrast, staining or fading (e.g., carbon copies, thermofax, documents with handwritten annotations or other markings), or that contain halftone illustrations or photographs.

- Color (24-bit RGB) must be scanned at 300-400 ppi. Scanning at 400 ppi is recommended. Color mode (if technically available) is appropriate for text containing color information important to interpretation or content.

Content Description	Minimum Scan Resolution	Format	Notes
High Resolution ***(NOTE: smallest character being larger than 1.00mm (or 2.835 point/font))***			
Bi-tonal (1–bit black and white)	600 ppi	Tiff JPEG2000 PDF/A	- Should be uncompressed - Should have OCR
Gray scale (8-bit)	400 ppi 8-bit grayscale or 24-bit color		
Color (24-bit RGB)	400 ppi where original is up to 34" x 55" 300 ppi where original is between 34"x55" and 46"x73"		
Mid Resolution			
Bi-tonal (1-bit black and white)	300 ppi for records with smallest significant character of 2.0 mm or larger	TIF PDF/ A PDF JPEG (high resolution) PNG	- Uncompressed or lossless compression - Should have OCR
Gray scale (8-bit)	300 ppi for records with smallest significant character of 1.5 mm or larger		
Color (24-bit RGB)	300 ppi for records with smallest significant character of 1.5 mm or larger		
Access Copies ***(NOTE: OCR works best at 300ppi or higher)***			
Bi-tonal (1–bit black and white)	150 ppi bitonal	TIF JPEG2000 PDF/A PDF JPEG PNG GIF	- Can be lossy - OCR recommended but depends on use
Gray scale (8-bit)	200 ppi bitonal or grayscale		
Color (24-bit RGB)	300 ppi 24-bit color		

Still Photographs-Prints

When converting analog materials, units should digitize to standards appropriate for the accurate preservation of the original image. Photograph digitization standards are more complex than paper scanning standards since the intent is to maintain the smallest detail of the photo. Size of the photo, quality, medium, and color all play a role in determining the best resolution for the scan.

300 ppi is generally considered the bare minimum to reproduce a photo well at the size of the original. As such, the lowest resolution for scanned images is listed at 400 ppi. The below recommendations assume you match the original size with no magnification or reduction of the physical photo. The minimum resolution section can be used by a unit to produce a decent quality image when responding to reference and reproduction requests. The access section is best used for web distribution of unit photos.

For the categories below, pixel array category defines the minimum resolution of pixels across the long dimension of the object to be scanned. Generally speaking the larger the image, the larger the pixel array.

Photo Size	Bit Depth	Pixel Array	Resolution	Format
High Resolution ***(unit copy)***				
Still photo prints (B&W, color, and monochrome) with a size range of 8" x 10" or smaller (square area smaller than or equal to 80 square inches)	B&W 8-bit grayscale Color & Monochrome 24-bit RGB	4000	Examples: 4" x 5" or 3.5"x 5" 800 ppi 5" x 7" 570 ppi Up to 8" x 10 " 400 ppi	TIF
Still photo prints (B&W, color, and monochrome) with a size range of larger than 8" x 10" and up to and including 11" x 14" (square area larger than 80 square inches or smaller than 154 square inches)		6000	Examples: 8" x 10 " 570 ppi Up to 11" x 14" 430 ppi	
Still photo prints (B&W, color, and monochrome) with a size range of larger than 11" x 14" (square area equal to or larger than 154 square inches)		8000	Examples: 11" x 14" 570 ppi Larger sizes will range down to around 400 ppi	
Minimum Resolution ***(reference and reproduction requests)***				
Still photo prints (B&W, color, and monochrome). Match original size with no magnification/reduction	B&W 8-bit grayscale Color & Monochrome 24-bit RGB	3000	Examples: 4" x 5" 600 ppi 8" x 10 " 300 ppi	TIF JPEG PNG
Access ***(web copy)***				
Still photo prints (B&W, color, and monochrome). Match original size with no magnification/reduction	B&W 8-bit grayscale Color & Monochrome 24-bit RGB	800 - 1200	72 ppi – 200 ppi	GIF (for smaller originals) JPEG (for larger originals, medium →high compression)

Architectural and Engineering Drawings

Records in this category cover a broad range of technical drawings that are intended for construction or mechanical purposes. Since these items are typically oversized, the scans of the originals have the potential to be very large in size. It is highly likely you will need a large format scanner and specialized handling to legibly capture small details which are prevalent in these materials.

NARA's scanning recommendations for Architectural and Engineering Drawings follow the same rules as the text documents in the above section.

Maps and Charts

Records in this category are graphic representations of selected geographical features and include such items as atlases, relief models, photomaps, hydrographic/nautical charts and cartograms. These items are typically oversized, and the scans of the originals have the potential to be very large in size. It is highly likely you will need a large format scanner and specialized handling to legibly capture small details which are prevalent in these materials.

NARA's scanning recommendations for Maps and Charts follow the same rules as the text documents in the above section.

Terminology

PPI and DPI

Pixels per inch (PPI) refers to the number of pixels captured in a given inch and is used when discussing scanning resolution and on-screen display. When referring to digital capture, PPI is the preferred term, as it more accurately describes the digital image. Dots per inch (DPI) is more often used when discussing the optical resolutions for images and hardware. This refers to how many dots per inch a printer puts on paper when printing it out or across a computer monitor. For the purposes of this document, PPI will be used for the recommended capture resolutions of various objects.

Resolution

Resolution refers to the quality of an image. The higher the PPI, the more accurate rendering of the original document is created. Images at low PPI may look fine on a computer screen, but will be illegible when printed out. High resolution images capture more information about the original document and therefore take up more server space in terms of storage. PPI will also vary depending on the object being scanned. An 8x10 photograph will most likely require a different resolution than an 8x10 text document.

Compression

Digitized image files have the potential to take up a lot of room on a drive. Uncompressed files generally take up the most space and are often considered the "raw" image that comes out of the scanning process or camera. Many scanning operations will perform some sort of image compression to reduce the file size for storage, processing and transmission. There are two key compression schemes that you will encounter during the digitization process but they have very different results.

Lossless compression

Lossless compression abbreviates the underlying binary code without discarding any of the information so when it opens, it can reconstruct the code and the image will be a perfect copy of the original. Using this technique, every single bit of data that was originally in the file remains after the file is uncompressed. While items stored with lossless compression tend to be larger than those stored with lossy compression, this is preferable for items that will be remain a long time at the unit or transferred to the archives for permanent retention.

Lossy compression

With lossy compression, the file size is reduced during the compression process by permanently eliminating information from the image. When the file is opened (uncompressed), only a part of the original information is still there. Although the discarded information may be invisible to the human eye when viewing the image, a loss of quality has occurred. Each time a lossy image is manipulated or edited, the quality of the image further decreases. Generational loss over time is the primary reason that lossy compression may be a good option for user copies or documents of a short-term duration, but not for an object that will be the official record for long periods of time or are of permanent duration.

Formats

The use of the digital images you are creating will determine which formats to choose when digitizing. While there are seemingly endless formats to choose from, only a small handful are currently viewed as being sustainable for preserving digital information as an authentic resource for future generations. These formats tend to be widely adopted; well documented, non-proprietary and has no external hardware or software dependencies. It is considered best practice to use these formats as they tend to be supported across most systems and and less likely to become obsolete over time. Choosing the correct format is part of a strong strategy for future preservation actions on the records such as the adoption of new technologies or the migration to new formats. A list of recommended formats for digitization projects and their uses are below.

GIF: GIFs are best used to store screen-quality images that do not contain many colors. GIF files are typically very small, but cannot reproduce the range of colors necessary to reproduce photographic images like JPEG. As such, this format is recommended for user copies of text records

JPEG: This lossy image file format is commonly used for photographs and other complex images. It is a great option for user copies on a website since you can create smaller sized images by reducing the image quality. This format is not recommended for text or line drawings.

JPEG 2000: JPEG 2000 is another lossless format that is increasingly being used as a long-term / archival format and has been adopted as such at the National Archives and has been published as an ISO standard. A downside to this format is that access to JPEG 2000 files may require a special reader on the user-end but access is becoming easier as JPEG 2000 becomes more ubiquitous. Also large in file size, this will take up more server space than most other formats. This format is acceptable for items being transferred to the State Archives for permanent preservation although you may prefer one of the other options for content that will stay with the unit. JPEG 2000 images can be lossy or lossless depending on the purpose of the final document. This format is not widely supported by web browsers and is not generally used on the internet.

PNG: PNG (Portable Network Graphics) is an open standard graphics file format that allows accurate rendering of greyscale and RGB color objects. PNG format can be used to store high-color images, which means it is also suitable for storing photographic content. This format is not widely implemented.

PDF: PDF (Portable Document Format) is best used to store vector-based graphics (i.e. graphics drawn using lines and curves rather than pixels). Vector graphics stored in PDF format will be much smaller, will read more cleanly, and any included text will be searchable. Equations, charts, and diagrams that combine text with vector-graphics are particularly appropriate to store in PDF format.

PDF/A: PDF/A is a special type of PDF format meant for documents needing to be preserved for long-periods of time. This format is an ISO standard which helps guarantee that it will be accessible for the foreseeable future as technology advances.

TIFF: TIFF (Tagged Image File Format) is a is a stable, well-documented, widely adopted, uncompressed file format and is a preferred format for archival and master images. This is an excellent choice for digitized content that will be replacing the paper records in a record schedule and will be preserved for a long-period of time → permanent. The downside, of course is that they will take up more server space than most of the other options. TIFF images can be lossy or lossless depending on the purpose of the final document. Its file extension is .TIF

Modes of Capture

Most imaging equipment offers four modes for capturing a digital image. The mode you choose is dependent on the content that will be scanned.

Mode of Capture	What it is	Best for
Bitonal	One bit per pixel representing black and white.	Best suited to high-contrast documents such as printed text
Greyscale	Multiple bits per pixel representing shades of gray. Grayscale has 1 color channel – typically black or white.	Grayscale is suited to continuous tone items, such as black and white photographs. Standard is either 8-bit or 16-bit grayscale images (more shades)
RGB (color mode)	Multiple bits per pixel made by combining 3 color channels: Red, green or blue.	Color capture is suited to items with continuous tone color information. Standard is either 24-bit color or 48 bit color images (more shades)
CMYK (color mode)	Short for Cyan-Magenta-Yellow-Black and is the standard model used in offset printing for full color documents. This mode is best for printing and should never be used for archival purposes.	Best when printing color images (not archival)

Bit Depth

Bit depth quantifies how many unique colors are available in an image's color palette in terms of the number of 0's and 1's, or "bits," which are used to specify each color. The lower the bit depth, the fewer colors are available to represent a scanned object. So for an object with a bit depth of 1, each pixel in the image will have two possible values – either white or black. As the bit depth increases for both grayscale and colored objects, more colors are added so highly detailed images with lots of color variations, gradients, shadows, etc. become clearer and a truer representation of the original object. A higher bit depth also means that more information is being captured and stored for the image which can lead to a larger file size.

Bit Depth	# of Colors	Colors Seen	How it works
1 bit Bitonal	2^1 or 2 colors	Black and White	An image with a bit depth of 1 has pixels with two possible values: black and white
8 bit Grayscale	2^8 or 256 colors	Black, white , gray	An image with a bit depth of 8 has pixels with 256 possible values: black, white and various shades of gray
24 bit Color	2^{24} or 16,777,216 colors	Colors across the Red, Green Blue Spectrum	RGB images are made of three color channels. An 8 bit per pixel RGB image has 256 possible values for each channel,

Resources referenced for this document

CDP Digital Imaging Best Practices Working Group. BCR's CDP Digital Imaging Best Practices V.2.0. 2008. http://sustainableheritagenetwork.org/digital-heritage/bcrs-collaborative-digitization-program-digital-imaging-best-practices-version-20 (accessed June 2017).

FADGI - Still Image Working Group. *Technical Guidelines for Digitizing Cultural Heritage Materials*. 2016. http://www.digitizationguidelines.gov/guidelines/digitize-technical.html (accessed June 2017).

Library of Congress. *Sustainability of Digital Formats: Planning for Library of Congress Collections*. March 2017. https://www.loc.gov/preservation/digital/formats/index.shtml (accessed June 2017).

National Archives and Records Administration (NARA). *Preservation -- Original Record Type*. 2017. https://www.archives.gov/preservation/products/definitions/original.html (accessed June 2017). (Congress 2017)

Storage and Maintenance

Storage and Maintenance

Where and how digital files will be stored is an important subject that needs to be discussed at a project's outset. Long term-storage capacity, file security and maintenance, number of copies, back-up procedures, and file and technology migration plans all become more important the longer the files are to be retained. It is essential that you discuss your post-digitization storage and preservation needs with whomever will be responsible for maintaining the digitized records **prior** to beginning any digitization project.

Agencies are responsible for ensuring that records remain "accessible, accurate, authentic, reliable, legible, and readable" per Administrative Rule 12. Addressing these key concepts ensures that digital records will be properly stored and maintained. Specific protective measures may include, but are not limited to:

- Establishment of security protocols, and approved administrators and users.
- Employment of system checks and error-checking utilities.
- Implementation of back-ups and disaster preparedness measures.

Planning for Storage of Digitized Records

While storage of the records comes at the end of a digitization project, this is a critical topic that needs to be established from the beginning to ensure your agency can commit the staff and resources to the long-term preservation of the digitized records. This is especially important for records that will be held permanently in digital format where the paper has been discarded. Your discussion should include:

- **Where records will be stored both during and after the digitization project.** It may make sense to have a limited, temporary storage area for the project while the records are being digitized and QA'd and then transfer to longer-term storage.
- **How long the records need to be kept.** This is important both to plan for storage space as well as to plan for the destruction or transfer of records from your organization. For records that will be stored in a content management system, the system should be help you manage the disposition of those records through technological reminders you place on a folder. For records that are stored on a network drive, you will need to establish some other way of tracking the disposition date of the records which may implement some combination of inventories and folder naming conventions to help you remember when the records need to be disposed of.
- **Plans to address management of files over time, including:**
    - File formats become obsolete over time. Today's formats will need to me monitored and migrated as needed
    - Storage systems become obsolete over time and digital materials should be migrated regularly between storage systems.
    - Migrations to new software platforms when current technology is no longer supported.

Storage Considerations

- What is the current storage capacity of your organization?
    - Is there currently enough space for the project?
    - Will there be space for future projects?
- What is the current file management system?
- Are all your files stored in one place or do you have geographically disperse storage locations?
    - Are you able to produce, manage, and store back-up copies of the files or will you need outside help?
    - How often are backups done?
    - Who is in charge of them?
    - How are they documented?
- Do you have a disaster recovery plan?

Discussing these topics and setting consistent expectations between IT, agency management, project staff, and content owners will go a long way toward making your project successful. The budget, staffing and resources of your organization will help you determine how you can best accomplish these goals.

Preservation vs. Backups

Today's IT systems can provide an excellent option for mass digitization projects. They are highly scalable, and can provide quick access to digitized content, however, traditional IT systems and backups are primarily geared toward active content. The transition to the long-term / permanent preservation of digital records requires a higher level of IT management that should be understood by those managing the digital files to help protect your organization in case of natural disasters, cyberattacks, computer hijacking, accidental deletion, or file corruption.

The National Digital Stewardship Alliance has developed a technology-neutral Levels of Preservation framework to provide guidance on preserving digital content at four progressive levels. It ranges from Level 1 which provides the bare minimum of requirements to minimally protect your data to Level 4 which provides guidelines that will provide the highest likelihood of successful long-term preservation. In reviewing this document, you will likely find your organization is at different points in each of the 6 categories, but you can use this framework to identify where you are currently operating and work toward Level 4 to the greatest extent possible. Of particular importance to the IT / Storage aspect are the sections related to "Storage and Geographic Location", "File Fixity and Data Integrity" and "Information Security".

Storage and Geographic Location

In a traditional IT paradigm, backups periodically perform either incremental or full scheduled backups of files. If at some point a file becomes corrupt, that corrupt file will become part of the backup and by the time it is discovered, there may not be an uncorrupted version available. This is particularly risky for files that are being held permanently by an organization and for this reason IT backups do not take the place of keeping multiple copies of digitized content.

The 3-2-1 rule is often cited as best practice for storing digital files. Ideally, your organization should have 3 copies of each master digital item scanned at the resolution described in the "*Digitization Guidelines*". Realistically, your organization needs to evaluate the costs, staffing and technological infrastructure to evaluate what your capabilities are. It is strongly recommended that the 3-2-1 rule is used for permanent digital records to minimize the risk of losing them over time.

- Copy of a Master Image
  - An exact duplicate of the file at the image's creation.
  - Can be kept in "dark storage" or otherwise not accessed unless there is an issue with the Master Image.
  - A copy of a master image should be migrated and refreshed the same as the Master Image.

- Backup of a Master Image (IT generated)
  - Does not take the place of records management; records should not be retrieved from back up tapes for use or to fulfil an open records request.
  - Backups periodically overwrite files, if a Master Image becomes corrupt, at some point, the backup will rewrite the "good" master with the corrupt file.

Traditional IT backups are appropriate for derivatives of master files that are created at a lower resolution for access copies or thumbnails.

File Fixity and Data Integrity

Every file is made up of bits and bytes that are arrange in a certain way that produces what you see on your computer screen. One of the greatest threats to digital files is the loss of those bits and bytes over time that lead to the corruption of the files so you can no longer access it. This is very different from traditional paper files. If a page in a paper document is damaged in some way, you can likely still read the other pages….there is some loss, but it is minimal. With a digital document, if you lose enough bits and bytes, the entire document is gone. Again, a high risk when you are holding the only record in digital format. One way to monitor file corruption is through performing integrity checks which monitor files over time. These programs calculate a checksum or digital signature of the original file. If over time, the file was to change in some way (corruption, manual change by a person, etc), the program will notify you that something has changed and you can replace it with one of your copies. If you have a robust IT network, you may already have system tools available to perform integrity checks on a periodic basis. Similarly, if one of you geographically distributed copies are with cloud providers, that may be a service they can provide you. If you need to perform integrity checks on your own, there are several free tools that can be implemented at your organization like AV Preserve's Fixity (https://www.avpreserve.com/tools/fixity/).

Types of Storage

Determining the best storage solution for your agency's digital assets involves evaluating likelihood of access, overall cost in maintaining them, and how access will be provided to the digitized records. Every storage type has advantages and disadvantages.

- **Online Storage:** Allows immediate access to records to anyone on your organization's network. Online storage maintains the greatest functionality but is more expensive than other storage options.

- **Near-line Storage:** Uses a system that is not a direct part of your network, but that can be accessed through your network. Files are accessed using an automated process that selects the correct disk/tape from a disk/tape library and makes it accessible.

  Near-line storage is less expensive than online storage, but requires extra time to manipulate both the files and media to access the records. Near-line storage is often used for backups as large quantities of data can be managed quickly.

- **Offline Storage:** Files are not accessible through your network. They may be saved on removable media like external hard drives or magnetic tape. Offline storage is a good option for records that do not need to be accessed frequently.

- **Storage with a Third-Party**: You may also consider using a third-party storage that can store, access, and deliver records to you. As part of their offerings, they may also have the infrastructure to perform integrity checks and store multiple copies of your files in geographically diverse locations. The Wisconsin Public Records Board has issued guidance on the Use of Contractors for Records Management Services and Guidance on the Use of Contractors For Records Management Services in Cloud Computing Environments. The State of Wisconsin Department of Technology (DET) has already vetted cloud service providers and may be a good resource for exploring this option.

A note about removable storage: Removable storage devices include CDs, DVDs, thumb drives, and other types of technology. These devices should NOT be used to store Master Image Files as they have the lowest life expectancy and highest fail rates.

Single external USB storage devices are not ideal being the sole copy of Master File images. Multiple external devices can be used as part of your overall storage plan for smaller institutions in conjunction with other options, but these must be rotated out and replaced at least every 5 years so that cost must be accounted for in long term plans.

Which Option Should I Choose?

Traditional IT procedures tend to backup all items, regardless of content, under a set time schedule. This is done by necessity given the amount digital content they are required to manage. Long-term preservation of digital records may require that different records may be treated differently over time. As stated earlier, it is strongly recommended that the 3-2-1 rule is used for long-term and permanent digital records to minimize the risk of losing them over time. Your organization may also want to implement a more robust plan for records that may not be permanent, but would result in negative public backlash should they be lost. This could be achieved through a combination of copies via on-line, near-line and dark storage with a cloud provider. Records that are digitized to provide access and the paper still remains could have a less robust storage plan since the records could be re-digitized. The storage of the records and any derivatives should be discussed with all relevant parties and documented as part of your digitization plan.

Maintenance

Long-term storage of digital objects demands greater planning and attention than the storage of paper records. Additionally, the expense of storing records electronically exceed the costs associated with storing paper records once the costs of multiple copies and derivatives are calculated.

- All digital media and hardware have a limited life expectancy based on factors such as manufacturing quality, age and condition, handling and maintenance, frequency of access, and storage conditions.
- Hardware and software may be subject to rapid advances in technology or changes in standards.

Due to the life expectancy of both software and hardware, no single digital storage medium or format can be considered "permanent" for the long term storage or preservation of records. The most generous estimate of physical obsolescence is thirty years, while technological obsolescence can be expected within five to ten years. Therefore, assume files will need to be migrated to a new storage medium at regular intervals and periodically to a new format.

Retention

Images must be stored, maintained, and remain accessible for the entire length of the required retention period. When designing your storage and maintenance plan for digitized records, you need to take into account the retention "trigger." Retention triggers can be easy to determine and implement like the end of a calendar year or the close of a fiscal year. Others are more difficult to determine and implement such as the close of a case or issuance of a final report.

Event based retentions like these can pose difficulties for systems monitoring retention and disposition. They add another layer of necessary metadata, and they require someone to designate the event date in the system so that records can be disposed of or transferred according to record schedules.

Disposition

The majority of records held by local governments will eventually reach a disposition date that involves either destroying the electronic records or transferring them to the Wisconsin Historical Society for permanent preservation. Disposition is a vital step in the records lifecycle and cannot be overlooked when planning and implementing a digitization project. Planning for the disposition of non-records that an agency decides to scan is equally important, and perhaps more challenging as this content does not have a schedule-driven disposition date. It is critical that the project plan includes a review date at a minimum to ensure the content is periodically reviewed and deleted when no longer of use to the agency.

Failure to plan for the disposition of digitized content (both records and non-records) will result in an ever increasing number of items, which will exponentially increase your agency's storage costs over time, make it more difficult to find records, and burden IT staff with the migration of digital content through storage and format migrations over time. Local governments need to remember that digital records are still public records and will need to go through the written notification process before they can be destroyed. Please consult with the Wisconsin Historical Society regarding the proper procedures on notification and destruction of public records.

Resources referenced for this document

Digital Preservation Coalition. Digital Preservation Handbook - Storage. 2017. http://www.dpconline.org/handbook/organisational-activities/storage (accessed June 2017).

LeBlanc, Suzannee. Digital Preservation and Technological Obsolescence: A Risk Assessment Strategy. 2011. http://fiq.ischool.utoronto.ca/index.php/fiq/article/view/15411/12442 (accessed June 2017).

Minnesota Historical Society. Electronic Management Guidelines. March 2012. http://www.mnhs.org/preserve/records/electronicrecords/erpreserve.php (accessed June 2017).

National Digital Stewardship Alliance (NDSA). Levels of Digital Preservation. 2013. http://ndsa.org/activities/levels-of-digital-preservation/ (accessed June 2017).

National Ditital Stewardship Alliance (NDSA). The NDSA Levels of Digital Preservation: An Explanation and Uses. 2013. http://www.digitalpreservation.gov/documents/NDSA_Levels_Archiving_2013.pdf (accessed June 2017).