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