

LOCAL HISTORY DIGITIZATION PROJECT MANUAL

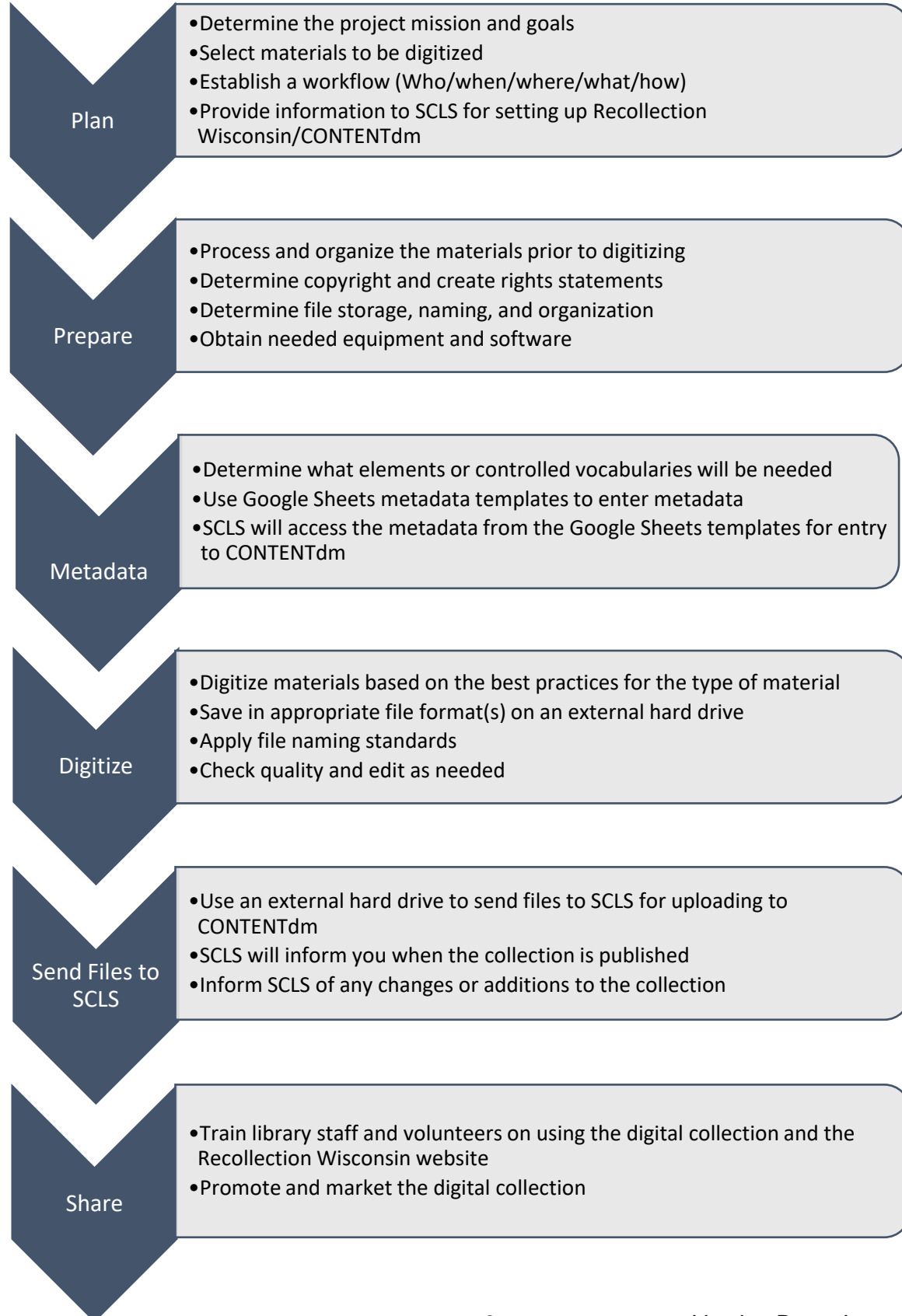
South Central Library System
4610 S. Biltmore Lane, Suite 101
Madison, WI 53718
608-242-4866
tramski@scls.info



Contents

Project Overview	3
Planning	4
Recollection Wisconsin	5
Selecting Materials	6
Preparing	7
Process and Organize Materials	7
Copyright and Rights Statements	7
File Storage, Formats, Directories, and Names	13
File Storage	13
File Formats	14
Directory Organization	14
Folder and File Names	17
Software	19
Metadata	19
Digitizing	24
Equipment	24
Indus BookScanner 9000s	24
Digitization Kits	25
Digitizing Print Materials	25
Digitizing Audio Materials	29
Digitizing Video Materials	29
Creating Transcripts for Digitized Materials	30
Provide Information to SCLS for Entry into CONTENTdm	32
Promote and Share the Digital Collection	33
File management	33

Project Overview



Planning

The first step in the digitization project will be to develop a vision of the project outcomes and to establish the mission and goals of the project. The Recollection Wisconsin Digital Project Planning Worksheet and/or Recollection Wisconsin Digital Project Planning Template can be useful for this.

<http://recollectionwisconsin.org/wp-content/uploads/2012/11/DigitalProjectPlanningWorksheet2016.pdf>

The following will need to be considered in the planning phase of a digitization project:

- Create a workflow
- Define a timeline/time dedicated to working on the project
- Determine who will fill key roles in the project
- Select the materials to be digitized
- Inventory the selected materials
- Plan for and acquire any needed equipment or materials (archival supplies)
- Metadata elements that will be needed to describe the items in the collection
- Digitizing best practices (based on format of item)
- File storage, directory organization, and file naming standards
- Technical support (hardware maintenance, software installation)
- Marketing and outreach related to new collections
- Ongoing collection management (quality checks, backups, archiving files, and migration of files)

The Digital Public Library of America (DPLA) has a series of training videos available that provide information about the steps involved digitization projects.

- Planning for Digitization <https://www.youtube.com/watch?v=l2eUVPB6X3g>
- Selecting Content for a Digitization Project <https://www.youtube.com/watch?v=01yQz4oztLo>
- Understanding Copyright <https://www.youtube.com/watch?v=AurzAmj4SvM&t=1s>
- Using Metadata to Describe Digital Content https://www.youtube.com/watch?v=RqgFLj_yspk
- Digital Reformatting and File Management <https://www.youtube.com/watch?v=CiTbXHqws7A>
- Promoting Use of Your Digital Content https://www.youtube.com/watch?v=_rzcamxx8b8
- Making Audio Collections Accessible <https://www.youtube.com/watch?v=1EESR5YqYMU&t=12s>

Recollection Wisconsin

For this project, digitized materials will be added to the Recollection Wisconsin website. The library will be considered a Content Provider to the Recollection Wisconsin website, which is a Service Hub of the Digital Public Library of America (DPLA). The collection will be hosted by the Milwaukee Public Library, using the digital asset management platform CONTENTdm. The Digitization Assistant at SCLS will upload the collection's image files and metadata to CONTENTdm. The collections will be harvested quarterly (in June, September, December, and March) by Recollection Wisconsin.

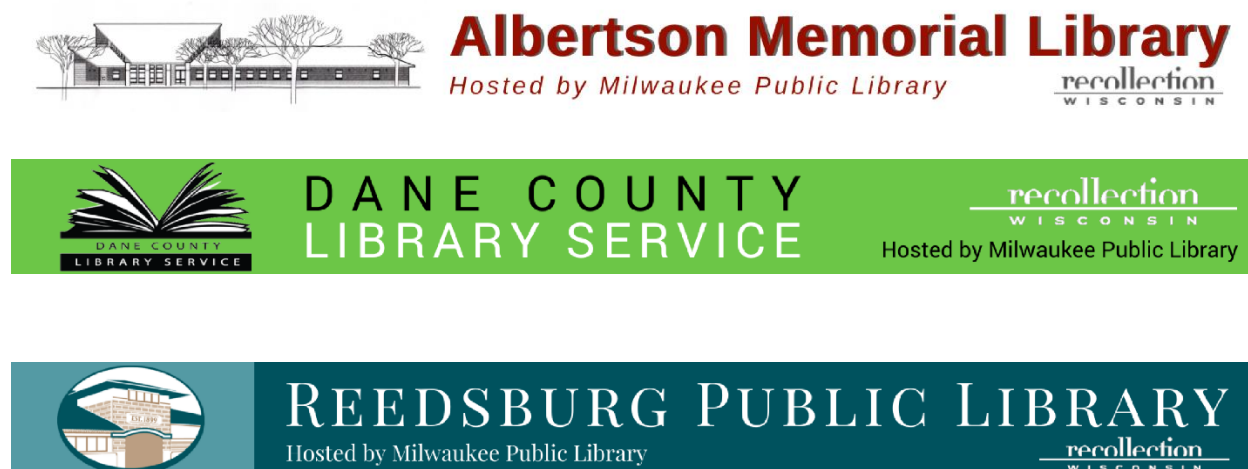
Review the following information from Recollection Wisconsin and be mindful of it throughout the project.

- Recollection Wisconsin Contributor Guidelines
<http://recollectionwisconsin.org/organizations/guidelines>
- Recollection Wisconsin Collection Policy
<http://recollectionwisconsin.org/wp-content/uploads/2016/07/Recollection-Wisconsin-Collection-Policy.pdf>

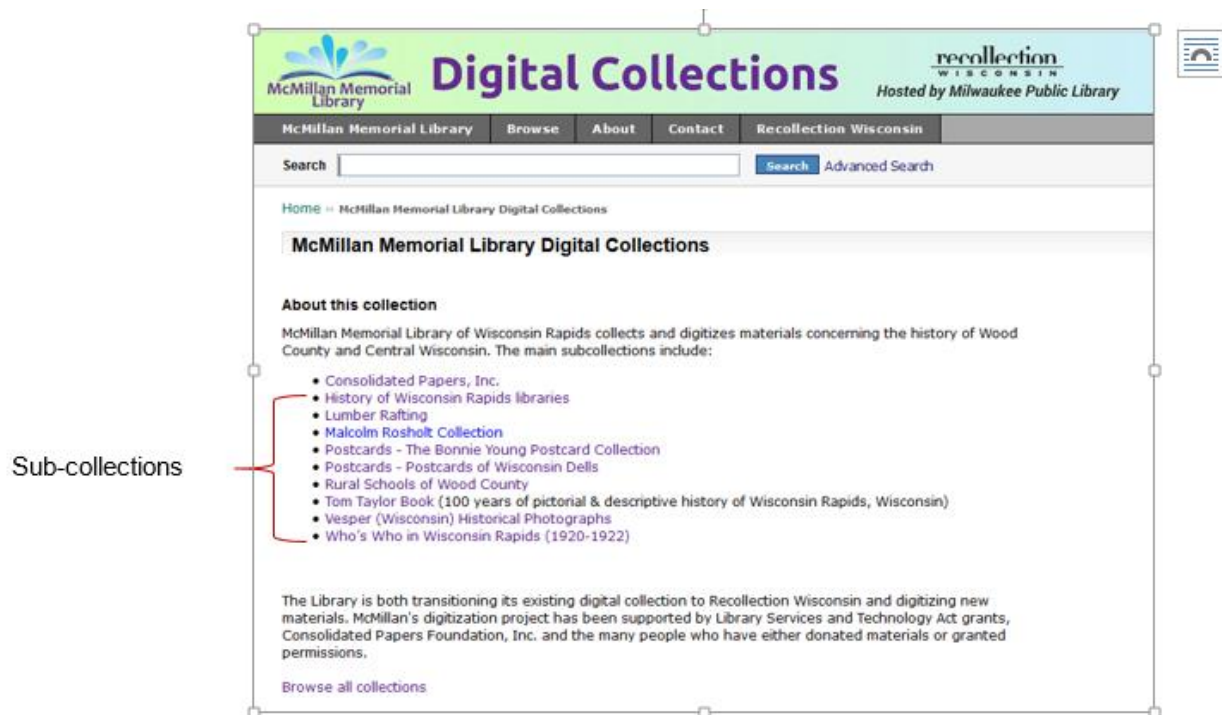
The following will need to be provided to the SCLS Digitization Assistant:

1. A nickname for the project that will be used in a short, permanent URL within the Milwaukee Public Library's CONTENTdm server.
2. A URL or email address for a "Contact Us" link that will appear on every page of the digital collection.
3. A brief summary about the library and collection(s).
4. A thumbnail image or logo file to be used for the banner section of the webpage.
5. Colors for the font and background of the collection's banner. (See examples below)

Examples of logos/fonts/background colors:



Items from the library's collection can be organized into "sub-collections" on the Recollection Wisconsin website. Think about what "sub-collections" the materials will be organized by and determine what the "sub-collections" will be named. The "sub-collections" will be defined by a metadata element named "collection". Below is an example:



Selecting Materials

There are several things that need to be taken into consideration when selecting materials to be digitized. A written selection policy should be developed that defines the scope, size, and capacity of the collection.

Recollection Wisconsin provides the following checklist as a guide for selecting materials. If any of these statements do not apply, it may not be good to select the item for digitization.

- The item is in the public domain or permission has been secured from the rights holder to make it available online.
- The item is rare or unique.
- The item holds a particular significance in the community.
- The item is frequently requested by patrons/visitors.
- The item or very similar items are not found anywhere else on the Internet.
- There is enough accurate information available about the item to add useful context.

- Appropriate equipment is available to create an accurate digital copy of this item (for example, item is not too large to fit on scanner), or resources are available to outsource digital imaging if necessary.
- The item is in stable condition and will not be damaged by scanning or other handling.

Source: Recollection Wisconsin. Contributor Guidelines. Retrieved from:

<http://recollectionwisconsin.org/organizations/guidelines>

Once materials have been selected, it is recommended to create an inventory of items that will be digitized. The inventory can be a general list or be very detailed.

Preparing

Process and Organize Materials

The collection should be processed prior to starting the digitization process. Evaluate the materials for any preservation and special handling needs. The materials need to be sorted and ordered before digitizing is started. Item-level metadata will need to be created for all items that are digitized. Information should be available about each item and any research on the materials should be completed prior to the digitization process so that the metadata can be created.

Determine a system for linking the digital file to the physical object and for indicating that the physical object has been digitized. Options for this include:

- Writing the digital file name (in pencil) on the physical object.
- Using stickers or labels to indicate that an object has been digitized (only apply to the object if they will not cause any damage).
- A hidden field can be used in the metadata template to indicate the location of the original object.

It may be necessary to review and evaluate materials that have previously been digitized to determine if they were digitized using best practices. Materials may need to be reformatted or re-digitized. Materials may also need to be renamed or reorganized to keep the collection consistent.

Copyright and Rights Statements

It will be necessary to consider copyright status, Creative Commons licensing, or fair use of the materials selected for the project. Possession of originals does not equal ownership or copyright. Deed of gifts should include a provision for digitization and online reuse of donated resources.

Copyright

Review the Recollection Wisconsin Copyright Policy (<http://recollectionwisconsin.org/wp-content/uploads/2016/07/Recollection-Wisconsin-Copyright-Policy.pdf>) and Contributor Guidelines (<http://recollectionwisconsin.org/organizations/guidelines>)

Materials in digital collections must:

- Be in the public domain. This means they are not in copyright and can be copied and disseminated without restrictions. They can be freely used without permission or attribution.

OR

- If protected by copyright, permission must be secured before uploading
- If protected by copyright, can be used if fair use provisions apply
- If protected by copyright, can be used if the digitizing institution owns the copyright

If copyright is undetermined, additional research will need to be completed.

Published Materials (likely to be books, posters, brochures, maps, pamphlets, and blank postcards) are in the public domain (based on digitizing dates in 2017) if:

- Copyright has expired (the work was published before 1923)
- Copyright owner published the work without a copyright notice (1923-1977)
- Copyright owner failed to renew copyright status (1923-1963)
- Copyright owner deliberately places/dedicates the work to the public domain using the Creative Commons Zero dedication
- Copyright does not protect the type of work (worked created by the Federal government and some states.)

Copyright guidelines:

Date of Publication	Conditions	Copyright Term
Before 1923	None	None. In the public domain.
1923-1977	Published without a copyright notice	None. In the public domain.
1978-March 1, 1989	Published without notice, and without subsequent registration within 5 years	None. In the public domain.
1978-March 1, 1989	Published without notice, but with subsequent registration within 5 years	70 years after the death of author. If corporate authorship, 95 years from publication or 120 years from creation, whichever expires first.

1923-1963	Published with notice but copyright was not renewed	None. In the public domain.
1923-1963	Published with notice and the copyright was renewed	95 years after publication date
1964-1977	Published with notice	95 years after publication date
1978-March 1, 1989	Created after 1977 and published with notice	70 years after the death of author. If corporate authorship, 95 years from publication or 120 years from creation, whichever expires first.
1978-March 1, 1989	Created before 1978 and first published with notice in the specified period	Greater of the term specified in the previous entry or December 31, 2047
March 1, 1989 – 2002	Created after 1977	70 years after the death of author. If corporate authorship, 95 years from publication or 120 years from creation, whichever expires first.
March 1, 1989-2002	Created before 1978 and first published in this period	Greater of the term specified in the previous entry or December 31, 2047
After 2002	None	70 years after the death of author. If corporate authorship, 95 years from publication or 120 years from creation, whichever expires first.
Anytime	Works prepared by an officer or employee of the United States Government as part of that person's official duties.	None. In the public domain.

Source: Hirtle, Peter B. for the Cornell Copyright Information Center. (2017, January 1). Copyright Term and the Public Domain in the United States. <http://copyright.cornell.edu/resources/publicdomain.cfm>
Pdf <http://copyright.cornell.edu/resources/docs/copyrightterm.pdf>

For unpublished materials (likely to be correspondence, photographs, postcards with writing, manuscripts, and diaries):

- Unpublished anonymous and pseudonymous works and works made for hire are in the public domain 120 years after the date of creation (=created before 1897)
- Unpublished works are in public domain 70 years after the author's death (=author died before 1947)
- Unpublished works when the death date of the author is unknown are in the public domain 120 years after the date of creation (=works created before 1897).

If a work is NOT in the public domain, it is in copyright and permission must be given by the creator or publisher. If that is not possible, steps of due diligence must be taken to try to obtain permission.

If an item is NOT in the public domain and the author or publisher cannot be identified, it is considered to be an orphan work, and digitization is done at the institution's discretion.

Copyright Resources:

- Copyright Genie <http://librarycopyright.net/resources/genie/>
- Public Domain Slider <http://librarycopyright.net/resources/digitalslider/>

Fair Use:

Fair use (Section 107 of the U.S. Copyright code) addresses the legal use of copyrighted material without the permission of the copyright holder under certain circumstances. Fair use may apply if copyrighted materials are used for purposes of criticism, comment, news reporting, teaching, scholarship, or research. There are four different factors that have to apply if fair use is to be applied:

- The **purpose** and character of use.
- The **nature** of the work.
- The **amount** of the work used in relation to the work as a whole.
- The effect of the use upon the potential **market** for or value of the copyrighted work.

It is fair use if it:

- Is for nonprofit educational purposes or is transformative
- Is factual
- Is a small amount of the material (relative to the original)
- Does not hurt the market for the original

It is not fair use if it is:

- Is used for commercial or duplication purposes
- Is creative
- Is a complete work or the heart of the work
- It hurts the market or if there is a potential market

In determining fair use, all four factors must apply. If the copyright owner disagrees with a fair use interpretation, or if it is not fair use, then the copyright holder's rights are infringed upon. Any resulting disputes may have to be resolved by a lawsuit or arbitration.

Source: Stanford University Libraries. Copyright and Fair Use. (2005-2017). retrieved from:

<http://fairuse.stanford.edu/overview/fair-use/>

Fair Use Resources:

- Checklist for Conducting a Fair Use Analysis before Using Copyrighted Materials:
http://copyright.cornell.edu/policies/docs/Fair_Use_Checklist.pdf
- Fair Use Evaluator Tool from the Office for Information Technology Policy of the American Library Association <http://librarycopyright.net/resources/fairuse/>

Creative Commons

Some materials may have a Creative Commons license that retains the material's copyright while allowing for certain uses. The Creative Commons license has to be agreed to by the copyright holder.

Creative Commons terms:

CC BY	Attribution	Can distribute, remix, tweak, and build upon the work as long as the original creation is credited.
CC BY-SA	Attribution-Share Alike	Can remix, tweak, and build upon the work for commercial purposes as long as the original creation is credited AND licensed under identical terms. All new works will carry the same license as the original, so any derivatives will allow commercial use.
CC BY-NC	Attribution-Noncommercial	Can remix, tweak, and build upon the work non-commercially. The new works must acknowledge the original and be non-commercial, but do not have to license the derivative works on the same terms.
CC BY-NC-SA	Attribution-Noncommercial-Share Alike	Can remix, tweak, and build upon the work non-commercially, as long as the original creation is created and the new creation is licensed under identical terms.
CC BY-NC-ND	Attribution-Noncommercial-NoDerivs	Works can be downloaded and shared with others as long as original is credited. Original cannot be changed in any way or used commercially.
CC BY-ND	Attribution-NoDerivs	Allows for redistribution, commercial and non-commercial, as long as it is passed along unchanged and in whole with credit to the original.

CC 0	"All rights granted"	Allows licensors to waive all rights and place a work in the public domain.
------	----------------------	---

Source: Creative Commons. About The Licenses. Retrieved from: <https://creativecommons.org/licenses>

If a take-down **notice** is received, the material must be expeditiously removed or have access to it disabled. If a take-down **request** is received, it may or may not be legally legitimate.

It is also important to be mindful of privacy or ethical issues.

Rights and Access Statements

A statement defining rights status and providing access information needs to be clearly and concisely stated in the metadata for every item in the collection. Recollection Wisconsin and the DPLA suggest using standardized rights statements and recommend modeling the rights statements after examples found at: <http://rightsstatements.org/en/>

The statement should include:

- A rights statement that indicates the rights status (copyright category, Creative Commons license, or fair use).
- An access statement that indicates how the item can be used.
- The library's contact information.

Examples of rights statements from Recollection Wisconsin, Contributor Guidelines: Retrieved from: (<http://recollectionwisconsin.org/organizations/guidelines>)

For an item presumed to be in the public domain:

This item is in the public domain and as such may be freely used without restriction.

There are no known restrictions on the use of this digital resource. Contact [your institution] for information regarding permissions and reproductions.

For an item under copyright; copyright holder has granted permission to put online:

This image has been made available with permission of the copyright holder and has been provided here for educational purposes only. Commercial use is prohibited without permission. Contact [your institution] for information regarding permissions and reproductions.

For an item in which copyright status is undetermined:

This material may be protected by copyright law. The user is responsible for all issues of copyright. Contact [your institution] for information regarding permissions and reproductions.

Additional Copyright Information:

- DPLA Webinar-Understanding Copyright
<https://www.youtube.com/watch?v=AurzAmj4SvM&t=1s>
- Recollection Wisconsin Guidelines on Copyright
<http://recollectionwisconsin.org/organizations/guidelines>
- Copyright Basics by the U. S. Copyright Office <http://www.copyright.gov/circs/circ01.pdf>
- How to Investigate the Copyright Status of a Work by the United States Copyright Office.
<https://www.copyright.gov/circs/circ22.pdf>
- Cornell University's Copyright Information Center (includes Peter Hirtle's Copyright Term and the Public Domain in the United States chart)
<http://copyright.cornell.edu/resources/publicdomain.cfm>
- Orphan Works: Statements of Best Practices published in 2009 by The Society of American Archivists (SAA) <http://www2.archivists.org/sites/all/files/OrphanWorks-June2009.pdf>

File Storage, Formats, Directories, and Names

File storage, formats, naming standards, and directory organization need to be determined at the outset of the project. They need to be consistent throughout the collection and documented so that they can be followed by anyone working on the project.

It is the library's responsibility to store master files and backups of access images. Recollection Wisconsin, CONTENTdm, and the Milwaukee Public Library are not responsible for the storage of master files or backups of access images.

File Storage

The library will be given three external hard drives. One will be used for working on the project, one will be transferred between SCLS and the library throughout the project (see Providing Digital Files to SCLS below), and the third should have image files copied to it and be stored (preferably offsite) in a dedicated and safe location.

To transfer or copy files to the external hard-drives, attach two hard drives to the computer and then drag and drop the desired folder from one external drive to the other.

Files should not be added to or kept on the library's local computers.

The external hard drives should only be used for this project, do not add any other files to the hard drives.

The Digitization Assistant at SCLS will add metadata and text files to the external hard drives (after reformatting the files from Google Sheets or Docs to Excel or text files) so that all image files, metadata files, and text files will be stored in a common place.

Once the project is completed, **ALL** image files, metadata files, and any text files should be on EACH of the three external hard drives. One of the external drives will be stored at the library, one should be stored off-site from the library, and one will be stored at SCLS.

File Formats

Two files of each image should be created:

- One will be an uncompressed master file, saved as a Tagged Image File Format (TIFF) file (for images) or a WAV/AIFF file (for audio files). The master file is considered an archival file and should not be accessed unless necessary. It is used to produce access (derivative) image files. The file name should indicate that it is a master file. For example: name_master_001.tif or name_mf_001.tif (master file is abbreviated with “mf”).
- A second file will be an access (derivative) file. Access files of images can be saved as a Joint Photographic Experts Group (JPEG) or a Portable Document Format (PDF). Audio files can be saved as MP3, QuickTime, Windows Media, or RealAudio file formats. The derivative file is a smaller and compressed file that is used for viewing, web access, and thumbnails. The derivative file can be an edited version. The file name should indicate that it is an access file. For example: name_access_001.jpeg or name_af_001.jpeg (access file is abbreviated with “af”).

The master and access files should be stored in separate folders. The access files will be used for uploading to CONTENTdm.

IrfanView can be used to batch convert file formats. Instructions for using IrfanView can be provided in a separate document.

Directory Organization

Digital files will need to be organized into directory folders in a specific way for uploading to CONTENTdm. It is best to have the folders created (on the external drives, as explained below) before starting the digitizing process. The directory organization will depend on whether the digital files are for simple objects, compound objects, or monographs.

-Simple objects are a single item (equal to one file).

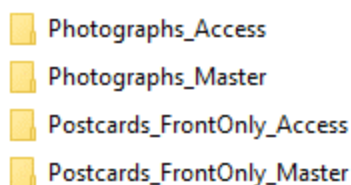
-Compound objects are two or more files bound together. They can be documents, monographs, picture cubes, or postcards.

-Monograph compound objects are documents that have a hierarchical structure (more than one level of items). For example, a book with chapters or a report with sections.

Depending on the project workflow, folders can be created (based on the steps described below) for batches of objects or based on sub-collections.

Simple Objects

Items that are one page are considered simple objects, and each simple object will have its own file. Simple objects that are part of the same collection can be saved together in one folder. However, there will need to be two folders for the simple objects: there will be a folder for the master files and a folder for the access files. The folder name should indicate that it holds either the master or access files. For example: filename_access and filename_master.



CONTENTdm will upload all of the files in the folder in ascending alphanumeric order.

Compound Objects

Two or more files that are bound together form a compound object. Examples of compound objects are: multi-page documents, books, audio files from a CD, or postcards with an image on the front and writing on the back. Every compound object will need to have a folder of its own that houses each of the object's files.

File names and directory structure are used to upload the files in the correct order and build the compound object. All of the files in the folder are imported in ascending alphanumeric order, determine file names accordingly. Page titles are used as navigation for end users.

Directory set-up for compound objects:

1. Create a parent folder for the object (required).
2. Create a **scans (or images) subdirectory** (required) that contains the image or audio files that make up the compound object.
3. Create a **transcripts subdirectory** (if needed). It will contain text files that are used with transcripts. The text files must be named with the same root name as the files that make up the compound object. For example, item.txt would be the transcript for a scanned file named item.tif.
4. Create a **display subdirectory** (if needed), that contains custom display images for full resolution scans. Display images must have the same root name as the full resolution scans. For instance, item.jpg is the display image for a scanned file named item.tif.

A **tab-delimited text file** is also required. It will define the metadata for the compound object. The tab-delimited text file must have the same name as the compound object directory. Optionally, the tab-delimited text file may also define the compound object's structure and contain page-level metadata. (The Digitization Assistant at SCLS will create the tab-delimited text file using the metadata entered to Google Sheets).

Every compound object has its own folder with a scans (or can be named images) subdirectory. If needed, subdirectories for transcripts (text files) or display images are created for the folder.

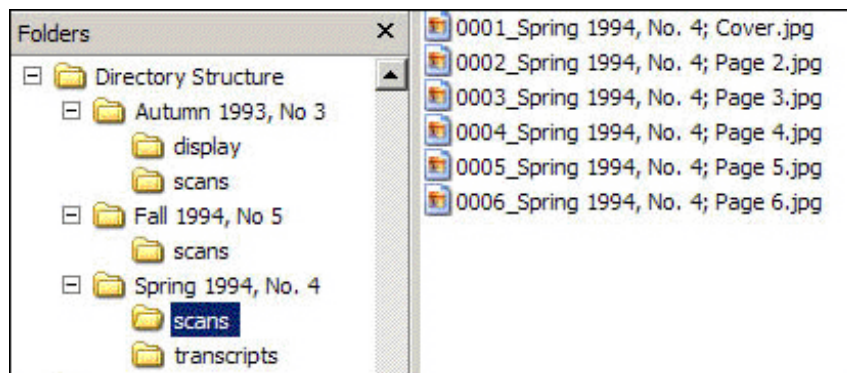


Image Source: OCLC. From a Directory Structure. (2017). retrieved from:
<https://www.oclc.org/support/services/contentdm/help/compound-objects-help/adding-multiple-compound-objects/directory-structure.en.html>

One parent folder should be created for the compound object's master files and a second parent folder should be created for the compound object's access files. The folder names should indicate that they are the master or access file.

An object, such as a postcard or photograph, that has writing on the back could have both sides scanned, and be considered a compound object. It would also work as a simple object, with only the (front) image scanned and the writing on the back typed into the description field.

Monograph Compound Objects

Monographs use directories to define document structure. Each item that is a monograph will have its own folder. Files are stored in subdirectories within the folder to create hierarchical organization. Subdirectories are used to establish chapters, appendices, sections, or other headings.

Files in the root directory folder are imported first, then any sub-directories. If there are files in the sub-directories, they will be imported before any directories in the sub-directory. They are imported in ascending alphanumeric order.

Monograph objects should have master and access files for all of the images and any text files. The files should be organized into folders (as described above and below). The subdirectories in the master files folder and in the access files folder should be organized in the same manner.

Monograph folders can be grouped together into a “Monographs” folder(s) (there should be one folder for the monograph master files and one folder for the monograph access files).

Directory set-up for monographs:

1. Create a folder with the name of the object.
2. Create subfolders for each chapter or section. These folders will hold all of the image files.
3. Create a subfolder for text files (if any). This can be done by the SCLS Digitization Assistant.
4. Images that are not part of a chapter or section (such as the cover page, table of contents, or prefaces) will be in the monograph folder, but not placed in a subdirectory.
5. The Digitization Assistant at SCLS will create the necessary metadata files using the information that the library enters into metadata templates on Google Sheets. Tab-delimited text files will be created and put into the appropriate directories on the external hard drives.

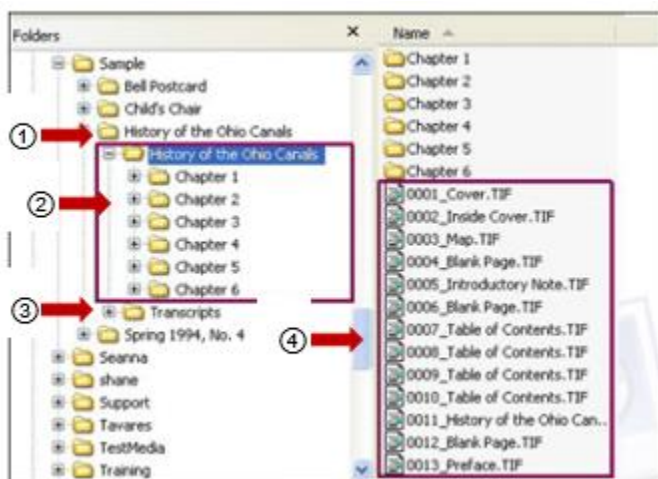


Image source: OCLC. Creating Compound Objects (Documents, Monographs, Postcards, and Picture Cubes). (2017, January, 18). Retrieved from:

<https://www.oclc.org/content/dam/training/CONTENTdm/pdf/Tutorials/Importing%20Items-Creating%20Objects/Creating-Compound-Objects-Documents-Monographs-Postcards-Picture-Cubes.pdf>

Folder and File Names

Folder and file naming standards should be established prior to starting the project.

Folder, and subfolder, names should:

- Be descriptive (collection name/compound object or monograph name/chapter number or section heading)
- Indicate if they hold master or access files

- If they are for batches of materials, they should indicate the extent (dates, item numbers) of their files

File names:

- Can use an existing identifier that is already associated with the item (this could be box/folder numbers, item-level accession numbers, or other ID numbers)
- Should be unique (no duplications)
- Should be consistently structured
- A file format extension should always be included (at the end of file names)
- Use at least as many digits as the highest number of materials expected in the collection
- Use only letters, numbers, dashes, underscores, or hyphens
- Not use any spaces or special characters
- The file name should indicate that it is the master file (example: name_master_001.tif or name_mf_001.tif) or the access file (example: name_access_001.jpeg or name_af_001.jpeg)
- For CONTENTdm, all of the files in a folder are imported in ascending alphanumeric order

Source: Recollection Wisconsin. Contributor Guidelines. Retrieved from:

<http://recollectionwisconsin.org/organizations/guidelines>

Examples:

- Simple objects: wrp-001.tif / wrp-002.tif / wrp-003.tif
- Compound objects: br035a.tif / br035b.tif / br035c.tif

For compound objects or monographs, the file name will appear in CONTENTdm as the page name or title. Files should be named so that their place in the structure of a compound object is clear.

For example:

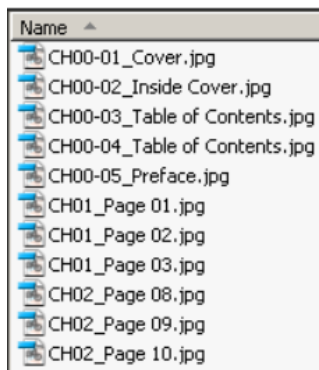


Image Sources: OCLC. Creating and Editing Compound Objects Using CONTENTdm Administration. (2017, January 22). Retrieved from: <https://www.oclc.org/content/dam/training/CONTENTdm/pdf/Tutorials/Importing%20Items-Creating%20Objects/Creating%20and%20Editing%20Compound%20Objects%20Using%20CONTENTdm%20Admin.pdf> and OCLC. From a Directory Structure. (2017). retrieved from: <https://www.oclc.org/support/services/contentdm/help/compound-objects-help/adding-multiple-compound-objects/directory-structure.en.html>

Files for transcripts should be saved with the same root file name as the image or audio file. For example, the transcript file for item.tif is item.txt. (Transcript files will be created in Google Docs, and then reformatted into text files by the Digitization Assistant at SCLS).

IrfanView can be used to batch rename files that had been previously digitized. Step-by-step instructions for using IrfanView will be provided in a separate document.

Software

Software that will be used includes:

- Scanning software will be installed on the computers that are provided with the Indus BookScanner 9000s and the digitization kits.
 - Virtual Library software with the Indus BookScanner.
 - Canon My Image Garden software with the digitization kits.
- Google Sheets will be used to enter metadata for the digitized objects.
- Google Docs will be used for any text files that are created for the image files (for example: transcripts).
- Audacity will be installed on the computers in the digitization kits for working with audio files.
- IrfanView will be used for batch renaming and batch conversions of file formats. It will be installed on the computer with the Indus and on the computers in the digitization kits. It can be added to additional computers in the library as needed.

Metadata

General

Metadata must be created for every item that is digitized. Recollection Wisconsin uses Simple and Qualified Dublin Core.

Guidelines and examples can be found in the Recollection Wisconsin Metadata Essentials (2016, November) document: <http://recollectionwisconsin.org/wp-content/uploads/2017/01/RecollectionWisconsinMetadataEssentials.pdf>

The Digitization Assistant at SCLS will assist with creating a metadata model that will identify how the items being digitized will be described. Metadata elements that are required by Recollection Wisconsin/ the Digital Public Library of America (DPLA) are: title, rights, type, and subject. Additional elements can be used based on the physical format or genre of the items in the collection. The additional elements have to be mapped to Dublin Core, but can be given descriptive display names. If controlled vocabularies will be used for certain elements, it will be indicated on the data model. A Google Sheets template will be created, based on the data model, for use in recording the metadata.

The DPLA will only display metadata mapped to: creator, date, description, language, place, publisher, rights, subject, title, and type elements. Only ONE date element will be displayed by the DPLA, and it will be labeled in DPLA as Date Created. Any of the other elements that are separated by descriptive display names in Recollection Wisconsin will be grouped together on the DPLA web page based on Dublin Core mapping.

Metadata Elements

Basic elements recommended by Recollection Wisconsin:

	Required	Description	Notes
Title	Required	A descriptive title for the item.	Suggested title= Subject/Activity, Location, Date
Type	Required	Broad category characterizing the type of item being described.	Use controlled vocabulary
Subject	Required	Search terms/keywords/tags that describe the content of the resource.	Can use a controlled vocabulary
Rights	Required	A statement describing the copyright status of the original item (e.g. in copyright, no copyright/public domain, or copyright undetermined), followed by the name of the organization providing the item.	
Date Created		Date the original item was created. If exact date is unknown, provide a year range, or use "ca." (Circa). Leave blank if unknown.	Use ISO 8601b Standard (YYYY-MM-DD)
Description		Briefly describe the item.	
Creator		Name of photographer, author, artist or other creator. Leave blank if unknown.	Use format Last Name, First Name
Place		Location described or represented in the item. Enter the most specific element of the location known, followed by the county, then state name.	
Format		Term describing the materials or physical characteristics of the original item.	Use controlled vocabulary
Submitter		The organization or entity submitting the digital object.	
Collection		(Relation-is part of).	
Transcript		Typed information taken from the item.	
File Name	Required (for upload)	Name of digital file.	

Examples of additional elements that may be useful include: language, date digitized, item number, accession id, shelving location, digitization specifications, extent, temporal coverage (a time range/period/date other than date created), publisher, latitude and longitude. Elements can be hidden if used only for internal purposes. For oral histories additional elements may include: interviewer or interviewee.

Controlled Vocabularies

A controlled vocabulary can be used for some elements. If a controlled vocabulary is to be used for an element, it will be indicated on the metadata template.

Tips:

- Do not mix terms from more than one controlled vocabulary in an element. Use a different element for each controlled vocabulary used.
- If a controlled vocabulary is assigned, a value from that source or list must be used.

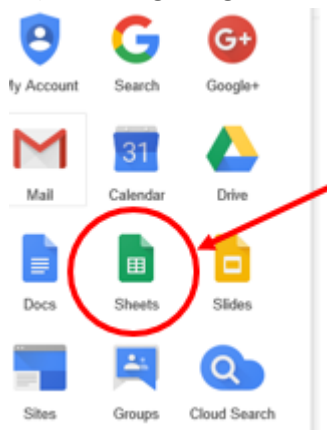
Controlled vocabularies that can be used for the basic Recollection Wisconsin elements:

Place (Coverage-Spatial)	Getty Thesaurus of Geographic Names (TGN)	http://www.getty.edu/research/tools/vocabularies/tgn/
Creator	Library of Congress Name Authority File (LCNAF)	http://id.loc.gov/authorities/names.html
Creator	Getty Union List of Artist Names (ULAN)	http://www.getty.edu/research/tools/vocabularies/ulan/
Date	ISO 8601b/ISO 639-3	http://www-01.sil.org/iso639-3/codes.asp
Format	Library of Congress Thesaurus for Graphic Materials (LCTGM)	http://www.loc.gov/pictures/collection/tgm/
Subject	Library of Congress Authorities (LCSH)	http://authorities.loc.gov/
Subject	Getty Art & Architecture Thesaurus (AAT)	http://www.getty.edu/research/tools/vocabularies/aat/
Type	MIME (TYPE) vocabulary	http://www.iana.org/assignments/media-types/media-types.xhtml

Metadata Template in Google Sheets

A metadata template(s) will be provided through Google Sheets for recording metadata for each of the objects in the collection. Please let the Digitization Assistant at SCLS know what Gmail account to use for sharing the Google Sheets.

The metadata template will be accessed by signing into Google (with the Gmail address that was provided to SCLS), selecting Google Sheets, and selecting the template file.



The metadata template will include the following sheets:

- Elements—a sheet describing the elements that will be used
- Controlled Vocabularies—controlled vocabulary information (if any are used)
- Copyright and Rights Statements—examples that can be used for the “Rights” Element
- Sheet(s) for metadata for simple objects
- Sheet(s) for metadata for multi-image objects
- A compound/monograph object master sheet (labeled CompoundMaster). Each compound object or monograph has to have its own sheet.
 - To create additional sheets (as needed for compound objects or monographs): Select the CompoundMaster sheet. Right-click on the tab. Select Duplicate. Select the newly created (Copy of...) sheet. Right-click on the tab. Select Rename. Type name of the object into the tab. Verify that any drop-down lists/controlled vocabularies are available in the new copy.

Notes about using Google Sheets:

- Information entered is automatically saved. Use caution when deleting, it generally cannot be undone.
- For fields with defined controlled vocabulary lists, click on the drop-down arrow in the lower right corner of the cell.
- There can be a maximum of 200 sheets in one Google Sheets file. Based on this, if additional files are needed, contact the Digitization Assistant at SCLS to have them created.

Simple and Compound Objects

For CONTENTdm, objects are considered either simple, compound, or monographs. There will be different metadata templates based on these types of objects.

A simple object is made up of one file. Metadata for all simple objects in a collection can be entered, row-by-row, into one sheet.

Compound objects and monographs are made up of multiple files. Examples are books, postcards with writing on the backside, pamphlets, letters, or documents. Compound objects and monographs need to have metadata that document the relationships between components (pages) of the object.

Depending on the number of files that make up a compound object or monograph and the desired amount of metadata to be used for every one of the files, there are two options for entering metadata.

If the compound object or monograph is made up of 10 or fewer files and each of the object's files (pages) use repeated metadata, it will be considered a multi-image object. The Google Sheets template will have one sheet for entering data for multi-image objects. Each multi-image object will have (only) one row in the multi-image sheet. The metadata will describe the image as a whole and will be repeated for every page of the object. There will be a "File Name" field, where EVERY file (from the compound object's folder) that makes up the object should be listed. All file names that make up the multi-image object will be in one cell, separated by commas.

Compound objects or monographs with more than 10 files will need a sheet for just that object. Objects with fewer than 10 files that require different metadata for each (page) file will also need to have sheets of their own. Each of these types of objects will have a sheet that holds metadata for just that object. The first row in the sheet will be a metadata record about the object as a whole. Subsequent rows will have metadata about each file (composite pages or items) that make up the object.

For monographs that are organized by chapters or sections there should be an element that indicates the chapter or section name and a second element used to indicate the page description (table of contents/Cover/Preface) or page number.

PDF files (that contain multiple pages) can be uploaded as simple objects, and viewed as a single object, or, they can be converted to compound objects, with each page of the object having its own metadata. If a PDF file is to be displayed as a single object, enter the metadata (for the object as a whole) on the simple-object tab of the metadata template. If the PDF is to be displayed as a compound object, create a compound-object sheet to enter metadata (for the object as a whole plus for each page of the object).

Metadata Formatting Rules and Tips

- Do not use an element for anything other than what it is intended.
- For elements with multiple values, they must be separated with a semicolon and a space
- Capitalize the first word and proper names.
- Only punctuate the end of complete sentences (not lists of terms).
- Do not put quotation marks around metadata.
- Avoid the use of abbreviations. Spell out the full names of communities and states. It is okay to use abbreviations that are prescribed by a controlled vocabulary for use in the subject field (ex: Madison (Wis.) from the Library of Congress Subject Headings).
- Fields for which there is no available information should be left blank.
- The elements can be in any order, can be repeated, and all of them do not have to be used.
- Metadata can be “copy/pasted” or “filled-down” for files with the repeated details.

Source: Recollection Wisconsin. Metadata Essentials. (2016, November). Retrieved from:

<http://recollectionwisconsin.org/wp-content/uploads/2017/01/RecollectionWisconsinMetadataEssentials.pdf>

Digitizing

As items are digitized, it will be important to apply the file storage, formatting, and naming standards established during the planning phase of the project. Also, as items are digitized, apply the system decided on during the planning phase to link the digitized item with the original object.

Equipment

SCLS has scanning and audio recording equipment available to loan to libraries. Training and support on using the equipment will be provided. Contact the Digitization Assistant at SCLS for more information about what is available and how to reserve it.

Indus BookScanner 9000s

The Indus BookScanner 9000s is an overhead scanner. It is ideal to use for scanning books, periodicals, and fragile documents. The Indus will be housed at the South Central Library System’s offices. Hands-on training will be provided on using the Indus.

Indus BookScanner 9000s Demo Video:

<https://www.youtube.com/watch?v=Ara99R6HwhA>

Indus Notes:

- It has a scanning area of 18.9" x 24.5"
- It is made with two adjustable plates that can cradle bound items
- It will scan at a resolution of 400 dpi

South Central Library System carries a service plan with Indus International, Inc. that provides for basic maintenance and repair services. If there are any issues with the Indus or Virtual Library software, please let the Digitization Assistant at SCLS know immediately. Be aware that the library may be held responsible for costs to repair any damage (chips in the glass, etc.) that may occur to the Indus while it is at the library.

Digitization Kits

SCSL has digitizing kits available. Training can be provided on using the kits and manuals will be included with the kits.

The kits include:

- Laptop and software for scanning and for creating/editing audio files
- Flatbed scanner (Canon CanoScan 9000f Mark II)
- Microphone
- Power strip
- Cleaning supplies

Digitization Kit Notes:

- The CanoScan has a scanning area of 8.5" x 11.7"
- 35mm film can be digitized with the CanoScan
- The CanoScan will scan at a resolution up to 600 dpi

Digitizing Print Materials

Scanning items will involve applying appropriate file formats, assigning appropriate file names, and saving to an appropriate directory/folder. (See the File Storage, Formats, Directories, and Names section above).

There should be a dedicated work area. Do not have food or drinks near the equipment or items being digitized. If wearing rings or other jewelry, be careful not to scratch the glass. Frequently clean the glass on the scanner to keep it free from dust or fingerprints.

When scanning, the following specifications need to be taken into consideration:

- Resolution: Measured in pixels per inch (ppi). Setting depends on the size and condition of the original and the intended uses of the digital object. Increased PPI means that more frequent samples are taken of the original. Higher resolution captures more information but results in larger file sizes.
- Bit Depth: Measures the number of colors available to represent the color/gray value in the original work. A bit depth of one = black and white. A bit depth of eight = 256 possible values. The higher the bit depth, the more colors an image can store.
- Histograms: Alter the light levels to bring out details in photographs
- Modes of capture:
 - Bitonal=one bit per pixel representing black and white. Best for high-contrast documents.
 - Grayscale=Multiple bits per pixel representing shades of gray. Best for black and white photographs.
 - RGB=Multiple bits per pixel representing color.
- Gamma: Adjusts the midtones from tonal scale, but keeps it black and white (optimizes contrast and brightness in the midtones).
- Luma: Brightness in an image.
- Chroma: Colorfulness relative to brightness.

Additional scanning tips:

- When possible, scan at a scale of 100%.
- Check “Unsharp Mask” for sharper images.
- When scanning several pages of a single document, use the same scanner settings for all pages, even if some have images and others do not.

Note: The Indus BookScanner will scan at a resolution of 200, 300, or 400 dpi. The Canon scanner will scan at a resolution of 75, 100, 150, 200, 300, 400, 600 dpi (or up to 4800 dpi for slides). Use the resolution that comes the closest to the recommended settings below.

The following table gives recommended settings for scanning by image type and size.

Images	Size of Original	Scanning Resolution (Pixels per square inch)	Bit Capture
	35mm	2100 ppi - 3000 ppi	8-bit grayscale for black-and-white 24-bit RGB for color
	3" x 5"	600 ppi - 1200 ppi	24-bit color
	4" x 2 1/2"	1200 ppi	24-bit color
	4" x 5"	800 ppi	24-bit color
	5" x 7"	400 ppi - 800 ppi (625 ppi=ideal)	24-bit color
	8" x 10"	300 ppi - 600 ppi (400 ppi=ideal)	24-bit color
	8 1/2" x 11"	300 ppi	24-bit color
	16" x 20"	300 ppi	24-bit color
	> 16" x 20"	200 ppi	24-bit color

Text	Print—No images	600 ppi	1 bit (bitonal)
	Print—with images	300 ppi	8-bit grayscale or 24-bit color
	Manuscript	300 ppi - 400 ppi	8-bit grayscale or 24-bit color

Maps	< 36"	600 ppi	
	> 36 "	300 ppi - 400 ppi	

Sources:

- Association for Library Collections & Technical Services (ALCTS) Minimum Digitization Capture Recommendations: <http://www.ala.org/alcts/resources/preserv/minimum-digitization-capture-recommendations>
- Federal Agencies Digitization Guidelines Initiative (FADGI), Technical Guidelines for Digitizing Cultural Heritage Materials as a Standard and Best Practice Guide to Quality Control and Quality Checklists http://www.digitizationguidelines.gov/guidelines/FADGI_Still_Image-Tech_Guidelines_2010-08-24.pdf
- U. S. National Archives and Records Administration: <https://www.archives.gov/files/preservation/technical/guidelines.pdf>
- Recollection Wisconsin Digital Project Planning Worksheet 2016 <http://recollectionwisconsin.org/wp-content/uploads/2012/11/DigitalProjectPlanningWorksheet2016.pdf>

While scanning, check the images for quality and consistency. Verify that the entire image has been scanned, that the image is not skewed, and that there are not any unwanted materials or digital artifacts (dust/hair) in the image. Images can be adjusted (straightened, rotated, sharpened, or cropped for

example), but try to keep the master file image as close to the original as possible. Additional editing can be done to access copies of the master file.

Resources for scanning best practices:

- BCR's CDP Digital Imaging Best Practices Working Group, BCR's CDP Digital Imaging Best Practices Version 2.0 http://mwdl.org/docs/digital-imaging-bp_2.0.pdf
- NARA 2004 Technical Guidelines for the Digitization of Archival Materials <https://www.archives.gov/files/preservation/technical/guidelines.pdf>
- Federal Agencies Digitization Guidelines Initiative (FADGI), Technical Guidelines for Digitizing Cultural Heritage Materials as a Standard and Best Practice Guide to Quality Control and Quality Checklists http://www.digitizationguidelines.gov/guidelines/FADGI_Still_Image-Tech_Guidelines_2010-08-24.pdf
- Recollection Wisconsin Digital Project Planning Worksheet 2016 <http://recollectionwisconsin.org/wp-content/uploads/2012/11/DigitalProjectPlanningWorksheet2016.pdf>
- The Sustainable Heritage Network. Guide to Quality Control and Quality Checklists. (2016, June 11). Retrieved from: <http://sustainableheritagenetwork.org/digital-heritage/guide-quality-control-and-quality-checklists>

As explained in the File Formats section above, there will need to be both master and access files for each of the image files created. How this is done will depend on what settings are used when images are scanned.

- For images that are scanned as TIFFs, the name of the folder that holds the file should indicate that it holds the master files. To create access files, use IrfanView* to batch rename and convert the files into JPEGs. The name of the folder with the access files should indicate that it holds access files.
- Ideally, all files should be scanned as TIFF files, but if files had previously been scanned using JPEG format, create a folder on the external hard drive, indicate in the folder name that it holds master files, and use IrfanView* to batch rename and convert (into TIFF format) the files. The original JPEG images can be used as access files and the folder where they are located should have a name that indicates they are access files.

*Step by step instructions for using IrfanView to batch rename and convert will be provided in a separate document.

Optical Character Recognition (OCR) can be used for some types of text-based materials to make them searchable. If it will be used, it can be applied by the Digitization Assistant at SCLS during the process of uploading the files to CONTENTdm.

Digitizing Audio Materials

Audio files should be named and organized using the guidelines described in the File Storage, Formats, Directories, and Names section above.

Audio files should be:

- Master files for audio materials should be in WAVEform audio format (WAV) or Audio Interchange File Format (AIFF). There is should be a 24-bit minimum bit depth and 96,000 kHz minimum resolution.
- Access (derivative) file formats should be MP3, QuickTime, Windows Media, or RealAudio. (192 Kbps for most audio/128 Kbps for oral history).

Audacity will be installed on the laptops that are provided with the digitization kits. The software can be used to create, edit, or reformat audio files. Audacity default quality settings=Sample Format 32-bit float (and Sample Rate 44100 Hz). Audacity can be used to combine files and convert them to mp3.

It is recommended to break long audio segments up into smaller segments. The smaller segments can be indexed as separate clips and searched or retrieved individually.

Transcripts of audio recordings can be typed out (following the guidelines in the Transcripts section below) and can include any extra information about the recording such as what, who, where, or when. In some cases, it may be appropriate to have the transcript of a recording as an object in CONTENTdm (without a link to the audio), but an indication that the library holds the recording.

Digitizing Video Materials

Video files should be named and organized using the guidelines described in the File Storage, Formats, Directories, and Names section above.

Video files should be:

Master Version	Derivative Version
•MPEG-4 AVC (H.264) or DV encoding (codec)	Adobe Flash is recommended for streaming video, while MPEG-4 is recommended for files that will be downloaded by users. •MPEG-4 AVC (H.264) encoding(codec)
640 x 480 resolution (assuming 4:3 original aspect ratio)	320 x 240 resolution (assuming 4:3 original aspect ratio)

30 bit sample size	
Progressive scanning	
30 MiB/s data rate	256-600 kbps data rate
AVI (.avi) or QuickTime (.mov) file format	Adobe Flash (.flv) or MPEG-4 (.mp4) file format

Source: CARLI:

https://www.carli.illinois.edu/sites/files/digital_collections/documentation/guidelines_for_video.pdf

For CONTENTdm, videos must be in the MP4/H.264 format and be enabled with Fast Start.

It is recommended to break long audio segments up into smaller segments. The smaller segments can be indexed as separate clips and searched or retrieved individually.

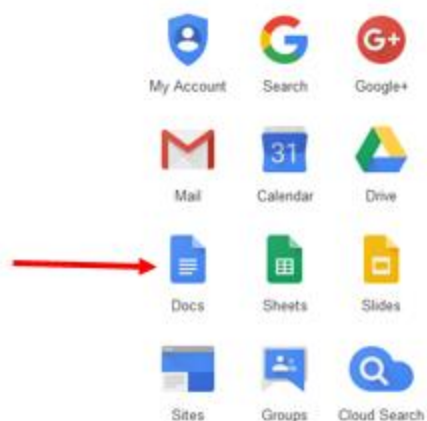
Transcripts of videorecordings can be typed out (following the guidelines in the Transcripts section below) and can include any extra information about the recording such as what, who, where, or when. In some cases, it may be appropriate to have the transcript of a recording as an object in CONTENTdm (without a link to the audio), but an indication that the library holds the recording.

Creating Transcripts for Digitized Materials

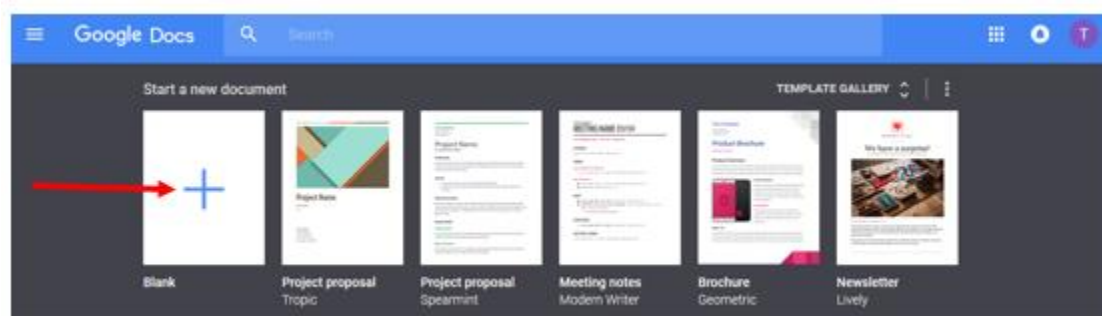
Typed transcripts of certain materials (such as hand-written letters, manuscripts, or oral histories) can be created. They should be typed into a Google Docs file using the file naming standards described in the Folder and File Names section above. Transcript files and their image files need to have identical root names. Transcribed text be entered into a searchable full-text metadata field. The Google Docs file name will need to be entered into the “Transcripts” element in the object’s row of the Google Sheets metadata template. Each file (page) of a compound object will have a separate file for its own transcript. All Google Docs that are created for transcripts will need to be shared with the Digitization Assistant at SCLS.

To create transcripts using Google Docs:

Log into the Google account and click on Docs



Select "Blank"



Type the file name in "Untitled document" field and then enter the text for the transcript.

Type the file name in "Untitled document" field and then enter the text for the transcript.




The file will be automatically saved.



To share the documents with the SCLS Digitization Assistant:

Select the blue "Share" button on the top right-hand side of the screen. Enter the SCLS Digitization Assistant's email address and select "Can edit". Select "Done."

Share with others

Get shareable link 

People



✓ Can edit

Can comment

Can view

Done

Provide Information to SCLS for Entry into CONTENTdm

The digital files on the external hard drives and metadata spreadsheets will be uploaded to CONTENTdm by the Digitization Assistant at SCLS so that they can be harvested by Recollection Wisconsin and the Digital Public Library of America.

The digitized image files will be transported to SCLS by means of one of the external digital hard drives, which can be sent via SCLS Delivery. They should be placed in the provided black box, and the black box should be placed into a red delivery box. The hard drives should be sent to SCLS as batches of items are digitized.

The Digitization Assistant at SCLS will access the metadata for the digital files from Google Sheets and any text-based transcript files from Google Docs and re-format them for uploading to CONTENTdm. The Digitization Assistant will also be responsible for Approving and Indexing the collection in CONTENTdm.

For items that were previously digitized:

- If possible, rename the files to match any new file naming standards established in planning the digitization project and organize them into appropriately named folders.
- While having the items in TIFF format is ideal, if they were scanned in JPEG format, it will not be necessary to rescan or reformat into TIFF files, as Recollection Wisconsin will use JPEG files. Add the files to the external hard drives so that all of the collections' files are in one place. Send the external hard drive to SCLS, as explained above.
- Use the templates in Google Sheets to provide metadata for the items.

The Digitization Assistant at SCLS will provide periodic summaries of what records have been uploaded. Let the Digitization Assistant know if any changes (such as edits, deletes, or updates) need to be completed.

Promote and Share the Digital Collection

Once the collection is available online, there will be a few more things to consider.

- Determine who is responsible for responding to inquiries about the collection.
- Determine how requests for copies will be handled?
 - How will requests be tracked?
 - Will there be fees for reproductions?
 - How will requested files be distributed?
- Library staff and volunteers should be trained on the scope and navigation of the collection(s).
- A link to Recollection Wisconsin should be added to the library's website.

For marketing ideas, view the DPLA YouTube video: Promoting Use of Your Digital Content

<https://www.youtube.com/watch?v=rzcamxx8b8&list=PL2kZcqArLFsxxjIGYZermtTuAnbEkn2g7&index=12>

Additional Marketing Ideas:

- Issue press releases about the collection(s)
- Create brochures/flyers
- Place announcements in newsletters or other publications
- Send staff with laptop into the community
- Share on social media sites
- Develop library programming around the collections

(Resource: Shipps, Bradley. InfoSoup Memory Project Guide for Contributors. Outagamie Waupaca Library System. (2016, July 6). Retrieved from:

<https://www.owlsweb.org/sites/default/files/owlsnet/2016IMPGuide.pdf>)

File management

Once the project is completed, ALL image files, metadata files, and any text files should be on EACH of the three external hard drives. One of the external drives should be stored at the library, one should be stored off-site from the library, and one can be stored at SCLS.

It is the library's responsibility to store master files and backups of access images. Recollection Wisconsin, CONTENTdm, and the Milwaukee Public Library are not responsible for the storage of master files or backups of access images. While a copy can be stored at SCLS, it is still the library's responsibility to maintain the master files and backups.

The digital collections will need to be managed throughout their lifecycle. Someone should be responsible for regularly reviewing the files on the external drive(s). Check that the files can be opened,

that file formats or names have not changed, and that the image quality is good. There should be a plan in place to evaluate migration needs every five years. It is recommended to move the content to new hard drives every five years, and then rotate them out on a regular basis after that. In the future, it may be necessary to migrate materials to different formats as new technology and preservation standards develop.