
InfoSoup Memory Project Guide for Contributors

Contact Information

Bradley Shipps
Assistant Director
Outagamie Waupaca Library System
225 N. Oneida St.
Appleton, WI 54911

phone: 920-832-6368
fax: 920-832-6422

bshipps@mail.owls.lib.wi.us

Table of Contents

Project Planning.....	3
Identify and Involve Stakeholders	3
Project Mission	3
Budget	4
Technology.....	4
Scanner	4
Workstation	5
External Hard Drive	5
Software	5
Selection	6
Copyright Clearance.....	7
Digital Asset Management	9
File Naming and Directory Structure	9
Storage and Backup	10
Long-term Preservation	10
Marketing.....	11
Impact on Public Services.....	12
Project Implementation.....	13
Workflow Overview.....	13
Standards for Scanning.....	15
Metadata.....	16

Project Planning

Is your organization ready to begin a digitization project? Consider the following topics and questions before submitting your New Project Profile form to OWLS.

Identify and Involve Stakeholders

All stakeholders should be involved at the beginning of the project. This includes staff, volunteers, administration, and any potential partners.

Conduct a user needs assessment. Is the project worth doing? Who is your audience? What are their access requirements?

Consider who will fill key roles in the project, such as:

- Project leader
- Selecting collections
- Inventorying selected objects
- Creating metadata
- Scanning
- Inputting metadata
- Uploading files
- Technical support (hardware maintenance, software installation)
- Marketing/public relations
- Public services related to new digital collections
- Ongoing collection maintenance (quality checks, backups, archiving files, refresh and migration of archive files)

Project Mission

Establish the mission and vision for your project. Set specific, measurable objectives and outcomes. How will you evaluate the project?

Budget

Consider how you will provide funding and staffing for this project in the short and long term. Potential costs include:

- Staff time
 - Scanning and metadata creation are very time-consuming
 - Time and travel for training
 - Time for quality control
- Staff costs (hiring costs, salary, and benefits)
- Consulting and outsourcing
- Software and hardware licenses/service contracts (if image processing software is purchased)
- Hardware
 - Windows PC with adequate memory and storage
 - Equipment upgrades (plan on new equipment every 3 years)
- Designated work space for digital projects (cubicles, chairs, tables, glare-free area, etc.)
- Project supplies
- Publicity/promotion costs
- Conference fees and travel
- Outsourcing costs (oversight and transportation of materials)

Technology

Scanner

- When choosing a scanner, consider the size and format of items to be scanned.
- Consider the volume of scanning you plan to do. Some scanners are faster than others.
- Consider the knowledge-sharing benefit if other InfoSoup Memory Project participants are using the same scanner and interface.
 - Contact OWLS regarding opportunities to use OWLS's or other participants' scanners for smaller projects.
- If you have a very large collection to scan or unusual objects, consider outsourcing the imaging rather than purchasing expensive equipment.

Workstation

- Check the minimum processing requirements for the scanner and the software you plan to use.
- Working with large image files requires good processing speed and lots of RAM.
- You will need significant hard drive space for saving images unless you plan to save them on an external hard drive. (Recommended for preservation purposes.)
- A larger monitor, 19-in or more, is ideal for working with multiple images.

External Hard Drive

- If possible, purchase two USB or firewire drives with 1 TB of storage. Save your master TIFF files on both drives, and store one off site.
- Serves as storage for large master files, conserving space on your working computers.
- If fast enough, can serve as swap disk for RAM
- Backup to a second medium if possible, such as tape or CD/DVD.

Software

- Scanning software that includes basic editing features, including ability to crop, rotate, descreen, sharpen, and adjust a histogram.
- Photoshop or Photoshop Elements
 - Image manipulation software might come packaged with your scanner.
 - Photoshop Elements allows batch processing of images, so that you can quickly create a set of access JPGs from your archival TIFFs.
- Microsoft Excel (or another spreadsheet program) for metadata collection, sorting, and uploading
- Optical Character Recognition (OCR) software to extract text from images. This may come with your scanner. OWLS also has OCR software.

Selection

For your first digitization project, consider starting with a small collection or a small portion of a larger collection. Photographs, postcards, and smaller monographs are ideal. Consider the following criteria when selecting resources to digitize:

- The collection chosen should be of high interest to your community. Give priority to items that are unique.
- The collection should have a defined target audience. There should be a collection development rationale for digitization.
- The collection should have a definable scope, so that a project timeline for completion and staffing levels can be established.
- The resources to be digitized must have clearly defined ownership. Rights and permissions for electronic distribution must be secured or securable. See the following section on Copyright Clearance.
- The originals must be in an appropriate format and physical condition for scanning. Books must be able to be handled. Photos must be non-degraded, with sufficient clarity of detail. Size and type of resources must be appropriate to available equipment. At this time, we are not set up to handle newspapers, microfilm, large books, and maps.
- The collection should already be organized and have some item-level metadata (e.g. title, creator, date, etc.) available or inherent. Image files uploaded to CONTENTdm must be accompanied by metadata. If necessary, staff should do research to provide usable access points such as pictured subjects and information on the circumstances surrounding the image (photographer, event, significance, timeframe, etc.).

Visit the Recollection Wisconsin guidelines page for more information on selection: <http://recollectionwisconsin.org/organizations/guidelines>

Copyright Clearance

OWLS and Recollection Wisconsin will be held harmless for all copyright challenges.

Possession of originals does not equal ownership of copyright. The depositor must determine the copyright status of the digital objects uploaded to your Memory Project. Document your project's copyright clearance process.

Your digital collections must fit into one of the following categories:

- Items are in the public domain and may be copied and disseminated without restrictions.
- Items are protected by copyright but permission to use them has been (or will be) secured before uploading.
- Items are protected by copyright but fair use provisions apply.
- Items are protected by copyright but your institution owns the copyright.

Review any existing deeds of gift. If the copyright owner is available, work with him or her to secure the appropriate permissions. Orphan works—those for which the copyright owner cannot be identified or found—may be considered for inclusion. Consult with OWLS before beginning a collection which includes orphan works.

You will need to choose the wording of the copyright statement that will appear with the images in CONTENTdm, and discuss how your organization and the copyright owner will handle requests for images.

Recollection Wisconsin, as a DPLA Service Hub, is moving toward requiring the use of the standardized rights statements currently under development by Europeana and the Digital Public Library of America (DPLA). We recommend that you model your rights statements on those found here:

<http://rightsstatements.org/en/>

We also recommend that you review these copyright sites:

- Copyright Basics by the U. S. Copyright Office:
<http://www.copyright.gov/circs/circ01.pdf>
- Peter Hirtle's chart helps you determine whether published or unpublished works are in the public domain:
<http://copyright.cornell.edu/resources/publicdomain.cfm>

- Mary Minow's article is a very thorough explanation of copyright as it relates to library digitization projects:
<http://www.llrx.com/features/digitization.htm>
- Interactive tool for evaluating Fair Use from the Office for Information Technology Policy of the American Library Association:
<http://librarycopyright.net/resources/fairuse/>
- The Society of American Archivists (SAA) has published best practices for orphan works (works for which the copyright holder cannot be reached):
<http://www2.archivists.org/sites/all/files/OrphanWorks-June2009.pdf>

Visit the Recollection Wisconsin guidelines page for more information on copyright: <http://recollectionwisconsin.org/organizations/guidelines>

Digital Asset Management

File Naming and Directory Structure

Simple objects consist of only one file plus metadata that is uploaded to CONTENTdm. A photograph and its metadata constitute a simple object in CONTENTdm.

To name simple object files, design a short, consistent, and descriptive naming convention and directory structure for your digital objects. Consider how you will differentiate the master image file from the access image file. Consider how you will match images to originals and how you will retrieve files when they are requested. Document your procedure so that future users will be able to navigate the files even if the creator is no longer available.

OWLS' recommended model for naming simple object files is described in the InfoSoup Memory Project Metadata Style Guide.

A **compound object** is an assemblage of files that is retrieved as a single file by end-users. Postcards, pamphlets, and books are compound objects.

Compound objects have different naming requirements because CONTENTdm uses the file names and directory structure to upload the files in the correct order and build the compound object. Please consult with OWLS to design an appropriate naming convention for your compound object(s) before you begin scanning. CONTENTdm uses four types of compound objects:

Object Type	Content
Document	An item with several pages, but no hierarchy; usually with matching searchable text files. Pamphlets and booklets are generally document-type compound objects.
Monograph	A document with multiple hierarchies, such as a book with chapters. Also usually includes searchable transcript files.
Picture Cube	Six image files, typically views of a 3D object
Postcard	Two image files

Storage and Backup

OWLS is not responsible for storage of master files or backups of access images. You will need significant storage space for Master files, which are large, uncompressed image files. Access images are smaller, compressed image files that might be used on your web site or shared with patrons. Both sets of files should be stored in a dedicated location and backed up. Consider whether this can be worked into your existing backup procedures or whether you will need to create new procedures.

Long-term Preservation

- A master image should be created using file formats and compression techniques that conform to industry standards such as TIFF, PNG or lossless JPEG2000. OWLS recommends TIFF.
- Store the master image file in a designated location on an appropriate long-term storage medium. The file should be used to produce derivative image files as needed for a variety of current and future user needs.
- Avoid proprietary formats such as camera RAW or JPEG as a master version.
- Create multiple backup copies of all files on a stable medium. We suggest acquiring two external hard drives, and storing one off site. CDs and DVDs have been found to degrade over a relatively short time. Store media in an appropriate environment.
- Monitor and recopy data as needed.
- Outline a migration strategy for transferring data across generations of technology.
- Anticipate and plan for future technological developments.

Marketing

OWLS works with Recollection Wisconsin to coordinate marketing efforts for the statewide collection and DPLA. Contributing organizations are responsible for marketing their local digital collections.

Marketing ideas:

- Issue a press release to local news outlets announcing the digitization project and digital collections.
- Create and distribute brochures and/or flyers.
- Announce the project in the organization's newsletter or other publications.
- Showcase new digital collections on your organization's home page.
- Send a staff member or volunteer with a laptop to popular local spots to demonstrate digital collections:
 - Visit the local coffee shop or farmers market.
 - Display digital collections at events held at local historic sites.
 - Hold a giveaway or raffle to get people's attention.
- Ask, "Where do people go first to look for this kind of information?" and then, market there!
 - Upload a few digitized images to Flickr with descriptions that point back to the library's related digital and physical collections.
 - Use historical resources to contribute to the local history pages on Wikipedia and include references pointing to the library and archives.
- Share digital collections on your organization's social media accounts. See Recollection Wisconsin's Tumblr account for some excellent examples: <http://wiscohisto.tumblr.com/>
- Request that the chamber of commerce and other relevant local organizations link to the new digital collections from their Web sites, and post announcements of the new digital collections.
- Work with local experts to develop library programming around your collections.

Impact on Public Services

Once images are online, contributing organizations should expect an increase in reference requests and requests for copies of those images, either as high resolution digital files or hard copies. It is important to have a policy and procedures in place to handle such requests. The policies will likely vary due to differences in the copyright status of various collections, and specific procedures should be recorded for each collection.

Questions to consider:

- How will your organization handle the increase in reference inquiries?
 - Often these requests will be from people outside of your community.
- Who will be the contact person for image requests?
- Is the image requested in the public domain?
 - If so, the library or donor should not charge licensing fees, but nominal service fees are reasonable.
- What are the wishes of the copyright holder?
 - If the library is in contact with the copyright holder, these issues should be discussed before publication.
- Will requests for commercial uses be treated differently than those for non-commercial purposes?
- How will fees be processed and tracked?
- How does your organization currently handle requests for images that have not been digitized?

See the sample forms at the end of this guide.

- Duplication Request Form
- Request to Publish or Display Form

Project Implementation

Workflow Overview

Please do not hesitate to contact OWLS for support at any stage in the process.

1. Read guidelines, review online resources, and attend training as available. OWLS will provide training as staff time allows.
2. Select your first collection and document copyright clearance.
3. Submit the Project Profile form to OWLS for approval.
4. Define your new collection – think about the items you have chosen, what you know about the items (their metadata), and what info is applicable to all the items in the set (perhaps a copyright statement or note about where you obtained the images, i.e., the source).
5. Determine your file naming convention for the items you'll be scanning.
6. Choose which optional Metadata Fields you want to use in addition to those that are *Required* and *Required if Available*.
 - a. If more than one person will be assigning metadata, create a Metadata Dictionary for your collection
 - b. Consult the appropriate controlled vocabularies to choose standard terms.
7. Contact the ContentDM Administrator at OWLS and provide to them
 - a. The Title of the collection (so it can be added to the server)
 - b. A one-paragraph description of the collection for use on the web site.
 - c. The Metadata Fields and local controlled vocabularies you want to use.
 - d. Sample records for 2-10 images.
8. Scan images according to Recollection Wisconsin guidelines, and assign them the appropriate file names as you save the files.
 - a. Have a clean, dedicated work area (no food or drink by photos).
 - b. 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.

- c. While scanning, regularly stop and check the images for quality and consistency. Be especially vigilant when more than one person is involved with scanning.
- d. File Types and Image Editing
 - i. Scan as TIFFs – archival version; these will be huge in file size
 - ii. Optional JPGs – The software will convert your TIFFs to access JPGs at upload; but you can create your own set of access JPGs at this time if you anticipate needing them.
 - iii. Keep editing to a minimum, but complete the following steps as needed for quality control.
 - 1. Straighten / Image Rotation
 - 2. Crop
 - 3. Adjust Histogram
 - 4. Sharpen
 - ii. Create transcript files in .txt format if applicable.
- 9. Complete the Excel metadata template appropriate to the item type you have scanned (simple or compound).
- 10. Send your image files, transcript files, and metadata file to OWLS for uploading.
- 11. The OWLS CONTENTdm Administrator will check your files for quality control and compliance with guidelines and will contact you with any clarifying questions.
- 12. The OWLS ContentDM Administrator will Upload, Approve, and Index the collection, and notify you when it is published.
- 13. Review the published collection and contact OWLS with any questions or corrections.
- 14. Train library staff on the scope and navigation of digital collections.
- 15. Link to your collection landing page from your library's web site and promote your collection to your community.

Standards for Scanning

General Rules:

- **Save all scans as TIFFs with NO compression.** JPGs use a form of lossy compression that is inappropriate for archival images.
- Scan at the highest resolution appropriate to the informational content of the originals.
- Scan at an appropriate level of quality to avoid rescanning and re-handling of the originals in the future.
- Straighten and crop images as needed. Keep other image manipulation to a minimum.

The following standards are based on RW guidelines. Scans that meet these recommendations will produce a print of reasonable quality up to 20" x 24" and should be sufficient for most publishing requirements. However, they are NOT intended to serve as preservation [archival] files to replace the original material. *(The last sentence means don't throw away the actual item just because you made a good digital version.)*

Original material	Scanning resolution	Bit depth	Approximate scanned dimensions	Approx. size of preview image
<i>Photographs</i>				
16" x 20" +	200 ppi	24-bit color	6400 x 8000 pixels	146 MB
8 ½" x 11"–16" x 20"	300 ppi	24-bit color	3200 x 4000 pixels	36 MB
8" x 10"	400 ppi	24-bit color	3200 x 4000 pixels	36 MB
5" x 7"	625 ppi	24-bit color	3200 x 4000 pixels	36 MB
4" x 5"	800 ppi	24-bit color	3200 x 4000 pixels	36 MB
4" x 2 ½"	1200 ppi	24-bit color	3200 x 4000 pixels	36 MB
<i>Text</i>				
Print—no images	600 ppi	1-bit bitonal	Varies	Varies
Print—with images	300 ppi	8-bit grayscale or 24-bit color	Varies	Varies
Manuscript	400 ppi	8-bit grayscale or 24-bit color	Varies	Varies

Visit the Recollection Wisconsin guidelines page for more information on scanning, including a video tutorial:

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

Metadata

Carefully review the **InfoSoup Memory Project Metadata Style Guide**. The Style Guide serves to ensure consistency across the various collections and will result in a better experience for the users. Consistent and complete metadata also makes it more likely that your digital objects will be found, especially when the data is integrated with other online digital collections.

List of Metadata Elements

	Element Name	Priority	Enforced Controlled Vocabulary	Recommended Controlled Vocabulary
1	Title	Required		
2	Creator	Required if available		Library of Congress Name Authority File (LCNAF)
3	Date	Required if available		
4	Description	Recommended		
5	Subject	Required	Thesaurus of Graphic Materials (TGM)	
6	Local subject	Optional		Local
7	Contributors	Optional		LCNAF
8	Publisher	Required if available		
9	Language	Optional	Languages	
10	Community	Required if available	Local	
11	County	Required if available	WHO Counties	
12	State	Required if available		
13	Type	Required	Dublin Core Metadata Initiative (DCMI) Type	
14	Format	Required	Internet Medium Type (IMT)	
15	Medium	Required	TGM	
16	Dimensions	Optional		
17	Submitter	Required	Local	
18	Digitization information	Optional		
19	Date digitized	Required		
20	File name	Required		
21	Digital collection	Required	Local	
22	Source	Recommended if applicable		
23	Rights	Required		Local
24	Contact	Required		
25	Hidden field	Optional		
26	Transcript	Recommended if applicable		

Enforced Vocabulary = a vocabulary displayed by CONTENTdm from which users choose values.

Recommended Vocabulary = a vocabulary (not displayed by CONTENTdm) from which users are advised to choose values.

Local vocabulary = a vocabulary set up for a collection to improve speed and consistency of data entry for fields with repetitive values, and to ensure consistent usage across projects and collections.

A **Data Dictionary** is a document that defines how metadata will be created and formatted for a specific collection. Create a Data Dictionary for your collection by starting with the InfoSoup Memory Project Metadata Style Guide and documenting the specific decisions you have made about each field for your collection. A Data Dictionary is especially important to ensure consistency when more than one person will be creating metadata for a collection.

Final advice: Please consult OWLS early and often to avoid costly mistakes. We're excited to help put local collections online.

Happy digitizing!