

The Creation and Improvement of a Digitization Workflow

Experience at the Milwaukee County Historical Society

Michael Barera



Project goals














- ▶ The goal is to create and successfully implement a modern but achievable digitization workflow for a county historical society
- ▶ Build on an older digitize-on-demand program that does not have a metadata component as well as an earlier attempt to create a digitization-with-metadata program that created policy documents but that was never implemented
- ▶ Create preservation and access files according to best practices in the field:
 - ▶ 600 dpi TIFFs for preservation
 - ▶ 600 dpi JPGs for high-resolution access for our patrons
 - ▶ 200 dpi JPG thumbnails for our website and low-resolution access for our patrons
- ▶ Create extensive and highly usable metadata in both spreadsheet form and embedded into the files themselves to provide modern retrieval capabilities by description, date, creator, subject, etc.
- ▶ Lay the groundwork for a future Internet platform for our photos by preparing a critical mass of digitized photos with metadata (we will begin working with Recollection Wisconsin, the state's DPLA affiliate, this fall)

Two-step process

- ▶ Step one, done by intern or volunteer:
 - ▶ Digitize print photographs as 600 dpi TIFFs
 - ▶ Create draft metadata in spreadsheet
- ▶ Step two, done by me as Assistant Archivist and Digitization Specialist:
 - ▶ Review and revise spreadsheet metadata
 - ▶ Embed metadata into 600 dip TIFFs
 - ▶ Create 600 dpi JPG and 200 dpi JPG thumbnail derivative files
 - ▶ Quality control on metadata embedded into JPGs

What went well the first time















- ▶ Digitizing the photos themselves and creating the following three files:
 - ▶ 600 dpi TIFFs
 - ▶ 600 dpi JPGs
 - ▶ 200 dpi JPG thumbnails
- ▶ Extensive and highly usable (and searchable) metadata in spreadsheet form
- ▶ Metadata successfully embedded into the files themselves (with two exceptions, as noted later)
- ▶ Tags (“subjects”) that enable effective keyword searching of the digitized photos

 Jones, John and Etta 6	7/23/2022 1:34 PM	TIFF File	44,884 KB
 Jones, John and Etta 7	7/23/2022 1:35 PM	TIFF File	41,084 KB
 Jones, John and Etta 8	7/23/2022 1:36 PM	TIFF File	41,383 KB
 Jones, John and Etta 9	7/23/2022 1:37 PM	TIFF File	37,912 KB
 Jones, John and Etta 10	7/23/2022 2:26 PM	TIFF File	41,802 KB
 Jones, John and Etta 11	7/23/2022 2:27 PM	TIFF File	42,692 KB
 Jones, John and Etta 12	7/23/2022 2:28 PM	TIFF File	18,996 KB
 Jones, John and Etta 13	7/23/2022 2:29 PM	TIFF File	19,515 KB
 Jones, John and Etta 14	7/23/2022 2:30 PM	TIFF File	38,851 KB
 Jones, John and Etta 15	7/23/2022 2:31 PM	TIFF File	11,602 KB
 Jones, John and Etta 16	7/23/2022 2:31 PM	TIFF File	18,352 KB
 Jones, John and Etta 17	7/23/2022 2:32 PM	TIFF File	32,043 KB
 Jones, John and Etta 18	7/23/2022 2:33 PM	TIFF File	36,289 KB

Original TIFF files (before renaming)

Jones, John and Etta 5	8/30/2022 2:27 PM	JPG File	1,081 KB
Jones, John and Etta 6	8/30/2022 2:28 PM	JPG File	4,612 KB
Jones, John and Etta 7	8/30/2022 2:29 PM	JPG File	3,932 KB
Jones, John and Etta 8	8/30/2022 2:29 PM	JPG File	4,064 KB
Jones, John and Etta 9	8/30/2022 2:29 PM	JPG File	3,600 KB
Jones, John and Etta 10	8/30/2022 2:30 PM	JPG File	4,623 KB
Jones, John and Etta 11	8/30/2022 2:31 PM	JPG File	4,126 KB
Jones, John and Etta 12	8/30/2022 2:31 PM	JPG File	1,975 KB
Jones, John and Etta 13	8/30/2022 2:31 PM	JPG File	2,188 KB
Jones, John and Etta 14	8/30/2022 2:32 PM	JPG File	3,577 KB
Jones, John and Etta 15	8/30/2022 2:32 PM	JPG File	1,116 KB
Jones, John and Etta 16	8/30/2022 2:33 PM	JPG File	1,948 KB
Jones, John and Etta 17	8/30/2022 2:33 PM	JPG File	2,644 KB
Jones, John and Etta 18	8/30/2022 2:34 PM	JPG File	2,286 KB

600 dpi JPG derivatives (before renaming)

 Jones, John and Etta THUMB 5	8/30/2022 4:40 PM	JPG File	117 KB
 Jones, John and Etta THUMB 6	8/30/2022 4:40 PM	JPG File	164 KB
 Jones, John and Etta THUMB 7	8/30/2022 4:41 PM	JPG File	154 KB
 Jones, John and Etta THUMB 8	8/30/2022 4:42 PM	JPG File	153 KB
 Jones, John and Etta THUMB 9	8/30/2022 4:43 PM	JPG File	139 KB
 Jones, John and Etta THUMB 10	8/30/2022 4:43 PM	JPG File	202 KB
 Jones, John and Etta THUMB 11	8/30/2022 4:44 PM	JPG File	139 KB
 Jones, John and Etta THUMB 12	8/30/2022 4:45 PM	JPG File	124 KB
 Jones, John and Etta THUMB 13	8/30/2022 4:45 PM	JPG File	125 KB
 Jones, John and Etta THUMB 14	9/10/2022 2:48 PM	JPG File	104 KB
 Jones, John and Etta THUMB 15	9/10/2022 2:49 PM	JPG File	109 KB
 Jones, John and Etta THUMB 16	9/10/2022 2:51 PM	JPG File	124 KB
 Jones, John and Etta THUMB 17	9/10/2022 2:52 PM	JPG File	124 KB
 Jones, John and Etta THUMB 18	9/10/2022 2:53 PM	JPG File	124 KB

200 dpi JPG thumbnails (before renaming)

	Entry/Object ID	Description	Type of Photograph	Subject	Subject Place
t Tables,	Jones, John and Etta 6	camera and others talking to each other. Mr. Dodye, Ms. La Nane, Ms. Ford's Roomer. By John and Etta Jones, Milwaukee, WI	N/A	banquet; table; group; people; eating; dining; smiling; talking; fancy; dodye; jones; ford; column; curtains	Milwaukee, WI
t Tables,	Jones, John and Etta 7	Many people sit at banquet tables, some facing the camera and others talking to each other. Ms. Ford, Mrs. Hartsfield. By John and Etta Jones, Milwaukee, WI	N/A	banquet; table; group; people; eating; dining; smiling; talking; fancy; hartsfield; jones; ford; column; curtains	Milwaukee, WI
ated	Jones, John and Etta 8	A group of women in formal attire pose for a photo. L-R: Unnamed, Mrs. G.L. Glover, Mrs. Orange Paine (Elvira), Mrs. Charlotte Ward, Mrs. Elliott Kiff, Sr., Unnamed, Unnamed. By John and Etta Jones, Milwaukee, WI	N/A	table; banquet; group; women; glover; pain; ward; kiff; jones; milwaukee; centerpiece; decoration; formal	Milwaukee, WI
t Tables,	Jones, John and Etta 9	A group of people sitting at banquet tables all face the camera. Photo by Reiersen Studio in Madison, WI	N/A	group; table; banquet; chandelier; formal; people; milwaukee	Milwaukee, WI
Undated	Jones, John and Etta 10	A grill is lit while a picnic table is set up. L-R: Mrs. Etta Jones, Mr. Alfred McClannahan, Mrs. Emma McClannahan. By John Jones, Milwaukee, WI	N/A	picnic; table; etta jones; alfred mccannahan; emma mccannahan; milwaukee; food; grill; eating; recreation; outside; park; trees	Milwaukee, WI
Hall,	Jones, John and Etta 11	Labor Temple Hall, 8th and Walnut. Large group of people in formal attire faces camera for a photo. By John and Etta Jones, Milwaukee, WI	N/A	banquet; hall; table; centerpiece; food; meal; cooking; milwaukee; formal; labor temple	Milwaukee, WI

Initial metadata in John and Etta Jones Collection spreadsheet

Iterative changes made after initial work

- ▶ Metadata embedded in the “Comments” field (name of the collection) and “Description” field of TIFFs did not transfer to JPGs, so this information had to be added manually to JPGs in a “cleanup” step after derivative file creation
- ▶ “Milwaukee County Historical Society” began being embedded into the “Copyright” field in the metadata as a sort of digital watermark

Iterative changes made after initial work

- ▶ A few changes to the metadata spreadsheet fields were implemented to produce improved results:
 - ▶ Renamed “Made/Created” to “Date” and moved it left to be one of the most important columns in the spreadsheet, then moved information about the date out of the “Description” field (this will better match the embedded metadata)
 - ▶ Created a new “Photographer/Creator” field, and moved information on the photographer to it from the “Description” field (this will also better match the embedded metadata)
 - ▶ Removed the “Type of Photograph” field as it is not useful to print scans
 - ▶ Renamed the “Subject” field “Tags” and the “Subject Place” field “Location” to improve clarity and understanding of the field

Iterative changes made after initial work

- ▶ A naming convention for files was created and implemented. It consists of four-letter codes for all photographers (usually the first four letters of their surname, with a few exceptions) followed by a four-digit serial number corresponding with existing assigned numbers. (This approach is designed to match an older digitization scheme that produced filenames such as “ps4083e” and “ns0469.”)
- ▶ Similarly, the naming convention appended a “t” to the end of the filename for 200-dpi thumbnails in the same way the old naming convention appended an “e” to mean “enhanced” (for a digitally altered file).
 - ▶ Initial filename: Jones, John and Etta THUMB 7.jpg
 - ▶ New filename: jone0007t.jpg
- ▶ The new filenames are then written on the back of the print photographs digitized in pencil or archival ink, matching the old digitization-on-demand approach and helping better connect physical prints to digitized photos.

Title	Man and Woman with Car, ...
Subject	A man and a woman, possi...
Rating	☆☆☆☆☆
Tags	group; people; women; man...
Comments	MSS 0616 Jones, John and...

Origin

Authors	Jones, John and Etta
Date taken	8/3/2022 12:43 PM
Program name	Windows Photo Editor 10.0...
Date acquired	
Copyright	Milwaukee County Historica...

Example of embedded metadata after revisions



jone0473



jone0471



jone0470



jone0469



jone0467



jone0466

Example of searching by tag: “automobile”



jone0340



jone0306



jone0305























Example of searching by tag: “church”

	A	B	C	D	E	F	G	H	I	J
1	Name/Title	Entry/Object ID	Description	Date	Photographer/ Creator	Tags	Location	Collection	Catalogued By	Typ File
2	Large Group at Picnic Table, Undated	jone0001	A group sits at a picnic table and smiles for a picture. L-R Back Row: Sonny Hurst, Caroline Dugas, Lester Dugas Sr., Etta Jones, Mildred Gilmer, Lilla Dugas, Delores Dugas. L-R Front Row: Pernell Omahundro, Geradene Goens, Lester Dugas Jr.	Undated	John Jones	picnic; table; bench; group; etta jones; john jones; milwaukee; group; recreation; dugas; hurst; gilmer; goens; food; eating; smiling	Milwaukee, WI	MSS 0616 Jones, John and Etta Photograph Collection	Lydia Prueher	TIF
3	Alfred McClannahan, Undated	jone0002	Alfred McClannahan lights a grill for an outing in a park.	Undated	John and Etta Jones	grill; alfred mcclannahan; cooking; outing; picnic; park; john jones; etta jones; milwaukee; recreation; food	Milwaukee, WI	MSS 0616 Jones, John and Etta Photograph Collection	Lydia Prueher	TIF

	A	B	C	D	E	F	G
1	Name/Title	Entry/Object ID	Description	Type of Photograph	Subject	Subject Place	Collection
2	Large Group at Picnic Table, Undated	Jones, John and Etta 1	A group sits at a picnic table and smiles for a picture. L-R Back Row: Sonny Hurst, Caroline Dugas, Lester Dugas Sr., Etta Jones, Mildred Gilmer, Lilla Dugas, Delores Dugas. L-R Front Row: Pernell Omahundro, Geradene Goens, Lester Dugas Jr. #12 Large Group. By John Jones, Milwaukee, WI	N/A	picnic; table; bench; group; etta jones; john jones; milwaukee; group; recreation; dugas; hurst; gilmer; goens; food; eating; smiling	Milwaukee, WI	MSS 0616 Jones, John and Etta Photograph Collectio

Revised (top) vs. original (bottom) spreadsheet fields

Name	Date modified	Type	Size
 hard0001t	11/16/2022 2:40 PM	JPG File	110 KB
 hard0002t	11/16/2022 2:41 PM	JPG File	80 KB
 hard0003t	11/16/2022 2:41 PM	JPG File	111 KB
 hard0004t	11/16/2022 2:42 PM	JPG File	214 KB
 hard0005t	11/16/2022 2:42 PM	JPG File	178 KB
 hard0006t	11/16/2022 2:42 PM	JPG File	186 KB
 hard0007t	11/16/2022 2:43 PM	JPG File	123 KB
 hard0008t	11/16/2022 2:44 PM	JPG File	107 KB
 hard0009t	11/16/2022 2:44 PM	JPG File	106 KB
 hard0010t	11/16/2022 2:45 PM	JPG File	180 KB
 hard0011t	11/16/2022 2:45 PM	JPG File	151 KB
 hard0012t	11/16/2022 2:46 PM	JPG File	172 KB
 hard0013t	11/16/2022 2:46 PM	JPG File	141 KB
 hard0014t	11/16/2022 2:47 PM	JPG File	142 KB
 hard0015t	11/16/2022 2:47 PM	JPG File	165 KB
 hard0016t	11/16/2022 2:47 PM	JPG File	118 KB
 hard0017t	11/16/2022 2:48 PM	JPG File	151 KB
 hard0018t	11/16/2022 2:48 PM	JPG File	124 KB
 hard0019t	11/16/2022 2:49 PM	JPG File	99 KB
 hard0020t	11/16/2022 2:50 PM	JPG File	128 KB

Revised appearance of filenames (Hardie collection thumbnails)

Policy documents created and progress made

1. [Digital file naming scheme \(by photographer\)](#)
2. [Digitization metadata crosswalk](#)
3. Digitization priority collections
4. [MCHS Digitization Workflow](#)

1	Last name	First name	Code	Example filename	Any deviation from standard format (first four letters of last name)
2	Boyd	Robert	boyd	boyd0001	
3	Breitwish	Alfred J.	brei	brei0002	
4	Bruhnke	Les	bruh	bruh0003	
5	Carberry	Fred W.	carb	carb0004	
6	Case-Knapp Family Photographs		cakn	cakn0005	First two letters of both last names.
7	Cassidy	Robert	cass	cass0006	
8	Hall	John S.	hall	hall0007	
9	Hardie	George	hard	hard0008	
10	Jones	John & Etta	jone	jone0009	
11	Koshollek		kosh	kosh0010	
12	Kuhli	L.F.	kuhl	kuhl0011	
13	Kwasniewski	Roman	kwas	kwas0012	
14	Mason	Harold	maso	maso0013	
15	Mazza	Theodore & Anna	mazz	mazz0014	
16	Millard	Gerald W.	mill	mill0015	
17	Minsah	James	mins	mins0016	
18	Oberwise	Lyle	ober	ober0017	
19	Otis	John P.	otis	otis0018	
20	Pfister/Vogel/Uihlein Family Photo Collection		pvuf	pvuf0019	First letter of each of the families' names plus the word "family."
21	Sonnemann	Henry	sonn	sonn0020	
22	Stanfield	Harold W.	stan	stan0021	

Digital file naming scheme (by photographer)

Digitization Metadata Crosswalk

Michael Barera's finalized conversions, Dec. 2022

Spreadsheet field	EXIF metadata field
Name/Title	Title
Description	Subject
Date	Date taken ["MM-DD-YYYY" format]*
Photographer/Creator	Authors
Tags	Tags
Collection	Comments
Repository	Copyright

* = unless undated; use "1-1-YYYY" if only year is known; use "MM-1-YYYY" if year and month, but no date, is known; leave unaltered if undated

Digitization metadata crosswalk (for embedding spreadsheet metadata into the files themselves)

Priority Collections for Digitization

Michael Barera, Apr. 2023

1. ~~John and Etta Jones Collection (Mss-0616)~~ [DONE]
2. George Hardie Papers photographs (Mss-0493) [PARTIALLY DONE, ON HOLD AS OF FEB. 2023]
3. ~~Miller Blatz and Miscellaneous Brewery photograph collection~~ [DONE]
4. Traffic Studies (M-95) [IN PROGRESS]
5. Socialist photographs (M-88/M-89)
6. Lawrence Peterson & Associates photo albums
7. African American Ethnic Groups (M-21A)
8. Schlitz photos
9. Berger, Philip K. (M-227)
10. Pabst photos
11. Allis-Chalmers Collection photographs
12. Parks (in A-Z collection)

Digitization priority collections (list of prioritized collections for digitization and progress made)

MCHS Digitization Workflow

Michael Barera, Feb. 2023

Phase one (intern or volunteer)

1. Scan the photograph or other item in question using the Plustek scanner and the VueScan software. Make sure that the following settings are selected in the VueScan software:
 - a. "Input" tab:
 - i. Source: "OpticPro A320E" (not "TOSHIBA")
 - ii. Media: "Color" for full-color photographs, "Gray" for black and white/grayscale photographs
 - iii. Media Size: "12 x 17 in." [the maximum size]
 - iv. Scan dpi: "600"
 - b. Output:
 - i. Default folder: [see step #2 for sample filepath]
 - ii. TIFF file: selected
 - iii. TIFF profile: selected [do *not* select any other filetypes, such as JPEG or PDF]
2. Save the resulting scan (a 600 dpi TIFF) in the TIFF subfolder of the folder named after the collection from which the photograph or other item is sourced. If no images from the collection in question have been digitized before, the folder and subfolder must be created.
 - a. Example filepath: Shared Folders\Data\Archives\Digital Archives\Photograph

MCHS Digitization Workflow (document used to guide interns' /volunteers' scanning/metadata creation and my revision, embedding, derivative creation, and quality control)