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 Collections\Jones, John and Etta\TIFFs
3. Name the file as [four-letter collection abbreviation][four-digital serial number].tiff. (See the document "Digital file naming scheme (by photographer).xls" in the [Digitization Workflow folder](#) for a full list of already-established four-letter codes.)
   a. Example: "hard0005.tiff"
4. Write the name of the digitized file on the back of the scanned photo in pencil.
   a. Example: "hard0005"
5. Fill out all the fields on the appropriate scans database as completely as possible. [See table below for the fields, their explanations, and examples.] This metadata will be reviewed by the Assistant Archivist and Digitization Specialist.
   a. Spreadsheet filepath: Shared Folders\Data\Archives\Digital Archives\Photograph Collections\Photograph scanning spreadsheets

Field names	Explanations of fields	Examples
Name/Title	A short description of the photograph, followed by a comma and the year of the photograph (if known) or the word "Undated" (if unknown).	Aerial View of Private Hangars on the East Side of Milwaukee County Airport, 1934
		Aerial View of General Mitchell Field, Undated
Entry/Object ID	A unique identifier assigned to each digitized photograph. It is	hard0208

	composed of a four-letter collection abbreviation followed by a four-digit serial number, given sequentially as photographs are digitized, followed by the file extension ".tiff." See the document "Digital file naming scheme (by photographer).xls" in the <u>Digitization Workflow folder</u> for a full list of already-established four-letter codes.	
		hard0459
Description	A thorough free-text description of the photograph. The first sentence should be similar to the Name/Title field, but typically with more detail. The first sentence or two of the Description should be written like a caption. It should be sourced from any captions or descriptions on the photo itself (if available), created by the intern/volunteer, or (ideally) a combination of the two. The last sentence or two should serve as "alt text," describing what is visible in the photograph for the benefit of visually impaired people using screen readers.	An aerial view of private hangars on the east side of Milwaukee County Airport. The large hangar on the right is occupied by Midwest Airways. "Midwest Airways" is visible on top of the large hangar on the right. A partial view of runways in the lower left corner is visible.
		USS Macon, a United States Navy dirigible, flying over Milwaukee. Clouds are visible. Written on the side of the airship is "U.S. Navy."
Date	The date the photograph was taken, if known. This should be entered with as much specificity as possibly, either as the year alone ("1920"), month and year ("June 1920"), or a specific date ("June 15, 1920"). For an approximate year, precede the year or date range with the word "Circa" ("Circa 1920" or	1934

		"Circa 1920s"). For an unknown date, use the word "Undated."	
			May 24, 1922
			May 1921
			Circa 1935
			Undated
Photographer/Creator	The photographer of the photograph, if known. This can be either a person or an organization. For a person, enter the name in a "Last name, First name" format ("Toepfer, Albert F."). If the photographer is unknown, use the word "Unknown."	Midwest Airways
			Toepfer, Albert F.
			Unknown
Tags	A group of free-text keywords describing the content of the photograph. These tags allow easier discovery of the photograph by both searching and browsing. Start creating these tags by using key words from the Description ("hangars," "Milwaukee County Airport," "USS Macon," "United States Navy," etc.). Then, supplement them with other descriptive terms relevant to the photograph that are not included in the Description, often because they are less important details or are synonyms to words used in the Description ("trees," "runway," "airship," "clouds," etc.).	Aerial view; hangars; Milwaukee County Airport; Milwaukee Mitchell International Airport; Midwest Airways; trees; field; runway; town of lake
			USS Macon; United States Navy; dirigible; flying; Milwaukee; clouds; airship
Location	The city where the photograph was taken, if know. Use the format "city, state abbreviation" ("Milwaukee, WI;" "West Allis, WI;" etc.). If only the state is known, use it instead ("Wisconsin," "Michigan," etc.).	Milwaukee, WI

	If the location is unknown, use the word "Unknown."	
		West Allis, WI
		Wisconsin
		Unknown
Collection	Name of the collection the photograph is from, starting with the manuscript number and including the full name of the collection ("MSS-0493 George Hardie Photography Collection," etc.). This field will be the same for every photograph from a particular collection.	MSS-0493 George Hardie Photography Collection
		MSS 0616 Jones, John and Etta Photograph Collection
Catalogued By	Name of the person who digitized the photograph and created the metadata. Entered in the "First name Last name" format ("Sam Hassold," "Michael Barera," etc.).	Sam Hassold
		Michael Barera
Type of File	Type of file created during digitization. This should always be "TIFF."	TIFF
General Notes	Any additional notes, such as the location of a photograph in a collection with multiple folders ("Box 3, folder 4," etc.). Most often, however, this field is left blank.	Box 3, folder 4
		[left blank]
Repository	The repository that physically holds the photograph in its collections and (most likely) has copyright to it unless it is in the public domain. This should always be "Milwaukee County Historical Society."	Milwaukee County Historical Society

Phase two (Assistant Archivist and Digitization Specialist)

1. Review the fields in the scan database, paying special attention to the description and tags fields, which typically require the most revision and improvement. Reviewing the scanned image and the database at the same time, add any relevant missing information and modify existing information in the database as appropriate.
2. Embed key metadata from the fields in the scan database into the TIFF file itself, using the "Properties" (and then "Details") options in File Explorer. Use the "Digitization metadata crosswalk" document in the Digitization Workflow folder for the correct source and destination fields for this metadata.
3. Open the TIFF file in GIMP (the GNU Image Manipulation Program, a free, open-source counterpart to Adobe Photoshop). Save a derivative 600 dpi JPG (for access) from the 600 dpi TIFF. Create a separate "JPGs" folder (parrel to the "TIFFs" folder) if one does not already exist. Name the file as [four-letter collection abbreviation][four-digital serial number].jpg. (In practice, this simply requires changing the file extension from ".tiff" to ".jpg."
   a. Example: "hard0005.jpg"
4. Using the "Scale" function in GIMP, reduce the resolution of the photograph to 200 dpi and the length of its longer side to 4". Save a derivative 200 dpi JPG thumbnail. Create a separate "Thumbnail JPGs" folder (parrel to the "TIFFs" and "JPGs" folders) if one does not already exist. Name the file as [four-letter collection abbreviation][four-digital serial number]t.jpg (the "t" stands for "thumbnail"). (In practice, this simply requires adding a "t" between the serial number and ".jpg" file extension when saving the file.)
   a. Example: "hard0005t.jpg"
5. Review the embedded metadata for both JPGs and thumbnails, adding any lost metadata from the spreadsheet. (In practice, the "Description" field is being replaced by the "Title" field and the "Comments" field is not saving at all, so both fields must be re-added to both JPGs and thumbnails.)
6. *NOTE*: After the "Description" and "Comments" fields are fixed, as described in step #5, the only outstanding technical limitation regarding EXIF metadata is the date field. Dates cannot be removed from the metadata [at least using the current workflow], so undated files retain their recent EXIF date, which corresponds with their date of digitization. Also, because the EXIF metadata for date is in the format MM-DD-YYYY, year-only and month-year dates are entered in a potentially confusing way: 1-1-YYYY ("1-1-1920," etc.) when only the year is known, and MM-1-YYYY ("6-1-1920," etc.) when the year and month, but not the exact date, are known. This causes potential confusion with dates entered for January 1 or the first of another month, but due to this technical limitation, this is the closest [known way] to correctly enter dates.