BENTLEY HISTORICAL LIBRARY
UNIVERSITY OF MICHIGAN

Digitization Guidelines for Photographs and Textual Documents

Published July 2012
Reviewed December 2018

The Bentley Historical Library (BHL) is pleased to provide these digitization guidelines for the use of individuals and organizations. These guidelines are for use in any digitization project, whether for personal use or professional-level projects. They have been developed by BHL staff and updated to incorporate technical guidelines used by the University of Michigan's Digital Library Production Service (DLPS)[1] and the 2010 Federal Agencies Digitization Guidelines Initiative (FADGI)[2].

Please read this entire document as it contains information about your scanning workflow and the documentation of your material.

---

[1] https://www.lib.umich.edu/digital-content-collections-dcc/about-dlps
[2] http://www.digitizationguidelines.gov/

Digitization and Description

Digitization is the process of transforming analog material into digital form. It is a lengthy process. Depending on the size of the collection at hand, it can take weeks or years to complete the technical process. Providing contextual information, or *metadata*, about the scanning process and original items takes a great deal of time and thoughtful effort.

Consider which parts of the collection have meaning and how this meaning can be conveyed. For example, when scanning a diary, the page order is essential. So might be the cover and binding of the diary, particularly if it contains imprints or artwork. Letters may include envelopes with information such as the addresses and postmarks. There may be relevant information written on the back of a photograph that could explain the scene or give names and dates.

Prepare to scan

To capture the clearest image possible, the scanner must be clean. Before scanning documents or photographs, use a soft cloth to clean the glass top of the scanner to remove debris and dirt, or follow the instructions in the owner's manual. Check the glass surface and platens regularly to make sure that dust from old paper and photographs is removed.

A workspace for digitization should have a controlled lighting source to maintain consistency and quality of images. Changes in room lighting can affect how images are represented on computer monitors and may affect the quality of the scan. Consider setting up a scanning station in a room with low or consistent lighting, with adequate space to lay out originals and keep track of process.

Select scanner settings

Scanner's settings are important because they control the quality of the digitized image.

Color Mode
Color Mode refers to color palettes that will be used to copy and display a scanned image.

- *Bitonal* is a simple two-tone black and white scan. It is typically referred to as "black and white" in scanner settings, and it does not pick up any details in shadings, such as those in photographs or hand-written documents. BHL does not recommend using bitonal color mode.

- *Grayscale* is a palette of 256 gray shades. Grayscale is typically used with black and white photographs or documents that are not scanned in a bitonal setting, such as pages with handwriting or stamped images that have shading details.

- *Color* is a palette of 256 to thousands of color tones, depending on your scanner. Color is typically used for scanning color, sepia photographs or documents with color (such as maps or illustrations).

Figure 1. Examples of a postcard digitized in different color modes. Detail of HS7997, Box 1, Postcard Collection, Bentley Historical Library, University of Michigan

Resolution

Resolution is the finely spaced detail of your scan in color depth and spatial resolution. Resolution is measured in bit-depth and dots per inch (DPI).

- *Bit-depth* refers to how many levels of shades or colors your scan will capture, roughly from 256 shades to over 65,000. Typically, black and white and grayscale images do well at 8-bit depth, and color requires 24-bit depth for extra-fine resolution.

- *DPI* refers to how finely spaced the detail of your scan will be by capturing a certain number of pixels per inch (PPI). Selecting a higher DPI means the detail of your scan will also be higher. It also means that the size of your digital file will be larger and your scanner will take longer to complete the digitized copy.

Figure 2. An example of different spatial resolutions of the same image, in this case the license plate of the leftmost car in the photograph. Illustration by Phil Michel from photograph by Jack Delano, "Cars of Migratory Tomato Wrappers. Kings Creek, Maryland," July 1940. Farm Security Administration Collection, LC-DIG-ppmsca-08927.

The following settings are recommended for common image sizes and types. Although large file sizes or longer scanning times may be beyond the capability or goals of some projects, it is highly encouraged to adhere to the minimum resolutions outlined in the table below and to scan all photographs regardless of whether they are sepia, black/white or color in 24-bit color mode.

Original Material		DPI (minimum / preferred)	Color Mode
Photographs	8" x 10" print	300 / 600	24-bit color
	4" x 6" print or smaller	400 / 800	24-bit color
Text	8.5" x 11" office paper	300 / 400	24-bit color
	4.5" x 6.5" cardstock size or smaller	300 / 400	24-bit color

Unusual sizes such as photographic negatives or panoramas need special consideration when scanning. To achieve the desired spatial resolution, you must measure the size of the original and obtain the size of the longest edge. It is recommended to capture at 5,000 pixels across the long dimension while maintaining a minimum of 300 DPI. These spatial resolutions can be particularly high, and may be outside of the reach of your scanning equipment's technical capacity or may be demanding on your computer's system.

Original Material		Calculation
Photographs	2" x 3" print	5000 pixels / 3 = 1667 approx DPI setting is 1650
	1" x 4" negative	5000 pixels / 4 = 1250 approx DPI setting 1250
	5" x 25" panorama	5000 pixels / 25 = 200 Approx. DPI setting 300 (minimum)

Enhancement of scanned objects

When digitizing, do not adjust the digitized copy with filters or scanning software to sharpen the image, remove scratches correct "red-eye," or adjust the size of the scanned image with reductions or enlargements. The material should be copied as is with no alterations. Additionally, please include all edges within the scan.

The below image on the left has been cropped to remove the edges, while the image on the right captures the edges of the photograph to give the viewer more information about the original.

Figure 3. BL002076, Class Work in Mammalian Anatomy and Histology, Bentley Image Bank

Scanning and saving

Set the object you are scanning against a solid piece of paper in a light color, such as white or light grey, to capture all edges of the original. Do not crop edges of documents.

While not required, the use of *color targets* provides more information about the color values of each photograph or document that is captured during digitizing. Color targets can range in price from $20-30 for the Kodak/Tiffen Color Separation Guide and Gray Scale (Q13) to over $200 from professional photography organizations.

Scan flat or loose pages of material

To scan flat or loose pages of material, such as photographs or letters, place the object on the scanner and copy the front, then the back. Keep the scan order in agreement with the order of the materials, such as page numbers or envelopes with corresponding letters.

Below is an example of a scanned letter with the front of the envelope scanned first, followed by the back of the envelope and the front of the first page.
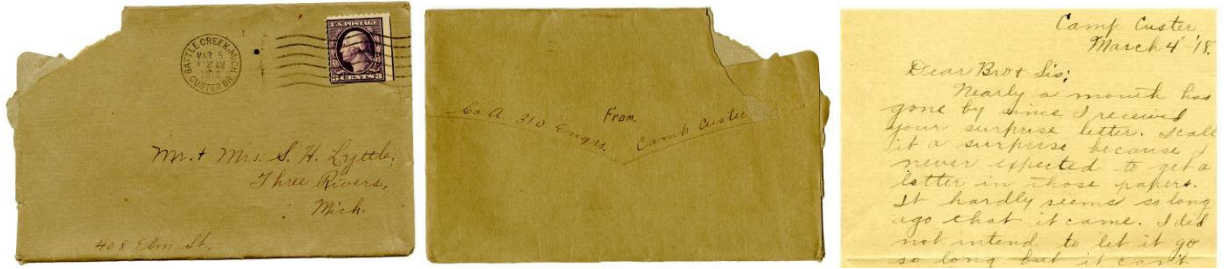

Figure 4. Alfred Lyttle papers, folder 1, box 1, Bentley Historical Library, University of Michigan

Scan bound materials

To scan bound materials such as books and diaries should be carefully placed on the scanning glass. Lightly press along the spine to capture all text or writing inside. Use cautious judgment about the amount of pressure applied to fragile materials. Be aware that moving the material while completing the scan will result in a blurred image that will be of little value. Once a preview scan has been created to satisfaction, scan the image and save it as a digital file.

Below is an example of a diary with the front cover scanned, followed by the inside cover and subsequent pages.

Figure 5. Poetry volume, Louise Willson Papers, Bentley Historical Library, University of Michigan

Save scanned images

Save scanned images in an uncompressed TIFF format (.tif or .tiff file extension). The TIFF format has become the de facto format for scanning high-quality master images of archival materials because it is open source and because it embeds information about the image in the file.

BHL highly recommends the use of TIFF as a file format. JPEG is the second best format. Aside from the loss of technical metadata that is captured in the TIFF file, JPEG files compress images using lossy compression. This means that specific information within the scanned image is discarded to make the file size smaller. This loss of quality means that the scanned image will not retain the same color or resolution quality that a TIFF will. While frequently used for providing access copies online, JPEG is not recommended for storing master image files because it also introduces a generational loss. Every time the image is edited or copied the quality of the image decreases.

Directory structure and file naming

How files are organized and named will have a big impact on your ability to find those files later and to understand what they contain. Be consistent and descriptive in naming and organizing files so that it is obvious where to find specific data and what the files contain.

The best method to maintain the relationship between files is to organize the files into directories (folders). Donors are encouraged to maintain the original order to individual items since that order may itself be informative. Directories should be created to help organize individual files and to provide descriptive information about the files.

```
C:\Documents and Settings
   📁Lowe
      📁 Photographs
         📁 France_October_1919
                  photo1.tif
                  photo2.tif
      📁 Papers
         📁 1918
            📁 Russia_Archangel_1918
                     letter1_envelope_front.tif
                     letter1_page1.tif
                     letter2_envelope_front.tif
                     letter2_page1.tif
            📁 France_Brest_1918
                     letter1_page1.tif
```

One option for organizing digitized files include creating folders by medium, such as Photographs or Papers. Within those folders include additional folders with explanatory names, such as Correspondence (such as letters or postcards, etc.), Diary (such as journals, bound notebooks), or you can group additional files into a Miscellaneous folder if there is no apparent order to subsequent items or if they do not fit into these categories (such as newspaper clippings).
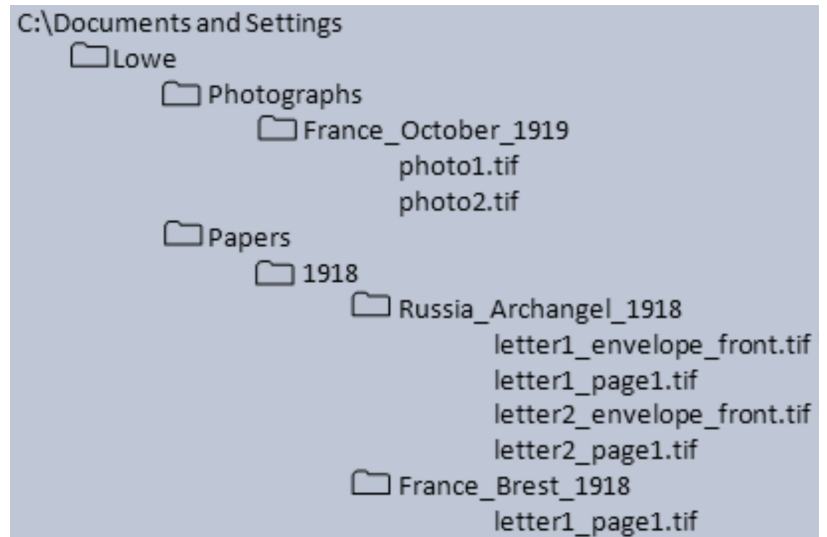
Metadata

Metadata, or information about a collection or file, is crucial for conveying information about your digitized material. To collect this information, BHL recommends using either a spreadsheet template or a text (.txt extension) file.

Share metadata is a spreadsheet (preferred)
Use of a spreadsheet requires spreadsheet-editing software, such as Microsoft Excel, Macintosh Numbers, or Open Office CALC.

All fields in the metadata spreadsheet marked with an asterisk (*) are highly recommended. Record information about scanning equipment, any software used, which format the saved files are in, and the general date of digitization.

Share metadata in a text file (minimum)
A text file can be used as an alternative to a spreadsheet. Note that file names are limited in the number of characters available, which is why it is preferred that you use a text file if you opt not to use the spreadsheet.

To share information about the scanned materials, create a file in a text editor such as Notepad in Windows or Text Edit on a Mac. Record information about scanning equipment, any software used, which format the saved files are in, and the general date of digitization.

Type any descriptive information that accompanies materials, such as names or dates written on a photograph. When thinking about descriptions, consider the important "who, what, where, when" questions that can be answered.

For example, titles provide context about what the item is, such as "William Lowe's Diary", demonstrates how it relates to other files, such as "page 3 of 27", or who and what the scanned image is about, such as "William Lowe in uniform, June 4, 1919, France".
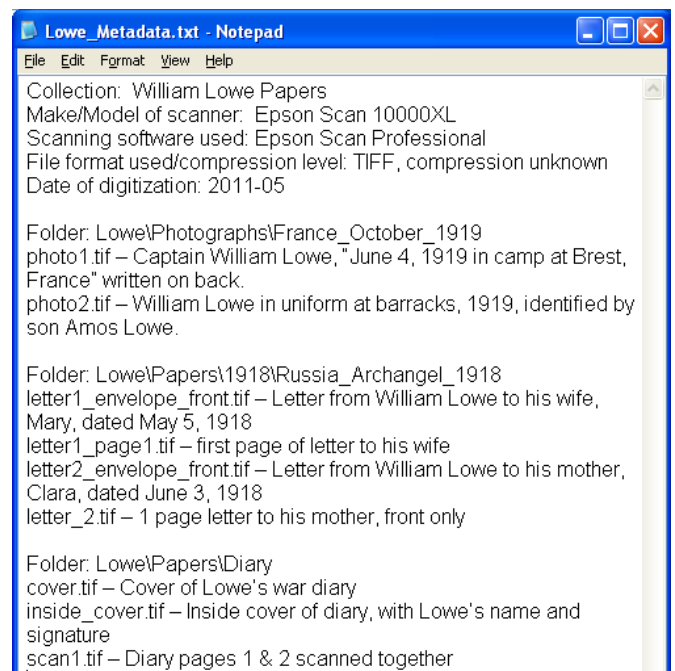
Elements of file description

Titles

Titles should be short, succinct and specific. Researchers might locate this item when searching for it using keywords, or use this information to determine what research value the object might contain when browsing through a list of items.

Creator

Creator is used to identify the creator of the original item, usually an individual such as the author or an organization.

Contributor

Contributor is closely tied to Creator, in that it is the person or company/corporation that played a secondary role in creating the item, such as book editors or publishers and photography studios.

Location

Location refers to the location of the scene that has been captured in photographs or the location where letters were written or sent from.

Description

Description, or subject, incorporates the keywords used in the title but gives more depth of information about the people, places, events, activities or themes depicted by the object. This field provides the best area to provide context for the event, activity or persons in the photograph.

Date of creation

Date of creation is the date of the original item. This could be stamped on the photograph by a film developer, written on the object by the author, or information that you know personally. If the date is known, record the year first, followed by the month and day, i.e., 1919/05/21 for May 21, 1919. If the exact date is unknown, use language such as circa or approximately, i.e., circa 1919.

Physical format

Physical format records format *of the original item* you are digitizing. The Getty Art and Architecture Thesaurus Online is a free tool that provides preferred names/terms and synonyms for people, places, and things[3].

Copyright

Copyright is for recording any known copyright information about the *original item* you are digitizing[4].

---

[3] http://www.getty.edu/research/tools/vocabularies/aat/

[4] https://www.copyright.gov/