Digitization Guidelines for Photographs and Textual Documents
Julianna Barrera-Gomez for the Digital Curation Division, Bentley Historical Library, University of Michigan
July 2012

Forward

The Digital Curation Division of the Bentley Historical Library is pleased to provide these digitization guidelines for the use of individuals and organizations. These guidelines may be used for any digitization project, either for home use or for projects through which content would be offered to an archive, library, or museum. The guidelines have been developed by Bentley Historical Library staff and updated to incorporate technical guidelines used by the University of Michigan’s Digital Library Production Service and the 2010 Federal Agencies Digitization Guidelines Initiative.

The Bentley Historical Library has built our collections through the generosity of over 10,000 donors. Digitized materials make up an increasingly important part of the collections. They enhance significantly the history of the University of Michigan and the state of Michigan. For a general overview of the Bentley and its collections see http://www.bentley.umich.edu.

If you would like your project to become a part of the Bentley collections, please consult with the acquisition staff by scheduling a consultation (contact coordinates are available at bentley.umich.edu or by phone at 734.764.3482). The acquisitions staff might not accept digitized material is suggested guidelines are not followed.

Contents

Digitization Guidelines 2

The Digitization Workflow 2

1. Preparing to scan 2

2. Selecting scanner settings 3

3. Enhancement of scanned object 6

4. Scanning and saving 7

5. Directory structure and File naming 9

6. Metadata 10

Questions 17

Works Cited 18
Digitization Guidelines

The following guidelines are intended for anyone interested in scanning photographs or textual documents and would like to follow the Bentley Library’s in-house digitization criteria. Digitizing is a lengthy process, which can take weeks or months to complete, depending on the size of the collection. Additionally, collecting metadata, or information about your scanning process and original items, will take time and thoughtful effort.

As a first step, please review the owner’s manual for your scanning equipment. In addition to the guidelines below, advanced users may wish to consult the Federal Agencies Digitization Guidelines Initiative (FADGI) Still Image Working Group’s *Technical Guidelines for Digitizing Cultural Heritage Materials* for further guidance and more detailed discussion about the technical aspects of archival grade scanning. For further reference, you can also refer to Matt Adair’s "*Bentley Historical Library Digital Reformatting Procedures Report*" on the 2011 Postcard Collection digitization project.

Context for your collection

Before scanning, consider which parts of your collection have meaning and how you can convey this. For example, if your collection contains a diary, you would want to scan it to preserve the page order, but you might also consider scanning the cover or binding of the diary as well, especially if it contains imprints or artwork. If you have letters, you can capture important information such as the addresses and postmark on the front of the envelope. If you have photographs, is there any information written on the back of the picture that could explain the scene or give names and dates? Contextual information can be captured by scanning the items however, some information may be captured in the metadata that you create for the scans.

The Digitization Workflow

1. Preparing to scan

Before scanning, please read this entire document, as it contains important information about your scanning workflow and the documentation of your material.

In order to capture the clearest image possible, your scanner must be clean! Prior to scanning your documents or photographs, use a soft cloth to clean the glass top of your scanner to remove debris and dirt, or follow the instructions in your owner’s manual. Check the glass surface regularly to make sure that dust from old paper and photographs is removed. Regularly clean the scanner platens, or the material on the scanner lid that presses against your documents, to ensure that it is clean and free of debris that can detract from the image quality of your scan.

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 your
workstation in a room with low or consistent lighting, with adequate room to lay out your originals and keep track of your process.

2. Selecting scanner settings
Your scanner’s settings are important because they control how well the image is copied. Listed below are some common terms that will help you determine the proper settings.

*Color Mode*: this refers to color palettes that will be used to copy and display your scanned image.

- **Bitonal**: a simple two-tone black and white scan. Bitonal is sometimes used with typewritten documents or printed texts that are crisp and easy to read. Bitonal is typically referred to as “black and white” in scanner settings, and it does not pick up any details in shading, such as those in photographs or hand-written documents. The BHL does not recommend using bitonal color mode.

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

- **Color**: a palette of 256 to thousands of color tones, depending on your scanner. Color is typically used for scanning color or 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](image-url)
Resolution: basically the finely spaced detail of your scan in color depth and spatial resolution. Resolution is measured in bit-depth and dots per inch, or DPI.

- Bit-depth: refers to how many levels of shades or colors your scan will capture, generally 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. This is commonly described in combination with color mode, such as scanning to “8-bit grayscale” or “24-bit color.”

- Dots per inch, or DPI: refers to how finely spaced the detail of your scan will be by capturing a certain number of pixels per inch (also referred to as pixels per inch or PPI).

Choosing a higher setting for the DPI means the detail of your scan will be higher, so that users can zoom in to look at finer details. A higher DPI also means 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.

For a more detailed discussion, see “Introduction to Basic Measures of a Digital Image for Pictorial Collections,” from the Library Congress’ Prints and Photographs Division.

Digitization Settings
The following settings are recommended for common image sizes and types. These have been developed to be accessible targets. It is realize that large file sizes or longer scanning times may be beyond the capability and/or goals of some projects. It is highly recommend 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.

Bentley Library recommended color mode and spatial resolution settings for archival materials

<table>
<thead>
<tr>
<th>Original Material</th>
<th>Scan Resolution &amp; Color Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>Photographs</td>
<td>24-bit color</td>
</tr>
<tr>
<td>8 x 10” print</td>
<td>600 dpi preferred, 300 dpi minimum</td>
</tr>
<tr>
<td>4 x 6” print or smaller</td>
<td>800 dpi preferred, 400 dpi minimum</td>
</tr>
<tr>
<td>Textual Documents</td>
<td>24-bit color, 8-bit grayscale minimum</td>
</tr>
<tr>
<td>8.5 x 11” office paper</td>
<td>400 dpi preferred, 300 dpi minimum</td>
</tr>
<tr>
<td>4.5 x 6.5” card stock size or smaller</td>
<td>400 dpi preferred, 300 dpi minimum</td>
</tr>
</tbody>
</table>

**Note:** Unusual sizes such as photographic negatives or panoramas, which are very rectangular, 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. **Special items such as these are recommended to capture at 5,000 pixels across the long dimension while maintaining a minimum of 300 dpi.**

**Example calculations:**
A 2” x 3” photograph: 5000 pixels divided by 3 = around 1667, thus the dpi setting would be 1600 dpi
A 1” x 4” negative: 5000 pixels divided by 4 = 1250 dpi
A 5” x 25” panorama: 5000 pixels divided by 25 = 200 dpi, *instead keep the 300 dpi minimum* or opt for 600 dpi.

![Image 3: Walter C. Matuszewski photograph collection, 1919, Bentley Historical Library, University of Michigan](image)

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.
3. Enhancement of the scanned object

The Bentley Library recommends that potential donors do not adjust the digitized copy with scanning software to sharpen the image or remove scratches or “red-eye”, or adjust the size of the scanned image with reductions or enlargements. The material should be copied “as is” with no alterations. Review your scanner owner’s manual for more information about this. In addition, please include all edges within the scan.

Figure 4: The 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. BL002076, Class Work in Mammalian Anatomy and Histology, Bentley Image Bank http://quod.lib.umich.edu/b/bhl/x-bl002076/bl002076
4. 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 clearly capture all edges of the original. **Do not crop edges of documents.**

![Sample scan setup](image)

Figure 5: Example set up of a preview scan. Note that the selected area for this scan (within the dotted lines) extends beyond the edges of the original, and the ruler captures the size of the original.

Capturing Original Size
To capture information about the actual size of the original you are scanning, we have created the following PDF of a printable ruler you may use to place your photographs or documents next to.

**NOTE:** Disable any “shrink to fit” options on your printer when printing. This document prints to a standard US letter size paper source (8.5” x 11”). Use of this PDF requires the freely available Adobe Reader. [Printable Ruler PDF](http://www.vendian.org/mncharity/dir3/paper_rulers/)

Color Targets
While not required, the use of color targets is welcome to give 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)](http://www.vendian.org/mncharity/dir3/paper_rulers/) to over $200 from professional photography organizations.

Scanning Flat/Loose Materials
To scan loose pages of material, such as photographs or letters, place the object on the scanner and copy the front, then the back (if it contains valuable information) keeping the scan order in
agreement with the order of the materials, such as page numbers or envelopes with corresponding letters.

Figure 6: 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. Alfred Lyttle papers, folder 1, box 1, Bentley Historical Library, University of Michigan

Scanning Bound Materials
Bound materials, such as books or diaries, should be carefully placed on the scanning glass and lightly pressed 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.
When you have adjusted your document on the glass and have created a preview scan you are satisfied with, you are ready to scan the image and save it as a digital file.

**Saving the Scan**

*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 (the *tiff header*). Refer to your scanner’s manual for more information about saving to specific file formats. If you are unsure about whether your scanning software can save to uncompressed TIFF (i.e. you only know it is saving to ‘.tif’) save files in the TIFF format and make a note in the metadata file you submit with your materials (see Metadata below for instructions).

The Bentley Library highly recommends that use of TIFF as a file format. If you are unable to save the file as a TIFF--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*, which means that certain information within the scanned image is discarded to make the file size smaller. This lost information might be invisible to you, but the 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.

**5. Directory structure and File naming**

When undertaking a scanning project, the Library recommends that you take the time to organize your digitized material.

*Creating a Folder Directory*

Create a folder for your collection of scans. Use your subject’s/creator’s last name or the University department the collection is associated with. If you are unsure or have questions about your subject or creator, contact the Bentley Library’s Digital Curation Division for more information.

*Naming and Organizing Digitized Files*

File naming and organization should follow a convention that keeps the order of the files related to each other if applicable. The best method to maintain the relationship between files is to organize the files into directories (folders). Donors are encouraged to maintain an "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. 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 post cards, 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).

Name each file with a short name to identify what dates or material it contains, bearing in mind that full descriptions may be provided in the metadata sheet we ask donors to complete. Should you choose not to use the metadata sheet, we ask that you provide as much information for the description of the file in the file name or in a text file that you can submit with your collection. See Metadata below for further instructions.

Example of a personal collection file directory:

```
C:\Documents and Settings
  | Lowe
  |   | Photographs
  |   |     | France_October_1919
  |   |     |     | photo1.tif
  |   |     |     | photo2.tif
  |   | Papers
  |   |     | 1918
  |   |     |     | Russia_Archipel_1918
  |   |     |     |     | letter1_envelope_front.tif
  |   |     |     |     | letter1_page1.tif
  |   |     |     |     | letter2_envelope_front.tif
  |   |     |     |     | letter2_page1.tif
  |   |     | France_Brest_1918
  |   |     |     | letter1_page1.tif
```

University collections:
See the Bentley Historical Library’s “Digital Recordkeeping Best Practices for Directory and File Naming” for information about file naming of university related content.

6. Metadata
Metadata, or information about a collection or file, is crucial for conveying information about your digitized material. To collect this information, it is recommended that you can use either a spreadsheet template (see below for example) or a text file.

Option 1: Sharing Metadata in a Spreadsheet (Preferred)

It is important to review and understand this document before you begin digitizing, because information is needed about your equipment, your process, and your collection materials and it is most convenient to supply as part of the process. If you have multiple collections, please duplicate this process for each collection.
**Note:** Use of this spreadsheet requires spreadsheet editing software, such as Microsoft Excel, Macintosh Numbers, or Open Office CALC. If you do not have this software, you can download a free spreadsheet editing program from Open Office. [Get Open Office CALC – Download Spreadsheet](#)

![Spreadsheet](#)

Figure 8: Screen shot of the recommended metadata spreadsheet with a sample collection and files entered, following the example file directory shown in Naming and Organizing Digitized Files above.

All fields in the metadata spreadsheet marked with an asterisk (*) are highly recommended. The first section of the document is intended to record information about your scanning equipment, any software you used, which format you saved your file as (make a note here if you are uncertain about your TIFF being uncompressed) and the general date of your digitization.

The Digitized Material section is intended to capture information about your individual scans. At a minimum, we ask that you complete the fields starting with the location of the file (Folder name), the file name, the title you choose to give it, and the creator of the original. See Creating File Descriptions below for more information about how to apply this to your collection.
Option 2: Sharing Metadata in a Text File (Alternative Minimum)

Should you choose not to use the metadata spreadsheet, you can record information in a text file. Note that file names are limited in the number of characters available, which is why we prefer 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 (a .txt file, using Notepad in Windows or Text Edit on a Macintosh). Record information about your scanning equipment, any software you used, which format you saved your file as (make a note here if you are uncertain about your TIFF being uncompressed) and the general date of your digitization. Download a template text file here.

Type any descriptive information that accompanies your 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 asked about it. Titles can provide context about what the item is (such as “William Lowe’s Diary”), demonstrate 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”). If any information is known by family members or other sources, include this in your description. See the Creating File Descriptions section below for more information to apply to your item descriptions.

```
Collection: William Lowe Papers
Make/Model of scanner: Epson Scan 10000XL
Scanning software used: Epson Scan Professional
File format used/compression level: TIFF, compression unknown
Date of digitization: 2011-05

Folder: Lowe\Photographs\France_October_1919
photo1.tif – Captain William Lowe, "June 4, 1919 in camp at Brest, France" written on back.
photo2.tif – William Lowe in uniform at barracks, 1919, identified by son Armos Lowe.

Folder: Lowe\Papers\1918\Russia_Archange_1918
letter1_envelope_front.tif – Letter from William Lowe to his wife, Mary, dated May 5, 1918
letter1_page1.tif – first page of letter to his wife
letter2_envelope_front.tif – Letter from William Lowe to his mother, Clara, dated June 3, 1918
letter_2.tif – 1 page letter to his mother, front only

Folder: Lowe\Papers\Diary
cover.tif – Cover of Lowe’s war diary
inside_cover.tif – Inside cover of diary, with Lowe’s name and signature
scan1.tif – Diary pages 1 & 2 scanned together
```
Creating File Descriptions

The Digitized Material section is intended to record information about the individual items you have scanned.

**Descriptive titles:**
These are used to inform users about what the item is and what it contains. This should be succinct and specific, because it will help users locate this unique object when searching for it using keywords, or help them determine what research value it might contain when browsing through a list of items. This is a short phrase, which can be easily searched, but does not go into the detailed description that the Description/Subject field does.

**Photographs of people**

- Identify people by their full names
- Give the location when available, being as specific as possible beginning with the city/state or country.
- Differentiate between portraits of people, which are taken by professional photographers in a studio setting, and casual photographs, which are usually taken by non-professionals and capture a larger scene—casual photographs will begin with the subject’s name or a description of the scene.

**Example of a portrait:**

<table>
<thead>
<tr>
<th>Title: Portrait of Harry Kipke, University of Michigan, Ann Arbor, Michigan</th>
</tr>
</thead>
<tbody>
<tr>
<td>This is a portrait, taken by a professional photographer in a studio. This title shows the subject’s first name, followed by his last name, followed by the location it was taken at (if known).</td>
</tr>
<tr>
<td>Any other information about the subject, such as positions held or job descriptions, should go into the more detailed Description field.</td>
</tr>
</tbody>
</table>

Figure 10: BL001300, Bentley Image Bank
http://quod.lib.umich.edu/b/bhl/x-blo01300/bl001300
Photographs of buildings, city/town views or general scenes

- Identify buildings with their names (if known) and give the location by address, city, state etc.
- Identify views by building, street, city or general location
- Append locations by any specific information the scene depicts, such as directional views, notable natural features, or specific events

<table>
<thead>
<tr>
<th>Title: Main Street looking south from 5th Street to 4th Street, Pentwater, Michigan</th>
</tr>
</thead>
<tbody>
<tr>
<td>This is a street view of Main Street, looking to the south, with location details given and the city and state of the view.</td>
</tr>
<tr>
<td>Other information about the view, such as “at sunrise” or “during a parade” that makes it more unique could be inserted before the town name.</td>
</tr>
</tbody>
</table>

For more ideas on how to create descriptive file names, see the Minnesota Digital Library’s helpful handout on creating descriptive titles for their Minnesota Reflections project.

Creator:
The Creator field is used to identify the creator of the original item you have scanned. There will only be one creator, which will either be an individual (the author or artist/photographer of your original item) or it can be an organization or corporate name for a company-produced publication.

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. This would include book editors or publishers and photography studios. If this information is known, enter it by name (last name, first name or company name) of the contributor followed by their location (city, state). If this is unknown, leave this field blank.

Location:
Note the location of the scene that has been captured in photographs or the location where letters were written or sent from. This will be the same location name you used in your descriptive title. Leave this field blank if you do not know what the location was.

Description/Subject:
The description is a longer, more detailed, narrative description of the item. It will incorporate the keywords used in the Title, but give more depth of information about the people, places,
events/activities or themes that can be seen in a photograph or that you may known about the object you have scanned.

The Description field will be used for keyword searches users may run to find items that have research value. This field provides the best area to provide context for the event, activity or persons in the photograph. For photographs, any information that has been written on the picture can be included here, such as “Verso:” (meaning ‘on the back’).

A. Example of a photograph scene:

| Title: Huron River Tug O' War, Ann Arbor, Michigan |
| Description: University of Michigan Spring Games, freshman and sophomores compete in a tug of war in the Huron River. Verso: E.T.H. Participated (Class of '29) |
| Date of Creation: ca. 1925 |

The Description provides more information about the scene this photo depicts.

The Date of Creation is an estimate, based on the phrase “Class of ‘29” found on the back of this photograph.

Figure 12: BL003706, Bentley Image Bank  
http://quod.lib.umich.edu/b/bhl/x-bi003706/bl003706
B. Examples of textual documents:

Title: Two Women Students In ‘86 Were Members of First Sanskrit Class At Michigan

Description: Michigan Alumnus article about University of Michigan alumni Sarah Satterthwaite Leslie and Chloe Daniels, who were members of the first Sanskrit class offered at the University.

Creator: Michigan Alumnus

Date of creation: 1941

Copyright: unknown [blank]

The Description gives more detail of what the article is about and names who it concerns, in this case the title was supplied by the editor.

The Creator in this case would be the author of the article, or if none is given, the publisher. If the author had been supplied, the publisher would become the Contributor.

---

Title: Envelope of letter from Frank Lauer to his sister, March 1919

Description: Envelope of a letter Frank Lauer wrote to his sister, Anna M. Lauer in March 1919 while serving with the 339th Infantry stationed in Archangel, Russia.

Date of creation: 1919/03

Creator: Frank Lauer

The Description gives information about this scanned envelope. The subsequent letter pages could have their own descriptions that discuss the contents of the individual letter.
Date of creation:
This is the date of creation of the original item you are scanning. 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 (1919/05/21 is May 21, 1919). If you are unsure what the year was but have some evidence that it may have occurred during a certain time, you may write “ca. 1919” or “ca. 1915-1920” to record this. If you are completely unsure of what the date is, leave this field blank.

Physical format:
This field records the physical format of the original item you are digitizing. Common format types include the following terms:

- Negative
- Painting
- Photograph
- Poster
- Diary
- Album
- Postcard
- Map
- Letter
- Published book
- Pamphlet
- Magazine
- Calendar
- Scrapbook
- Stereograph
- Sketch
- Architectural drawing
- Envelope

The size of the original object can also be included, such as “3x5 postcard.” If you are unsure of how to describe your item’s physical format, contact Digital Curation Services for assistance.

Copyright:
This field is for recording any known copyright information about the original item you are digitizing. If the copyright status is unknown, you may leave this field blank.

Questions
Should you have questions or issues regarding these guidelines, please contact Digital Curation Division (Nancy Deromedi, Michael Shallcross) at the Bentley Historical Library, 1150 Beal Avenue Ann Arbor, MI 48109-2113 U.S.A. or contact us by phone at 734-764-3482.
Works Cited:


State Library of North Carolina (http://digitalpreservation.ncdcr.gov/)
Tutorial on changing file names: http://www.youtube.com/watch?v=8ogTKWpq6is&feature=youtu.be

2007  University of Michigan University Library Digital Library Production Services Digital Conversion Unit for the University of Michigan Libraries University Digitization Specifications www.hathitrust.org/documents/UMDigitizationSpecs20100827.pdf

Tutorials: http://www.mndigital.org/training/tutorials/