Bentley Historical Library

Digital Reformatting Procedures Report
Summer 2011

Matthew L. Adair

Institute of Museum and Library Services
Digital Preservation Intern

Table of Contents

Overview

During the summer of 2011, the Bentley Historical Library participated in the University of Michigan's School of Information program titled *Engaging Communities - Fostering Internships for Preservation and Digital Curation*.[1] Funded by Institute of Museum and Library Services, a student intern was selected to examine how well the Bentley's existing digital reformatting practices meet digital preservation best practices as articulated in the *Technical Guidelines for Digitizing Cultural Heritage Materials* developed by the Federal Agencies Digitization Initiative (FADGI).

Scope

The purpose of this project is to examine the Bentley Historical Library's in-house digital reformatting procedures with the goal of implementing as many of the specifications recommended by the Federal Agencies Digitization Guidelines as practically possible. The included recommendations are for the creation of archival master images – the creation of derivative images is not covered.

It should be noted that the view taken during this project is that the resulting digital images are for more than simply providing another means of access. The creation of archival master images need to be capable of accurately representing the original object, and, if necessary, replacing it, while at the same time balancing the needs of preservation with those of time and financial resources. Due to time limitations associated with the project, the recommendations will focus primarily on the Bentley Historical Library's Postcard Collection, though some recommendations will be very general in nature.

Summary of Findings and Recommendations

- The FADGI *Guidelines* are a work in progress and will continue to evolve.
- Best practices have moved away from "one-size-fits-all" approaches such as blanket "600dpi, 24-bit color" statements
    - Scanning parameters need to be set according to project and format specific goals, and a variety of factors can affect these decisions
    - Knowledge of image science is becoming increasingly important in order to make informed decisions
    - Using specific tools, certain image quality parameters can be measured to determine how well parameters and workflow are
- It is acceptable to make certain informed modifications to digital images.
- The software which ships with scanners does not necessarily provide the best image possible

[1] *Engaging Communities - Fostering Internships for Preservation and Digital Curation:* http://preservation.cms.si.umich.edu/

General Recommendations

- Whenever practical, use high quality scan targets during the scanning process
- Fit the scanning parameters to the needs of the project
- Do not rely on the scanning software for image processing – better results are obtained by post processing using Adobe Photoshop
- In order to ensure long-term preservation, more technical metadata needs to be captured
- Structure metadata using the METS as an outline – more than one image with one object
- If plans are to expand digitization programs, then implement a quality control program
- There is a need to improve the general cleanliness of scanners and work areas:
    - Clean scanners before scanning
    - Watch clutter
    - Consider purchasing new document mats
- Provide better lighting in areas surrounding scanner workstations so the that the light quality better matches "natural" light to assist with making visual adjustments to scans
- Consider some sort of network attached storage to assist with workflow and in house storage and retrieval
- Continue to use the TIFF file format, but be aware that JPEG2000 is increasingly becoming popular as a master file format

General Scanning Steps

Detailed processing instructions are shown in Appendix C – Detailed Digitization Steps. A general workflow for scanning would be:

Scan group of images ⇨ Process in Photoshop ⇨ Enter metadata into database

Scanning Steps

- Clean scanner as necessary
- Verify general scanner software settings, such as file naming and location
- Turn off all scanner color correction options
- Set the desired resolution
- Set bit-dept depth – 48 bit color or 16 bit gray scale preferred
- Scan images

Post Processing Steps

- Convert to Adobe RGB color space for color images
- Correct color and tone using preset curves to match target values
- Apply sharpening to Luminosity
- Convert to 8-bits per channel – either 24-bit RGB or 8-bit grayscale.
- Save the file

History

Digitization efforts at the Bentley began about 1999 with the concept of placing a few hundred images online which represented the "Best of the Bentley." A number of collections were selected, metadata entered into a FileMaker database, and the images and metadata were delivered to the University of Michigan's Digital Library Production Service (DLPS). DLPS scanned the selected images and ingested them into the University's system for online display and access. Since then, additional images have been added as the result of various projects. There are currently a little over 5,500 images in the Bentley Image Bank, hosted by DLPS. In addition to scanning images for use in the Image Bank, scans are created on an as needed basis to fulfill patron reproduction requests.

FADGI Technical Guidelines Overview

The *Technical Guidelines for Digitizing Cultural Heritage Materials* was published by the Federal Agencies Digitization Initiative in August of 2010. The guidelines are based heavily upon the National Archives and Records Administration's *Technical Guidelines for Digitizing Archival Records for Electronic Access: Creation of Production Master Files – Raster Images* but have been updated to reflect the most recent changes in the field. They reflect the increasing recognition that digital reformatting is more than just "color vs. grayscale" and "300 dpi vs 600 dpi." Carl Fleischhauer and Michael Stelmach of the Library of Congress, and participants in the Still Image Working Group, noted this in a 2009 presentation:

> "This has brought us to consider what are the relevant quality factors. We need to move away from things like color vs. monochrome, pixel density, and bit depth, all of which are usually referenced (without saying so) in terms of outputs. Instead we want to be attentive to more appropriate ways to specify tonality, spatial resolution, color, uniformity, and noise."[2]

Burns and Williams note that all too often, "digitizing requirements are over-specified. Except for the most demanding spatial detail, true 600 dpi scanning is rarely required for most reflection work. Yet this requirement is often rubber-stamped into digitizing requirements simply because it is a safe, albeit expensive, and incomplete choice." There is a tendency to think of imaging performance in terms of boundaries, or, more simply, "more is better." For example, if 300 dpi is good, then 600 dpi is better. In fact, just the opposite can be true. In the case of some black and white negatives, digitizing at too high a resolution can actually introduce noise patterns that will detract from the final image. Continuing on, Burns and Williams state:

[2] Fleischhauer and Stelmach, "Federal Digitization: Moving to Common Guidelines," 17.

> This "more is better" thinking can actually detract from high fidelity imaging of certain objects. In faithfully digitizing 19th century photographs the real challenge is to sufficiently capture the subtle and finely incremented tones and near neutral colors. Wide color-gamuts and large dynamic ranges may actually detract from such a goal. It is logical, in fact, to desire a minimized color gamut so that the available digital count levels may be efficiently assigned just to those limited colors in the collection content. An insufficient color gamut in this example is not the problem. [3]

"The *Guidelines* are intended to be informative, not prescriptive."[4] It is also important to realize that the authors of the *Guidelines* state they do not currently cover all aspects of digital imaging parameters and therefore the *Guidelines* will continue to evolve. Much of the information presented is highly technical in nature and to fully understand and implement it requires a strong background in imaging science and technology. This level of expertise is likely outside the range of many institutions. There is, however, a lot of important information which can be used to formulate steps to improve an institution's digital reformatting procedures. What follows is a summary of each of the major sections of the *Guidelines.*

The FADGI Technical Guidelines cover the following primary topics:

- Technical Overview
- Imaging Workflow
- Digitization Specifications for Record Types
- File Format Comparison
- Metadata
- Storage Recommendations
- Quality Management

A majority of the *Guidelines'* content focus on the Technical Overview, Imaging Workflow, and Digitization Specifications for Record Types.

Technical Overview

This section begins with a description of the basic characteristics of a raster image. The three primary characteristics they define are the Spatial Resolution, Signal Resolution, and Color Mode. Spatial resolution is commonly termed in dots per inch (DPI), but should be more properly described as pixels per inch (PPI).[5] It is "the ability of an imaging component or system to distinguish finely spaced detail. Specifically, the

[3] Williams and Burns, "Preparing for the Image Literate Decade," 126.

[4] Federal Agencies Digitization Initiative (FADGI) - Still Image Working Group, "Technical Guidelines for Digitizing Cultural Heritage Materials: Creation of Raster Image Master Files," 2.

[5] Pixels per centimeter and pixels per millimeter are valid measurements and in use.

ability to maintain the relative contrast of finely spaced detail."[6] Signal resolution, most often called bit-depth, defines the maximum number of shades and/or colors in an image. This is commonly discussed in terms like "8-bit grayscale" or "24-bit color." At its barest, the Color Mode is an abstract mathematical model of color representation.[7] Digital systems encode color into varying numerical values and the Color Mode refers to the system used to encode these values.

Following image characteristics is a section with recommendations regarding the environment in which the actual scanning takes place. These recommendations are based upon the fact that how the human eye perceives an object can vary greatly depending on lighting conditions and equipment. The *Guidelines* recommend creating a neutral environment to minimize these variations. Another factor that affects whether an image is represented properly is the quality of the monitor and whether it has been calibrated to display colors properly. Their recommendations are based on ISO standards and basically state that:

- High quality, graphic-arts monitors
- Rooms painted a neutral gray
- Low levels of illumination
- Lighting color temperature at around 5000K

The next section, Quantifying Scanner/Digital Camera Performance, is probably the most complex and difficult to understand section of the Guidelines. It is also one of the most important in that it reflects the current standards for measuring image / device performance beyond characteristics such as DPI and Bit-depth.

What is important to recognize is that some aspects of image quality can be objectively measured. While the examples provided in the *Guidelines* use a software package from Image Science Associates called Golden Thread, the measurements are based upon a series of ISO standards. In general terms, the metrics which can be measured are:

- Tone, exposure
- White balance, neutrality
- Color encoding accuracy
- Sampling rate
- Resolution
- Sharpening
- Noise
- Aliasing

[6] Conway, "SI 675 2011 Week 4 Encoding and Measurement.pptx."
[7] Ibid.

For a number these metrics, the *Guidelines* have developed a four-tiered performance rating system. For the example below, if the white balance is within ±3 points of the desired aim point of 0, then it would receive a 4 star rating.

White Balance Error		
PERFORMANCE LEVEL	AIM	TOLERANCE - 8-bit equivalent (applies to G-B, G-R difference for all neutral density levels)
****	0	± 3 count levels
***	0	± 4 count levels
**	0	± 6 count levels
*	0	± 8 count levels

Table 1

These ratings can be combined as needed depending on end use and imaging objectives. For a particular project, it might be acceptable to have Sharpening at a 2 or 3 star level, but require Tone at a 4 star level. The tighter the tolerance level, the better the performance, and generally, the higher the cost. The Categories and Objectives sub-group of the FADGI Still Image Working Group is developing a matrix that will map to the very type of multi-tier model used in the Quantitative Performance Guidelines. The work of this sub- group can be viewed here - http://www.digitizationguidelines.gov/guidelines/ccdo-subcat.html

Measuring performance is also important to determine whether scanning equipment is performing as expected, or even capable of delivering the required results. The *Guidelines* note that scanner performance can change over time, and even with software updates. Therefore, it is important to frequently measure the performance of the equipment.

The final portion of the Technical Overview discusses the use of reference targets during the scanning process. The *Guidelines* recommend "including reference targets in each image of originals being scanned, including, at a minimum, a photographic gray scale as a tone and color reference and an accurate dimensional scale."[8]

There are several different kinds of targets that serve various purposes. What the guidelines describe as Object-Level and Device-Level targets are used to provide the data for the performance levels described above.

> The Device target [Figure 1] is designed for benchmarking of imaging devices as well as to validate requirements for a project or class of content being digitized. This target is imaged on a regular basis for cameras and scanners, but the target is not imaged along with the actual content.

[8] Federal Agencies Digitization Initiative (FADGI) - Still Image Working Group, "Technical Guidelines for Digitizing Cultural Heritage Materials: Creation of Raster Image Master Files," 34.

The Image target [Figure 2] is designed to be imaged along side (top, bottom, or either side) the object being digitized. The inclusion of the image target with the item **results is a self-describing image file** with regard to image performance [emphasis added].[9]

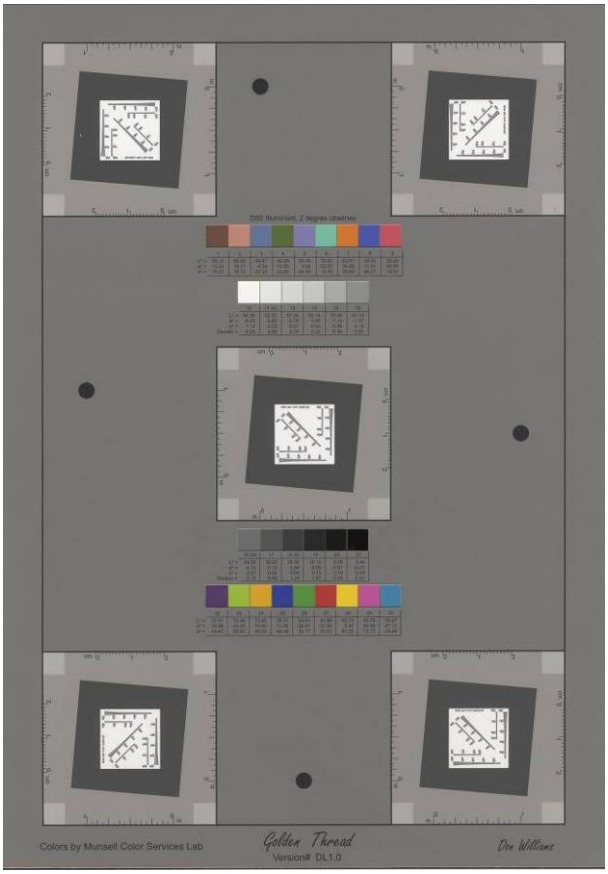


Figure 1 - Device Level Target

[9] Ibid., 35.

Figure 2 - Object Level Target

Besides the Object and Device-Level targets, the *Guidelines* also discuss the use of dimensional scales and targets for tone and color reproduction. Targets used for adjusting tone and color are generally composed of a series of gray scale patches in known density increments. One of the most popular of these is the Kodak Q-13 (now made by Tiffen - Figure 3). Targets are also available for use during transmission scanning of positive or negative transparencies.

Figure 3 - Q-13 Gray Scale Target

Imaging Workflow

The next major section of the Technical Guidelines document covers the imaging workflow, including post scan adjustment of image files, scanning aimpoints, color management, and other post processing techniques.

Contrary to popular belief, it is entirely acceptable, and often required to perform post scan adjustments on images. The fact of the matter is that unless you are using "raw"

images from a capture device, there is always a certain amount of image processing applied in order to produce images of good quality. The only time that the *Guidelines* "recommend saving unadjusted files is if they meet the exact tone and color reproduction, sharpness, and other image quality parameters" required (which is rare).[10]

One of the more important adjustments is to use aimpoints within the image to adjust for neutral color and tone. This is where the use of a gray scale target is of prime importance. Using the gray scale target, three points are located which correspond to the white point, black point, and mid point. The color values are measured and the adjusted so they fall within a specified range of values. Figure 4 shows an architectural drawing captured using a digital camera. Figure 5 shows the same image, this time corrected using the aimpoints procedure detail in the *Guidelines*.

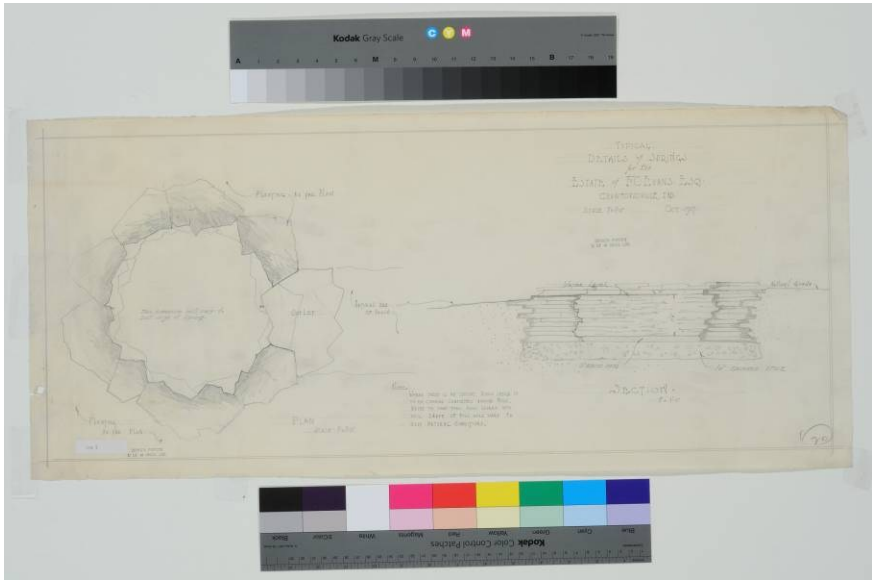


Figure 4
Jens Jensen drawing, Bentley House Scan 7528, uncorrected

[10] Ibid., 39.

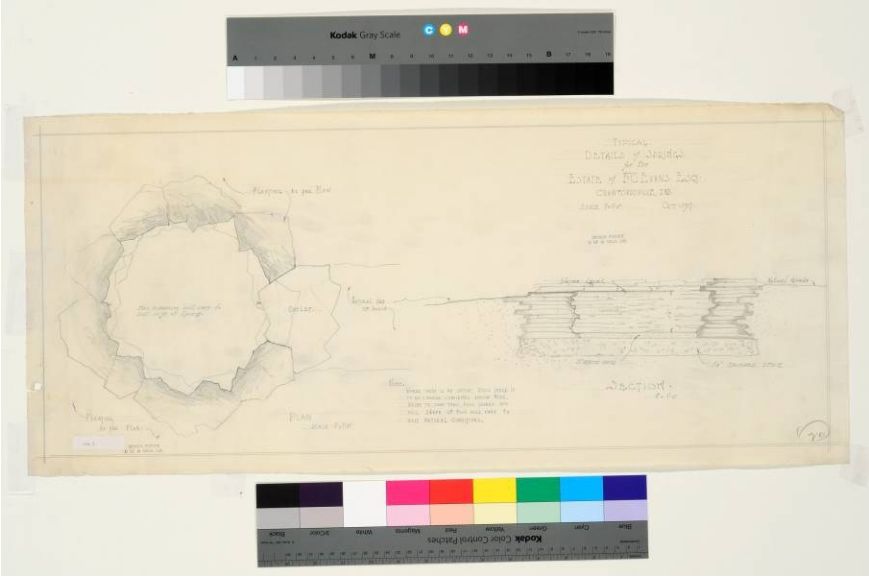


Figure 5
Jens Jensen drawing, Bentley House Scan 7528, adjusted for aimpoints

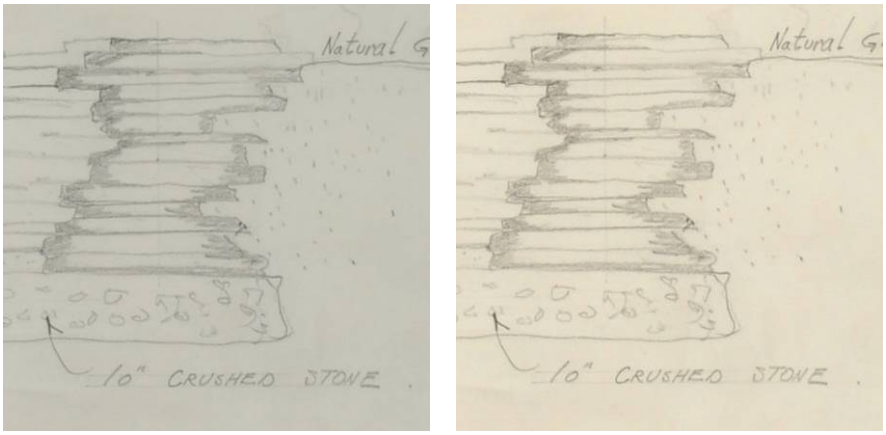


Figure 6 - Detail of the uncorrected (left), and corrected (right) image

Following the discussion on aimpoints, the *Guidelines* move into an overview of color management. Color Management helps to ensure that colors are accurately rendered across input and output devices. Without proper color management, what appears as "red" on one device may look more "maroon" on another.

The next portion of the Imaging Workflow section provides recommendation regarding post capture image processing, specifically regarding color correction, tonal adjustment, and sharpening. Again the *Guidelines* note that it is generally necessary to perform post processing to produce the best rendition possible of the digital file. Any post processing steps, however, should not be used as a substitute for poor capture techniques in the first place. This section on image processing does not detail steps, but

broad recommendations. It is up to the reader to search out the specific techniques using other resources.

Digitization Specifications for Record Types

This section of the FADGI document provides broad guidelines regarding digitization parameters for specific types of records. Records are broken down in into several broad categories as follows:

- Textual Documents, Graphic Illustrations/Artwork, Maps, Plans, and Oversized
- Photographs - Film / Camera Originals - Black-and-White and Color - Transmission Scanning
- Photographs - Prints - Black-and-White, Monochrome, and Color - Reflection Scanning
- Aerial - Transmission Scanning
- Aerial - Reflection Scanning
- Objects and Artifacts

Again, it is important to note that the recommendations have moved away from a one-size-fits-all method of capturing images. Each category has specific recommendations not only based upon the category, but on the important characteristics of the original that need to be captured, which may vary from project to project or even item to item. The Digitization Specifications for Record Types section also provides guidance when dealing with items such as encapsulated documents and copy negatives.

File Format Comparison

This portion of the *Guidelines* provides a brief comparison of file formats common in digital preservation. It also has discussion some of the common factors to take into consideration when choosing a file format to ensure the longevity of captured images.

Metadata

> Although there are many technical parameters discussed in these *Guidelines* that define a high-quality master image file, we do not consider an image to be of high quality unless metadata is associated with the file. Metadata makes possible several key functions – the identification, management, access, use, and preservation of a digital resource – and is therefore directly associated with most of the steps in a digital imaging project workflow: file naming, capture, processing, quality control, production tracking, search and retrieval design, storage, and long-term management. Although it can be costly and time-consuming to produce, metadata adds value to master

image files: images without sufficient metadata are at greater risk of being lost.[11]

The Guidelines note that no one metadata standard may be appropriate for all projects or collections. Therefore, it is becoming increasingly popular to create "application profiles" for projects which incorporate the necessary elements from a variety of standards. Included is a description of common metadata types and their uses, including Descriptive, Administrative, Rights, Technical, Structural, Behavior, Preservation, and Tracking. Also given consideration are file naming conventions, and the management of metadata.

Storage Recommendations

The *Guidelines'* discussion on storage recommendations is quite brief. They do not recommend the use of optical media for long term storage, but, at a minimum, a redundant hard drive system. Their ideal recommendation is a managed digital repository.

> "Digitization of archival records and creation of metadata represent a significant investment in terms of time and money. Is it important to realize the protection of these investments will require the active management of both the image files and the associated metadata."[12]

Quality Management

The last major section of the *Guidelines* document deals with quality control and quality assurance of images parameters and metadata. Concerns such as file format consistency, image quality, existence of appropriate metadata, and whether the content of that metadata matches industry or institutional standards are addressed. While the goal is 100% compliance, they recognize the impracticality of this, especially for large projects. Instead, they recommend a random sample with a 1% cutoff level for pass / fail.

[11] Ibid., 70.
[12] Ibid., 82.

Current Procedures

Creation of Master Images

There are currently no formal procedures or guidelines in place for the digital reformatting procedures at the Bentley Historical Library. There exists a short, one page document dated November 2000 titled "Bentley Library Image Bank, Creating Master Images and Derivatives: Guidelines for Digitizing Photographs In-House." It is not entirely clear whether these guidelines are actually followed. A quick summary of recommendations in the document are:

- Do not digitally retouch or make changes to the master image file.
- Tonal scale and color balance adjustments should be controlled with the scanner software rather than to image after scanning.
- Slightly sharpen image files as needed with unsharp mask algorithm to achieve approximate appearance of original.
- Scan at 8 bits grayscale or 24 bits color, RGB mode
- 3000 to 5000 pixels on long side
- Uncompress, TIFF file format

In practice, most in house digitization varies somewhat from these steps. For reflective materials, most items are scanned at resolutions ranging from 300 dpi to 1200 dpi, with 600 dpi being the norm. The same holds true for larger sized transparent media, such as medium format or glass plate negatives. Transparent media such as 35mm slides or negatives are typically scanned in the range of 3200 dpi. Images are created in either 8-bit gray scale or 24-bit color. Most color and tonal corrections are typically handled using the scanner's automatic color correction modes. Any other color and tone corrections are typically made to a "what looks right" level of specification. Files are saved in uncompressed, TIFF format.

Metadata

Metadata for images is managed and stored in a FileMaker Pro database, in version 5 format. One metadata record is created for each individual image. As is quite common, all metadata is manually entered for each record, though the capability of FileMaker to create a new duplicate record speeds the process up. The information captured seems have been developed on an "as needed" basis with no formal metadata specification used.

Information captured includes mostly descriptive plus some technical / administrative metadata:[13]

Current Metadata		
Descriptive Metadata		**Technical**
Filename	Box No.	Scansource
Negative	Folder	Scanner
CD No	Image creator	Resolution
Image Title	Work Creator	Scantype
Title source	Cataloger	Scan Date
Caption	Catalog Date	File Type
Image date	Edit Date	
Date certainty	BHL Copyright statement	
Notes	Subjects	
Collection creator	Personal Names	
Collection title	Corporate Names	
Mirlyn	Geographic Names	
Daohref	Genres	
Eadlink	Project	

Table 2 - Current Metadata

Storage

Items scanned in house are stored on external media and drives. The general procedure is that a scan is stored first on the local machine. On a periodic basis, these scans are moved to an external hard disk. When the hard disk approaches capacity, the older scans are transferred off onto optical media. Files included in the image bank are delivered to DLPS, where they are stored on a system that creates derivatives on and as needed basis. It is not clear whether these files would be accessible in the event of any sort of local data loss.

Recommended Procedures

Description of the Postcard Collection

The postcard collection consists of 10.5 linear feet (12 boxes) and 1 oversize folder. It dates from the 1890's onward and has been assembled from a variety of different sources. While there some rare examples, a majority of the postcards have been commercially or mass produced. This collection was chosen for a test bed due to the

[13] These fields are based upon the blimage database used for the Image Bank. Not all fields from the database are included here – many of which are calculated based upon values in the listed fields. Additionally, these fields have been renamed to make them slightly more human readable.

wide variety of formats available in the collection, and the desirability of providing online access to the collection through the Image Bank.

As noted, the collection consists of a variety of formats. Most are two sided, the front consisting of a black and white or color halftone image, the reverse containing the space for address and message. They typically range in size from about 3 ½" x 5" to 4"x6". Additionally, the collection includes a number of true photographic postcards, large format postcards up to about 8"x10", panoramic postcards, and postcard folders consisting of multiple postcards printed on long, folded stock with an integrated envelope (see Figure 7 below). One box (Box 5), that is not quite full, contains 600 post cards. Of these, 148 (about 25%) were real photo postcards or folded postcards. Each box likely contains 600-700 postcards. Taking an average of 650, there are about 7150 postcards in the eleven small boxes. In addition, there is one oversize box and one oversize folder

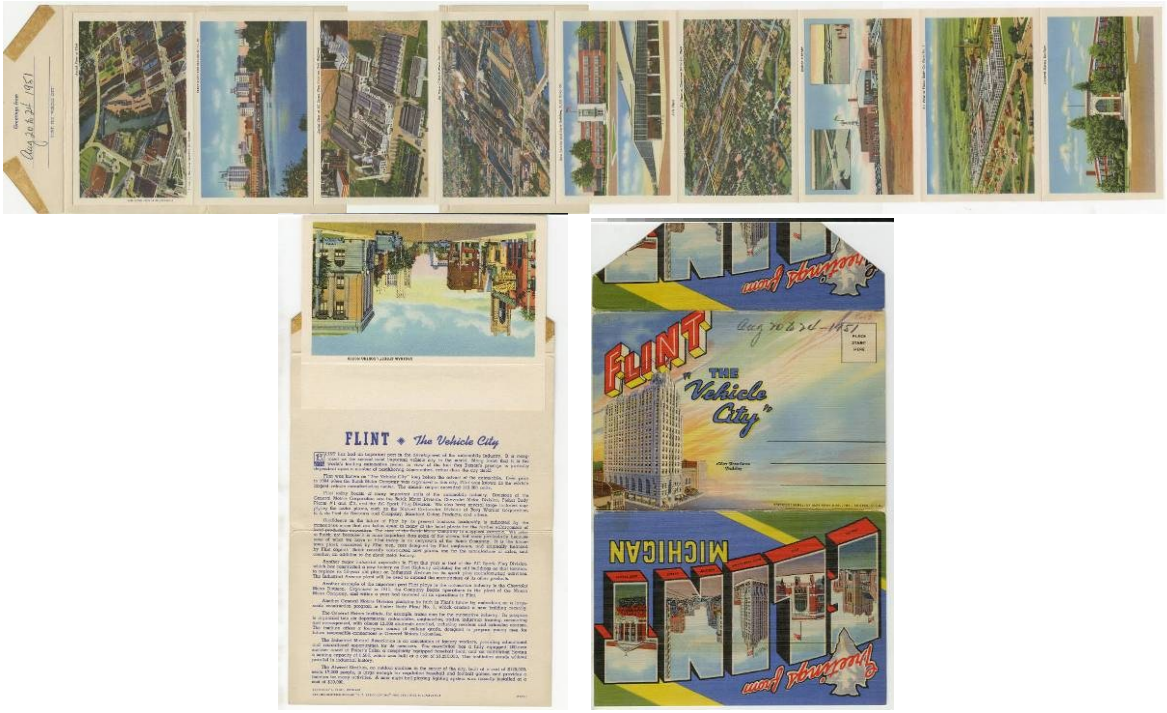


Figure 7 - Example of a folded postcard envelope

Scanning Specifications

Adjust Image Parameters based on record type

As previously described, there is no longer accepted practice to perform a "one size fits all" method of scanning. Numerous factors should be taken into account when making methodology decisions. Paul Conway describes some of these issues in his Imaging Product Model:

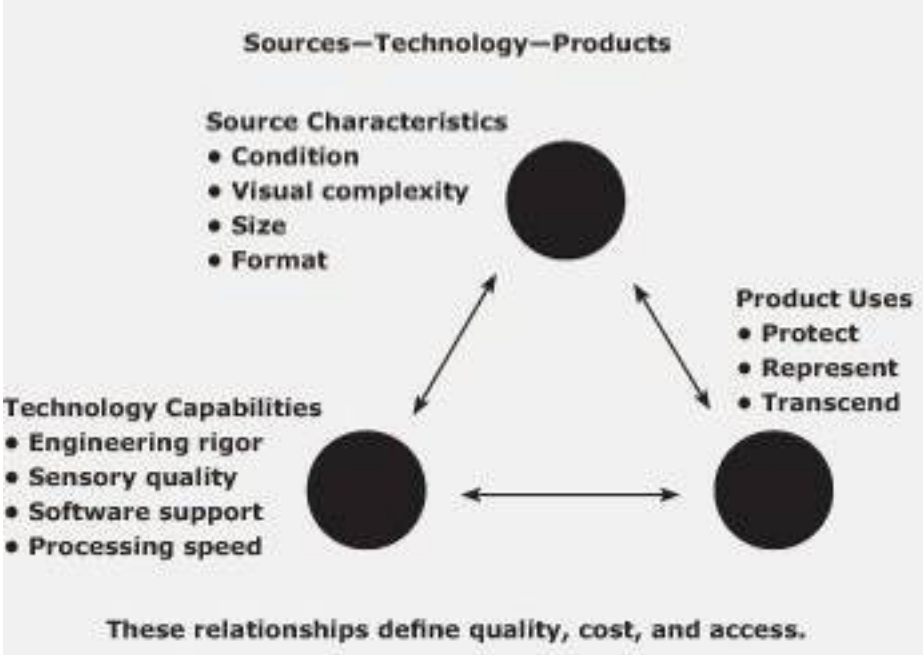


Figure 8 - Conway's Imaging Product Model

"The digital imaging process results in a product with its own characteristics that are distinct from the characteristics of the original sources. The biggest challenge in building an image product is to balance three issues: the characteristics of the source; the capabilities of the technologies of digital conversion; and the purposes or expected uses of the end product."[14]

An imaging project can be approached from any one of the points illustrated in Figure 8. A project may also be limited by the any one of the points as well. The available technology, for example, may not be capable of capturing the important characteristics of the source object. These issues help frame the project's goals and outcomes. It may be a small scale project (occasionally called a "boutique project") where color /tone or the capture of minute detail is of extreme importance, as is often the case with 19th century glass plate negatives and important works of art. Or, it could be a large scale project where it is simply the capture of the information on the page, as in the case of many book digitization projects. Projects may also fall within many points in between.

With some slight variation, the post card collection consists of a fairly homogeneous set of materials in a good state of preservation. The postcards generally fall within the *Guidelines* category of "Textual Documents, Graphic Illustrations/Artwork…" As noted, some of the postcards would also be categorized under the Photographic / Prints category of the *Guidelines*. While they do not have a high intrinsic or artifactual value, there is the potential for great interest in the part of the end user.

[14] Conway, "Digital Technology Made Simpler."

Considering the postcard collection, it can be concluded that the characteristics that are important to capture are:

- Information on the postcards – ie, the picture, captions, notes. The structure of the postcard, such as paper grain, the makeup of the halftone dots, is not critical
- Color – in this case color is important, but not nearly as critical as a rare piece of artwork by a famous artist
- Structure – as in the case of the folded postcard envelope

Textual Documents, Graphic Illustrations/Artwork, Maps, Plans, and Oversized

Document Character - Original	Recommended Image Parameters	Alternative Minimum
Documents as described for grayscale scanning and/or where color is important to the interpretation of the information or content, or desire to produce the most accurate representation	24-bit color mode - adjust scan resolution to produce a QI of 8 for smallest significant character or 24-bit RGB mode - 400 ppi for documents with smallest significant character of 1.0 mm or larger **NOTE:** Regardless of approach used, adjust scan resolution to produce a minimum pixel measurement across the long dimension of 4,000 lines for 24-bit files	24-bit RGB mode - 300 ppi for documents with smallest significant character of 1.5 mm or larger

Table 3 - Extracted from Textual Documents, Graphic Illustrations/Artwork, Maps, Plans, and Oversized category, page 58 of the Guidelines

If the recommended image parameters shown in table three are strictly followed (4000 ppi along the long dimension), then most postcards would be scanned in the range from 700-800 ppi. A test scan performed at 800 ppi took a full two minutes to complete. Calculating this out from flat figure of 7000 postcards, and not including metadata entry and variations due to structure (folded postcards for example), it would take over 460 hours to scan both sides of all of the items. A resolution of 400ppi took approximately 30 seconds, or about 120 hours for the entire collection.

Recent work in imaging describes methods for adapting scanning specifications to the materials being scanned. In his paper "300ppi? How to Determine your Digitization Project Needs," Matt Pearson of Stanford outlines a method for determining a minimum resolution based upon size of the smallest feature to be captured.[15] This technique involves using a magnifying device with a precision scale to find the size of the smallest line on the image and calculate the appropriate resolution from there. It was not possible to perform these sorts of calculations since the necessary equipment was not readily available. However, by scanning and visually inspecting the results, it is possible to obtain similar findings.

[15] "300ppi? How to Determine your Digitization Project Needs – matpearson.com."

The postcard collection was examined to find a representative sample that contained the smallest amount of information that was desired to be captured. This postcard was then scanned at 100, 200, 300, 400, and 600 ppi and the results visually inspected.

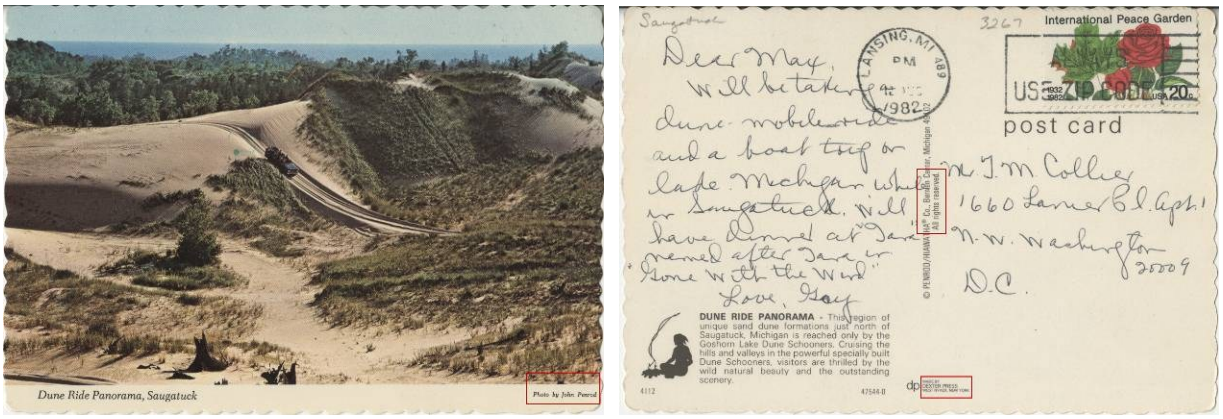
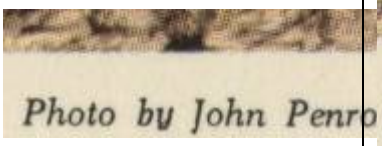
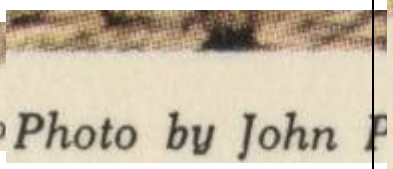
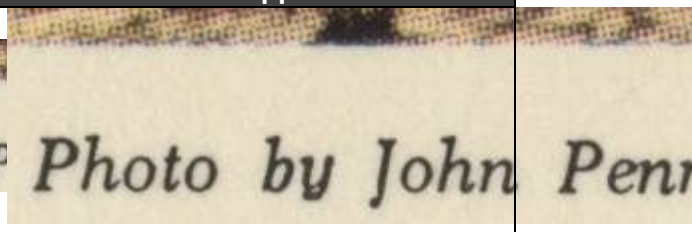
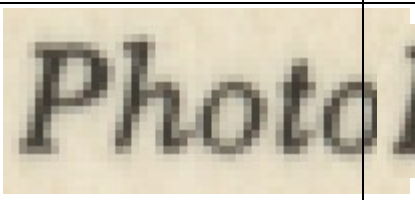
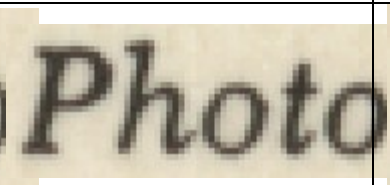
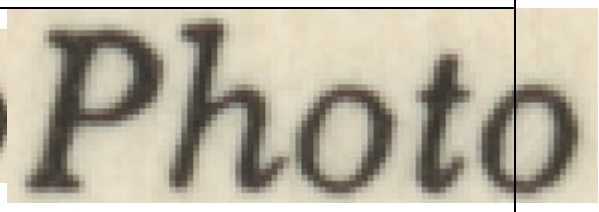
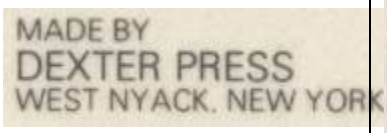
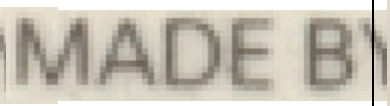
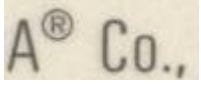
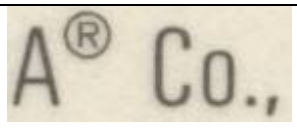


Figure 9 - Postcard showing areas sampled for resolution check

300 ppi	400 ppi	600 ppi
 100%	 100%	 100%
 400%	 300%	 300%
 100%	 100%	 100%
 400%	 300%	 300%
 100%	 100%	 100%

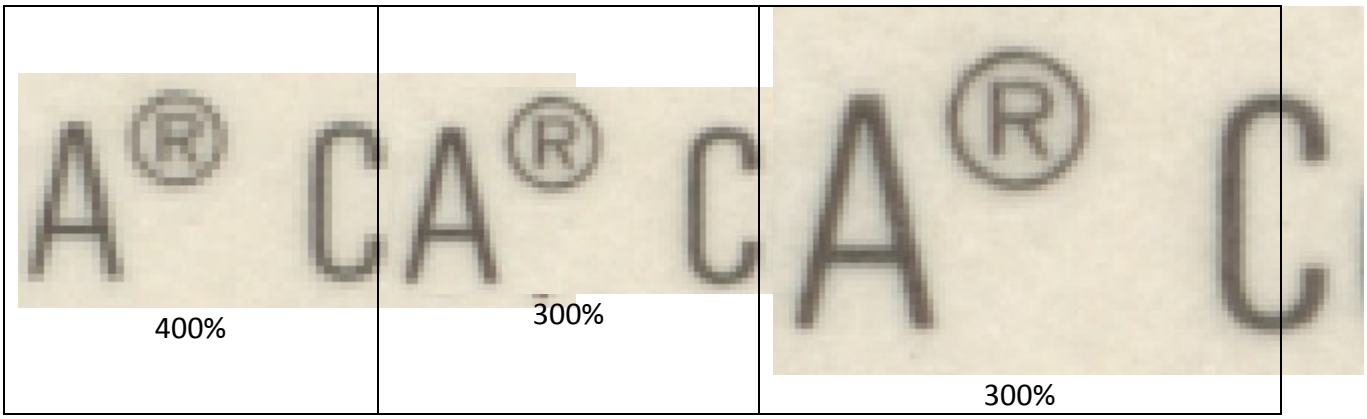


Table 4 - Details of postcard at various scan resolutions

Table 4 shows the results of scanning at 300, 400, and 600 ppi. The 100 and 200 ppi scans are not shown as it was readily apparent they did not adequately capture enough detail. When examining the 300ppi scan at 100%, the finer details are readable, but at magnification, it is noticeable that it does not capture the fine detail. Pearson notes that a practical sampling frequency will capture the finest detail in approximately four pixels. Looking at the details of the registered trademark symbol (®), the 400 ppi scan shows 2-3 pixels of capture detail, and the 600 ppi scan shows 4-5 pixels of detail.

Given the above characteristics, the size of the collection and that most of the collection is composed of halftone prints, scanning at a high resolution is impractical and, in most instances, is capturing more information than is necessary. Taking this into account, the postcard collection is a good candidate for creating batch process that can work at a relatively high volume. Balancing the needs of quality and speed of production, it is recommended that a sampling frequency of 400 ppi be used for scanning the collection.

Image Processing and Workflow

Paul Conway, who teaches the Digitization for Preservation course at the University of Michigan's School of Information, kindly donated the use of the Golden Thread scanning targets and software to allow for the comparison of various scanner settings. The results of the Golden Thread analysis are given in Appendix D – Scanner Analysis Workflow Results.

A baseline device target was scanned using the Epson 1680 scanner using no color correction or sharpening, essentially delivering as close to a "raw" output as possible. This was then analyzed using the Golden Thread software. The same process was performed using a device target scanned with the current workflow settings. Finally, the "raw" baseline device image was processed in Adobe Photoshop to produce an image that was sharpened and color corrected according to values on the targets.

The results show that post-processing using Photoshop produces a superior result compared to that of the "raw" scan or the adjustments made by the scanning software. In fact, the images produced by the scanning software produced the worst overall performance. As a result, any color correction, sharpening, and other image adjustments should be performed on "unprocessed" images using Adobe Photoshop

Detailed processing instructions are shown in Appendix C – Detailed Digitization Steps.

Scanning Steps

- Clean scanner if necessary
- Verify general scanner software settings, such as file naming and location
- Turn off all scanner color correction options
- Set the desired resolution
- Set bit-dept depth – 48 bit color and 16 bit gray scale preferred
- Scan images

Photoshop Post Processing Steps

These steps can be saved in a series of actions which can then be used in a batch process.

- Convert to Adobe RGB color space for color images
- Correct color and tone using preset curves to match target values
- Apply sharpening to Luminosity
- Convert to 8-bits per channel – either 24-bit RGB or 8-bit grayscale.
- Save the file

Reference Targets and Scanner Performance

There is a fair amount of debate on the use of reference targets during the scanning process. The degree's of thought varies from never include them, to sometimes including them, to always include them. The *Guidelines* recommend that, at a minimum, a grayscale and dimensional target be included in every scan. As was demonstrated in Figures 4-6, the use of reference targets is particularly important for adjusting an image's tone and color balance.

Reference targets, such as those shown in Figure 1 and Figure 2, can also be used with software for determining scanner performance. Vendors and institutions which implement this level of quality control typically include targets with every image, which are cropped out for access, or at critical points during the conversion process. This allows them to ensure they are delivering a high quality image that faithfully represents the original item.

It may be outside the infrastructure capabilities of Bentley Historical Library to implement such a high level quality control program considering the resources

available and technical expertise required. Using a self-describing, object level target with every image would still provide a level of quality assurance. The "burden" of ensuring color and tone quality would be shifted to the end user. Figure 10 shows an enlargement of a portion of the Golden Thread object level target.

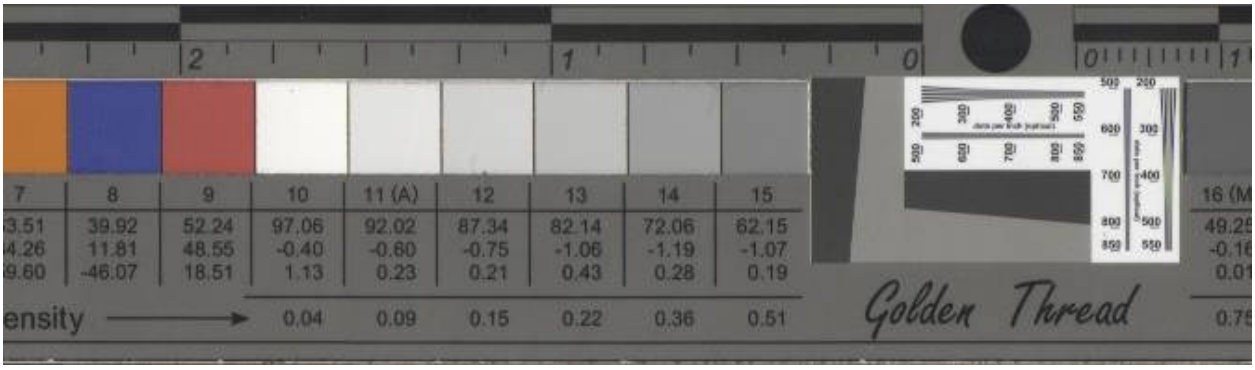


Figure 10 - Golden Thread object level target enlargement

Below each color swatch is a set of values that accurately describe the color levels for that swatch. If desired, the user would be able to use a software package such as Photoshop to adjust the image to match the values of the swatches. It is not necessary refer to outside sources, or generic terms such as "red," because accurate information is included.

Barring physical constraints (such as lack of room on the scanner platen), it is recommended that all scans where color is important include a self describing reference target with a dimensional scale, such as those from Image Science Associates. While expensive, these targets have many advantages, including:

- High quality: Color patches are made to exacting standards using Munsell color sheets
- Self describing: values are included for all color patches directly on the scale
- Includes dimensional scale
- Includes grey scale: can function in place of the Q-13 for adjusting aimpoints
- Can be used in the future if quantifying imaging performance becomes important
- The .5 target is about half the size (4-5/8" x 5/8" x 3/32") of the Q-13 grayscale, therefore allowing use in more situations and reducing file size.

Image Science Associates (http://www.imagescienceassociates.com) sells three versions of the Object Level Target: Full size, .5x size, and 2x size. The full size target is $249. If desired, all three targets can be order as a package for $799 providing a more flexible configuration which more closely matches the target to the size of the media.

Scanning Environment

The *Guidelines* give some fairly specific recommendations regarding the environment in which scanning takes place. Strictly meeting these guidelines would require a significant reconfiguration of current spaces and is impractical given the resource requirements. Some basic improvements could be made, though that could have an impact in image quality.

Though certainly not horrible, the areas around the scanning stations are a bit crowded and disorganized. Removing extraneous equipment would help keep the area clean, assist with project flow, and protect original items.

The scanner platens and document mats need to be cleaned on a more regular basis. They should be cleaned before any scanning starts for the day. If the scanner is being used heavily, then it should also be cleaned at least one additional time during the day as well. The document mats, particularly for the Epson 10000XL, are beginning to show signs of wear. Dirty and worn mats can not only affect image quality, but also detract from the aesthetic quality of images as well. Consider purchasing new document mats for both scanners. Cleanliness should especially be watched at the station in the Audio/Visual room. During the course of the project, it was noticed that red wall covering is leaving red dust particles on the equipment.

Lighting significantly affects how objects are viewed and therefore steps that might be taken to adjust for color and tone. The *Guidelines* recommend a natural, relatively dark environment, lit at a specific color temperature. This would be difficult to achieve at the current Epson 10000XL scanning station, but providing task lighting with 5000K bulbs would help. It is much more achievable in the front Audio/Visual room. Simply replacing the fluorescent bulbs with new bulbs capable of delivering a 5000K color temperature would help tremendously. This especially includes the bulbs in the perimeter lighting.

Color Management and Color Calibration

Color management is one of the more costly and challenging aspects of digitization. Each piece of hardware in a digitization workflow, from scanner, to display, to printer, introduces certain biases of color and tone. Without proper color management, what appears as "red" on one device may look more "maroon" on another. Attempting to make digitized materials "look good" on un-calibrated equipment can significantly introduce these color biases into master images.

To perform proper color management, special equipment and software is required to calibrate each part of the workflow. This specific kind equipment was not available for testing and creating specific recommendations. Therefore, not using color calibration

and management makes the use of reference targets in the scanning workflow more critical.

If future projects are planned where color reproduction is critical, then consideration should be given to implementing color management into the workflow. This would involve the purchase of additional equipment, possibly including monitors with better color control features such as gamma and color temperature adjustments.

File Naming

Any number of file naming conventions could be implemented. Many institutions embed a certain amount of metadata into file names and directory structures with naming conventions often varying between projects. One possible example could be:

- Postcards
  - Box 10
    - Saginaw Schools
      - postcard1.tif
      - postcard2.tif
    - Saginaw Street Scenes

The limitations of the scanner software make the above sort of implementation impractical. The person performing the scanning would be constantly changing software settings, thereby slowing down the process. Additionally, the structure of the above scheme would be recorded in the database.

Because of the scanner software's limitations for naming images, it is recommended that a simple sequential numbering scheme be used: pc000000.tif In order to facilitate data entry and management, the ODD numbered files will be the front of the postcard, the EVEN numbered files will be the back of the postcard. In the case of folded or unusual postcards, numbers will be sequentially used until all the necessary views are captured. Numbering will then resume with the ODD/EVEN schema for front and back. As scanning and post processing is completed, the files can be moved in batches of 100 into named folders to facilitate local browsing:

- Postcards
  - 0001-0100
    - pc000001.tif
    - pc000002.tif
    - …
  - 0101-0200
    - pc000101.tif
    - pc000102.tif
    - …

Storage

Detailed storage recommendations, such as digital repository recommended by the *Guidelines,* are beyond the scope of this report. It is of some concern that the data is currently only stored on non-redundant hard disks and optical media. Having some sort of network attached storage (NAS) with basic redundancy would be a better solution. This NAS device would not serve as the digital repository, but serve primarily to improve internal workflow and access within the Bentley.

A practical example of this would be scanning of the postcards. Due to scheduling and availability of the workstation with scanners, work will likely be conducted on several machines. In either case, data must be transferred from one station to another in order to centralize storage. Right now, there is no convenient method for transferring data. Common methods include using jump drives, or uploading and downloading data via personal IFS storage. This will start to become impractical with the amount of data involved with the postcard collection.[16] Having a storage space mounted as a network drive on the workstation would allow the person scanning so save the files directly to the intended location.

File Format

The Bentley Historical Library currently uses the TIFF file format for storing master images. TIFF has the advantage of being a proven format with a relatively long record of stability. It is also widely supported in various software packages and is currently considered the de-facto standard for storing master image files.

JPEG2000 is another format that bears consideration. It is increasingly becoming an accepted format for master files – U of M DLPS specifies JPEG2000 for its vendors which deliver continuous tone images. Some of the advantages include:

- Open standard
- Lossless compression (typically about 50% space savings compared to uncompressed TIFF)
- Extensive metadata support
- Multiple resolutions in one file allowing one master to replace multiple derivatives

Additionally, using a lossy compression, significantly higher compression rates are also possible (as much as 90-95%). JPEG2000, however, uses a different compression method compared to that of the traditional JPEG image and therefore does not exhibit the same visual loss of image quality. Currently, the primary disadvantage for JPEG2000 is that it is not widely supported by as many software packages.[17]

---

[16] A rough estimate indicates there would be over 130 gigabytes of data for the postcard collection.

[17] The Library of Congress recently hosted a JPEG 2000 Summit on May 12-13, 2011. Presentations and information from the Summit are located on the FADGI website at http://www.digitizationguidelines.gov/resources/jpeg2000.html

It is recommended that the Bentley Historical Library continue using TIFF as the format for storing its master images, but keeping an eye to the future with regards to JPEG2000.

MetaData

It was decided that more value would be gained from the collection if both sides of the postcards were scanned. This, along with items such as the folded postcard envelopes, presented new issues with regards to tracking metadata. In the past, it was assumed there would be one record that held all metadata for one object. The situation is now such that one object could have multiple images associated with it. Without some method of linking these images together, then it becomes difficult to "reassemble" the structure the original object in digital form.

The Metadata Encoding and Transmission Standard (METS) is increasingly becoming the norm for creating this kind of structural metadata. While the Bentley Historical Library does not have an infrastructure for managing individual digital images and METS documents, it can provide a framework for structuring the data.

Keeping this in mind, a new database was designed using FileMaker Pro version 9 that allows each record to have multiple images attached to it. If desired, at some point in the future, it should be possible to export and convert the database into METS records associated with the image files. However, one should be aware that this would not necessarily be a simple process and would need somebody with strong XML and programming knowledge.

For our purposes, a METS document describing an object has three primary containers that hold metadata:

1. Descriptive Metadata
2. Administrative Metadata, consisting of:
    a. Rights Metadata
    b. Technical Metadata
    c. Preservation Metadata
3. Structural Metadata, consisting of:
    a. Files Section
    b. Structural Map

Descriptive Metadata

METS can hold descriptive metadata in any number of formats, such as Dublin Core, MODS, MARC, etc. The data can either be internal, or linked to an external resource. Dublin Core is what most institutions use for a minimum data set as well as being

recommended by the *Guidelines*.[18] The database fields currently in use for descriptive metadata meet the minimum required for the Dublin Core. Not all of these fields are necessarily required for the postcards project and trying to capture them all could significantly slow the project down.

Items marked with ✳ are recommended for use with the postcards project.

BHL Term	**Dublin Core Equiv.**	**Typical values**	**Notes**
✳ Image Title	title		Take title from item. Include specific names of buildings, bridges, parks, hotels, etc. If information is not available, make description (e.g., bridge over Huron River) and location. When no information available, enter "unknown location".
Title Source		Supplied, verso, image	When you supply title, enter "Supplied." Otherwise, use terms from the drop down menu (e.g., verso)
Caption	description		
Image Date	date or dateCopyrighted?	1917, [19--?], [196-?]	use date displayed on item. When available, use publication, copyright date ("c. 1956), postmarked date, or date contained in message, etc. Alternatively, use description on card to identify approximate date and use "ca.", or "after" in date certainty field (e.g., 1917 (ca.); 1936 (after). If no dates available and you can approximate the decade, use the following format: [196-?] If you can't approximate a decade, use the following format: [19--?] If you can't approximate whether item is from the 19th or 20th century, enter "undated"
Date Certainty		ca., after, before	(see above)
Notes	description		additional valuable information or anything in need of explanation
✳ Collection Creator	creator?	Bentley Historical Library	Can be duplicated automatically on record creation
✳ Collection Title	Source?	Postcard Collection 1890s-[Ongoing]	Can be duplicated automatically on record creation
✳ Box No.	Source?	4	Can be duplicated automatically on record creation

[18] Federal Agencies Digitization Initiative (FADGI) - Still Image Working Group, "Technical Guidelines for Digitizing Cultural Heritage Materials: Creation of Raster Image Master Files," 71.

✳ Folder	Source?	Detroit	Can be duplicated automatically on record creation
Image Creator	contributor		Manufacturer, and photographer or artist – record the copyright holder (ie publisher) under the Rights Metadata section
Work Creator	contributor		If architect or builder name is available, record it in the "Work creator" field
Project			Should probably change to a relational project ID…
✳ EAD Link		umich-bhl-93347	The ID to the online finding aid; Can be duplicated automatically on record creation
✳ Daohref			Link to the series / folder ID where this item is located in the collection; Can be duplicated automatically on record creation
✳ Mirlyn			The collection record ID from ALEPH
✳ Repository	publisher	Bentley Historical Library	Auto Entered
✳ Cataloger		MLA	Initials of the person entering the record; Can be duplicated automatically on record creation
✳ Persname	subject		Can be duplicated automatically on record creation
✳ Corpname	subject		Can be duplicated automatically on record creation
✳ Geogname	coverage		Can be duplicated automatically on record creation
✳ Genre	type	Postcards	Can be duplicated automatically on record creation
✳ Subject	subject		Can be duplicated automatically on record creation
ScanSource / Type	format		Equivalent to BHL REF Type, for Imagebank DB, it is part of ScanSource – should probably make this a controlled list – The Dublin Core equivalent would be made of several other fields of the Tech Metadata

Rights Metadata

In METS, Rights metadata is included as a sub-section under the Administrative Metadata section. A MetsRights XML Schema as been developed, but it is generally too in depth for our needs at this time.

Current BHL Term	Dublin Core Equiv.	Typical values	notes
Copyright	rights	text	The standard BHL copyright notice
✳ Original	rightsHolder	text	Text describing copyright info recorded on object –

Copyright			ie., publisher, etc.
✳ Copyright flag			A controlled list / flag to make searching easier. Examples: Copyrighted Public Domain Unknown Copyright transferred to BHL copyright released

Preservation Metadata

At this time, it is probably outside of the infrastructure capabilities for the Bentley to start tracking in depth preservation metadata as defined by the PREMIS Data Dictionary. It is recommended that some PREMIS kinds of information be tracked in the Technical Metadata section, such as when a master image is assembled using multiple scans.

Technical and other Administrative Metadata

The most accepted standard for Technical Metadata is the "Data Dictionary - Technical Metadata for Digital Still Images (ANSI/NISO Z39.87-2006)." This standard provides a hierarchical structure, based upon the specifications for TIFF header tags, for capturing still image metadata and has been encoded into XML using the MIX schema managed through the Library of Congress. Technical Metadata specified in the "University of Michigan Digitization Specifications" is based on a 2002 draft of this document.

In regards to recording metadata, the technical metadata is the area in which the current procedures could use the most improvement. Very little of the technical metadata is formally captured. It is also manually entered, increasing the chance of human error and inconsistent data entry. Tools are available (such as JHOVE and ExifTool) that can assist with the automation of this process and cut down on the amount of manual data entry required.

Table 5 (below) details the recommended metadata tags that should be captured for digitization projects in general.

Table 4 Column Descriptions

Auto Capture:

This value is typically already included in the TIFF headers when the file is created and can be capture using external utilities. In the new database written using FileMake 9, fields marked as Auto Capture are saved by recording the output of JHOVE and ExifTool into two separate text fields. Any value that is **not** marked as Auto Capture would need to be manually recorded in the database.

Write to file:

This field would need to be written to the file from manually entered database values using an external utility such as ExifTool.

Tiff Tag:
The TIFF Header ID number used internally within the file.

***Z39.87*:**
The value recorded here is the relevant section number in the "Data Dictionary - Technical Metadata for Digital Still Images (ANSI/NISO Z39.87-2006)" document.

Items in the Z39.87 column marked with an asterisk (*) indicate that the data there is further broken down into other subfields. For example, 8.3.5.* ScanningSoftware is broken into:

8.3.5.1: scanningSoftwareName
8.3.5.2: scanningSoftwareVersonNo

Auto Capture	Write to file	TIFF Tag	Z39.87	DLPS	Name (s)	Example Values	Notes
X		256	7.1.1	X	ImageWidth	3349	Recorded in pixels
X		257	7.1.2	X	ImageLength	2165	Recorded in pixels
X		258	9.2.1.*	X	BitsPerSample	8 8 8	Corresponds to the BHL Scantype – 8 8 8 would correspond to “24 bit color”
X		259	6.6.*	X	Compression / CompressionScheme	1= Uncompressed	Tiff Std. uses an integer, Z39.87 uses a lists of strings.
X		262	7.1.3.*	X	Photometric Interpretation	RGB	May need to include more detail info on colorspaces, color profile, etc
	X	269	6.1.*	X	DocumentName / ObjectIdentifier	pc0000001.tif	Unique filename
	X	271	8.3.1	X	Make/ ScannerManufacturer	Epson	
	X	272	8.3.2.*	X	Model/ ScannerModel	10000XL	Z37.89 specifies the Model Name, Number, & Serial Number
X		273		X	StripOffsets		“The number of components per pixel.”
X		274	8.5	X	Orientation	1	1 = Horizontal (normal) 2 = Mirror horizontal 3 = Rotate 180 4 = Mirror vertical 5 = Mirror horizontal and rotate 270 CW 6 = Rotate 90 CW 7 = Mirror horizontal and rotate 90 CW 8 = Rotate 270 CW
X		277	9.2.2	X	SamplesPerPixel	3	
X		278		X	RowsPerStrip		
X		279		X	StripByteCounts		
X		282	9.1.2.1	X	XResolution / XSamplingFrequency	600	ppi
X		283	9.2.2.2	X	YResolution / YSamplingFrequency	600	ppi
X		284		X	Planar Configuration	Chunky	
X		296	9.1.2	X	ResolutionUnit / SamplingFrequencyUnit	Inches	Units which the X and Y resolution values are recorded in. For our purposes, Inches

							will almost always be the normal value
	X	305	8.3.5.*	X	Software / ScanningSoftware	Epson Scan 3.04	
		306	8.2.1	X	DateTime / DateTimeCreated	2004:09:08 00:00:00	ISO8601 format YYYY:MM:DD HH:MM:SS
	X	315	8.2.2	X	Artist / ImageProducer	Bentley Historical Library, DLPS, etc.	[scanning contractor company name]
	X	270			ImageDescription	String that describes the subject	LOC Considers this a baseline TAG. Use the ImageDescription from the descriptive data for the value.
X			6.2		Filesize		
			6.3.*		FormatDesignation	image/tiff	Corresponds to the current BHL Filetype – need to use mimetype as the actual values here
X			6.5		ByteOrder	little_endian	Required by the NISO Z37.89 standard
X			6.7.*	X	Fixity	786fd62b33b65560d9993d0f5f4235d1	The MD5 checksum of the file – should only be calculated AFTER all modifcations and TIFF headers have been written to the file
			8.1.1		SourceType	Postcard, Photoprint, copy negative, etc.	Equivalent to BHL reference datbase Type, for Imagebank database, it is part of ScanSource. Us a controlled vocabulary list
			8.1.3.*		SourceSize	8 inches x 10 inches	Equivalent to BHL Scansource – need to record it as separate X and Y dimensions with units - Note: Z37.89 std give instructions when you are only scanning partial of an original
			8.2.3		captureDevice	Transmission scanner, reflection scannner	May not be necessary, but could be recorded using a related table associated with basic scanner information
			8.3.3		maximumOpticalResolution	4800 dpi	May not be necessary, but could be recorded using a related table associated with basic scanner information
			10.1.*		ImageProcessing		Z39.87 allows for detailed recording of

							image processing steps. This is also traditionally recorded in PREMIS metadata. If necessary, basic information should be recorded. For example, if an image is created by assembling multiple scans in Photoshop.

Table 5 - Technical Metadata

Additional information in tag descriptions and values can be found at the Aware Systems website http://www.awaresystems.be/imaging/tiff/tifftags.html and the ExifTool website http://www.sno.phy.queensu.ca/~phil/exiftool/TagNames/

Structural Metadata

The METS standard is highly flexibly with regards to how individual images can be associated with a record. The Structural Metadata section of a METS document consists of two parts:

- Files Section
- Structural Map

The Files Section of the METS document lists all of the individual files associated with a record. Depending on the institution, often a file might have a variety of versions: master, thumbnail, hi-res use, etc. The list is really endless and depends upon the needs of the institution or even the particular object. Within the File Section, it is possible to create File Groups to separate these items. Roughly outlined, it might look like the following:

- Files Section
    - File Group = Master Images
        - Image1.tiff
        - Image2.tiff
        - Image3.tiff
    - File Group = Thumbnails
        - Thumb1.jpg
        - Thumb2.jpg
        - Thumb3.jpg
    - File Group = Use Copy
        - Image1.jpg
        - Image2.jpg
        - Image3.jpg

The Structural Map (or <structMap>) is one of the core sections of a METS record and it used to describe the physical layout of an item.

> It provides a means for organizing the digital content represented by the [FileSection] of the METS document into a coherent hierarchical structure. … Such a hierarchical structure can be presented to users to facilitate their comprehension and navigation of the digital content. It can further be applied to any purpose requiring an understanding of the structural relationship of the content files or parts of the content files. The organization may be specified to any level of granularity (intellectual and or physical) that is desired. Since the <structMap> element is repeatable, more than one organization can be applied to the digital content represented by the METS document.

> The organization provided by the <structMap> may be purely intellectual or logical (such as a book divided into chapters), purely physical (a book divided into sequences of pages), or a mixture of logical and physical (a book sub-divided into chapters and subsequently divided into a sequence of pages). [19]

Given the above list files from our Files Section, one Structural Map could look roughly like the following:

- Structural Map
    - Physical Layout
        - Page 1
            - Image1.tiff
            - Thumbnail1.jpg
            - Image1.jpg
        - Page 2
            - Image2.tiff
            - Thumbnail2.jpg
            - Image2.jpg
        - Page 3
            - Image3.tiff
            - Thumbnail3.jpg
            - Image3.jpg

In the case of the Bentley, most materials that will be digitized consist of individual objects. An object might have multiple images associated with it, for example:

- Front and back of a post card
- Parts of a folded postcard set
- Pages of a journal

Attempting to create a database that would incorporate the large number of possibilities available within a Structural Metadata section is impractical given the resources available.

Reviewing the kinds of items that might be digitized, it was felt that the kinds of images (essentially, the "File Groups" from above), could be divided into two groups:

- **Archival master(s):** Typically containing a color target, and represents the object as accurately as possible.

[19] "Metadata Encoding And Transmission Standard: Primer And Reference Manual," 58.

- **Derivate image(s):** These are for facilitating access and may not 100% represent the original. Decisions may need to be made as to what are the important characteristics that need to be captured.

A record may have any number of master and derivatives. In fact, it is felt that for most scenarios, a record would most likely have only master images associated with it, from which derivatives are automatically created on an as needed bases, as in the case of the Bentley Image Bank.

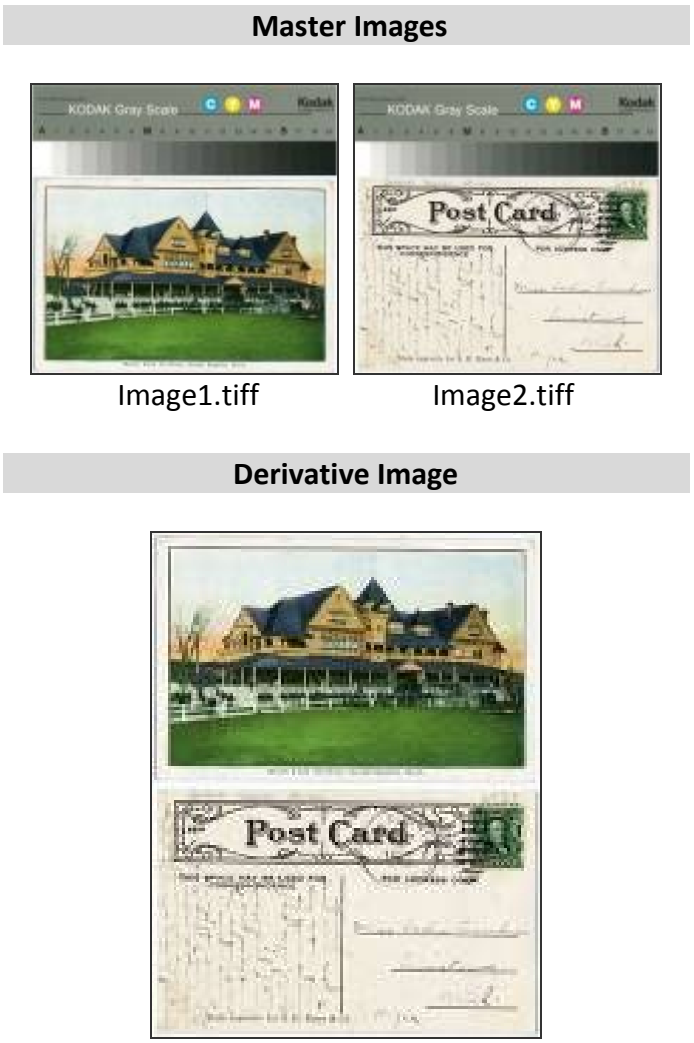


Figure 11 - Post Card Sample, master and derivative images

The structure for Figure 11 could conceivably be:

- Files Section
  - File Group = Master Images
    - Image1.tiff
    - Image2.tiff
  - File Group = Derivative

        - Image1.jpg
- Structural Map
    - Physical Layout "Masters"
        - Image1.tiff
        - Image2.tiff
    - Physical Layout "Derivatives"
        - Image1.jpg

The important thing to realize from all of this is that the **structure would be automatically handled in the database.** An image would simply have to be marked as a "master" or a "derivative" when it is attached to a record.

Quality Management

The *Guidelines* recognize quality management / quality control as an important part of any digitization workflow. Currently, some quality control measures are built into process for recording metadata in the new database:

- Using JHOVE and ExifTool to validate files
- Specific required and unique fields
- Ability to duplicate descriptive metadata when creating new records (though this could be a problem if the metadata was incorrectly entered in the first place).

A full quality control plan should include inspection of the digital files for meeting technical standards, the information captured "looks right," missing or incorrect metadata, and recording of device performance. Pages 82-85 of the *Guidelines* provide more details on recommended quality management procedures.

Implementation of such a program would take thoughtful planning and likely require additional resources to create a comprehensive system. For example, in order to track quality control on imaging devices, software such as the Golden Thread package from Image Science Associates would need to be purchased.

If the Bentley Historical Library intends to expand its digitization programs, then consideration should be given to creating a more comprehensive quality management program.

Equipment and Software

Beyond what has been previously mentioned, the only recommendation for purchase is the Troi File Plug-in for FileMaker Pro. A trial version of this plug-in was used in the design of the new FileMaker 9 database. Among its features, it allows FileMaker to select image files for inclusion in the database and also run tools such as JHOVE and

ExifTool. Without it, the automated capture of much of the technical metadata would not be possible. The cost is $89 per seat, or a 10 user pack for $399 and is available from www.troi.com/software/fileplugin.html

Appendix A – Resources

General

Glossary of terms: http://www.digitizationguidelines.gov/glossary.php

Websites

Metadata Encoding and Transmission Standard (METS) Official Web Site: http://www.loc.gov/standards/mets

The Dublin Core® Metadata Initiative: http://dublincore.org

Federal Agencies Digitization Guidelines Initiative: http://www.digitizationguidelines.gov/

Guidelines: Content Categories & Digitization Objectives: Reformatting Historical Printed Matter, Documents and Manuscripts, and Pictoral Materials — Content Categories and Subcategories table - Federal Agencies Digitization Guidelines Initiative http://www.digitizationguidelines.gov/guidelines/ccdo-subcat.html

NARA Special Media Products and Services, Reformatting Approaches Based on Original Record Type: http://www.archives.gov/preservation/products/

File Formats

Florida Digital Archive: http://www.fcla.edu/digitalArchive/

Sustainability of Digital Formats: Planning for Library of Congress Collections at: http://www.digitalpreservation.gov/formats

Janosky, James S., and Rutherford W. Witthus. "Using JPEG2000 for enhanced preservation and web access of digital archives-A case study." In *IS&T's 2004 Archiving Conference*, 1:145–149, n.d.

Federal Agencies Digitization Initiative (FADGI) - Still Image Working Group. "Technical Guidelines for Digitizing Cultural Heritage Materials: Creation of Raster Image Master Files", August 2010. (Pages 67-69)

"JPEG 2000 - Resources - Federal Agencies Digitization Guidelines Initiative", http://www.digitizationguidelines.gov/resources/jpeg2000.html.

Metadata

BCR's CDP Digital Imaging Best Practices Working Group. "BCR's CDP Digital Imaging Best Practices, Version 2.0", June 2008.

California Digital Library. "CDL Guidelines for Digital Objects: Version 2.0". Regents of the University of California, January 2011. http://www.cdlib.org/services/dsc/contribute/docs/GDO.pdf.

"Data Dictionary - Technical Metadata for Digital Still Images (ANSI/NISO Z39.87)". NISO Press, 2006.

Federal Agencies Digitization Initiative (FADGI) - Still Image Working Group. "Guidelines for TIFF Metadata, Recommended Elements and Format, Version 1.0", February 10, 2009. http://www.digitizationguidelines.gov/guidelines/TIFF_Metadata_Final.pdf.

———. "Technical Guidelines for Digitizing Cultural Heritage Materials: Creation of Raster Image Master Files", August 2010. http://www.digitizationguidelines.gov/guidelines/digitize-technical.html.

"Harvard University Library's Administrative Metadata for Digital Still Images data dictionary", 2008. http://hul.harvard.edu/ois/systems/drs/imagemetadata.pdf.

"Metadata Encoding And Transmission Standard: Primer And Reference Manual". Digital Library Federation, 2010. http://www.loc.gov/standards/mets/METSPrimerRevised.pdf.

The University of Michigan, University Library Digital Library Production Services, Digital Conversion Unit. "[University of Michigan Digitization Specifications]", n.d. http://www.hathitrust.org/documents/UMDigitizationSpecs20100827.pdf.

Appendix B – Records Handling for Digitization[20]

See NARA's *Preservation Guidelines for Vendors Handling Records and Historical Materials* at http://archives.gov/preservation/technical/vendor-training.html; as well as Appendix E, "*Records Handling for Digitization*," in NARA's 2004 publication of *Technical Guidelines for Digitizing Archival Records for Electronic Access: Creation of Production Master Files – Raster Images*, available at http://archives.gov/preservation/technical/guidelines.pdf

[20] Federal Agencies Digitization Initiative (FADGI) - Still Image Working Group, "Technical Guidelines for Digitizing Cultural Heritage Materials: Creation of Raster Image Master Files."

Appendix C – Detailed Digitization Steps

Note that some of the scanning steps here should not be considered "absolutes" and are meant as general guidelines. These steps have also been optimized for the use with the Postcard Collection, though they can, in principle, be transferred to other projects.

The scanning process should be broken up into small, manageable batches of postcards. A good dividing point for each batch would be the dividers within the collection. This will primarily allow for easier metadata entry into the database. The general workflow for scanning would be:

Scan group of postcards ⇨ Process in Photoshop ⇨ Enter metadata into database

It is very important that any post scan image processing in Photoshop take place **before** the files are entered into the database. If the files are changed after entry in the database, then the metadata, such as the MD5 Hash, will be different.

Scanning Steps – for two sided postcards

- Clean scanner if necessary
- Set the basic scanner settings (See Figure 12):
    - Document Type: reflective
    - Auto Exposure Type: Photo
    - Image Type: 48 bit color
    - Scanning Quality: Best
    - Resolution: 400 dpi
    - Target Size: Original (100%)
- Open up the File Save Settings and check (See Figure 12):
    - File Name pattern is correct,
    - Start Number is correct (should be an odd number if scanning the front of a postcard)
    - Image Format is TIFF format, (Do not embed profile, if applicable)
    - Folder Location is correct
- Turn off all scanner color correction options (See Figure 12)
- Place target on scanner, allowing a little room around target. (The scanners tend to clip edges off, especially the Epson 10000xl)
- Select the largest postcard from the batch to be scanned
- Place the postcard face down in the scanner bed, centered on the target
- Perform a preview scan
- Adjust the selection marquee to fit the largest postcard and target
- Place first postcard face down in the scanner bed, centered on the target
- Scan
- Flip the postcard, center on the target, and scan rear
- Repeat for batch

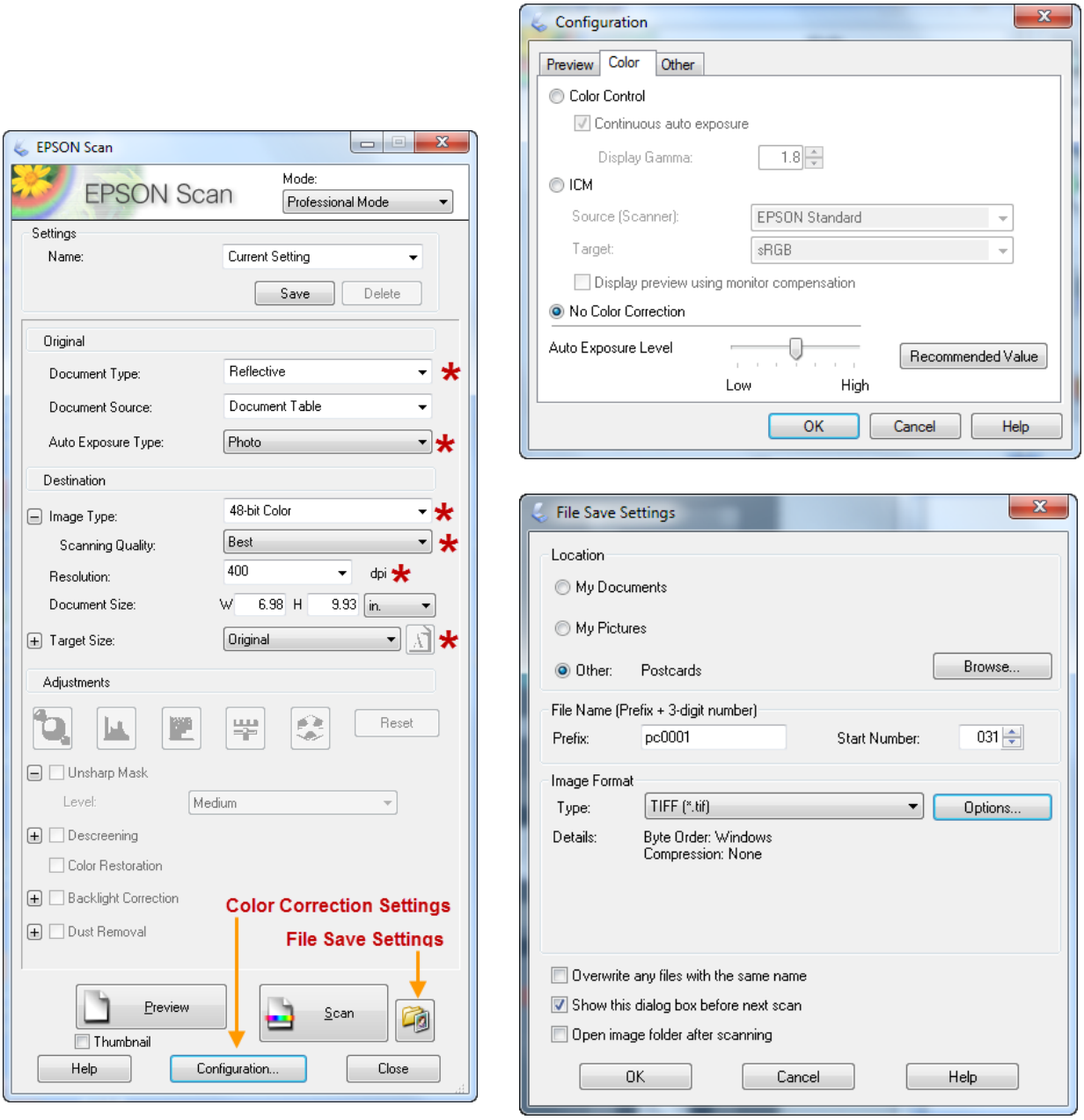


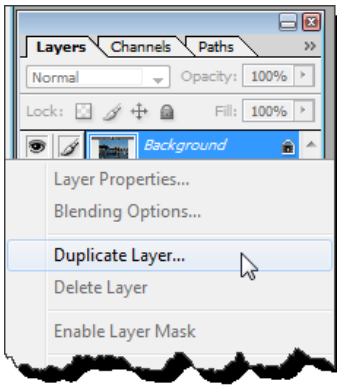
Figure 12 – Recommended Scanner Settings

Photoshop Post Processing Steps

These steps can be saved in a series of actions which can then be used in a batch process. Steps described are for Photoshop CS. They may need to be slightly modified for later versions of Photoshop.

- Convert to Adobe RGB color space for color images:
  Image Menu ⇨ Mode ⇨ Convert To Profile…
- Correct color and tone using preset curves for the scanner used:
  Image Menu ⇨ Adjustments ⇨ Curves…
  Click the Load button to load the curve profile saved for that scanner (See below for creating a preset color correction curve)
- Apply Unsharp mask to a Luminosity layer (See below)
- Convert to 8-bits per channel – either 24-bit RGB or 8-bit grayscale.
- Save the file
- Move to "permanent location"

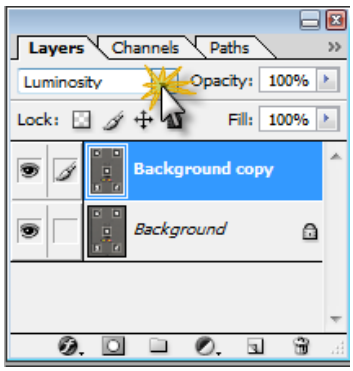
Applying Unsharp Mask in Photoshop



- Duplicate the current layer (typically background). One method to do this is by right-clicking it in the layers palette and selecting Duplicate Layer… The default value for the layer names is acceptable since the name of the layer is not critical.

- Set the mode of the new layer to Luminosity

- Apply Unsharp Mask Filter to new layer:
  Filter Menu ⇨ Sharpen ⇨ Unsharp Mask…
  - Recommended settings are
    Amount: 50%
    Radius: 1.5 pixels

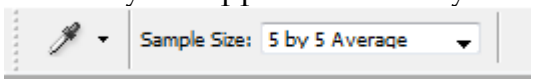


- Flatten the results:
  Layer Menu ⇨ Flatten Image

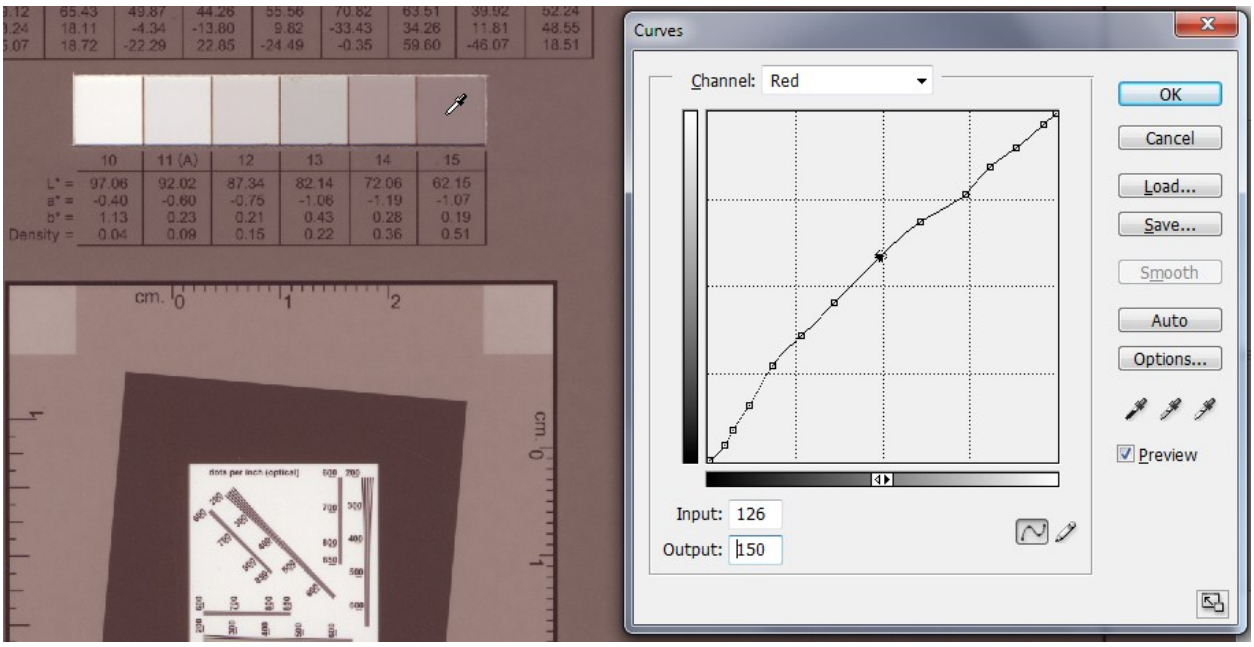
Creating a preset color correction curve in Photoshop

Every scanner will have some variation its color encoding accuracy. Therefore, this needs to be done for each individual scanner using an object or device target.

- Open an image with a target
- Select the Eye Dropper tool from the tool palette. Ensure that the Sample Size for the Eye Dropper is set to 5 by 5 averages in the options toolbar.



- Zoom in on the image so all the grayscale patches are visible.
- Open the Curves dialog by going to Image Menu ⇨ Adjustments ⇨ Curves… When you hover over the image, the Eye Dropper tool should be active.



- Set the Channel to Red
- Using the eye Dropper, Control-Click in first Gray Scale patch (ie #10)
- A point representing this sample will appear on the curve. Notice the Input and Output values. The Input is what the current value of the sample is.
- Change the Output value to the corresponding value from the table below. (ie, Red for Patch #10 is 247)
- Repeat the process for each of the 12 gray scale patches on the target.
- Once the process is complete for the Red channel, repeat for the Green and Blue Channels
- Click the Save button and save the resulting settings file to a commonly accessible location.

Grayscale Patch	Red	Green	Blue
10	247	246	244
11 (A)	230	233	232
12	216	218	218
13	202	204	204
14	175	177	176
15	148	150	150
16 (M)	116	116	116
17	92	92	92
18	70	68	67
19	40	40	38
20	22	24	23
21	11	11	11

Table 6 -Output values for Golden Thread grayscale patches based on L*a*b* values in AdobeRGB color space

Grayscale Patch	RGB Values
10	242
11 (A)	227
12	210
13	192
14	161
15	133
16 (M)	98
17	73
18	52
19	30
20	19
21	12

Table 7 - Output values for Golden Thread grayscale patches, based on an ideal 1.8 gamma curve

Appendix D – Scanner Analysis Workflow Results

The Golden Thread device level target allows for the analysis of spatial frequency response across five different patches (see Figure 1 - Device Level Target), and across both the horizontal and vertical axis for each patch. This gives a total of ten different graphs for examining the spatial frequency response of the device. For the sake of space, only the Center Vertical results examined here.

Uncorrected, "raw" image

- Scanner: Epson 1680
- Document Type: reflective
- Auto Exposure Type: Photo
- Image Type: 24 bit color
- Scanning Quality: Best
- Resolution: 300 dpi
- Target Size: Original (100%)
- Color Correction: None
- Sharpening: None

Spatial Frequency Response

Spatial Frequency Response objective measures a devices limiting resolution and also quantifying other imaging characteristics such as sharpening, flare, aliasing, and overall image quality. The curves for a scanner which delivers a perfect SFR would align along the dashed yellow line. For a target scanned at 300 dpi, the SFR curve should cross the 0.1 Amplitude at the 300 dpi line.

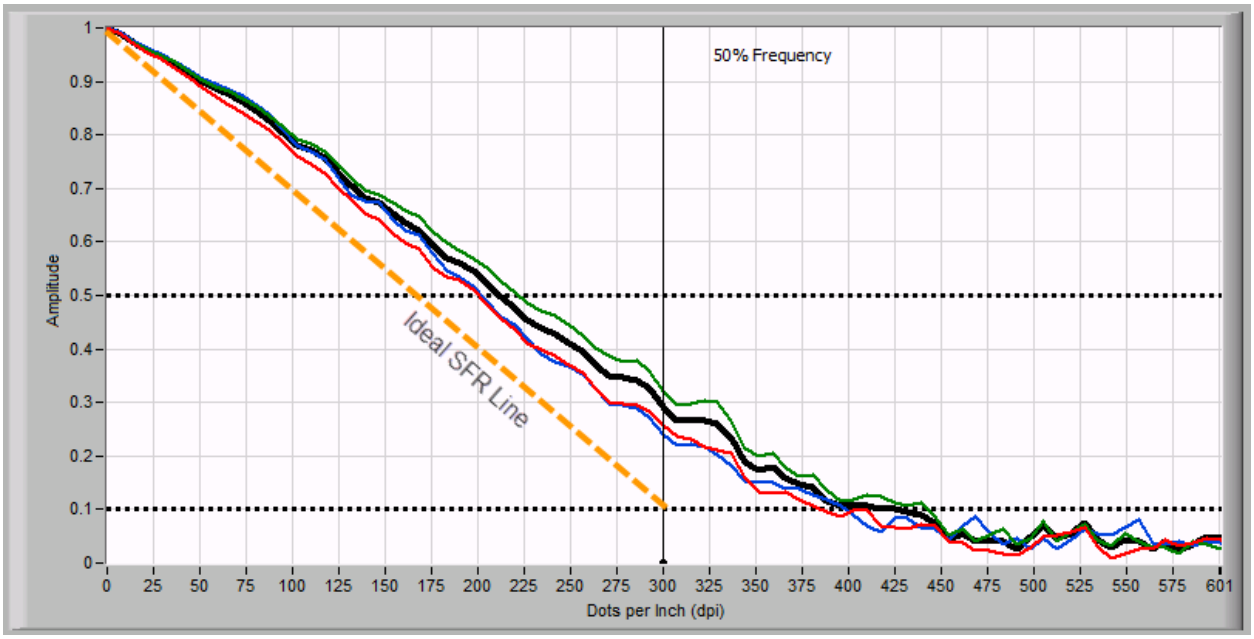


Figure 13- Epson 1680 "RAW" SFR Curve

Examining the result of the "raw" SFR analysis shows some moderate issues. The scanner appears to be delivering a resolution higher than the requested resolution of 300 dpi. The curves for each color channel cross about the .3 amplitude line at 300 dpi, which gives a borderline two-three star performance level according to the G*uidelines*. Alignment between the red and blue channels is good, though the green channel shows some misalignment.

Tone and Color Response

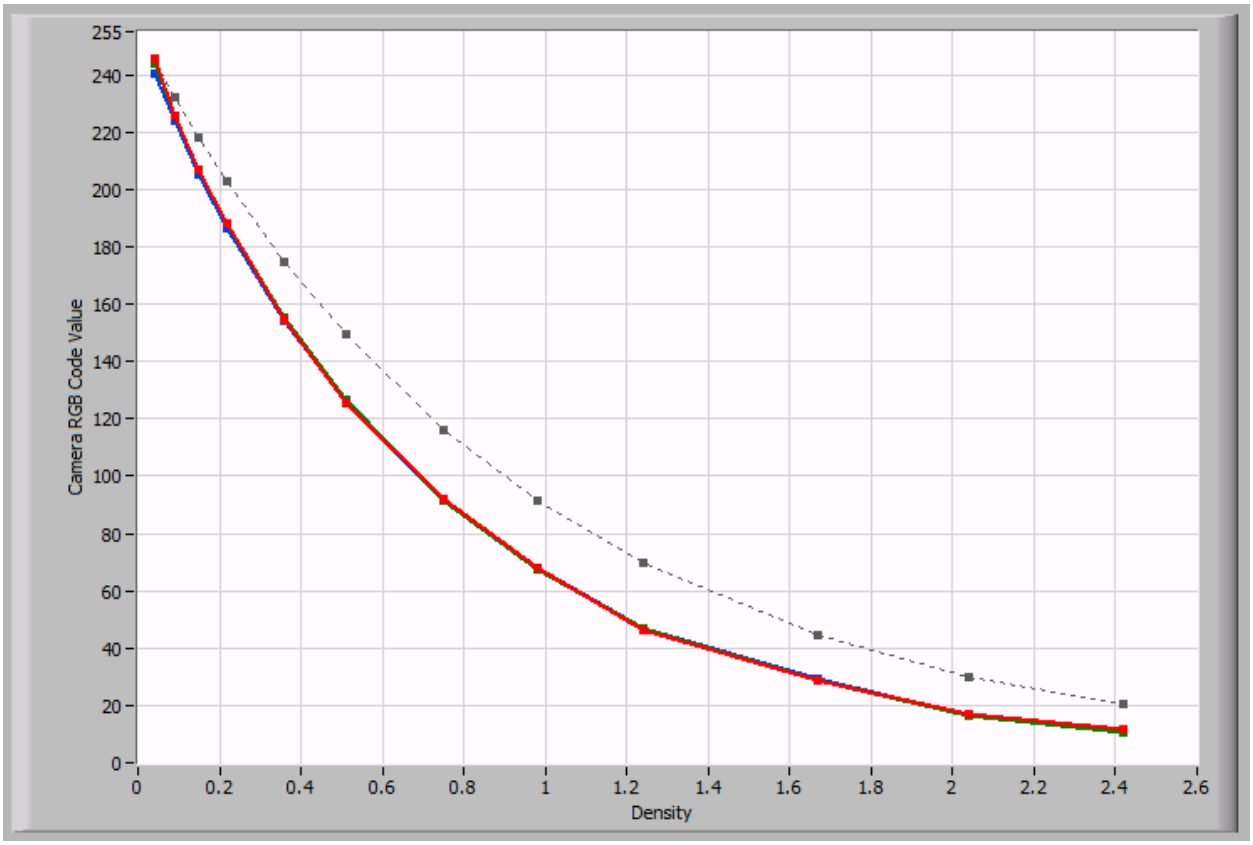


Figure 14 - Epson 1680 "RAW" OECF Curve

The "RAW" OECF Curve shows a fairly good response. The three Red, Green, and Blue channels align on top of each other very well. The entire curve is shifted a bit below the 2.2 gamma ideal curve, indicating that the image might be a little dark.

Tonscale Delta From Aim

For each density patch, the average digital value in each color channel (red, green, blue) is subtracted from its aim value (as defined in the active profile). The difference is compared to the difference between the aim and limit (both upper and lower). If an image is captured perfectly, the measured digital value would be equal to aim and the difference between the two would be zero. Upper and lower calculated limits are shown as red dashed lines.

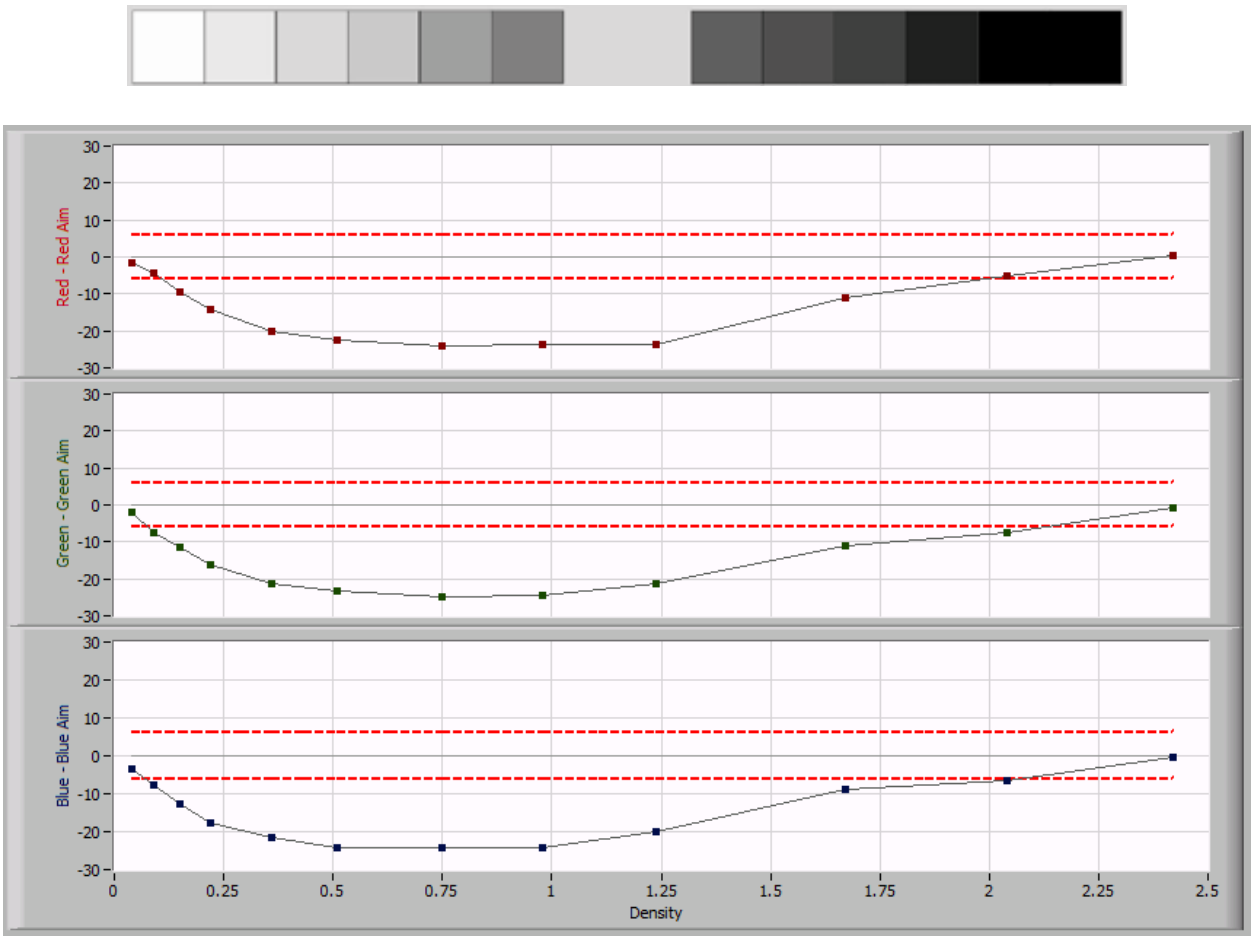


Figure 15 - Epson 1680 "RAW" Delta from Aim

The red lines show the limits for a three star rating as specified by the *Guidelines*. As can be see in the image, the values fall well outside of these limits and overall would rate nothing better than a one star according to the *Guidelines*.

White Balance / Gray balance/ Neutrality

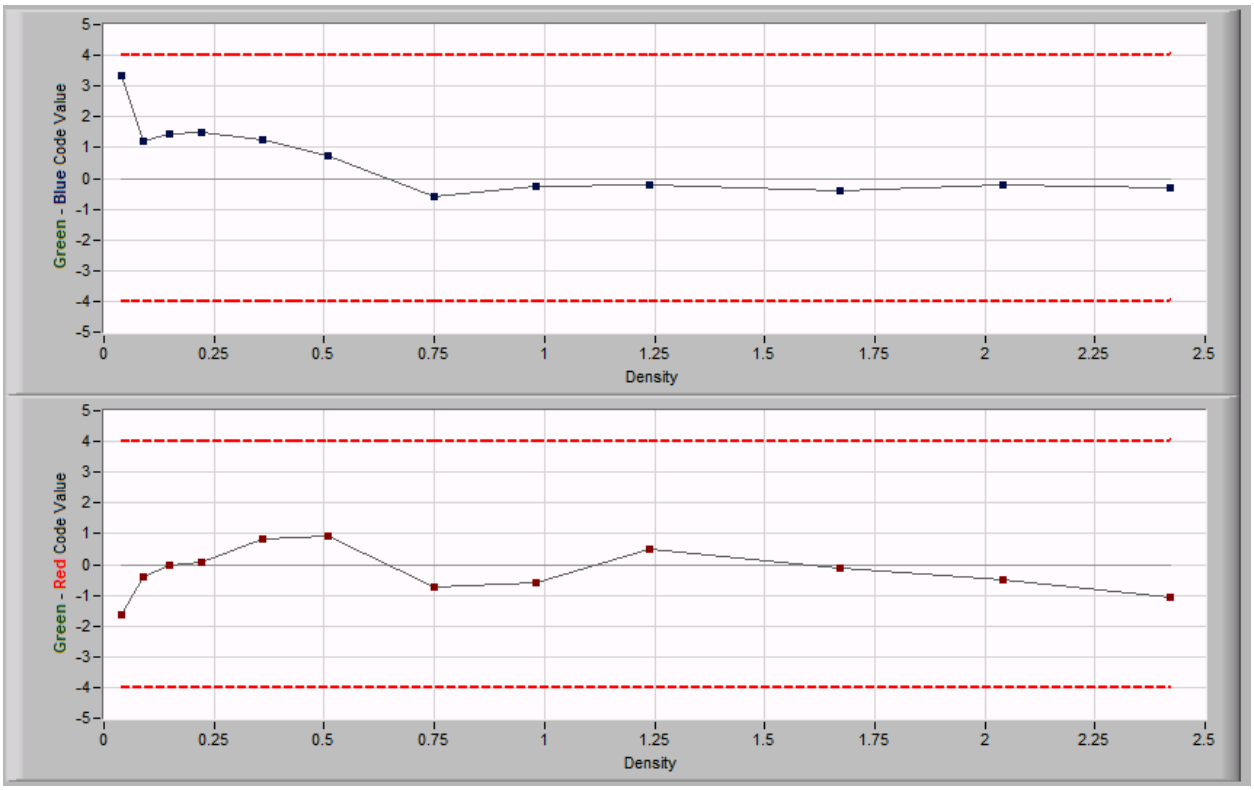


Figure 16 - Epson 1680 "RAW" Color Neutrality

In an image with a perfect White/Gray balance, the lines would be close to horizontal. While not horizontal, the line does fall within a "three star" rating as described in the *Guidelines*.

Current Workflow Settings

- Scanner: Epson 1680
- Document Type: reflective
- Auto Exposure Type: Photo
- Image Type: 24 bit color
- Scanning Quality: Best
- Resolution: 300 dpi
- Target Size: Original (100%)
- Color Correction: Continuous Auto Exposure, 1.8 Gamma
- Sharpening: Medium

Spatial Frequency Response

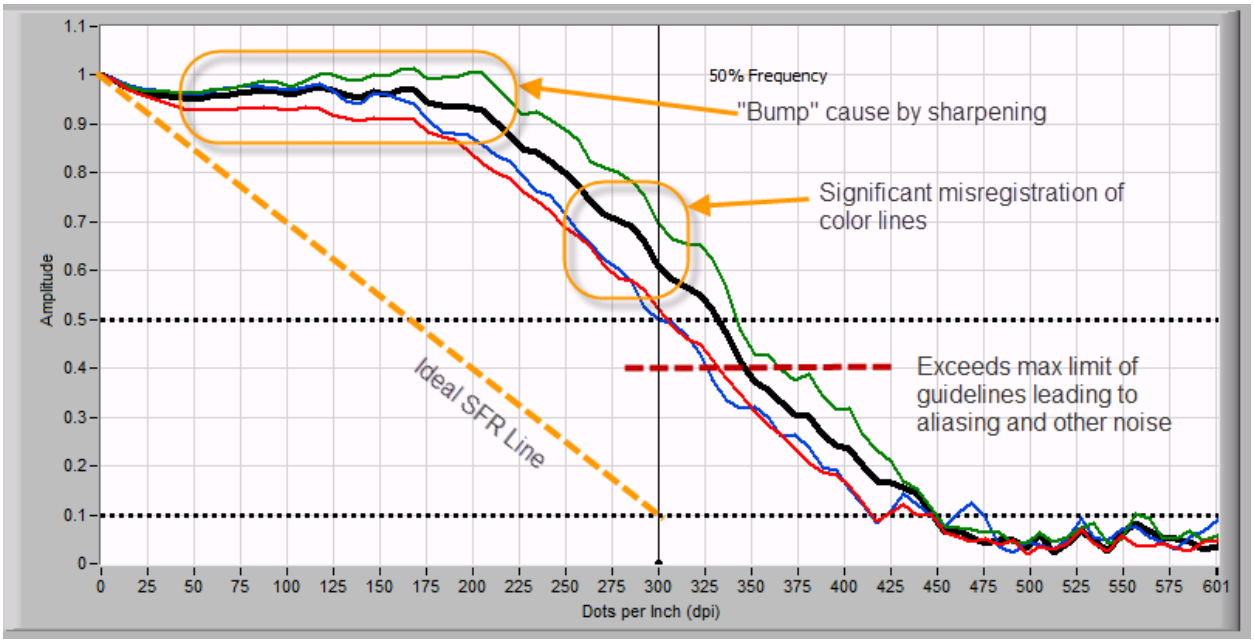


Figure 17 - SFR Curves of the current workflow settings

The SFR curves produced by the settings used in the current workflow note several areas of concern. The "bump" caused by sharpening is not significant and actually falls within the current "four star" rating of the *Guidelines*. Of greater concern is the separation of the color channels, especially between the Red/Blue and Green channels. Additionally, the curves should fall below where the dashed red line intersects the solid vertical line, in order to fall with even the lowest of the *Guidelines* ratings. This indicates that there could be significant aliasing in the image which can degrade quality.

Tone and Color Response

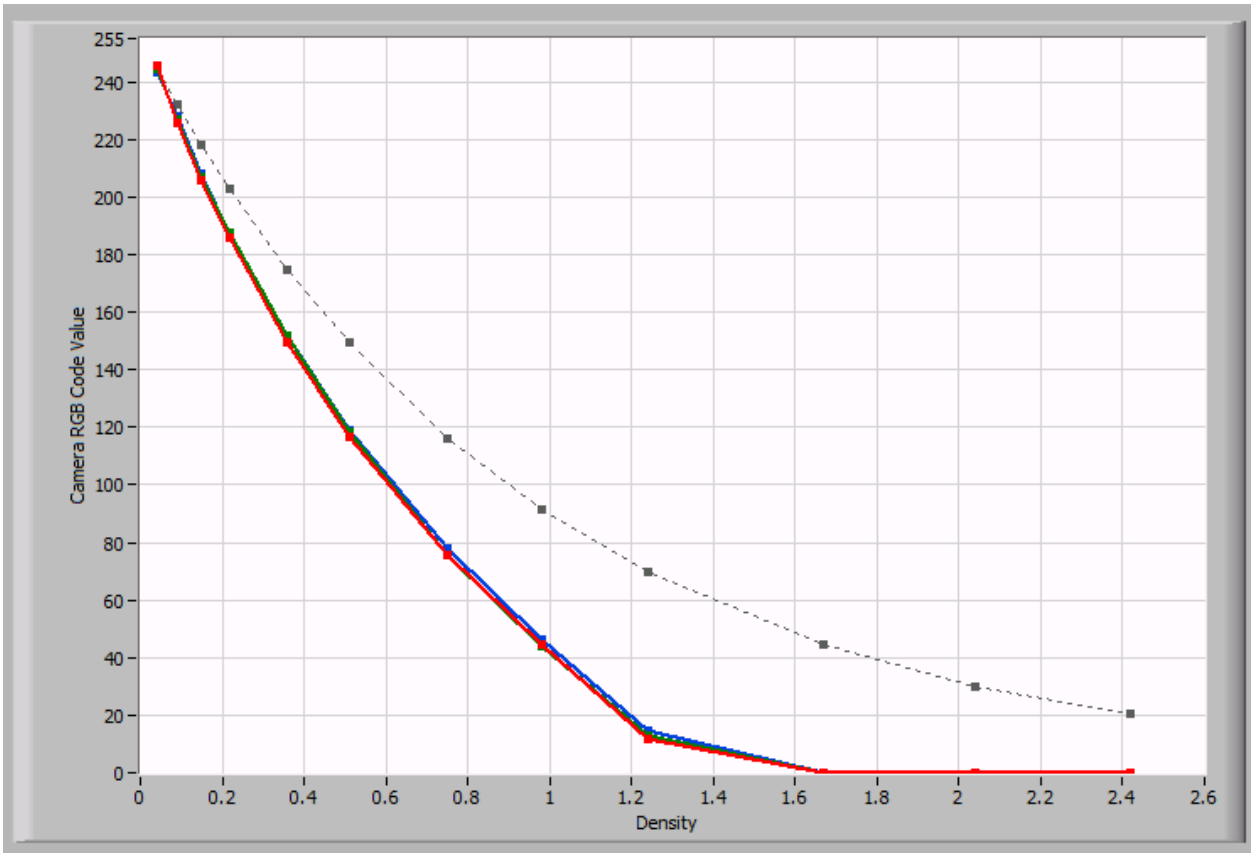


Figure 18 - OECF Curves of the current workflow settings

The OECF curve of the current workflow settings show a slightly greater deviation in the color channels from that of the "RAW" OECF curve, but not enough to cause great concern. The most significant issue is the flattening of the curve in the lower right side of the image. This indicates clipping of the dark areas of the image where these regions would appear as one uniform black color.

Tonescale Delta from Aim

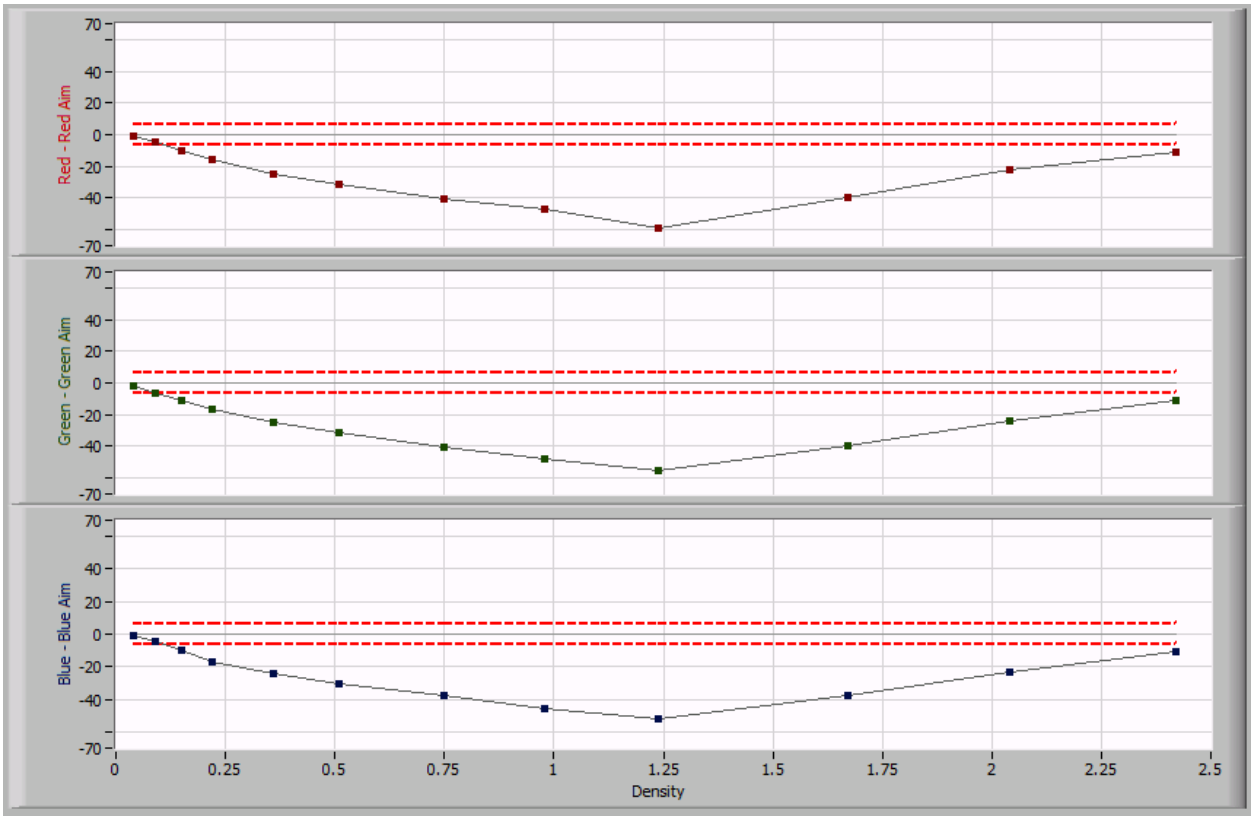


Figure 19 - Delta from Aim for current workflow settings

The red lines show the limits for a three star rating as specified by the *Guidelines*. The values here are much worse than the "RAW" image analyzed in Figure 15, and fall far outside of these limits and overall would rate nothing better than a one star according to the *Guidelines*.

White Balance / Gray balance/ Neutrality

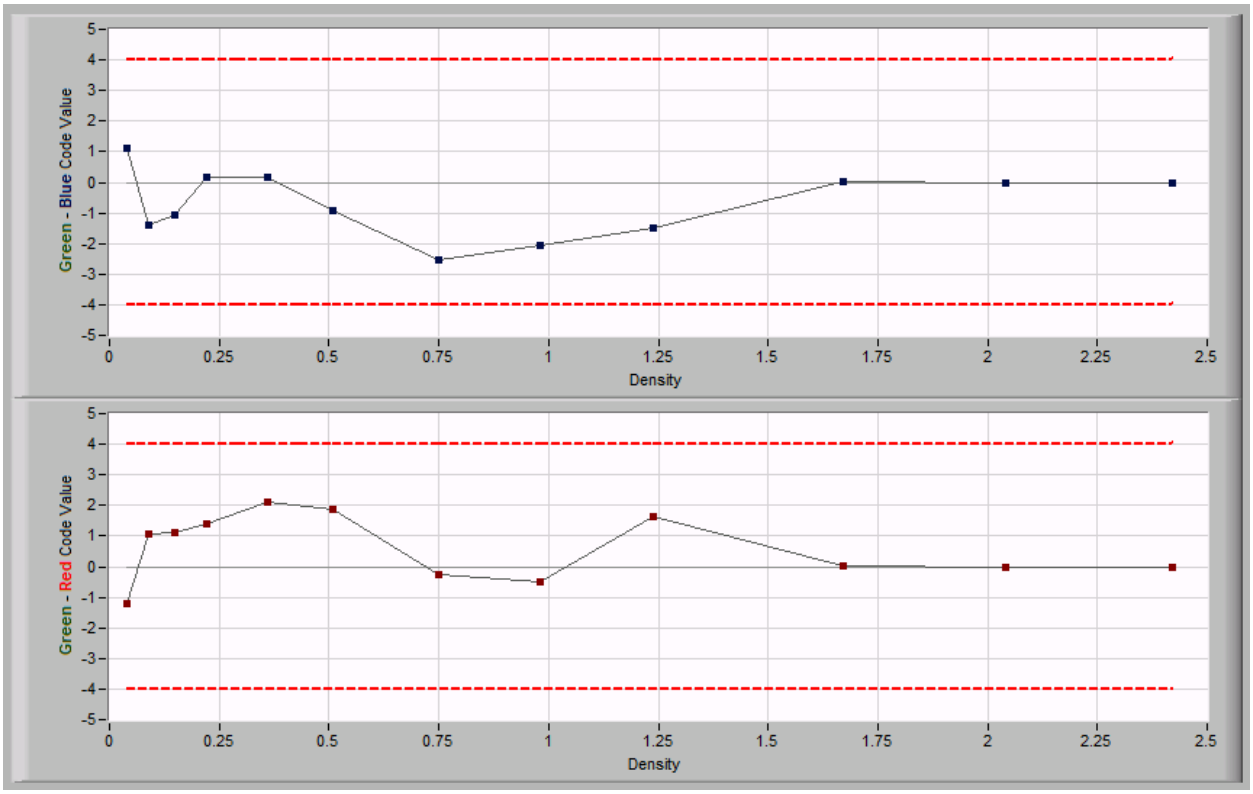


Figure 20- Color Neutrality results of the current workflow settings

In an image with a perfect White/Gray balance, the lines would be close to horizontal. As with the "RAW" output, while not horizontal, the lines do fall within a "three star" rating as described in the *Guidelines*.

Recommended Workflow Settings

- Scanner: Epson 1680
- Document Type: reflective
- Auto Exposure Type: Photo
- Image Type: 24 bit color
- Scanning Quality: Best
- Resolution: 300 dpi
- Target Size: Original (100%)
- Color Correction: None
- Sharpening: None

All image processing performed in Adobe Photoshop CS.

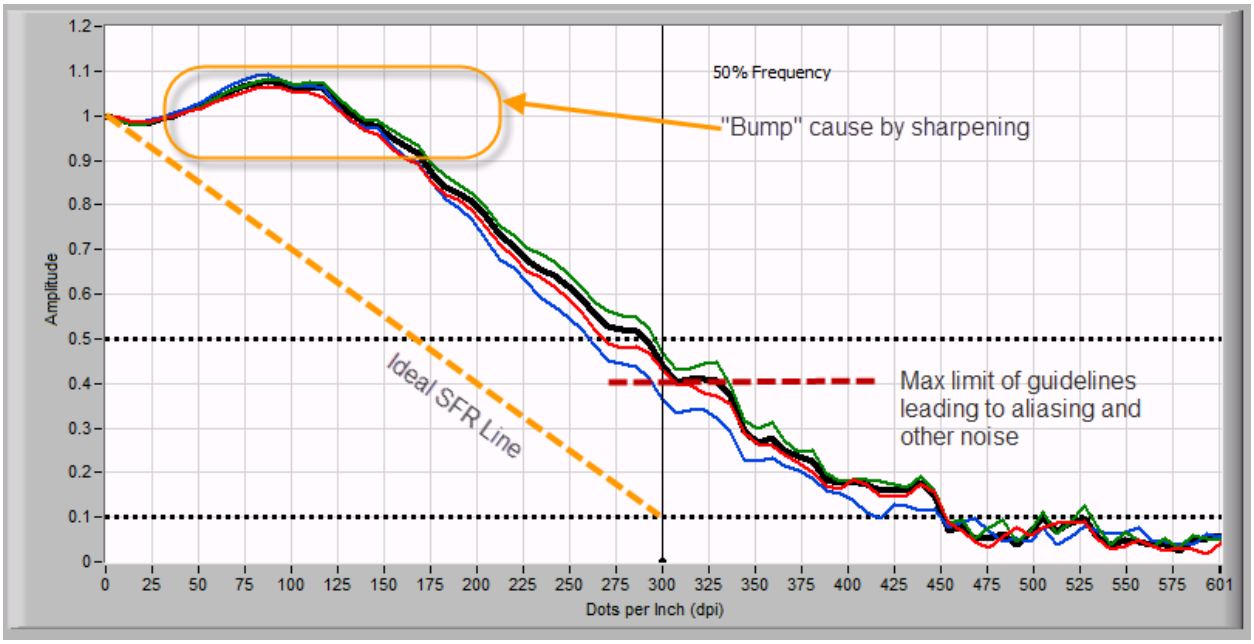


Figure 21 - SFR Curves for Photoshop Corrected Workflow

The SFR curves produced by the image created by post processing in Photoshop exhibits better performance that that of the processed by the software included with the scanner (see Figure 17 for comparison).

The sharpening "bump" still falls within the four star rating of the *Guidelines*. More importantly, the separation between the color channels is far less than that of the current workflow. Additionally, the curves cross the solid vertical line closer to the maximum deviation recommended by the *Guidelines*.

Tone and Color Response

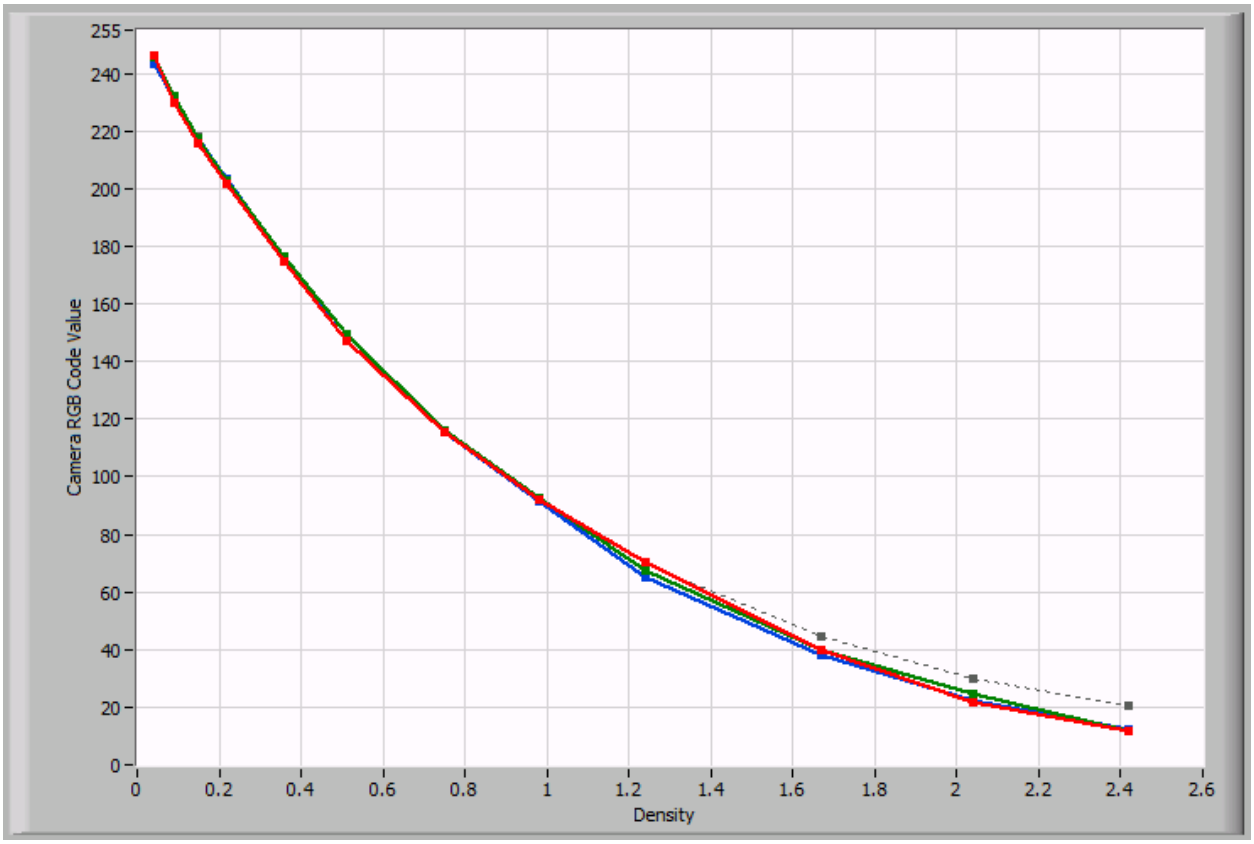


Figure 22 - OECF Curve for Photoshop Corrected Workflow

The Photoshop corrected OECF Curve shows a fairly good response. The three Red, Green, and Blue channels align on top of each other fairly well, though there is a slight mis-alignment in the area around 1.25 density patch. The lower density values curve is shifted a slightly below the 2.2 gamma ideal curve, indicating that the image might be a little dark in the shadows, but not terribly so. It is very probable that these could be fixed with minor adjustments in the corrections made by Photoshop.

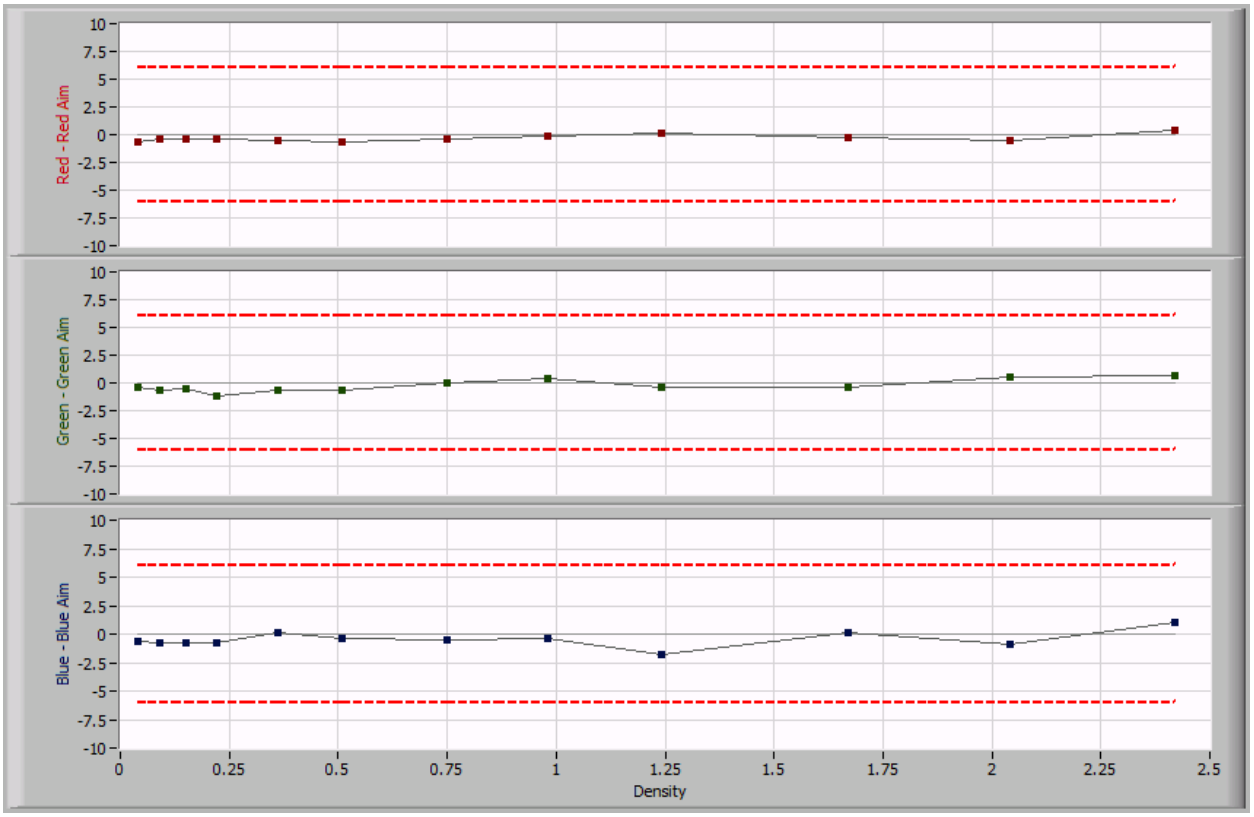


Figure 23 - Delta from Aim for Photoshop Corrected Workflow

The Delta from Aim show excellent alignment and fall very close to the target values. The red lines indicated a *Guidelines* rating of three stars, but the values shown fall well within values the four star rating.

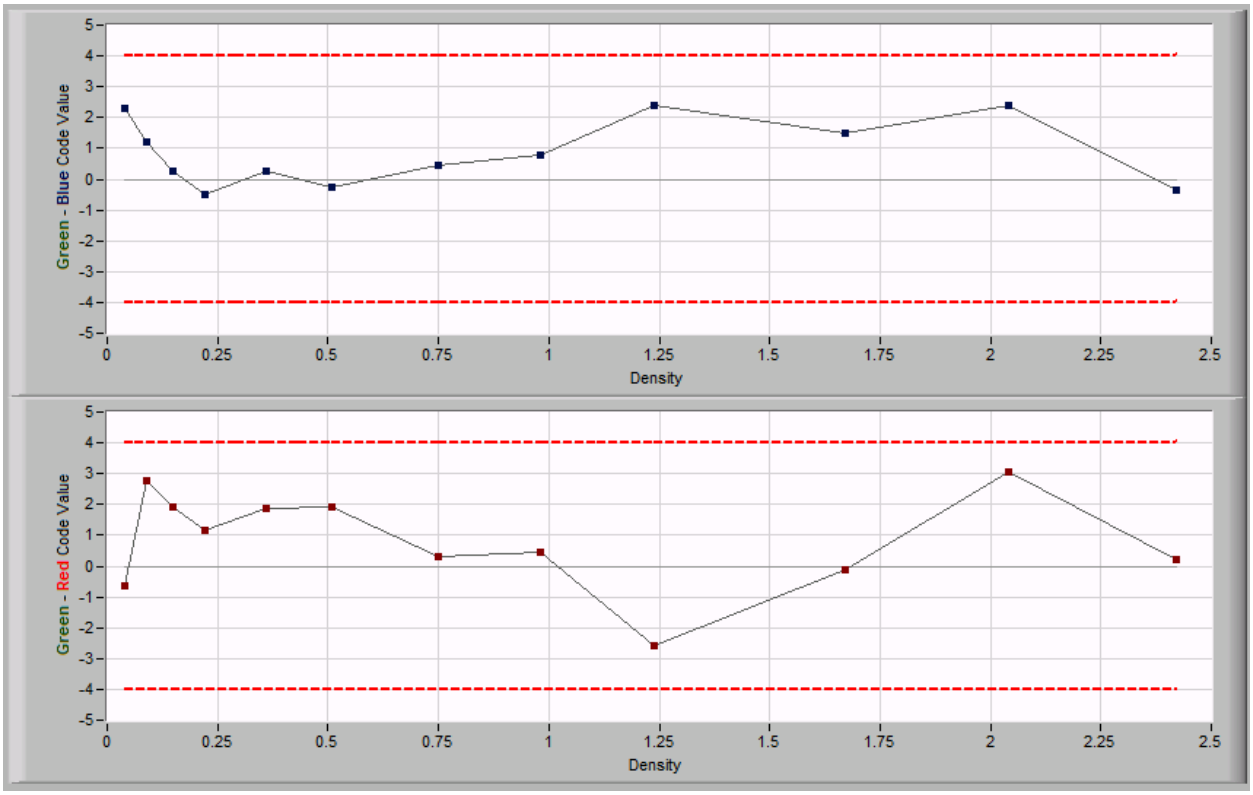
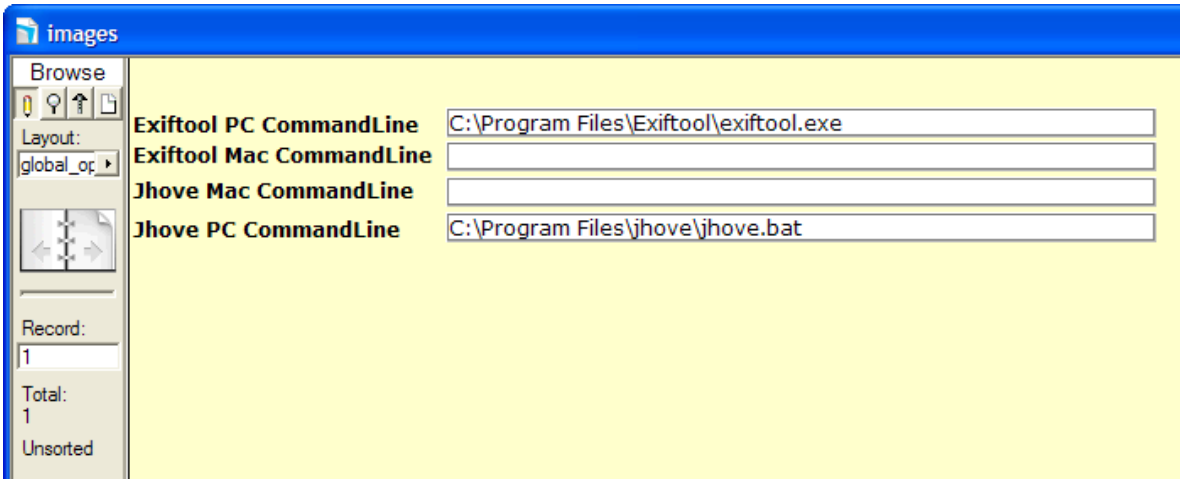


Figure 24 - Color Neutrality for Photoshop Corrected Workflow

As noted previously, an image with a perfect White/Gray balance, the lines would be close to horizontal. While not horizontal, the values fall within a "three star" rating as described in the *Guidelines*.

Appendix E – JHOVE and ExifTool installation instructions

The new FileMaker database for recording metadata has the capability to use JHOVE and ExifTool for metadata extraction. In order to operate properly, these tools must be installed in the same paths on all machines that will use them. The paths to the executables are then set using the global_options layout in FileMaker.



Recommended locations for installed are shown in the image, but are not critical, as long as it is consistent across workstations.

JHOVE Configuration

The FileMaker database was designed to work with JHOVE 1.6. Basic instructions for installing JHOVE can be found at http://hul.harvard.edu/jhove/using.html

For proper operation, the `<jhoveHome>` element in the configuration file, `jhove/conf/jhove.conf`, must be edited to point to the absolute pathname of the JHOVE installation, or home, directory and the temporary directory (in which temporary files are created):

```
<jhoveHome>C:\Program Files\jhove\conf</jhoveHome>
<tempDirectory>C:\temp</tempDirectory>
```

The `jhove.bat` file should be edited to read similar to the following (Lines in **RED** should be edited; lines in **BLUE** should be added):

```
REM Change JHOVE_HOME if location of JHOVE is different from the
following

SET JHOVE_HOME="C:\Program Files\jhove"
SET JHOVE_CONF=%JHOVE_HOME%\conf\jhove.conf

REM Change JAVA_HOME to your installed Java directory
SET JAVA_HOME="C:\Program Files\Java\jre6"
SET JAVA=%JAVA_HOME%\bin\java.exe

SET EXTRA_JARS=

REM NOTE: Nothing below this line should be edited
REM ############################################################


SET CP=%JHOVE_HOME%\bin\JhoveApp.jar
IF "%EXTRA_JARS%"=="" GOTO FI
  SET CP=%CP%:%EXTRA_JARS
:FI

REM Retrieve a copy of all command line arguments to pass to the
application

SET ARGS=
:WHILE
IF %1x==x GOTO LOOP
  SET ARGS=%ARGS% %1
  SHIFT
  GOTO WHILE
:LOOP

REM Set the CLASSPATH and invoke the Java loader
%JAVA% -classpath %CP% Jhove -c %JHOVE_CONF% %ARGS%
```

Works Cited

"300ppi? How to Determine your Digitization Project Needs – matpearson.com", n.d. http://matpearson.com/blog/?p=21.

California Digital Library. "CDL Guidelines for Digital Images". Regents of the University of California, January 2011. http://www.cdlib.org/services/dsc/tools/docs/cdl_gdi_v2.pdf.

———. "CDL Guidelines for Digital Objects: Version 2.0". Regents of the University of California, January 2011. http://www.cdlib.org/services/dsc/contribute/docs/GDO.pdf.

Conway, Paul. "Digital Technology Made Simpler". Northeast Document Conservation Center, n.d. http://www.nedcc.org/resources/leaflets/6Reformatting/03DigitalTechnology.php.

———. "SI 675 2011 Week 4 Encoding and Measurement.pptx", n.d.

"Data Dictionary - Technical Metadata for Digital Still Images (ANSI/NISO Z39.87)". NISO Press, 2006.

Federal Agencies Digitization Initiative (FADGI) - Still Image Working Group. "Technical Guidelines for Digitizing Cultural Heritage Materials: Creation of Raster Image Master Files", August 2010. http://www.digitizationguidelines.gov/guidelines/digitize-technical.html.

Fleischhauer, Carl, and Michael Stelmach. "Federal Digitization: Moving to Common Guidelines" presented at the IS&T Archiving 2009 Conference, May 2009. http://digitizationguidelines.gov/guidelines/IS&T_MS_CF.pdf.

Janosky, James S., and Rutherford W. Witthus. "Using JPEG2000 for enhanced preservation and web access of digital archives-A case study." In *IS&T's 2004 Archiving Conference*, 1:145–149, n.d.

"Metadata Encoding And Transmission Standard: Primer And Reference Manual". Digital Library Federation, 2010. http://www.loc.gov/standards/mets/METSPrimerRevised.pdf.

The University of Michigan, University Library Digital Library Production Services, Digital Conversion Unit. "[University of Michigan Digitization Specifications]", n.d. http://www.hathitrust.org/documents/UMDigitizationSpecs20100827.pdf.

Williams, Don, and Peter D. Burns. "Preparing for the Image Literate Decade." In *IS&T Archiving 2009 Proceedings*, 124-127. Arlington, VA: Society for Imaging Science and Technology, 2009. http://www.imagescienceassociates.com/mm5/pubs/ImageLiterateDecade_2009.pdf.