

The Basics for Scanner Setup Station

The information that this document will provide is the standard setup for an archival scanning station. My intention is to explain in a general way how to digitize archival material in a way that will be user-friendly and also meeting industry standards.

Setting up a digitization station will not only keep the archival material safe but will provide a controlled environment for meeting standards. Limiting the usage to trained staff or volunteers will be extremely helpful as far as the quality but also with the archival, cataloging process.

Please review the following and contact me with any questions. I would also be willing to setup a time for a basic training session after the equipment and software is purchased.

EQUIPMENT

1. Computer: [DELL Inspiron 660](#) – (Average \$600.00)

- Reason: comes standard with the software you WILL need for scanning (Adobe, Photoshop, and Microsoft Office).
- You can buy this model rebuilt but DO NOT buy used.

2. Scanning Equipment:

- Scanner for manuscripts, photographs (negative and positive):
 - [EPSON Perfection V500 Photo Scanner](#) – **GOOD** (\$179.00)
 - [EPSON Perfection V600 Photo Scanner](#) – **BETTER** (\$199.00)
 - [EPSON Perfection V700 Photo Scanner](#) - **BEST** (\$599.00)

Note: It is important to buy a relatively good scanner as these EPSON are user friendly and the software is easy to follow.

- You CAN scan slides, film etc... on the above listed scanners, however, the specific scanner the OHS uses for all slides, film and the like are as follows:
 - [NIKON Super Coolscan 5000](#) – (Average **new: \$5,179.00** and **used: \$2,500**)

Note: As you can see this type of equipment is very expensive but for digital quality it is imperative to purchase the right equipment. However, as I stated previously you CAN use

the EPSON scanners as long as you make sure the settings are correct for each scan. I am going to attach our manual for reference.

Overview of Scanning Photos

1. Use the scanner to scan, not to fix

For starters, if you really want a good result, separate this into two processes: (a) scanning the photo, and (b) tweaking the scanned image. Specifically:

- Don't try to make the scanner "improve" the image, by adding sharpen, contrast, color or other filters. Just scan it.
- Plan to use a photo-manipulation software application, like Photoshop, to do the work of tweaking the scanned image.

Yes, this means you need to have an imaging application like **Photoshop**. But you really do need that, if you want good digitals of old photos.

In fact, I receive a lot of scans of interesting old photos, from all over the world, and unfortunately, most of them aren't very good scans, so they are of greatly diminished utility. There are two typical reasons: the person "fixed" the scan, by applying sharpen and contrast filters, or the person scanned at a low resolution.

On the other hand, if you send me a scan as described below, even if it looks washed out or faded to you, I bet I can make it look a lot better in **Photoshop** than if you fiddle with the scanner settings to try to improve it.

2. Plan to scan the photos at **high resolution**, in an **uncompressed** format.

"Resolution" is a matter of "dots per inch," or "dpi." This means, literally, how many dots per inch the scanner will record. More dots per inch = more fine detail is captured. Fewer dots per inch = less detail is captured.

An "uncompressed format" is an image type like **TIFF (.tiff)**, which contains all the data the scanner picked up. A "compressed" image, like **JPG (.jpg)** takes the data from the scanner and reduces it, so the file size is smaller and the picture looks okay, but really there is much less detail.

For images in a web site, the typical resolution is 72 dpi. This is the highest resolution the typical monitor is capable of displaying. So for a web site, a 300 dpi image won't look any better than a 72 dpi image. This is why people make 72 dpi JPG images for web sites.

If you want to look at the image on your computer or print it, higher resolution makes a **HUGE difference**. For on-screen viewing, you can "zoom in" on the photo, if the scan is high-res., and

see fine detail that you can't even see in the original photo with the naked eye, in some cases. Literally, you can pick up the facial features on some person way in the distance, in some of these old silver-nitrate prints.

For printing, 72 dpi simply isn't the standard. It's not nearly good enough to make a decent print. Typical recommendations for printing copies are to scan at 300 dpi for documents, and even higher for photos if you want to make a blow-up.

3. Always scan in color, even with black and white photos

The toning in old photos isn't just nice looking. Those subtle shades of sepia, and the various imperfections on the surface of the print - flecks of red or blue, or whatever other artifacts there may be - carry a lot of useful image data. When you scan in grayscale, you take away all that data. The blue mark becomes a black mark; the red mark is a smudgy gray mark, and the fiber or texture of the paper is flattened into the image.

The result is much less clarity, a flat-looking scan of the photo. And not only are the artifacts flattened into the image, but now you can't fix them in Photoshop, or some other image application, because you can't tell if that gray blot is a red pen mark, or if it's defect on the negative, or if it's really a part of the actual scene.

Always scan in color, even if you plan to make a black-and-white print. You can convert it to grayscale later on, if you really want to; but if you scan in grayscale, that's the end of the story. If you want to see the color detail later on, you'll have to re-scan it.

My recommendations:

For documents: Scan in color, no filters, 300 dpi, as tiff images.

For photos: Scan in color, no filters, 600 dpi to 1200 dpi*, as tiff images.

* 1200 dpi will make a huge, almost unreasonably huge, file -- say, 100 Megs. I use this sparingly, for cases where there is some interesting fine detail I want to bring out. Normally I scan at 600 dpi, or if it's an exceptional photo, 1200 dpi.

- Negatives and transparencies should be scanned on a scanner with a transparency unit
- The Nikon slide scanner should be used for slides and filmstrips

<http://www.subchaser.org/photographing-documents-02>

Overview of indexing scanned images:

1. The main goal is to effectively identify the images, make a catalogue of the identified images and to make sure they are easily located within your museum.

- a. Organizing a numerical system for each series of photographs, collections by name or by location. This is what we will refer to as identifiable number.

Example: The Kerr-McGee Corporation was donated by Kerr-McGee and given the accession number **23138**.

1. Collection Heading: **23138**.
2. Subject (if needed): **1**.
3. Image Number: **001**



This image is now named: **23138.1.001** (it is the first negative scanned in subject#1 which we have decided is **Advertisements**).

2. Open a new Excel file worksheet and list this as **23138.1.001** and describe the picture in the next cell (Kerr-McGee Employee with Blue Velvet Ride oil can on June 6, 1952). The date is not always listed on the photo or negative but if it is then please list for ease of locating the image later. If ANYTHING is written or typed use this cell to enter the information.
3. Also, when we embed the information into the digital image we will use the **Photoshop** software. The **Photoshop** software will provide a proper format for completing this step.
 - a. Open **Photoshop: Adobe Bridge**
 - b. Open the image you are working on: **23138.1.001**
 - c. On the top right hand bar will be **METADATA** click this tab
 - d. After clicking **METADATA** a new pane will open on the far left with a highlighted tab titled **METADATA**
 - e. Fill in everything we know about this image in the list focusing on the File Properties and IPTC Core. More information on **METADATA** is provided in the

embedded PDF documents in the last few pages of this document. There will be a lot of headings to fill in but for the future all will be necessary to find the image and to know exactly what the image is with all the information following the image forever.

4. Now that the image has been scanned, numbered, indexed and the metadata has been set, make sure we have housed the image in the proper manner.
 - a. Acid free envelopes
 - b. Penciled number, description etc... on the envelope and boxed in the proper archival material. All of this information is attached to this document as well written by the Photograph Archivist at the Oklahoma Historical Society.
 - c. Generally, the written portion is not all of the metadata that we listed in Bridge but is a good summary of what is inside of the envelope.

Overview of digital storage:

1. External hard drive:

- [LaCie Network Space 2](#) – GOOD (\$179.99)
- [LaCie Network Space MAX](#) – BETTER (\$499.99)
- [LaCie 5big Network 2](#) – BEST – 15TB (\$1,499.00)

DOUBLE CLICK THE IMAGE FOR ENTIRE PDF: (document by Rachel Mosman)

DOUBLE CLICK THE IMAGE FOR ENTIRE PDF: (document by LOC)

NARA Guidelines for Digitizing Archival Materials for Electronic Access



National Archives and Records Administration
8601 Adelphi Road, College Park, MD 20740, USA

Date: January 1998

Researched, Written, and Compiled By: Steven Puglia, Preservation and Imaging Specialist,
and Barry Roginski, Computer Specialist

NOTE: The *NARA Guidelines for Digitizing Archival Materials for Electronic Access* presented here have been developed solely for internal NARA use in our pilot Electronic Access Project. We are posting the *Guidelines* on this site in response to requests. We welcome your comments and suggestions. In reviewing this document, please keep in mind the following:

- The *Guidelines* are not intended to be promulgated as a standard for digital imaging. The *Guidelines* cover only the process of digitizing archival materials for on-line access through the Electronic Access Project. Other issues that must be considered in conducting digital imaging projects, including preservation, are not addressed here.
- The *Guidelines* do not constitute, in any way, guidance to Federal agencies on records creation or transfer to the National Archives of the United States.

Digital Imaging Requirements

The goal of the National Archives and Records Administration's Electronic Access Project (EAP) is to produce an on-line catalog that will provide information about NARA holdings and a core collection of digital copies of selected high-interest documents.

These digitizing specifications have been developed to provide a method of evaluating the quality of images produced, to estimate the data storage for access files (on-line) and master files (off-line), and to assist in determining upgrades to the NARA network infrastructure. Inherent differences in document types dictate different approaches to scanning. Due to inconsistent legibility throughout NARA's holdings, due to the desire to image documents in a consistent manner, and due to the limited number of records to be digitized (approximately 120,000 to 200,000 in the initial phase), all scanning will be 8-bit grayscale or 24-bit color.

The tone reproduction approach for digitizing images has been geared toward providing good monitor representation, based on an "average" computer monitor determined by the visual evaluation of a selection and variety of monitors. The tone reproduction is to be done consistently for all images to facilitate the batch processing of derivative files or future adjustments to the distribution of the tones.

Digital Standards

Microfilm

- Images should be scanned in grayscale at the proper reduction ratio
- All images should be scanned at a minimum of 300dpi
- Project master images should be outputted in JPEG2000 or TIFF
 - JPEG2000 must be set to virtually lossless (ie, quality: 100)
 - Patron order images may be outputted in JPEG, out of necessity
- Two-up pages should always be split
- Images should be de-skewed if they are beyond an angle of 3 degrees

Photographs

- Images should be scanned at a **minimum** of 3000 pixels on the longest side of the image
 - 8x10 = 300dpi
 - 5x7: 428dpi
 - 4x5, 5x5: 600dpi
 - 4x4: 750dpi
 - 3x3: 1000dpi
 - 2x2: 1500dpi
 - 1x1, slides, filmstrip: 3000-4000dpi
- TIFF is the only acceptable master format for photographs
- Grayscale images may be scanned as 8-bit grayscale TIFF
- Color images should be scanned as 24-bit Color TIFF
- If there is informational info on the reverse of a print, it should be scanned at 200dpi grayscale
 - Certain reverse side designs deemed to be of intrinsic value may be scanned at the normal standards

- Negatives and transparencies should be scanned on a scanner with a transparency unit
- The Nikon slide scanner should be used for slides and filmstrips

Audio

- Audio should be transferred using the best available equipment
 - Never use an aux-in on a sound card
- Tapes should be cleaned and play-wound before transfer. Phono records should be cleaned using an appropriate disc cleaner
- Audio should be saved as .wav or broadcast wave (.bwav)
 - Access copies may be in mp3
 - Mp3 standard should be 256kbit/44.1khz
 - Id3 tags may be used in mp3, but should not be considered preservation metadata
- Wav files should be digitized as 96khz/24bit
- Editing and noise reduction are not appropriate for preservation masters, but may be used on access copies
- Metadata may accompany audio files in xml, or may be embedded into broadcast wave files
- A note on the distinction between spoken word and music:
 - If space is a major constraint, spoken word may be digitized at 48khz/24bit.
 - Music should be 96/24

Manuscripts and Maps

- Manuscripts and Maps should be digitized at a minimum of 300dpi
- All manuscripts and maps should be scanned in 24-bit color
- TIFF should always be used for maps
- TIFF is the ideal master for manuscripts
 - PDF/A may be used as well

- PDF and JPEG are acceptable for access copies
- May be accompanied by xml metadata

Film and Video

- DVD is generally not an acceptable master format
 - An exception may be made for oral history video. The purpose of this is to ensure that the tape master is not used as an access copy.
 - All oral history video **interviews** should have audio captured at a minimum of 48khz/24-bit, wav or bwav
 - High-value oral histories should be digitized to a file-container video format (see below)
 - High-quality DVDs should be used to make a DVD “master.” Low quality DVD may be used for patron access copies. Do not use the DVD master for access.
 - DVD may be used for other video projects, but should be avoided if possible.
- Video should be digitized as .avi
 - Uncompressed is ideal, if possible
 - Otherwise, compressed at 100 quality
 - 48khz audio for all video files
 - For standard definition: NTSC DV Codec
 - For high definition: NTSC HDV Codec
- Video that comes in as a compressed format other than .avi may be maintained as such. They should not be compressed further
- Eventually video should be migrated to an uncompressed format such as MXF with JPEG2000 Codec
- Accompanying xml metadata is ideal, to eventually be embedded in a future preservation master

Considerations for Selection for Digitization

- Should we do it?
 - ✓ Why do you want to digitize?
 - ✓ Does it support your mission?
 - ✓ What do you want to digitize?
 - ✓ Who is your audience?

- Do we have the permissions to digitize?
 - ✓ Who owns the material?
 - ✓ What use has been negotiated in the past?

What can be negotiated now and in the future?

- Do we have the resources, relationships, and knowledge to digitize properly?
 - ✓ Where will the digitization be done?
 - ✓ Who will do the digitization work?
 - ✓ What equipment will be needed?
 - ✓ How will it be supported financially?

- How will you care for the digital objects?
 - ✓ What will the objects be stored on?
 - ✓ How big will the collection be?
 - ✓ Who will maintain the collection?
 - ✓ Where will the funding come from?
 - ✓ Where will the files be stored?
 - ✓ How long are the objects to be kept?

The answer to ALL of the arrowed questions should, ideally, be “YES” before you proceed...



All dots per inch (dpi) specified for scanning are equivalent dpi for the size of the original documents, unless otherwise specified. All scanning from photographic intermediates or microfilm shall be done at the dpi needed to achieve the cited resolution on the original documents. Due to inconsistent legibility of documents and the desire to image documents in a consistent manner, all scanning will be 8-bit greyscale or 24-bit color. The tone reproduction and aimpoint values are intended to provide good monitor representation of the images on an "average" computer monitor based on the visual evaluation of a selection and variety of monitors. Scanning and quality control workstation monitors shall be set to 24 bits (millions of colors), set to a gamma of 2.2, set to a color temperature of 6500°K, calibrated for color neutrality, and the contrast and brightness adjusted for the aimpoint values.

Scanning Aimpoints- Using the Kodak greyscale, the tonal controls of the scanner shall be adjusted to place step #19 at RGB levels of 8, 8, and 8 or at % black of 97%, to place step "M" at RGB levels of 105, 105, and 105 or at % black of 59%, and to place step "A" at RGB levels of 247, 247, and 247 or at % black of 3%. Using the Kodak color bars, the tonal controls of the scanner shall be adjusted to place single color black patch at RGB levels of 19, 19, and 19 or at % black of 93%, to place the grey background at RGB levels of 101, 101, and 101 or % black of 60%, and to place the white patch at RGB levels of 241, 241, and 241 or % black of 5%. Variability: no more than ± 3 RGB level variance from aimpoints and no more than 3 RGB level difference in the individual channels within a patch for RGB scanning and no more than ± 1% level variance from the aimpoints in % black for greyscale scanning.

TYPES OF DOCUMENTS

Table with 5 columns: FILES, Textual, Photographs, Maps and Drawings, Graphic. Rows include MASTER FILES, ACCESS FILES, and THUMBNAIL FILES, detailing scan resolution, file format, and resolution for each document type.