

CHAPTER 4 DIGITAL PRODUCTION

Digital production is probably one of the easiest functions of creating a digital project --- and the most fun! It is exciting to see long-stored items, fragile materials, and negatives come to life on a screen. While scanning or photographing is easy, it can be deceptively so. And if you are looking for a high quality product, it can get complicated. Once the initial thrilled wears off, the processes in production become repetitive rapidly and even boring, causing mistakes to be made. Aside from the physical handling of original materials, digital production is time-consuming and redundant work, and therefore the imaging should be done correctly the first time. It is very difficult and expensive to go back and scan or photograph again or to recover documentation that did not accompany the original production image.

To this end, NC ECHO follows the "scan once methodology" (this covers scanning, digital photography, and digital audio). In basic terms, this means that in creating digital objects, the digital production should also be done at the highest level of quality that an institution can afford. The higher the quality, the longer the life of the object and the more versatile its uses will be.

As you plan for digital production and are determining the level of quality your institution can support, consider the future uses of the digital images. Do not anticipate returning to re-digitize. Many originals could suffer from the handling and exposure to the bright light required by digitization. For instance, it is reported that scanning exposes a document to four times the destructive light as one photocopy. Therefore it is best to simply "scan once," to create a master image, and make any future duplicates from that master image.

This chapter of the *Guidelines* specifies the equipment, standards, and techniques required to conduct the digital production portion of a digital project. It discusses digital production stations, the "scan once methodology," purchasing decisions for hardware and software, imaging and audio standards, basic steps to follow, documentation, and quality assurance.

Setting up a Digital Production Station

One of the major considerations in beginning a digitization project is to make best use of workspace. While some considerations listed here may not be practicable for your institution, we provide the best possible scenario and reasoning so that you can make the appropriate decisions according to your physical configurations. In an ideal world, institutions would have the resources to have designated digital production station that will not only be perfect for scanning or digital imaging with a camera but also not deprive you of space already devoted to other activities. This ideal is rare, so institutions are often faced with multi-purposing space or reconfigurations that put additional strains on an already limited resource.

In order to make the best use of existing space, begin with the following questions:

1. What will the space be used for?

This can include not only the physical work of producing digital images, but the creation of metadata and the preservation of originals. Will the space be someone's office? Can other staff members use the station where it is located? What other work will have to be accommodated besides digital production?

2. What type of materials will be captured?

Having a clear vision of the formats of the original material intended for digital capture is important in terms of space because those formats may have space requirements themselves or their capture may have physical implications. Large flat surfaces in the production area are a must, as is the ability to secure the digitization area or a portion of it.

3. What sort of equipment will be used?

It is important to recognize that different capture methods require variations in the physical configuration of your workspace. Scanners require table-top locations; digital cameras require space for tripods and other set-up requirements similar to a photography studio to optimize the capture.

4. How many staff will be working at one time?

The number of people working in one space simultaneously effectively will be determined by the configuration of the workspace. It is important to consider the maximum capacity for a workspace prior to establishing schedules for your digital production staff and catalogers.

Once you have considered the above questions and surveyed your existing physical facilities, consider the following as "best practice" for digital production workspace configurations:

- **Be healthy, be safe:** This includes not only the physical surroundings but the workstations, chairs, and equipment selected for digitization. Also be aware that such things as cords stretched across well-traveled areas or overloaded on outlets present safety risks and are violations of safety regulations. Scanning and metadata capture also require long periods of sitting so an ergonomic chair and work table are essential.
- **Lighting the area:** Digitization of images requires a great deal of viewing and visual discrimination and manipulation. Lighting should be standardized so that visual judgments are made consistently. Techniques to help with this include painting the room a neutral color, eliminating extraneous lighting from outside the work area, eliminating overhead lighting to decrease glare but using enough light to assure that eye-strain does not become a health factor, and a background desktop color that is mid-gray to balance your images across the monitor screen. You will also want to use the same lighting when you calibrate the equipment as you will have while scanning or photographing.
- **Contentment of personnel:** The people working on the project form the core of it. Check in with them to make sure that they are comfortable and happy. Encourage regular breaks from the workstation, comradery, and teamwork. Ease boredom through radio or music. Consider the use of headphones if conflicts arising personal taste become an issue. Set achievable goals and be sure that personnel feel proud of

those achievements. Hold regular team meetings to ensure open communication and ask questions about their comfort. Pay attention to any ergonomic complaints they might have. Above all, accommodate their individual work patterns as much as possible while maintaining cohesive project team.

- **Stability:** Try to maintain a stable environment for project team members as much as possible. Not only will this improve the digital capture of your materials, but will provide less apprehension or chaos among project staff.

Establishing an environment that is functional and pleasant will make the experience of a digital project all the more rewarding and will increase your efficiency in producing the digital images

The “Scan Once Methodology”

It is expensive for institutions to go back and re-digitize their holdings. Few ever do so. In addition, many originals could suffer from the handling and exposure to bright light required by digitization. Therefore, it is best to simply “scan once,” create a master image, and make any future duplicates from it.

Step One – Create a Master Image

The highest quality copy of a digital image, often called the master image, is expected to be a quality surrogate of the original. As such, it should represent the *un-manipulated* original and be created at a high resolution and stored in an uncompressed format (usually TIFF). High resolution equals large amounts of information captured, and large amounts of information captured usually equal a higher quality digital image. The higher the quality, the longer the life of the digital copy and the more versatile its uses. It is the master image that holds the promise of versatility and longevity. From it, high quality prints or publications might be made as well as derivatives for a variety of uses.

Step Two – Create an Access Image

Access images are lower resolution copies taken from the master by using a “save as” function and changing the storage format and resolution. Access images may be of varying quality and are generally manipulated for better display upon the screen or page (cropping, re-sizing, etc.) Additional images, such as “thumbnails” (even lower resolution copies) may also be created from the master or access image. These thumbnails allow for even quicker downloads of pages, and faster retrieval of large numbers of images. Suggested resolutions, bit depths, and storage formats of each of these types of digital reproductions (master, access, and thumbnail) are outlined below. Images created from the master are often referred to as derivative images.

Step Three – Storing the Master Image

The master image is the copy to be maintained for the long-term. As such, it should be stored appropriately. Master images take up a great deal of space, and most institutions will not wish to store them for the long-term on computer hard-drives. Some institutions maintaining large amounts of digital images will wish to work with a form of tape or server backup, while those institutions engaged in more modest digital products may choose to store master images on CDs. If an institution decides

to use CDs as a storage medium, it is suggested that two copies of each CD be prepared and stored separately. One will serve as the "master" CD and the other will be the "use" CD from which access images, copies for users, etc. may be prepared. CDs used in this way should be "refreshed" regularly, that is copied from the old CD to a new CD (approximately every 5 years).

Getting the Equipment

Selection of the necessary equipment can have the greatest impact on the quality of images for a digital project. The development of scanning and digital camera technology has led to a proliferation of equipment varying in quality and availability. This section provides the necessary information to make an effective decision for your institution.

Before any equipment is purchased, consider the following overall questions:

1. What can your staff and your physical environment accommodate?
2. What can your current technology support?
3. What type of material are you digitizing (photos, documentation, art images, artifacts, etc.)?
4. What financial restrictions do you have?
5. How will you provide storage for you project?

Hardware: Digital Capture

There are basically six types of digital production devices.

Flatbed scanner - The most commonly used type, it accepts a broad range of formats and varies in quality and price. Flatbed scanners are typically modeled for a scan area of 8" x 11," but larger flatbed scanners are available. They can be purchased with transparency adapters which handle negatives and slides very easily. High end scanners have less problem with "flare" and now come with front side USB and fire wire connectors which are much easier to use, especially with digital cameras.

Sheet-fed scanner - Similar to the flat-bed scanner. It is used for batch work and should never be employed with originals because of potential jamming which could damage or destroy the originals.

Drum scanner – The drum scanner produces high quality images but is quite expensive. Because materials are affixed to a rotating drum, they are not recommended for cultural heritage materials but are suitable for surrogate negatives and transparencies. There are now drum scanners that are sometimes called roll scanners and instead of a rotating drum, they utilize a conveyor belt arrangement which is less damaging to the original materials. Again, they are quite expensive.

Reprographic stand scanner – Also known as an overhead scanner, these scanners are quite expensive but allow for digitization of books and oversized materials with minimal damage to the original. The reprographic stand scanner has the a camera mounted over the scanning area, decreasing the amount of pressure placed on a book spine or allowing for a large scanning area.

Digital camera - Good for 3-dimensional objects, digital cameras vary widely in quality and price. It also has a problem with "flare," or bright patches on the images.

Lens are geared toward the capture of 3-dimensional scenes and may introduce distortions to flat materials. If a digital camera is necessary, it works best in a controlled, studio-type environment.

Film scanner – Specifically designed to digitize transparent materials such as 35 mm film, the film scanner is particularly good for roll film, but less productive for slides. It too has a problem with "flare."

Pros and Cons of Digital Capture Devices

Scanner type	Pros	Cons
Flatbed scanner	<ul style="list-style-type: none"> • Highly addressable • Inexpensive • Many units can handle both transmission and reflection materials • Flexible software drivers • Most good up to 600 dpi of real resolution. • Low learning curve 	<ul style="list-style-type: none"> • Low productivity, frequent document handling • Tendency toward streaking and color misregistration • Prone to inflated marketing claims
Sheet-fed scanner	<ul style="list-style-type: none"> • High productivity • As good as or better than flatbed scanners • Many automatic features 	<ul style="list-style-type: none"> • Unsuitable for fragile, bound, wrinkled, 3-D, or inflexible objects • More expensive than flatbed scanners • May not handle all sizes of documents
Drum scanner	<ul style="list-style-type: none"> • Very high image quality • High resolution • Low noise • High dynamic range • Good tone/color fidelity • Few artifacts • Very flexible software drivers • Variable sampling rate 	<ul style="list-style-type: none"> • Expensive • Low productivity • Frequent handling • High operator skill level • Handles limited document types; must be mountable on drum
Reprographic stand scanner	<ul style="list-style-type: none"> • Very high image quality • High resolution • Low noise • High dynamic range • Good tone/color fidelity • Few artifacts • Flexible software drivers 	<ul style="list-style-type: none"> • Expensive • High operator skill level • Frequent document handling, although minimized impact on document.

<p>Camera</p>	<ul style="list-style-type: none"> • Can handle a variety of document/ object types (3-D, bound, glass plates, non-flat, oversized) • Unlimited field size • User-controlled lighting. • Rapid capture for area arrays • Non-contact capture • May have interchangeable lenses • Generally good image quality 	<ul style="list-style-type: none"> - Good models expensive. • Limited sensor size • Low productivity for linear array types • Nonuniformity artifacts common • Area array devices prone to low dynamic range due to flare • Moderate skill level required
<p>Film scanner</p>	<ul style="list-style-type: none"> • Highly productive for roll film • Low flare/ good dynamic range for linear arrays 	<ul style="list-style-type: none"> • Low productivity for sheet film or slides • Potential for high flare in area-array devices • Dust/scratch artifacts common • Image quality characterization difficult due to lack of targets

For detailed information about selecting the appropriate scanner, see Don Williams, "Selecting a Scanner," *Guides to Quality in Visual Resource Imaging*, available at: <http://www.rlg.org/legacy/visguides/visguide2.html>. The above table was adapted from his documentation.

Hardware: Computers

Select the computer that will be used in the digital production. It is recommended to devote one computer to this and below are outlined some guidelines on the best selection for this.

- has as much Random Access Memory (RAM) as possible (at least 512 mb). More memory allows the computer to process large amounts of image data more quickly.
- has a processor that is optimized for image manipulation.
- supports high-speed data input through serial connections USB 2.0, or IEEE 1394 "Firewire."
- has an ISO 9660 compliant CD-RW burner to create archival storage CD-ROMs of your digital images.

If you are going to be purchasing a new computer to act as your digitization station, it is recommended that you review trade publications such as *PC Magazine* to help make an informed decision. In making these decisions, it is recommended that you involve your technology support as much as possible. Not only can technology personnel provide help in making decisions, but they will be better able to perpetuate their support throughout your digitization project. Digital camera reviews can be found at <http://www.dpreview.com/>.

Hardware: Purchasing

In purchasing hardware, consider these issues: what are the resolution capabilities? Is the scan bed large enough to handle your originals? How long does it take to scan one image at your master image specifications? Does the manufacturer have a good reputation for service and durability?

Optics quality is important. Manufacturers' claims sometime may be unreliable, especially relating to the number of pages scanned per minute and the maximum possible resolutions. Look for reviews, ask those using the equipment, and play close attention to *actual* rather than *interpolated* resolution. A scanner's speed is directly related to the associated computer's capabilities. The higher or faster the RAM, Hard disk space and CPU speed, the better.

Software

Some kind of software usually accompanies the digital production device. For a scanner, this is the scanning software and for a digital camera, this is the software that provides the interface to download images from the camera to the computer. A second kind of software is used to manipulate the scanned image. This is image manipulation software. It may come with the scanner, but it will usually allow for only the very basic editing of an image. Remember, you get what you pay for. Manipulation software is mounted on the hard drive of a computer and is used to orient the image; crop it; adjust brightness, contrast, and resolution; transform; flip; or otherwise manipulate the image.

The *de facto* standard for image manipulation is the software package, Adobe PhotoShop. It can import the scanning software so that you are able to scan and manipulate the image within the PhotoShop umbrella application. There are several versions of PhotoShop, ranging from PhotoShop Elements (about \$40.00) to PhotoShop Creative Suite Premium (about \$1,200). Other imaging software is adequate for basic tasks (Paint Shop Pro, Deskscan II, etc.). It is recommended that you look for software that allows you some flexibility for advanced manipulation and saves the image in all the common formats (i.e., TIFF, JPEG, GIF). It is also recommended that the software allows conversion from one format to another. If the project will require the processing of a large volume of images, it is best to consider additional software that allows batch processing such as PhotoShop, Debabelizer or ImageMagic that will enable the automatic processing of files and the standardization of compression.

When selecting image manipulation software, institutions should look for

- ✓ Ability to work directly with scanner software through TWAIN or other plug-ins
- ✓ Support for a wide variety of file formats
- ✓ Tools for controllable image optimization (i.e, color adjustment or color spaces)
- ✓ Usable documentation and reliable technical support
- ✓ Extensibility
- ✓ Ability to create macros for frequently applied functions
- ✓ Batch processing

Software: Purchasing

Manipulation software. How versatile is the software? What storage formats does it support? What are the options for manipulating the image? Can you turn off some of the options or does the software force you to

"improve" the image?

Scan software. What are its resolution capabilities? What are the save file options? Can you set the default? Does it allow you to change the default settings or must you change them each time you scan or state a scanning session?

Purchasing Equipment

The main factors to consider in purchase:

Cost

Scanners and digital cameras can range anywhere from \$100.00 (or less) to thousands of dollars. Generally you get what you pay for. Scanners in the mid-range of several hundred dollars are likely to be adequate for most scanning projects. Look carefully at warranties, maintenance reputation, reliability, good documentation, flexibility of the scanning platform and non-proprietary interface cards.

Installation

Installing the scanner should be very direct. With only a few exceptions, the scanner is a plug and play peripheral. Be very careful to purchase a scanner that does not require a proprietary interface card, as this card may create incompatibility in other computer functions. A USB interface has become the standard (although SCSI2 is still better and fire wire connectors are almost as popular as they are faster). SCSI 2 allows attachment of other devices to the computer with few complications (tape-drives, Zip-drives, CD Rom drives, etc.) but also requires a special hardware card. The other devices may be required for storage and for transport of large files as the institution's digital collections grow. Installation does not impact digital cameras, although you will want to be assured that accompanying hardware will work on your computer platform.

Destination of the image

Web? File? Print? If the use will be for Web images alone, an inexpensive capture device may suffice. If archiving or migration is of concern, aim for the higher-end machines. Since it is recommended to "scan once," most institutions, no matter the size will want to factor in both master and access images.

Resolution needed

A 4 x 5 photograph will be fine on a 600 dpi scanner. A 1 x 2 contact print will need a higher resolution, more in the range of 1200 dpi, and will require a more expensive scanner.

Number of items to be scanned

If you plan to process large collections, the 30 seconds or more needed to scan one image can add up to an enormous drain on resources. Consider buying a faster scanner or buy two scanners (this won't help if your staff is small!). A "single pass" scanner is the faster scanner but may not capture all the information.

Format of items to be digitized

Slides, photographs, color, grayscale, half-tone print, graphics, text, three-dimensional objects, etc. will all need to be treated differently for best results. Can the scanner handle a variety of formats? If there are three-dimensional objects or large oversized flat materials to be digitized, a digital camera will need to be purchased. Slides and film require more sophisticated scanners, and the purchase price will be higher if a stand-alone system is

purchased.

Additional Tools

Some tools will come with the scanner. These often include masks for transparencies and negatives. These are strongly recommended, as a dark surrounding field for transparencies and negatives produces the best scanned image. Compressed air, and/or a soft brush will be useful for photographs and to keep the bed of the scanner free of lint. Tripods and other equipment are necessary for a digital camera to create a stable digitization station. These would be items that would have to be purchased in addition to your camera. And of course, add to this list of tools cotton gloves for those handling originals.

If you are purchasing an expensive capture device, company representatives should demonstrate its capabilities. You should also negotiate a trial period in which you can evaluate the results of digitizing a full range of materials.

What Can be Digitized and How

Below is a table showing major types of materials that can be digitized: the type of file (master, access, and thumbnail) and suggestions for corresponding resolution, storage format, and bit-depth. These suggestions are based upon standards and best practices being followed by some of the nation's major digitization projects. The resolutions, abbreviations, bit types etc. mentioned in this table are discussed later in this section.

FORMAT TYPE	MASTER IMAGE	ACCESS IMAGE	THUMBNAI L IMAGE
TEXT (printed documents)	Scan at 200-300 dpi grayscale. Uncompressed TIFF Intel (IBM) byte order Bit depth 8	8 bit grayscale JPEG, 4-6 on 1/10 scale (medium) File resolution 200 dpi Unaltered image size	Generally not used for text files.
PHOTOGRAPHS	Scan at 4000 pixels on long side OR 600 dpi Uncompressed TIFF Intel (IBM) type order Color scan RGB color 24 bit; Black and white scan 8 bit grayscale	8 bit grayscale, 24 bit color JPEG 8-10 on a 1/10 scale (high) File resolution 300 dpi unaltered image size	4 bit grayscale, 8 bit color JPEG 4-5 on a 1/20 scale (medium) 72 dpi
DOCUMENTS (manuscript materials)	Scan at 4000 pixel on long side OR 600 dpi Uncompressed TIFF Intel (IBM) byte order Color scan RGB color 24 bit Black and white scan 8 bit grayscale	8 bit grayscale, 24 bit color JPEG 8-10 on a 1/10 scale (high) File resolution 300 dpi unaltered image size	4 bit grayscale, 8 bit color JPEG 4-6 on a 1/10 scale (medium) 72 dpi

MAPS, DRAWINGS, BI-TONAL	Scan at 300 dpi Intel (IBM) byte order RGB color, bit depth 24	8 bit grayscale, 24 bit color JPEG 8-10 on a 1/10 scale (high) File resolution 200-300 dpi unaltered image size OR reduced to equivalent of 8 x10"	Optional for bi-tonal maps & drawings 4 bit grayscale, 8 bit color JPEG 4-6 on a 1/10 scale (medium) 72 dpi
OBJECTS	Use a digital camera at 300-600 dpi Uncompressed TIFF, RGB color, bit-depth 24	8 bit grayscale, 24 bit color JPEG 8-10 on a 1/10 scale (high) File resolution 300 dpi Unaltered image size	4 bit grayscale, 8 bit color JPEG 4-5 on a 1/10 scale (medium) 72 dpi

Digital Audio standards

Many digitization projects are interested in including digital audio, whether digitizing analog audio media or creating new digital media. Audio files provide depth and variety to digital projects. Transferring analog audio to a digital media is a relatively simple process. The conversion involves four devices: an analog audio playback device, an analog-to-digital converter, a computer to process the digital signal, and a device for digital file storage. Other devices can include a mixing device. There are several audio software programs available to allow manipulation of the audio, including volume adjustments, tracking, equalization, noise reduction, and compression. For master files, these methods are used sparingly, but for derivative files can help to provide enhanced access to the audio file.

Digital audio files can be recorded in many formats, such as WAV, AIF, and MP3. The most important aspect in selecting a file format is to choose one that is non-proprietary, with a high potential for future readability. Uncompressed formats will provide maximum audio fidelity.

The WAV file was developed by Microsoft and is in widespread use. WAV is readable by virtually all audio software programs. AIF file type was developed by Apple Computer and is also used widely. Both WAV and AIF are uncompressed and accepted for long-term file storage. MP3 file format has emerged as the file type of choice for many applications. This file format is highly compressed for electronic transfer. It is recommended that institutions use WAV for master files but can use MP3 for access files and delivery on the web.

Audio File Storage Requirements

Requirements	Sample Rate	Bit Depth	Pros	Cons
Minimum	44.1 kHz	16-bit	Maximizes storage space	Concerns over migration quality
			Lowest level of processing time	Limits ability to enhance source file for delivery

Recommended	44.1 kHz	24-bit	Accurate reproduction of source material	Requires 50% additional storage space
			Increased dynamic range	Requires additional processing time
			Increased ability to enhance source file for delivery	
			Current professional audio standards	
Optimal	96 kHz	24-bit	Increased frequency range	Dramatic increased storage space and processing time
			Further increased ability for enhanced source file for delivery	May require compression for delivery
			Highest recommended current quality	

The above table and text are based upon the *Digital Audio Best Practices*, version 2.0 published by the CDP Digital Audio Working Group in November 2005 (available at: http://www.cdpheritage.org/digital/audio/documents/CDPDABP_1-2.pdf) For detailed information about audio digitization, please see their guidelines.

Elements of a Digital Object

3 Types of Scan

Scanners generally support three types of scans and present these options to their users.

- **Bi-tonal** - also known as line art or "black and white," it is best for printed text and high contrast graphics. While once a popular type of scan, it is not used as often today.
- **Grayscale** - provides a range of shades of gray in an image and delivers a better quality of scan than black and white, it is best for continuous tone documents and black and white negatives
- **Color** - duplicates the range of possible colors in an image with the higher the range the more accurate the scan at duplication, it is best for photographs and any document with color. Digital cameras produce color only.

File Formats

Digital images are stored in five major types of formats. It is the type of format and level of resolution, which is the difference between the "level" of scans. In order to save space and "move faster" over the Internet, some formats drop information from an image. Later the software analyzes what it did not drop, infers what must have been discarded and partially reconstructs the original image. This process is called "compression." Digitizers of fragile originals do not compress master images but attempt to maintain as much original information as possible.

- **TIFF** (Tagged Image File Format) is a storage format that does not compress the images and thus does not "drop" or lose information from an original digital capture. It is used for master images. TIFF is the preferred file format because it is designed for all platforms and is ubiquitous. Any image editing program produced in the last 10 years can open TIFFs so it will be around for a long time.
- **JPEG** (Joint Photographic Experts Group) is a compressible storage format and does drop some pixel information so that images might be stored in less space and be retrieved faster. It is often used for access or thumbnail images that are presented by the Web.
- **GIF** (Graphic Image File Format) is a compressible storage format and does drop some pixel information so that images might be stored in less space and be retrieved faster. It is sometimes used for access and thumbnail images that are presented by the Web. It is best for images with large areas of one or more colors. It is a proprietary format.
- **PNG** (Portable Network Graphic) is a compressible storage format, but does not drop some of the pixel information. It does 24-bit color but it does not allow saving the metadata as TIFF and JPEG do. It is still relatively new and not supported by all image viewers yet.
- **JPEG 2000**. This format is not related to the regular JPEG. It uses compression algorithms with an option to not lose pixel information and to communicate metadata and structure within the code stream. For more information see, <http://www.jpeg.org/jpeg2000/>

LEVELS OF SCAN	FILE FORMAT	USED FOR	ALTER ?
Master image	TIFF	Long-term storage or print	Do not alter, or resize, or compress
Access image	JPEG or JPEG2000	Screen display or print	Taken from the master, it is altered for presentation over the Web or other uses.
Thumbnail	JPEG or GIF	Screen display	Taken from access, reduced size but not altered otherwise.

Master images must be of the highest quality. Web images need not require such stringent quality controls. But, before compromising on image quality, consider the cost of migrating the image. Because migration is costly, it is far sounder to migrate a high quality (master) image than one of lesser quality. All digital images will have to be migrated, if kept long enough. Their caretakers will have no choice. While the primary use of images in North Carolina ECHO is focused on Web access, repositories need to be mindful of future use, remembering the fragile nature of the originals. Publishing on the Web will result in requests for high quality copies of the images, so consider all possible needs before you produce your digital master image. Remember the advice, it is better to "Scan Once, Save Twice!"

Basic Production Steps

Scanning

The basic steps in scanning an image will be determined by the format of the material to be scanned, but all formats (color, B&W, and Bi-tonal) share common scanning techniques. While a full scanning manual is beyond the scope of this document, scanning an individual image might look like this:

1. Align material on the clean scanner bed, mask if necessary. (Because old documents "flake," cleaning after each image may be required.)
2. Preview the scan.
3. Crop the image, leaving sufficient margins (white space).
4. Using the scanner software, set the resolution (dpi/ppi) and/or printer scale (dpi/lpi).
5. Scan.
6. Save at high resolution a raw image using the TIFF format.
7. Transfer the TIFF master to a file on your computer (with accompanying documentation).
8. Pull up the image in the manipulation software.
9. Using the manipulation software, crop carefully. CAUTION: do not over-crop. (Master files should maintain margin, showing to future users that the whole image has been digitized.)
10. Adjust histogram. (The graph of brightness values vs. number of pixels having that value, histograms are included in the image manipulation software; be careful this procedure reduces the amount of original information.)
11. Set gray mid-point (if used).
12. Adjust image size, if needed.
13. Make adjustments (tone, sharpness, noise, etc.) for the clearest image possible.
14. Adjust resolution needed for access
15. Check for quality against original.
16. Write second (derivative) file to TIFF or to JPEG. (This is the access or Web image.)
17. Change resolution (dpi/ppi) and write third (derivative) file to GIF or JPEG. (This is the thumbnail image.)
18. Store the master image in a secure format. (See **Digital Preservation** of this guide)
19. Add the access and thumbnail images to your Web site. (See **Metadata** and **Presenting your Digital Project** of this guide)

Digital Cameras

1. Align objects on copy stand directly under the camera.
2. Clean the camera lens.
3. Preview the image through the viewfinder or the LCD viewer on the camera before you take the picture. This will give you a more accurate idea of what will be captured in your digital image. If your camera is wired directly to your computer, you may be able to preview the image on the computer screen.
4. Check the lighting in the room to be sure that it is correct. Turn off overhead lights and close curtains.
5. Check your camera settings for appropriateness of the object you are capturing.
6. Save the image as a high resolution TIFF.
7. Transfer the TIFF master to a file on your computer (with accompanying documentation).

8. Pull up the image in the manipulation software.
9. Using the manipulation software, crop carefully. CAUTION: do not over-crop. (Master files should maintain margin, showing to future users that the whole image has been digitized.)
10. Adjust histogram. (The graph of brightness values vs. number of pixels having that value, histograms are included in the image manipulation software; be careful this procedure reduces the amount of original information.)
11. Set gray mid-point (if used).
12. Adjust image size, if needed.
13. Make adjustments (tone, sharpness, noise, etc.) for the clearest image possible.
14. Check for quality against original.
15. Write second (derivative) file to TIFF or to JPEG. (This is the access or Web image.) Change resolution (dpi/ppi) and write third (derivative) file to GIF or JPEG. (This is the thumbnail image.)
16. Store the master image in a secure format. (See **Digital Preservation** of this guide)
17. Add the access and thumbnail images to your Web site. (See **Metadata** and **Presenting your Digital Project** of this guide)

Digital Audio

1. Check the quality of the analog audio to ensure that they will not be damaged through conversion (stickiness or shedding require consultation with a conservator before conversion).
2. Hook up devices (audio playback device, analog-to-digital converter, and computer with audio software)
3. Choose WAV file at maximum allowable kHz and bit depth and write to digital file.
4. Make adjustments and changes to digital audio file using audio software.

Image Size and Proportion

When trying to determine the size of the image as it will appear on a monitor, confusion often arises from the method of measurement. What is the difference in ***dpi***, ***ppi***, and ***lpi***?

The ***original image*** may be measured by inches, centimeters or millimeters. The ***screen image*** is measured by ***ppi*** (pixels per inch) or ***dpi*** (dots per inch).

- **DPI** - Early manufacturers of laser printers devised the convention of **DOTS PER INCH** to suggest the quality of the print of an image. The term ***dpi*** is used when preparing an image for printing.
- **PPI - PIXELS PER INCH** is a more accurate measurement for describing the image in its digital form on the monitor screen. The word pixel comes from "picture elements." Ppi is used to indicate the resolution of a photograph. Pixels do not have a size in a computer nor do they have a digital size. The computer converts pixels to numbers, and this array of numbers is recognized as an image. The size of the image or the scanned object determines the pixels. The horizontal measurement is always given first. The pixels in the image should not exceed the size of the screen. If it does, then users must scroll up and down to see the image, never being able to see all of it at once.
- **LPI - LINES PER INCH** is sometimes used interchangeably with ***dpi***. For example, FotoLook software, which comes with some Agfa scanners, uses ***lpi*** instead of ***dpi*** as a measurement.



LANDSCAPE VIEW
640 x 480 pixels



PORTRAIT VIEW
480 X 640 pixels

A longer horizontal dimension indicates a "landscape" view. A longer vertical dimension indicates a "portrait" view. In the example above, if you want the entire image to show on the 640 x 480 pixel computer screen, you would have to resize the portrait view to reduce the vertical dimension to 480 pixels or less.

The information in a pixel is fixed and does not change. What can change is the array of numbers of pixels. Increasing the numbers of pixels increases the size and resolution of an image. All scanning is a "sampling" of portions of the original. The higher the resolution of this sampling, the more real information you have to work with. However, **the higher the *ppi*, the longer it will take the computer to load the image to the screen.**

Consider the following examples when figuring out the proportion of computer screen image to original.



If the original image is 4 inches high, and you want a screen image which is 1 inch high, to get an image that is 1/4 of your original, **scan the image at 25%**. To get good detail select 300 **ppi** or higher.



When you are ready to place the image on the Web page, use a photo application such as Photoshop to reduce the resolution to 72 **ppi**, the standard Web image resolution. Your image will be 1 inch high and will read well on the Web. It will also load quickly, as the lower **ppi** allows for more rapid loading.

Let's say you want a print of your image, and you capture it in at 600 **dpi** (*remember, dots per inch for printing*). The image will print satisfactorily, but the screen resolution will be over-kill. Typically, the resolution of the access image should be about 1/2 the resolution you want to make a print. Note that the 300 **ppi** (*remember pixels per inch for screen*) used for the image above is 1/2 the print resolution of 600 dpi.

Keeping your **image resolution at 1/2 your print resolution** is a good rule of thumb and will lessen the confusion often found between print resolution and image resolution. To re-cap: the print size of an image is important only for printing. **If you go by the print pixel size and carry that image to the screen, it will usually exceed the screen size.** For example, a printed image of 3 x 2 inches scanned at 300 **ppi** will be 900 x 600 pixels on the screen and the same image scanned at 600 **ppi** will be 1800 x 1200 pixels! Therefore, the image intended for a printer is far too large to fit most screens and must be re-sized with your photo application by lowering the resolution.

COLOR information in an image is dependent on the number of colors or shades of gray that can be carried by a pixel. This carrying capacity is referred to as the pixel's dynamic range or its bit-depth. The standard minimum carrying capacity is 8 bit and currently the maximum (today) is 42 bits. However, most image viewing software cannot display 42 bits yet.

A BIT is the smallest storage unit in a computer. It is the unit of measurement for determining the range of color or shades of gray found in an image. The greater the dynamic range or bit-depth, the greater the subtlety of color or gray. Remember the trinity of three types of scan: bi-tonal, grayscale, and color? There is a preferred bit-depth for scanning each of these.

Bit Depth

TYPES OF SCAN	PREFERRED BIT-DEPTH	THIS MEANS
bi-tonal	1 bit	each pixel is either black or white
grayscale	8 bit	each pixel can be 1 of 256 shades of gray
color	8 bit or 24 bit	8 bit: each pixel can be 1 of 256 shades of color or 24 bit: each pixel can be 1 of 16.8 million color possibilities

One bit bi-tonal is obviously more suited to line drawings and text and 8 bit and 24 bit color more suited to images where the full range of colors are needed. Eight bit grayscale is the *de facto* standard for black and white photography and many graphics. One bit bi-tonal may not be satisfactory for some text and line drawings because it does not capture enough information. Experimentation with grayscale may also be necessary. Likewise, some badly faded color photographs may be better scanned in 8 bit grayscale.

Some will wish to know why their institution should digitize at a higher resolution than a computer screen is able to present. The answer lies in potential and multiple uses. High quality printers (the type used in books and magazines, for example) use higher resolutions than do computer screens. If a publisher spots an image in a digital collection, he or she may wish to use it in a printed work. If the original has only been scanned for access (low resolution or compressed), then it will necessarily need to be scanned again at higher resolution and an uncompressed format. This means re-handling and another exposure to a bright light source. More handling and more light means more damage to the original. And who knows? Higher resolution computer screens and delivery systems that can handle the correspondingly large files may be a common part of homes and offices in the not-too-distant future.

Signal to Noise or Image Quality Features

The signal to noise ratio is the primary measure of image quality. If the signal to noise (S/N) ratio is high, the image quality will be high. However, measuring signal to noise is largely a subjective process. "Noise," or degradation of the image, is not generally a good thing. The more noise, the poorer the image. The aim in a good scan is to decrease the noise. There are several ways to do this that use standard image quality features found in most image manipulation software. The most common noise reduction features are:

Despeckling

Despeckling a photo makes global changes to the entire photo, smoothing color transitions in an image, which can help remove graininess.

Descreening

Often scanned images of half-tone printed objects end up with ripples or patterned "stars." These are called a *moire* effect. Descreening reduces this effect by defocusing the original scanned image.

Low pass filtering

Low pass filtering "averages" pixel information along lines of great contrast,

"smoothing" or "softening" an image.

Decreasing contrast or balancing gamma (generally lowering it)

"Gamma" is the contrast affecting the mid-level grays or midtones of an image. Adjusting the gamma of an image allows you to change brightness values of the middle range of gray tones without dramatically altering the shadows and highlights. Decreasing gamma "drops" the brightness or contrast.

Adjusting and Optimizing Color Palette

Some screen images will require color manipulation to match the original. If a change to a master image is required, it should be saved as a part of the metadata associated with the digital copy.

Be careful that features are not used that can increase noise. The following features should be used with caution as they increase the noise of a scan:

Increasing the contrast or increasing the gamma

"Gamma" is the contrast affecting the mid-level grays or midtones of an image. Adjusting the gamma of an image allows you to change brightness values of the middle range of gray tones without dramatically altering the shadows and highlights. Increasing gamma "raises" the brightness or contrast.

Aggressive color management and manipulation

Color is generated for scanned images by "mixing" pixels of various basic color values, such as red, green and blue. By changing these values of individual pixels, the color mix of the overall image can be altered.

Sharpening

Accentuates the differences between adjoining areas of significantly different hue or tone.

Reminder:

Image quality is generally measured by evaluating some or all of the following features:

- Tone Reproduction
- Resolution
- Color Reproduction
- Noise
- Detail and edge reproduction
- Artifacts including, nonuniformity, dust and scratches, streaks, color misregistration, aliasing, contouring/quantization

Monitor Display

When images are viewed on a monitor, the monitor resolution will often vary with the type of monitor used. Standard monitor resolutions follow:

Monitor	No. of pixels x No. of lines	Quality
VGA	640 x 480	Low
SVGA	800 x 600	Medium

XGA	1024 x 768	High
SXGA	1280 x 1024	High

Monitors should also be calibrated for best results. Software packages like Adobe Photoshop often include a basic monitor-calibration tool.

Quality Assurance

Quality assurance needs to be performed throughout the creation of your digital images. While it may seem like a daunting task, digital images should be checked to assure that they are at the level of quality specified by your institution. Documents can be skewed and not noticed by the staff member creating the digital images. Always use more than one set of eyes to assure that your images are consistent and of high quality.

It is recommended that you establish a system early for quality assurance and do the work on a regular basis. Digital objects tend to pile up, and you don't want to have a large number to go through at one time. That situation can lead to sloppy checking. Some institutions choose to look at a sampling of the images once initial production set up has been established.

A good quality assurance program will help you to stick to the "Scan Once Methodology" by assuring that the master image is accurate.

Optical Character Recognition

When a computer scans a text, all it duplicates are graphical bits on a virtual page. In other words, it creates a digital image or copy of the page. A user can not edit or search in the newly-created document. If that image is passed through an Optical Character Recognition program (OCR), the software converts the shapes it recognizes into individual letters, creating a text document. However, OCR recognizes and converts few documents perfectly. It makes frequent errors, especially if the original image is blurred, faded, or otherwise unclear.

Work with OCR is even more labor intensive than straightforward imaging of pages, requiring a great deal of editing and quality control. It does, however, produce a much more versatile digital product. OCR'd documents must be proofed, word by word. Such things as unusual proper names, blemishes on the document, uneven light, tables, borders around text, offset fonts, superscripts and subscripts can throw the word recognition off.

There are various devices available for capturing images for OCR, but a desktop scanner will also work. Because a desktop scanner digitizes the image by dividing it into hundreds of pixel-sized boxes per inch and represents each box with either a 1 or a 0, the OCR program organizes the patterns of dots into characters. This allows for the computer to translate character images into editable text.

In addition to OCR, there is the PDF (Portable Document Format) which is an open and universal file format that preserves the fonts, images, graphics, and layout of any source document, regardless of the application and platform used to create it. Governments and enterprises around the world have adopted PDF to streamline document management, increase productivity, and reduce reliance on paper. PDF is available to anyone who wants

to develop tools to create, view, or manipulate PDF documents. A good place to research PDF's is <http://www.pdfzone.com>.

Documenting the Digital Object

Many institutions are now realizing the importance of documenting the process of digital production. There are many reasons documentation is wise. One of the most critical reasons is that as technology changes, migration of earlier formats is a reality. The more documentation that is available, the easier and less costly the migration will be when it occurs. And *it will occur*.

NC ECHO has created a preservation metadata standard that helps you document your digital production by encouraging you to record information about each digital object you create (<http://www.ncecho.org/presmet/index.htm>). Some aspects of the preservation metadata remain relatively stable while others will change for each image. But preservation metadata is not the only documentation you should consider. Other documentation that can inform your digital product and management include:

- **Document planning decisions**
Were all images in a collection scanned? If not, what was the thinking behind the selection?
- **Document capture, editing, and processing decisions** How were images manipulated for presentation? Were images batch processed, etc.?
- **Keep institutional decision memos.** These memos will allow you to revisit decisions made previously.
- **Document revisions of institutional decisions.** As with documenting initial decisions, any revisions need to be monitored and should be available for review.
- **Administrative guidelines.** Administrative guidelines and training materials provide a valuable framework for this digital project as well as future ones. Learn things once and benefit from those lessons in the long term.
- **Workflow guidelines/ workflow revisions.** This should clear up and issues with responsibilities and help you track particular images.

The Historic American Sheet Music Project from the Rare Book, Manuscript, and Special Collections Library at Duke University has created some documentation as an example: (<http://scriptorium.lib.duke.edu/sheetmusic/procedure.html>)

Conclusion

Digital production is the fun part of a digitization project, and it can be done relatively easily. Unfortunately, the ease can be deceptive. To produce a "down and dirty" quick digital image takes almost no time and effort at all. With just a bit more time, effort, and storage space, an image can be created that would

- Better preserve the original by reducing the amount of times it would need to be handled in the future
- Be more easily migrated into new technologies and storage media
- Serve multiple purposes such as providing sources for publishing as well as Web access.

To support the long-term viability of a digitization project, the production process should be thoroughly documented to help future caretakers and users of the images. Once there is a documented digital master, a variety of processes (including some that are automated) can be used to manipulate the access and thumbnail images for better presentation over the Web. These images are not expected to be surrogates of the originals, but "overviews" or "quick views" of them, pointing the way to either the original or master digital image.

The scanner or camera awaits -- and it only takes a little more time and effort to give those images a better chance at varied use and long-term viability.

Further reading

Besser, Howard. *Procedures and Practices for Scanning*.
<http://sunsite.berkeley.edu/Imaging/Databases/Scanning/>

CDP Digital Audio Working Group, *Digital Audio Best Practices*, version 2.0, November 2005,
http://www.cdpheritage.org/digital/audio/documents/CDPDABP_1-2.pdf

Research Libraries Group and Digital Library Federation. *Guides to Quality in Visual Resource Imaging*, <http://www.rlg.org/legacy/visguides/visguide6.html>

Kenney, Anne and Steven Chapman. *Digital Imaging for Libraries and Archives*. Ithaca: New York, Department of Preservation and Conservation, Cornell University Library, June 1996.

Kenny, Anne R. and Oya Y. Rieger. *Moving Theory into Practice*. Research Libraries Group, 2000. See also:
<http://www.library.cornell.edu/preservation/tutorial/index.html>

Williams, Don. "Selecting a Scanner," *Guides to Quality in Visual Resource Imaging*, <http://www.rlg.org/visguides/visguide2.html>