DIGITAL IMAGING FOR THE SMALL ORGANIZATION - VERSION 1.4 2016

CONTENTS

PART 1: PURPOSE

Digital Imaging for the Small Organization offers guidance to provide digital imaging recommendations for smaller institutions and collections that are planning/involved in digitization projects. It is a guide, not de facto standard, for digital imaging, image capture, presentation, storage and preservation.

This guide which is based primarily on the document: *Western States Digital Imaging Best Practices Version 1.0*. However with 75% of the museums in the United States being small museums this document deviates significantly as this is written with the small museum/organization in mind with an all volunteer staff or a paid staff of 1-3 individuals and with very limited access to technical resources.

Western States' original document has been both slimmed down and technical language simplified to make reading it more realistic for the small museum staffer with a limited time budget. Whole sections as well as most of the Appendixes have been left out.

The most significant deviation from other "Best Practice" documents on digitization is the recommended bit depth for color images/docs/maps for master files which has been reduced to 24 bit color from the recommended 48 bit color. The increase in the file size by scanning at 48 bit color does not result in a significant enough quality gain to justify the extra costs in storage space to a small organization. As storage costs continue to drop this may be investigated again at some future time.

In sections, like the Book Scanner section we recommend an open source solution. Where we have done so, it is to assist the small organization in finding low cost solutions or provide more flexible solutions. Controversially sometimes the low cost solution is not the best solution. However, it is important to realize smaller organizations often have very limited resources and compromises sometimes have to be made. You may note that "Best Practices" is not included in the document title. The money, time and resources for preserving the *Declaration of Independence* are generally not available to preserve or provide access to your community's founding charter.

Western States Digital Imaging Best Practices Version 1.0 is an important document and should be looked at as the authoritative source and while this document might be an access point for learning about digital preservation, please refer to their document for much more in depth information.

Lastly this is a work in progress and will always be, that is simply the nature of digital technology, never to be set in stone. Technologies AND even archival practices change and so will this document. This is the first run at this, **there will be updates, modifications and corrections.** Please feel free to provide any feedback and suggestions to: joe.hoover@mnhs.org

If you are a small historical organization we would also like to direct you to the American Association for State and Local History's StEPs program. StEPs is a voluntary assessment program for small- and mid-sized history organizations which uses assessment questions and performance indicators (basic, good, better) to rate the organization's performance in six standards sections.

GENERAL PRINCIPLES & SUMMARY

- Digitize at the highest resolution appropriate to the source material
- Digitize at an appropriate level of quality to avoid redigitizing and rehandling of the originals in the future

- Digitize an original or first generation (i.e., negative rather than print) of the source material to achieve the best quality image possible. In the case of art prints, the developed print is considered the original piece.
- Create and store a master image file that can be used to produce surrogate image files and serve a variety of current and future user needs
- Use system components that are nonproprietary
- Use image file formats and compression techniques that conform to standards within the cultural heritage community
- Create backup copies of all files on servers and have an off-site backup strategy
- Create meaningful metadata for image files or collections
- Store digital files in an appropriate environment
- Monitor data as necessary
- Document a migration strategy for transferring data across generations of technology
- Plan for future technological developments

PART 2: LAYING THE GROUNDWORK

GENERAL PRINCIPLES

- Digitize the original source material at the highest appropriate resolution to avoid having to re-digitize.
- Digitize an original or first generation (i.e., negative rather than print) of the source material to achieve the best quality image possible. In the case of art prints, the developed print is considered the original piece.
- Avoid media and file formats that are proprietary (ie: patented technologies a single company)
- Describe your images with metadata so people can find them later. Use a spreadsheet or database.
- When archiving digital files create backup copies with one set being on a file server at ta different locations. Technology will change - you need to plan accordingly.

QUESTIONS TO ASK BEFORE STARTING A DIGITIZATION PROJECT

- A little preparation goes a long way to making your project a success
- What is your purpose? Archival Preservation? Public access? Both?
- Does the project support the institution's mission?
- Who is your audience?
- Who owns it?
- What are the physical characteristics of the collection?
- What is your time frame?
- How is the project being funded?

- Who will be responsible at different stages of the project?

DOCUMENTATION

Good concise documentation will aid those that follow you in understanding your guidelines and procedures for storage archiving file naming and metadata.

STAFFING

In a small organization you will need to wear many hats. It would help to review the following to aid for digitization project.

- Project management skills
- Knowledge of cataloging, registration methods and metadata tags
- Understanding of photographic techniques and methods
- Subject matter specialists
- Database development and administration skills.
- Computer programming skills
- Web design and development skills
- Artistic/graphic design skills

IN-HOUSE VS. OUTSOURCING

The following are some points to consider for both strategies

IN-HOUSE PROS:

- Learning
- Control over process
- Flexibility
- Security

IN-HOUSE CONS:

- Large costs in time and money for training and equipment
- Limited expertise may result in lower quality
- Institution is responsible for quality controls

OUTSOURCING PROS:

- Pay for cost of scanning the image only, not equipment or staffing

- High production levels
- On-site expertise
- Less risk
- Vendor absorbs costs and responsibility of quality control

OUTSOURCING CONS:

- Less control
- Communication with vendor must be clear
- Vendor may understand image process but not library/archiving/museum issues
- Risk to original images if transported

COSTS

- Costs vary wildly digitization is only a third of the total cost, cataloging description and indexing make up the other two thirds
- Look to share resources and collaborate with other organizations
- Ongoing costs of maintaining data and systems over time

RIGHTS MANAGEMENT

As appropriate, projects must be careful to obtain copyright permissions from repositories or copyright holders prior to distribution. Before beginning a digitization project establish which objects will require permission from the copyright holder. If collection has been chosen for digitization and copyright is not yet cleared, locating the copyright holder and obtaining permission can be a lengthy and costly process

PART 4: CREATION BASICS

PPI/DPI

The term "**pixels per inch**" (PPI) refers to the number of pixels captured in a given inch and is used when discussing scanning resolution and on-screen display. The term "**dots per inch**" (DPI) while the term is used in digital creation its application has more meaning for print rather than digital. When referring to digital capture, DPI can be used and often is, but PPI is the preferred term, as it more accurately describes the digital image.

MODES OF CAPTURE

Most imaging equipment offer four modes for capturing a digital image:

- **Bi-tonal** — black and white only, no color, no grays. Best suited to high-contrast documents such as printed text
- **Grayscale** — Shades of gray. Best suited for black and white photographs
- **RGB** — Short for Red Green Blue. Best suited for color photographs and objects were capturing color information is important
- **CMYK** — Short for Cyan-Magenta-Yellow-Black. Used for color off-set printing and should never be used for archival purposes

There is no clear mode of capture and require some subjective decisions. A typed document may have red annotations in ink. A black and white photograph may have sepia tones that should be considered to be captured in color. Images that may have started out bi-tonal like illustrations and advertisements in a newspaper may be more accurately captured in gray scale.

BIT (COLOR) DEPTH

Because bit depth is one of those things you hear over and over in digital imaging but no one every seems to make it clear just what Bit Depth is and the reason is we need to get very basic when talking about it.

BIT

A "bit" is the smallest unit of data and each bit has two states. It can be 1 or 0, black or white, on or off. Think of a piece of paper as a bit with a black side and a white side. Just like there are 8 ounces in a cup, 4 cups in a quart, 4 quarts in a gallon - 8 bits comprise a "byte".

Now a byte or 8 bits, with each bit possessing two possible values, can therefore in combination can represent 256 different possible gray values. In math terms: 28

DEPTH

What about “Depth”? Well think about it this way. A digital image is made up of vertical and horizontal pixels with each pixel occupying its own space on the image. So Depth represents the possible values (or shades) of gray that each pixel in that image can represent.

If you ever had sanded woodwork that has been painted with many coats of different paint you might understand this better. The top layer is ivory and the layer under that is light gray, under that is dark gray and under that black. So now you have four different colors and depending on where you sand and how deep you can create your own piece of art with dark gray exposed on one area and ivory on another and black exposed over other areas.

Now the 256 different possible gray values while they are not physically lying in a row under each other, they in a sense really do take up space which is why the greater the Bit Depth the larger the file size of the image.

However, you might not want to scan an "black and white" image in gray scale. The scan would only be able to lay down 256 shades of gray, from black to white not nearly enough for a decent image that can give the richness of tones in that black and white image. Instead you might want to scan a black and image in color, which means that all three channel colors - Red, Green and Blue (RGB) will be mixed together to create 16 million shades of gray or

256X256X256. More than enough.(site note: RGB is not the primary colors of Red, Blue and Yellow you learned about in school) .

SPEAKING OF COLOR

RGB images are made of three 8 bit color channels. So an an 8 bit color image is actually 24 bit when we're dealing with color. Each channel has 256 possible values (remember 2^8). So the Red channel as 256 shades of red, the Green channel has 256 shades of green and the blue channel 256 has shades of blue. An so just as an 8 bit color image is actually 24 bit a 16 bit color image is actually called 48 bit (16 bits X 3 channels). That 48 bit image is capable of billions of colors.

WHERE IS THE BLACK AND WHITE CHANNEL?

There is no black/white channel. The RGB colors when absent or combined in total create black and white.

COLOR SPACE

Again, it might be good to get very basic and keep it simple when talking about the confusing term color space and to use appropriate metaphors to describe it.

Color space consists of a color model along with a specific mapping of that model onto an absolute color space. There are a large number of color spaces in use in the world today. Their design allows you to edit images in a controlled, consistent manner.

Choosing a wide gamut space, such as **ProPhoto**, **RG**B or **Adobe RGB**, allows for greater color information to be captured about the original item and will allow you to convert to a narrow space later which is especially good when thinking about digital images that are to be archival. Narrow gamut space such as **sRGB** captures less information and is intended for images used on the web.

The best to wrap your head around it is to go to Home Depot and go to their paint department. Their Martha Stewart line of paints offer quote: "Over 280 custom interior and exterior colors to open up dazzling new possibilities for your home" however, the Behr line of paints offer 2,000 PREMIUM PLUS® paint colors - a far larger color space which I think offers even more “dazzling possibilities”. However, a ginormous selection like that may be unnecessary or over-welling so a selection like Martha Stewart's (aka: the sRGB color space if you are working web images) may do just fine.

RESOLUTION

Resolution determines the quality of an image. It is described either by pixel dimensions (height and width) for on-screen use or by physical size and PPI. Increased PPI take more frequent samples of the original and contain a more accurate representation of the original. Since higher resolutions are capturing more information, file sizes also increase. There is no one “perfect” resolution to scan all collection materials. Resolution should be adjusted based on the size, quality, condition and uses of the digital object. The combination of PPI and size of the original object determine the resolution needed to accurately capture as much information about the original object as is available.

There is a point at which adding more pixels per inch no longer adds content, because the original source object has a finite amount of information available based on the way that it was produced.

TONAL DYNAMIC RANGE

In a very simple nutshell are your whites white and your darks dark? Or does your image look flat and washed out?

FILE FORMATS FOR PRESERVATION

There are proprietary and non proprietary formats for image files. The recommendation for master file image capture is to use a non-proprietary format.

TIFF — Tagged Image File Format (TIFF) is the format of choice for archival and master images. It is a flexible, highly portable, widely accepted, open standard image format and considered the professional image standard.. TIFF images are not suitable for web delivery.

JPEG 2000 — JPEG 2000 is being used increasingly as a repository and archival image format. Many repositories are storing "visually lossless" JPEG 2000 files: the compression is lossy and irreversible, but the artifacts are not noticeable and do not interfere with the performance of applications. While still a lossy, compression technique, but may have potential for becoming the file format of choice for archival master images in the near future.

PDF and PDF/A — In April 2008, the United Kingdom's Digital Preservation Coalition (DPC), named Portable Document Format (PDF) as one of the best file formats to preserve electronic documents and ensure their survival for the future. The report suggests adopting PDF/Archival (PDF/A) for archiving electronic documents as the standard that will help preservation and retrieval in the future. It concludes that it can only be done when combined with a comprehensive records management program and formally established records procedures. Be aware there are many flavors of PDF, many use proprietary features and/or not suited for use in archival preservation.

DNG — In response to concerns over the future support for proprietary RAW files (image file that contains unprocessed data in a format that is unique to each manufacturer) , Adobe has produced the DNG (Digital NeGative) format. DNG is supported by a growing number of camera and software manufacturers and, as of this writing, has been submitted to the ISO as a vendor-independent standard. A free Adobe DNG Converter is available to translate many different proprietary RAW formats to DNG. "DNG is likely to be readable long after the original RAW format becomes obsolete. Additionally, DNG offers significant file-size savings through a lossless compression that can reduce the file size by up to one third.

Proprietary Formats — Proprietary formats are controlled or owned by a particular entity that licenses the format for use by others. These formats often require special plug-ins or software for viewing. Proprietary formats are not recommended for master images because licensing requirements may prevent the long-term access and preservation of images. One example of a proprietary file format is Encapsulated PostScript (eps).

FILE FORMATS FOR WEB ACCESS

JPEG — Joint Photographic Experts Group. A compression algorithm for condensing the size of image files. JPEG image files allow online access to full screen image files because they require less storage and are therefore

quicker to download into a web page. The JPEG format is most frequently used for access images requiring lossy compression.

PNG — Unlike JPEG, PNG is lossless, the png format is a great alternative to jpeg especially when delivering high quality images over the web. PNG was designed for transferring images on the Internet, not for professional-quality print graphics, and therefore does not support non-RGB color spaces such as CMYK. Remember, using this format is more about access than preservation.

COMPRESSION

Compression can greatly reduce the size of a digital file. There are two kinds of compression Lossy and Lossless. In general, it is usually recommended to void compression all together but because compression and greatly reduce storage costs for an organization if compression is used go with Lossless.

Lossy — In lossy compression, a certain amount of information is discarded during the compression process. Although the discarded information may be invisible to the human eye, a loss of quality occurs. Lossy compression formats also introduce generational loss each time a lossy image is manipulated or edited, the quality of the image decreases. Generational loss is one of the reasons master and service master images are not stored using compression.

Converting images from a bit depth of 16 to a bit depth of 8 is also considered a lossy compression method as color information is discarded.

Lossless — Lossless compression results in a file similar to the original image, with no loss of information. The Tagged Image File Format (TIFF) supports lossless compression such as the LZW algorithm. Compared to uncompressed TIFF, visually lossless JPEG 2000 compression can reduce the amount of storage by an order of magnitude or more.

Typically even a lossless compression is discouraged because it introduces another layer which could be an access barrier for future generations. However, for a small organization with a limited budget it can reduce storage costs significantly.

PART 5: HARDWARE CONSIDERATIONS

When creating digital content consider your workflow process. From creating the digital content, preserving it by cataloging and archiving or finally creating public access to the content, to be successful each of these aspects - creation, preservation, and access must be well planned and organized. The number of objects, their format and condition and the timeline to completion, will determine the scale and rate of the project.

COMPUTERS

Projects that require the purchase of computer hardware should consider the following:

- Purchase a computer with the primary duty to work with digitization initiatives. These computers will require more power and speed than a computer simply used for data or word processing
- Purchase as much Random Access Memory (RAM) as your budget allows
- More memory allows the computer to more quickly process large amounts of image data

- Purchase computers with processors optimized for image manipulation
- Slow processors may result in a slow redraw rate
- Purchase computers that support high-speed data input through serial connections, USB 2.0 or Firewire as you will need to connect your camera/scanner to computer
- Purchase as much hard drive space as possible. External hard drives that attach via the USB port can also be used to supplement the hard drive workspace

MONITORS/DISPLAY

Most computers now use liquid crystal display (LCD) as opposed to the older cathode ray tube (CRT) technology. Avoid using CRT screens.

CAPTURING IMAGES

There is definitely more than one way to digitize a book or photograph. What devices you use and how you go about doing it will be influenced by:

- Project goals
- Format
- Size of collection to be scanned
- Condition of materials

OTHER FACTORS THAT MAY ALSO PLAY A ROLE ARE

- Skill level of person/s involved
- Desired speed
- Size of scan area
- Budget

PART 6: WHAT DO I NEED?

SCANNERS

Flatbed Scanners — This is the most common type of scanner - and the least expensive. They can be used to scan materials such as papers, photographs and other printed material. They should not be used for original art, including art prints. Some flatbed scanners also have the capacity to scan negatives and transparencies. While it is a better practice to use a dedicated slide/film scanner, the prohibitive cost of purchasing a separate and expensive slide/film scanner may make it acceptable to buy an multi-use flatbed scanner model for smaller organizations.

Some scanners have automatic document feeders which are acceptable when scanning multiple pages of contemporary material, but cannot be used for historic materials or photographs because of danger of damage.

Slide/Film Scanners — These scanners are specifically designed to digitize slides and film. Although a flatbed scanner with a transparency lid can be used for this purpose, a dedicated film scanner has much higher quality

scanning capabilities. Optional slide feeders can be purchased to allow a batch scan of up to 50 slides. Slide scanners are highly recommended for projects with large numbers of slides.

Large Format Scanners — These scanners can be useful when scanning maps, blueprints, architectural drawings, site plans, posters, etc. They look and operate like a flatbed scanner, but are much larger. However, they remain potentially cost prohibitive for many and organizations may want to consider outsourcing to a digital imaging vendor or simply using a digital camera for image capture when dealing with large material.

Book Scanners — Book copiers or scanners allow for overhead copying of bound books, newspapers and oversized and/or fragile materials that cannot be placed on a flatbed scanner. Book scanners include software which compensates for any distortion caused by the curve of the page when digitizing complete books. These scanners are more complex and, consequently, much higher priced than flatbed scanners. However, costs can be dramatically reduced if an organization builds its own book copy stand (there are many plans available on the internet) and use open source software:

http://www.openplanetsfoundation.org

http://wiki.fluidproject.org/display/fluid/Decapod

DIGITAL CAMERAS

Point and Shoot — Consumer-oriented "point and shoot" cameras are not suitable for digitization projects. The lens quality is limited, and there is no studio flash synchronization.

Digital Single Lens Reflex — For many digitization projects especially when dealing with oversized and delicate printed material, a digital single lens reflex (DSLR) type camera may be the best capture device. The advantages of this type of camera include: high image quality; appropriate image resolution; flexibility in shooting situations; flash synchronization; the ability to connect the computer and camera via USB or Firewire (tethered shooting*).

> *Tethered shooting allows for technical and aesthetic decisions to be made immediately. This insures the final capture will be archival quality.

CAMERA STANDS AND LIGHTING

Copy stand with lights — A copy stand is a stand for mounting a camera, complete with base board and arms to hold lights. The stand enables you to take photographs of flat art, books and three-dimensional objects. The lamps cast even lighting on the work being photographed. A copy stand with lights offers a controlled environment.

Book cradle — Book cradles are special supports for scanning bound documents. A book cradle allows a book to be fully supported and properly aligned for safe and accurate digital capture.

Lighting — The key to taking quality images with a digital camera is adequate, proper and even lighting. Another important consideration is to avoid light sources that raise the temperature of the physical item being digitized. Because the camera synchronizes with studio lights, greater control of lighting is possible, resulting in a better image.

ADDITIONAL EQUIPMENT FOR SYSTEM:

- Color/grayscale target
- Studio flash system and camera stand
- Portable computer optimized for digital imaging
- Tabletop copy stand with proper lighting
- Book cradle
- Studio space (no windows)

SOFTWARE CONSIDERATIONS

SCANNING SOFTWARE

Most scanners come equipped with software to manage this transfer for you. However opt to use the "advanced or custom settings mode over the software's automatic settings. This gives you more control over both the resolution and size of image captured and the file formats generated by your scanner.

IMAGE EDITING SOFTWARE

Once you have captured a high resolution image from your scanner and have saved an unaltered master file to your storage media, you will may need to manipulate surrogate copies of that image using image editing software. Adobe Photoshop is the industry standard for the creation of surrogates which can be delivered via the web, print publication or for in-house uses such as exhibits. However, there are free and low cost image editing software options available; 1. GIMP – Free and open source. 2. Pixlr - free online image editor. 3. Affinity Photo - Low cost and powerful photo editor.

PART 7: THE RIGHT EQUIPMENT FOR THE RIGHT TASK

SMALL PHOTOS/DOCUMENTS

Two-dimensional objects that fit on the surface of the scanner; will require no external studio lighting; most fit on a table top; and they can easily be used by someone with only a basic understanding of the digitization process.

BOOKS/NEWSPAPERS*

If you are scanning complete books and newspapers, a dedicated book scanner may be the ideal choice.

NEGATIVES/FILM TRANSPARENCIES

If your collection has been photographed on film, then a either a dedicated transparency/film scanner (which is becoming increasing rare) or a scanner with capabilities to do transparencies/film will be needed.

OVERSIZED MATERIALS, LARGE MAPS, ORIGINAL ART*

You will need a digital camera that can be used to capture any object, two- or three-dimensional, small or large.

*Keep in mind for both a Camera and Book scanner will need to have an external lighting source which will require both additional skill and money.

PART 8: WORKSPACE

Proper climate control and security are important if collections will stay in the workspace for long periods of time. A workspace for digitization should offer a controlled lighting source to maintain consistency and quality of images. Rooms with windows should be avoided or at least the windows should be covered up so no light from the outside can get in. Changes in room lighting can affect how images are represented while shooting and while viewing the images on computer monitors. Ideally walls should be neutral-gray; room light levels should be low, but consistent.

If at all possible, the camera system, lighting and equipment should be mobile. There are times when it is better to move the studio to the collection rather than the other way around.

Providing a comfortable, safe and secure workspace for a digitization project can increase productivity and quality of images by reducing operator fatigue and potential damage to collections.

PART 9: GUIDELINES FOR CREATING DIGITAL IMAGES

The quality and condition of the original and nature of the materials you are scanning (such as the quality of the shooting or processing technique in the case of photographs) impacts on the resolution at which you capture and the resulting quality of the digital image. These are not hard and fast rules for every collection and every institution. The key to quality imaging is not to capture at the highest resolution possible, but to scan at a level that matches the informational content of the original.

The following guidelines contain the recommended settings for the creation of master, access and preview images. Large files may be difficult and/or costly to manage. Because of storage concerns and time considerations, it may be necessary to reduce the recommended resolution and/or bit depth when creating master images. Carefully consider institutional needs and capabilities, however, before departing from recommended best practices.

MASTER FILES

Good digital imaging projects begin with a high-quality master or archival image and then derive multiple versions in smaller sizes or alternative formats for a variety of uses.

A high-quality uncompressed master image will make the investment in the image capture process worthwhile. A digital master must be available with all the original information to accommodate future needs and applications. The master image should be the highest quality you can afford; it should not be edited or processed for any specific output. Quality control should be applied during the creation of master image files. Any errors made during this time will mean going back to the scanner or camera to capture another image.

With the proper image editing software, it is not necessary to subject source materials to multiple scans.

Master digital images should be stored in a nonproprietary / open source file format that supports long-term preservation of the image. The recommended format most frequently used for master digital images is TIFF or JPEG 2000 (not to be confused with standard "JPEG" format).

SERVICE MASTER FILES

The service master is an optimized working copy of the master file which can be used as a source for all subsequent derivatives. They are also used to create print publications. Creation of a service master from a master will depend on the source file and the resources and time of the organization. Typically all scans of photos and most scans of historic documentation are more likely to be subject of needing rework and the creation of a service master file. Newer documents may need no rework so need no master service file if resources are slim.

DERIVATIVE FILES

Derivative files are created from the service master or master file and are used for general Internet or network access. Derivative files typically include an preview image, which is sized to fit within the screen of an average monitor or other delivery mechanism and a often a thumbnail image, which is small enough to load quickly and linked to the larger preview image.

FILE NAMING CONVENTIONS

Good file naming is important for management of the digital asset. File names need to be unique, follow an established naming convention.

FILE NAMING RECOMMENDATIONS INCLUDE:

- Use lowercase letters of the Latin alphabet and the numerals 0 through 9.
- Use underscores or dashes in place of spaces
- Characters to avoid: ¬ ! " £ $ % ^ & * ( ) + = { [}] : ; @ ~ # ? < > , | \ ` '
- Begin each file name with a two- to three-character acronym representing the organization's name
- Follow the organization's acronym with an object ID. The object ID consists of any unique numbering scheme already in use to represent the object or, if no such number exists, a short description representing the item
- Include a part designator after the object ID, if it is part of series when applicable
- File names should be limited to 31 characters, including the three character file extension Remember
- Think long-term
- Select a system that will outlast staff involved in the current project
- Consider the number of files your institution will ultimately be managing
- How simple or easy will it be to make a mistake?

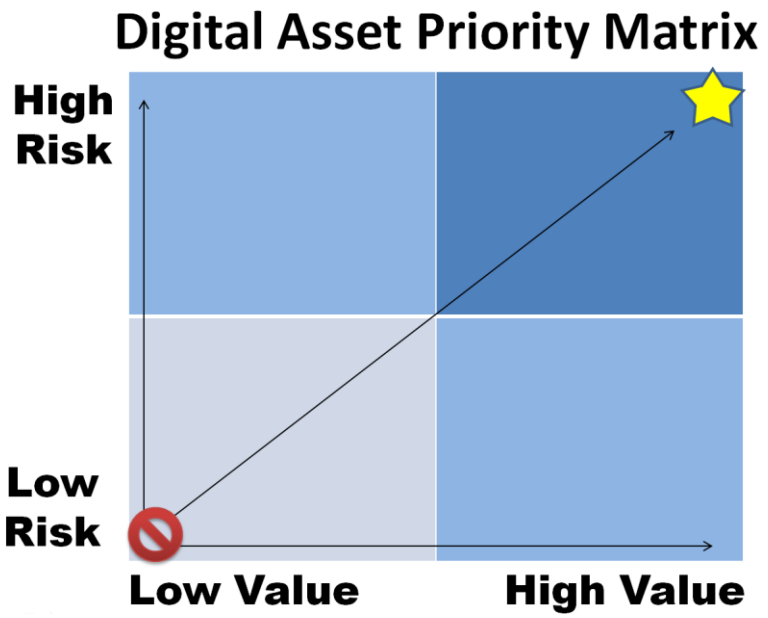
- File names do not take the place of metadata
- Keep them simple and straightforward.

PART 10: PRIORITIZING DIGITAL PRESERVATION TASKS

Resources in an organization are always limited and every action cannot be accomplished. It is crucial to determine which actions are the most important so that those receive consideration first. Prioritizing is the process of deciding which actions will have the most significant impact, which are the most important, and which are the most feasible.

- Material in your collection that could not be replaced if it were lost or damaged.
- Material in your collection that is frequently requested or accessed and has high research value
- Materials that are deteriorating or fading in the case of color images.

Also: Consider the prioritization of Analog/magnetic media such as video or cassette tapes when developing your preservation plans. Often that media is incredibly ephemeral as well as requiring rare/outmoded devices for playback when looking digitization of documents and images. A tintype or black and white photograph can last centuries when properly stored but VHS tapes of WW II veteran oral histories could already be unplayable.



PART 11: DIGITIZING STANDARDS

PHOTOGRAPHS

- Digitize from the negative (or the earliest generation of the photograph) to yield a higher-quality image.
- In the case of photographs developed according to artist specifications, the photograph itself should be digitized rather than the negative.
- Digitize sepia-tone as color images to create a more accurate image.
- Digitize the backs of photographs as separate image files if there is significant information on the back of the photo (which may be of interest to users)

	Master	Web Access	Thumbnail
File Format	TIFF	PNG or JPEG	JPEG
Bit Depth	16 bit grayscale 24 bit color	8 bit grayscale 24 bit color	8 bit grayscale 24 bit color
Spatial Resolution	400 to 800 ppi	72ppi	72ppi
Spatial Dimensions	4000 to 8000 pixels across the long dimension, depending on size of original, excluding mounts and borders	600 pixels across the long dimension	150 to 200 pixels across the long dimension

MAPS

- Scanning maps may involve items that vary widely in size, condition and amount of detail.
- Small maps may fit easily onto a flatbed scanner, while large plat maps may need captured by a camera.
- Size of the image can become a problem for storage, but also for viewing, serving over the web or processing.
- Smaller maps (less than 36 inches on the longest dimension) should be digitized at 600 PPI, 24-bit color or 16-bit grayscale if possible.
- Larger maps, 300-400 PPI may be more practical.
- If it becomes necessary to digitize a map in sections and stitch the image together in Photoshop, keep both the original images of the sections as well as the combined image.

	Master	Web Access	Thumbnail
File Format	TIFF	PNG or JPEG	JPEG
Bit Depth	16 bit grayscale 24 bit color	8 bit grayscale 24 bit color	8 bit grayscale 24 bit color
Spatial Resolution	400 to 800 ppi	72ppi	72ppi
Spatial Dimensions	4000 to 8000 pixels across the long dimension, depending on size of original, excluding mounts and borders	600 pixels across the long dimension	150 to 200 pixels across the long dimension

GRAPHICS

- Graphics include the various techniques used to reproduce words and images from originals such as engraving, lithography, line art, graphs, diagrams, illustrations, technical drawings and other visual representations.
- Nearly all graphics will be two dimensional and should be scanned using the following guidelines.

	Master	Web Access	Thumbnail
File Format	TIFF	PNG or JPEG	JPEG
Bit Depth	16 bit grayscale 24 bit color	8 bit grayscale 24 bit color	8 bit grayscale 24 bit color
Spatial Resolution	600 to 800 ppi	72ppi	72ppi
Spatial Dimensions	6000 to 8000 pixels across the long dimension, excluding mounts and borders	600 pixels across the long dimension	150 to 200 pixels across the long dimension

TEXT

- Documents with smaller printed text may require higher resolutions and bit depths than documents that use large typefaces.
- Images that produce the best results for OCR may not be pleasing to the eye and may require separate scans for OCR and human display. Test pages at several resolutions to find the most satisfactory results.
- Projects with large amounts of textual materials, particularly hard-to-read materials such as manuscripts, should provide transcriptions of the materials in addition to the digital image.
- As rekeying text can be cost prohibitive, projects considering transcriptions should investigate including Optical Character Recognition (OCR) software in their toolkit as well as using WikiSource to crowd source transcribing public domain documents http://en.wikisource.org
- Access to textual material can be further enhanced through SGML/XML markup schemes such as the Text Encoding Initiative (TEI) http://www.tei-c.org.

	Master	Web Access	Thumbnail
File Format	TIFF	PDF-PDF/A	JPEG
Bit Depth	16 bit grayscale 24 bit color	8 bit grayscale 24 bit color	8 bit grayscale 24 bit color
Spatial Resolution	Adjust scan resolution to produce a minimum pixel measurement across the long dimension of 6,000 lines for 1bit files and 4,000 lines for 8 to 16 bit files	72 to 200ppi	72ppi
Spatial Dimensions	4000 to 6000 pixels across the long dimension, excluding mounts and borders	600 pixels across the long dimension	150 to 200 pixels across the long dimension

FILM NEGATIVES

Master scans of black and white camera originals may be captured and saved in RGB, particularly those negatives that contain color information as a result of staining, degradation or intentional color casts. Derivative files could later be reduced to grayscale in the scanning software or during post-processing editing.

	Master	Web Access	Thumbnail
File Format	TIFF	JPEG	JPEG
Bit Depth	16 bit grayscale 24 bit color	8 bit grayscale 24 bit color	8 bit grayscale 24 bit color
Spatial Resolution	Resolution to be calculated from actual image format and/or dimensions - approx. 2800 PPI for 35mm originals, ranging to approx. 600 PPI for 8x10 originals	72ppi	72ppi
Spatial Dimensions	4000 to 8000 pixels across the long dimension, depending on size of original, excluding mounts and borders	600 pixels across the long dimension	150 to 200 pixels across the long dimension

PART 12: QUALITY CONTROL

Quality control should be conducted throughout all phases of the digital conversion process. Inspection of final digital image files should be incorporated into your project workflow. Typically, master image files are inspected online for a variety of defects. Depending on your project, you may want to inspect 100 percent of the master images or 10 percent of the files randomly.

- Implement and document quality control procedures and clearly define the specific defects that you find unacceptable in an image.
- Images should be inspected while viewing at a 1:1 pixel ratio or at 100 percent magnification or higher.
- Quality is evaluated subjectively, by project staff (scanner operator, image editors, etc.) through visual inspection
- Quality is evaluated objectively, by the imaging software (by using targets, histograms, etc.).

MASTER DIGITAL IMAGE VISUAL INSPECTION CHECK LIST

Things to look for during archival master visual inspection may include:

- Image is the correct size
- Image is the correct resolution
- File name is correct
- File format is correct
- Image is in correct bit depth and color mode (i.e., color image has been scaled as grayscale)
- No loss of detail in highlight or shadows
- No excessive noise especially in dark areas or shadows
- Even tonal values, no flare
- Correct focus
- Not pixilated
- Excessive dust spots or other objects
- No digital artifacts (such as very regular, straight lines across picture)
- Image not cropped
- Image not rotated or reversed
- Correct color balance
- Histogram:
    - No spikes or clipping
    - No tonal values lower than 9 or higher than 247

SERVICE MASTER VISUAL INSPECTION CHECK LIST

In addition to the items on the Master Digital Check List, here are additional things to look for during service master visual inspection:

- No digital artifacts (such as very regular, straight lines across picture)
- No moire patterns (wavy lines or swirls, usually found in areas where there are repeated patterns, such as half-tone dots)
- Image rotated correctly and not reversed
- Image centered and not skewed

PART 13: DESCRIBING DIGITAL ASSETS

METADATA THEORY AND PRACTICE

Often overlooked by time and cash strapped organizations, metadata must be given a high priority in the creation of digital assets and at the planning stage of the project to ensure information about the collections can be found by users, shared with other organizations and managed over the life of the digital asset.

Metadata creation can be from scratch or it reuse preexisting descriptive information from its physical counterpart, such as accession records, finding aids or catalog records. It can also be created automatically and embedded in the file with the creation of the digital asset.

There are numerous metadata schema; therefore, the institution will need to select the schema most appropriate to the collection, audience and institution. The most common metadata schemes include the following sets of information, with some overlap:

The term metadata is inclusive of the five different types of metadata information — descriptive, technical, administrative, structural and preservation.

- **Descriptive Metadata** describes the intellectual content of a resource and used for the indexing, discovery and identification of a digital resource.
- **Administrative Metadata** includes management information about the digital resource, such as ownership and rights management.
- **Structural Metadata** is used to display and navigate digital resources and describes relationships between multiple digital files, such as page order in a digitized book or diary.
- **Technical Metadata** describes the features of the digital file, such as resolution, pixel dimensions and hardware. The information is critical for migration and long-term sustainability of the digital resource.
- **Preservation Metadata** is information that supports and documents activities related to digital preservation. It is information that is used in supporting the processes of ensuring the core preservation

processes of availability, identity, understandability, authenticity, viability and renderability. While some of these activities require descriptive and structural metadata, the majority of it is administrative metadata.

PART 14: BACKUP AND STORAGE

3 -2-1 BACKUP STANDARD

Good practices for storage involve Redundancy, Scalability and Long-term viability

Have the backup storage plan follow the **3 -2-1 standard**:

- Make **3** copies
- Have at least **2** different kinds of storage media mediums. (Cloud storage, hard drives*, tape backup etc...)
- **1** needs to be stored off site away from the other two.

* If using only hard drives for storage, make sure to use at least two different brands to reduce risk of loss in case of hardware failure due to a manufacture's defect.

MEDIA STORAGE VS. NETWORKED STORAGE VS. PARTNERSHIPS

MEDIA STORAGE

Storage on a DVD or a CD ROM is strongly discouraged for all but the tiniest of collections. Since DVDs and CDs are short lived, the data on them has to be moved every few years. This can be problematic if the collection is large and is stored on multiple disks. Another better option is a storing them on external hard drives. The costs of drives are dropping rapidly and Terabyte drives are now becoming affordable and common. It is critical that copies of the data are created and stored off-site.

NETWORKED STORAGE

STORAGE AREA NETWORK (SAN)

Sometimes referred to as a 'file server' a SAN is a dedicated storage network that provides access to consolidated, block level storage. Costs have dropped significantly so they are accessible to small and medium sized organizations. However, the organization is responsible for the upkeep and security of the network.

Some benefits to on-site disaster protection for your information is the instant access provided, reliability of getting to the data and the inherent security of the data.

Unlike off-site storage which would require requesting copies of information to be transported back to your location and the time delay while that happens, on-site recovery can begin immediately. With on-line backup, a copy of your information is moved over a network.

NETWORK-ATTACHED STORAGE (NAS)

Similar to a SAN but the data is stored off-site (in the cloud – also referred to as “Vaulting”). The amount of affordable "live storage" one can purchase has increased while the price has come down. Smaller organizations can take advantage of some expertise of the highly, well-trained technical staff at these large scale data centers which is unavailable for the organization running their own SAN.

Simply put, offsite storage gives you protection from having something happen to your local data. On the other hand, local onsite data should be faster and avoids the cost of accessing data over a network. A hybrid approach which combines local onsite data storage with offsite to store backups for data protection, support disaster recovery has worked well for many companies.

PARTNERSHIPS

Partner with a larger organization that is able to take their "work product" (digital objects and descriptive metadata) and place the content in a data center setup to manage data backups and has a plan to preserve the digital objects and metadata.

CHECKSUMS

A checksum is a method/program for detecting file/data corruption or tampering and helps monitor the condition of digital files in storage and when moved from one system or media to another system or media. In an over simplified sense, a checksum is something like a fingerprint of the file. The checksum of a newly created file is checked against the checksum created from the master file.

Checksum data can be stored as part of the technical metadata of the image in a collections management program and/or in a spread sheet for easy batch processing to audit the integrity of digital assets. Checksums can also be embedded in the digital image, specifically in EXIF metadata in TIFF files to help establish if that data is altered or stripped during processing.

AV Preserve has a free open source Checksum program for both Macs and PCs, **Fixity**, and offers both a user guide and video tutorial.