Minimum Digitization Capture Recommendations

Minimum Digitization Capture Recommendations

The Association for Library Collections and Technical Services
Preservation and Reformatting Section
June 2013

Submitted by:

- **Ian Bogus**, University of Pennsylvania
- **George Blood,** George Blood Audio & Video
- **Robin L. Dale,** LYRASIS
- **Robin Leech,** Oklahoma State University
- **David Mathews,** The Image Collective

Contents

Preface

While many cringe at drawing similarities between microfilm and digitization, the foundations for digitization lay firmly in the lessons learned from microfilm. Microfilm was originally seen as a cheap and compact alternative to print materials. Institutions could store their rapidly growing collections in a compact format that was easy to reproduce and share. The first standard for microform production was published in 1979 around the time when many academic libraries were taking action on the brittle book problem by establishing preservation programs. Before this date, microforms were made on nitrate or acetate film neither of which have long shelf lives. Libraries could work together on non-duplicative microfilming projects once the standard was established. If a book was microfilmed according to the standard then another library could purchase a copy rather than re-film it. Items that were microfilmed before 1979 frequently needed to be re-filmed because of poor quality images or materials.

The language to describe the benefits of digitization is an echo from the microfilm era: it is touted as a cheap, easily distributed format that can be stored in a compact space. At this point there is no official standard for digitization, but institutions are discussing how they can collaborate and share digitized content. Collaborative projects such as HathiTrust and the Internet Archive raise the question of whether an already digitized book can be a surrogate without re-digitizing it. Some institutions have struggled with the differences between “digitizing for access” and “digitizing for preservation.” Ostensibly, content that is “digitized for access” is expected to have a relatively short shelf life, and content that is digitized for preservation should be sustainable for decades, if not longer.

This document was created as a guideline for libraries digitizing content with the objective of producing a sustainable product that will not need to be re-digitized. Institutions can feel secure that if an item has been digitized at or above these specifications, they can depend on it to continue to be viable in the future. In some cases, institutions may want to request a digital copy to preserve themselves, further safeguarding materials by preserving them in multiple locations.

There have been numerous studies exploring the technical side of how items should be digitized and most institutions engaged in digitization have decided on specifications for themselves. It was not the authors’ intention to duplicate this important work but rather to build on it. The authors reviewed past research, the practices at almost 50 organizations including the recent Federal Agencies Digitization Initiative (FADGI) and other guidelines from governments, universities, and other institutions. The authors also examined samples of digitized works to determine a recommendation of minimum specifications for sustainable digitized content some of which are included in this document. The guidelines given here are in line with most previous recommendations as well as requirements for recent collaborative projects like HathiTrust. The intent is to codify a set of recommendations as accepted minimums for libraries as a whole and not individual institutions or projects.

The scope of this document is narrow. It only speaks to the technical specifications of the digitized content itself and not to the larger issue of digitally preserving said content. Issues such as light spectrums, equipment calibration, staffing expertise, file types, compression schemes and other subjects about producing good digital objects is also out of scope of this document. Citations to other resources are provided throughout; most explore topics at a deeper level than is intended here.

[top]

Basis for Recommendations

An adequate surrogate needs to replicate characteristics of the original that users require. When digitizing objects one must be aware that different types of materials are used in different ways and that sometimes requires variations in the digitization specification. The recommendations stated here should fulfill most needs, but may need to be adjusted to a higher specification in instances where the expected use is different than is described in this document.

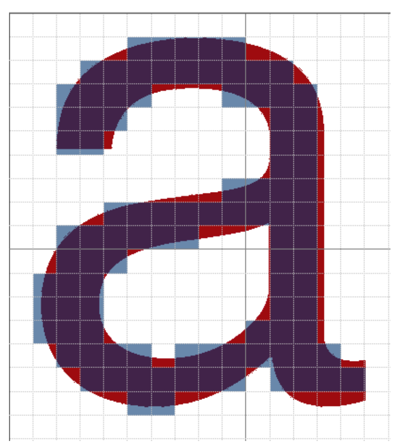
There are two general categories of media discussed below: static and time-based media. Static media is a term that encompasses common library collections such as books, photographs, maps, and microfilm that can usually be represented by image surrogates. Many resources differentiate between reflective media, where light bounces off the surface, and transmissive media, where light passes through the object. For the purposes of this document, both reflective and transmissive media are covered in the static media section. The nature of the various formats, playback mechanisms, the requirement of specific orders and timings must be captured to represent the object correctly make digitization of time-based media challenging. The section on digitizing time-based media follows the section on Static Media.

For further reference, please see a compendium of institutional guidelines in **Appendix IV** that includes types of original materials and links to the guidelines.

[top]

Static Media

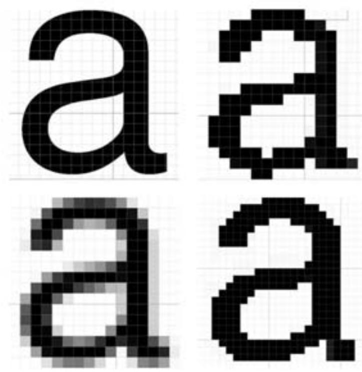
Initially in digitization practice, it was common to capture only in black and white, otherwise known as bitonal, for textual documents or line drawings. This practice is now infrequently used for various reasons. When a digital image is created it is essentially made of ordered blocks of color called pixels. Software assigns a color for each individual pixel. It is almost impossible for the scanner's sensors to align with the edges of every physical, so the system must perform some interpretation during the digitization process. Bitonal images appear more pixelated than grayscale even when the size of each pixel is the same. Grayscale images actually depict the shape better because a lighter shade can be chosen for a pixel that does not fully cover an area that has color in the original. To account for this, many practices included increasing the bitonal image capture resolution by 50%.



A bitonal example of how pixels relate to an object.

In the microfilm era, there was a system of determining legible quality called the Quality Index (QI) formula. This formula was updated for digital materials in an AIIM standard:

Resolution as it Relates to Photographic and Electronic Imaging. Essentially, the index delineates how many pixels are required to represent the smallest significant character in the source (usually a lowercase "e" in text scenarios), which is essentially twice the QI number. Naturally, an acceptable resolution to render an object depends on the size of the characteristics. Color and grayscale images require at least 16-pixels to depict the smallest character with excellent detail resolution (QI-8, or half the amount of pixels). Bitonal images require 24-pixels. As a frame of reference, 400-ppi[1] will capture a 1mm artifact with 16-pixels and 600-ppi will capture that same artifact with 24-pixels. This means that 400-ppi is sufficient to adequately capture a document with 1mm characters (or a 6 pt font) in grayscale or color, and 600-ppi is sufficient to adequately capture the same document in bitonal. This is where some of the early recommendations for using 400-ppi or 600-ppi began. The figure to the right depicts a representation of the QI and differences between bitonal and grayscale representations.



Clockwise from upper left: original image, bitonal 16-pixels, bitonal 24-pixels, grayscale 16-pixels.

As resolution increases so does the file size. One must weigh this increase with the functional advantages of additional resolution. Increasing the resolution makes sharper details when the image is blown up. This can be helpful, but there comes a point where increasing the resolution does not yield any gains. For example, a three inch square of uniform color will not be better represented at 3000-ppi than 300-ppi, but it will be ten times the size. There are other issues too. Without perfect focus, the details in an image with very high resolution can be lost in the blur. Most cameras were not intended to take images with microscopic resolution from a distance of feet.

The bit-depth also affects the file size. The color of each pixel is recorded as a binary series, usually in multiples of eight (1, 8, 16, 24, 48…). The higher the number, the more colors are available. For example, there are 256 combinations for a binary 8-digit number and therefore an 8-bit grayscale image has 256 possible shades of gray for each pixel (Black=00000000, White=11111111). A 16-bit grayscale image will be twice the size because it requires 16 instead of 8 bits to record each pixel, but there are 65,536 combinations to represent 65,536 shades of gray. Color images work similarly, but are frequently broken down further into channels. The RGB color model with 24-bits is most common when digitizing library and archives materials. At times this is described as RGB 8-bits per channel (3 channels * 8-bits per channel = 24-bits total). What this means is that each pixel is encoded with a possible 256 shades of red, green, and blue, which in combination allows for 16,777,216 individual colors. The amount of total bits per pixel is relative to the size of the file. A 24-bit color image will be three times the size of an 8-bit grayscale or color image.

The examples used[2] show differences in resolutions between 200-700-ppi at the original size and blown up ten times, and then fifty times[3]. Individual examples are included in most of the sections, but they are all available for download **here**[4]. Most images were not originally intended to be viewed under magnification[5]. Extreme magnification can sometimes be helpful for special collection materials in limited cases. Generally, these materials should be available physically and extreme magnification is usually more helpful on an actual object than digitally.

A good reference for line detail is a twenty dollar bill, which is adequately resolved at 400-ppi. If an object has elements with finer lines than currency engravings then the resolution may need to be higher than 400-ppi. Conversely, an object with artifacts significantly larger than currency engravings may be adequately digitized at a lower resolution.

The resolution value itself may not always be an adequate gauge of a good digital image because the image may not be at the same size as the original, the image may be out of focus, the image may be interpolated, the equipment may not be calibrated, the equipment may not be functioning properly, or other potential issues. These issues can be difficult to assess using production images along. Targets allow the user to have a standard and consistent manner to evaluate the equipment and images easily and objectively. Targets usually have several components such as rulers, to ensure that the size of the image is captured correctly, as well as tools to evaluate image quality factors such as resolution, noise, artifacts and tone/color reproduction. Adding some basic information about the object, such as a title, creator, or an identifier, makes the digital object self-identifying in case it gets separated or becomes unlinked from its metadata.

Suggested reading for more details on digitizing static collections:

- Association for Information and Image Management, “Resolution as it Relates to Photographic and Electronic Imaging,” Technical Report for Information and Image Management 26-1993 (1993): 18.
- Burns, Peter, and Don Williams. “Ten Tips for Maintaining Digital Image Quality.” Trans. *Array Archiving 2007*. Arlington: Society for Imaging Science and Technology, 2007. 16-22. Print. **http://www.imagescienceassociates.com/mm5/pubs/50Arch07BurnsWilliams.pdf**.
- Berns, Roy. “The Science of Digitizing Paintings for Color-Accurate Image Archives: A Review.” *Journal of Imaging Science and Technology*. 45.4 (2001): 203-325. Web. 14 Feb. 2013. **http://art-si.org/PDFs/Acquisition/JIST_45_4_305_325Berns.pdf**.
- “JISC Digital Media.” Still Images: creating new digital images. Joint Information System Committee, 2012. Web. 24 May 2012. **http://www.jiscdigitalmedia.ac.uk/stillimages/docs/category/creating-new-digital-media/**.
- Kenney, Anne, and Stephen Chapman. *Tutorial Digital Resolution Requirements for Replacing Text-Based Material: methods for benchmarking image quality*. Washington DC: Commission on Preservation and Access, 1995. Web. **http://www.clir.org/pubs/reports/pub53/pub53.pdf**.
- Peterson, Kit. *Standards Related to Digital Imaging of Pictorial Materials*. Library of Congress, Sep 2004. Web. 24 May 2012. **http://www.loc.gov/rr/print/tp/DigitizationStandardsPictorial.pdf**.
- *Specifications and metrics for Converted Content – a functional solution of the Future Digital System (FDsys)*. Washington DC: U.S. Government Printing Office, 2006. Web. **http://www.fdlp.gov/home/repository/doc_download/821-gpos-digitization-specification-33-final**.
- “Still Image Working Group.” *Federal Agencies Digitization Guidelines Initiative*. FADGI, 27 Mar 2012. Web. 8 Jun 2012. **http://www.digitizationguidelines.gov/still-image/**.
- “Universal Photographic Digital Image Guidelines.” N.p., 22 Sep 2008. Web. 24 May 2012. **http://www.updig.org/index.html**.
- Wheeler, Barry. “What Resolution Should I Use? Part 1.” *The Signal*. Library of Congress, 5 Dec 2012. Web. 14 Feb 2013. **http://blogs.loc.gov/digitalpreservation/2012/12/what-resolution-should-i-use-part-1/**.
- Wheeler, Barry. “What Resolution Should I Use? Part 2.” *The Signal*. Library of

Congress, 11 Jan 2013. Web. 14 Feb 2013. **http://blogs.loc.gov/digitalpreservation/2013/01/what-resolution-should-i-use-part-2/**.

[top]

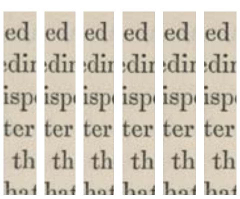
Books and Textual Based Materials Without Images (non-rare)

Min Resolution	Min Color Space	Min Bit Depth	Notes
300	Gray	8	Capture in color (24 bit) whenever possible. 300-ppi will capture details larger than 1.4 mm at a QI of 8. The resolution may be adequately adjusted according to the largest detail to be represented.

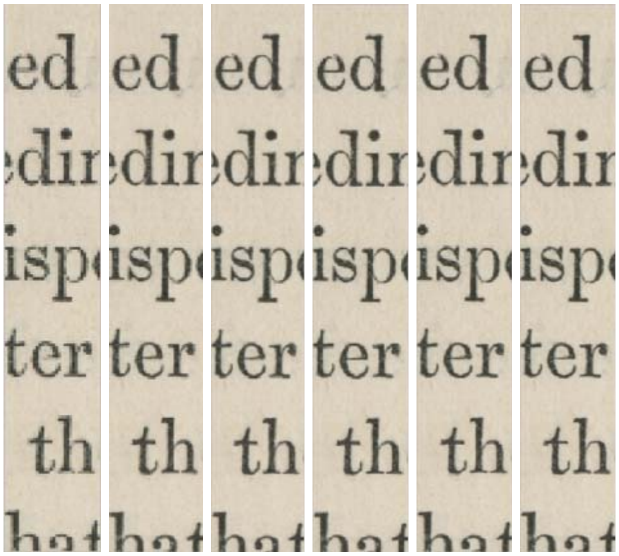
The research value in most textual based materials is in the content itself. These images must be easily legible and processed through OCR or other mining software. Determining acceptable resolution depends on the size of the characters. Text larger than 1.4mm is represented by the requisite 16-pixels for excellent detail representation with 300-ppi or more. Objects that have smaller text should be imaged at a higher resolution. Grayscale images should be sufficient, but color is becoming more common and should be used whenever possible. The small “e”s in the example are approximately 1.7mm, a similar size as common 11 pt fonts.

For more information on digitizing textual based materials with images, see the **suggested reading for digitizing static media**.

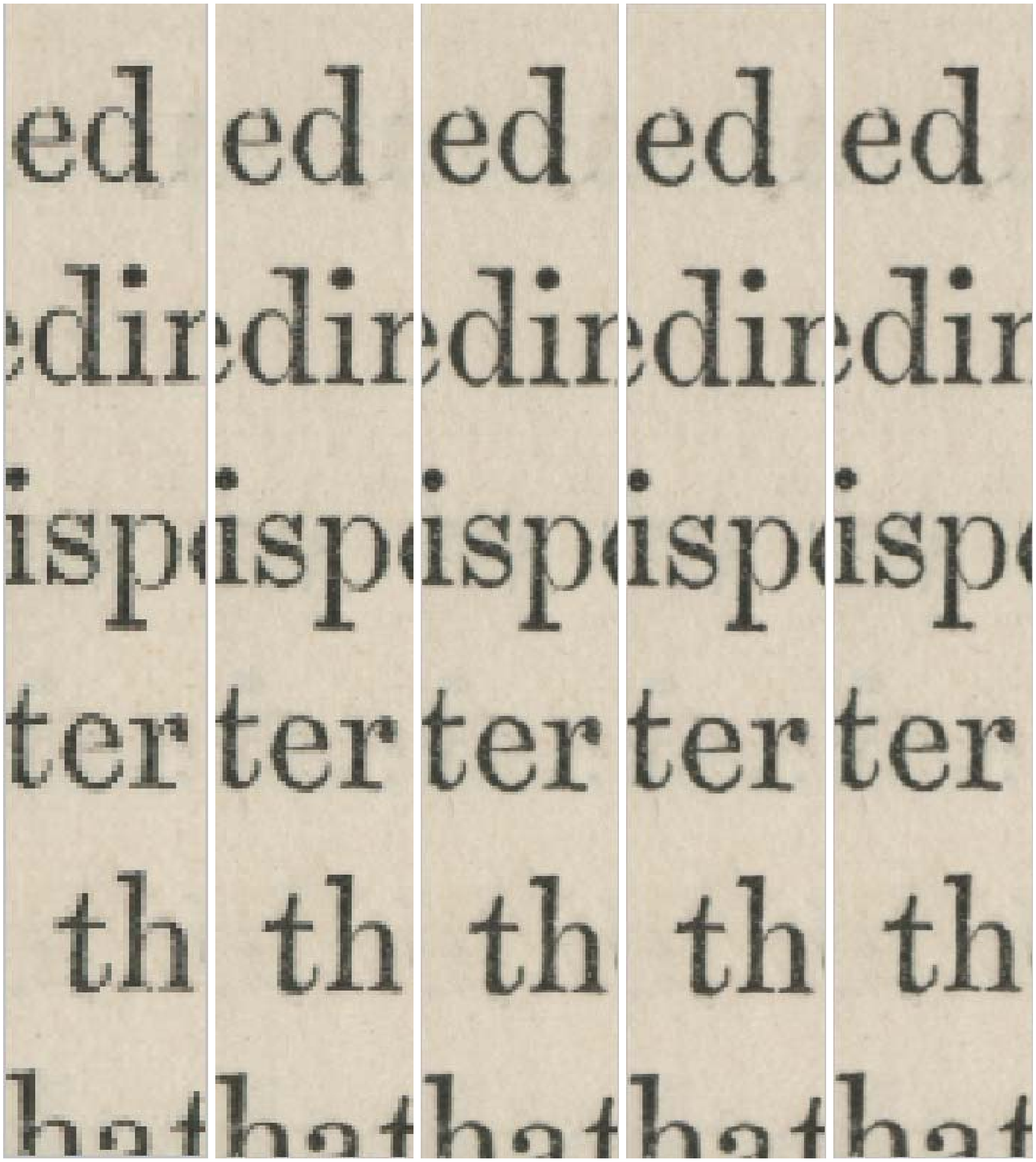
Example images (**download full set here**):

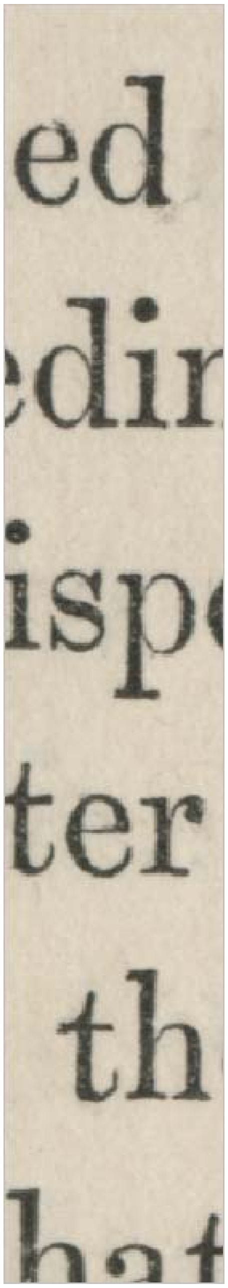


Above: Image increasing resolution from 200-ppi on the left to 700-ppi on the right.



Above: Image blown up 9x increasing resolution from 200-ppi on the left to 700-ppi on the right.





Above: Image blown up 50x increasing resolution from 200-ppi on the left to 700-ppi on the right.

[top]

Books and Textual Based Documents with Images (non-rare)

Min Resolution	Min Color Space	Min Bit Depth	Notes
400	Gray	8	Capture in color (24 bit) whenever possible. Fine lines in etchings or other pictorial elements require more definition than text alone. 400-ppi will capture details larger than 1 mm at a QI of 8. The

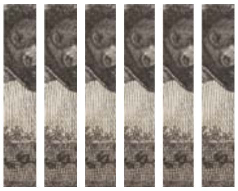
Min Resolution	Min Color Space	Min Bit Depth	Notes
			resolution may be adequately adjusted according to the largest detail to be represented.

Images that are contained within books with textual based documents usually have details that can be important to users. The primary requirement is that the image is clear. Some users are interested in the image production method, requiring the visibility of the individual lines or dots that comprise the image. Moiré patterning may present problems, which can be overcome in the digitization process by changing the angle of capture, software, resolution or a combination of all three.

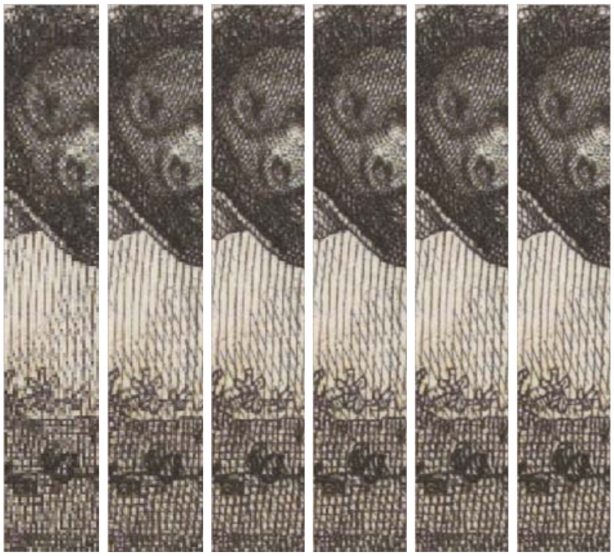
In most cases, 400-ppi will adequately capture necessary details. As a reference, the hash marks in the background behind the man in the example are approximately 0.25mm apart. While they can be seen even at 200-ppi, increasing the resolution allows greater clarity. One can also see color artifacts at the bottom of the lower resolution images. Grayscale is sufficient for most images, but color can be helpful and should be used whenever possible.

For more information on digitizing textual based materials with images, see the **suggested reading for digitizing static media**.

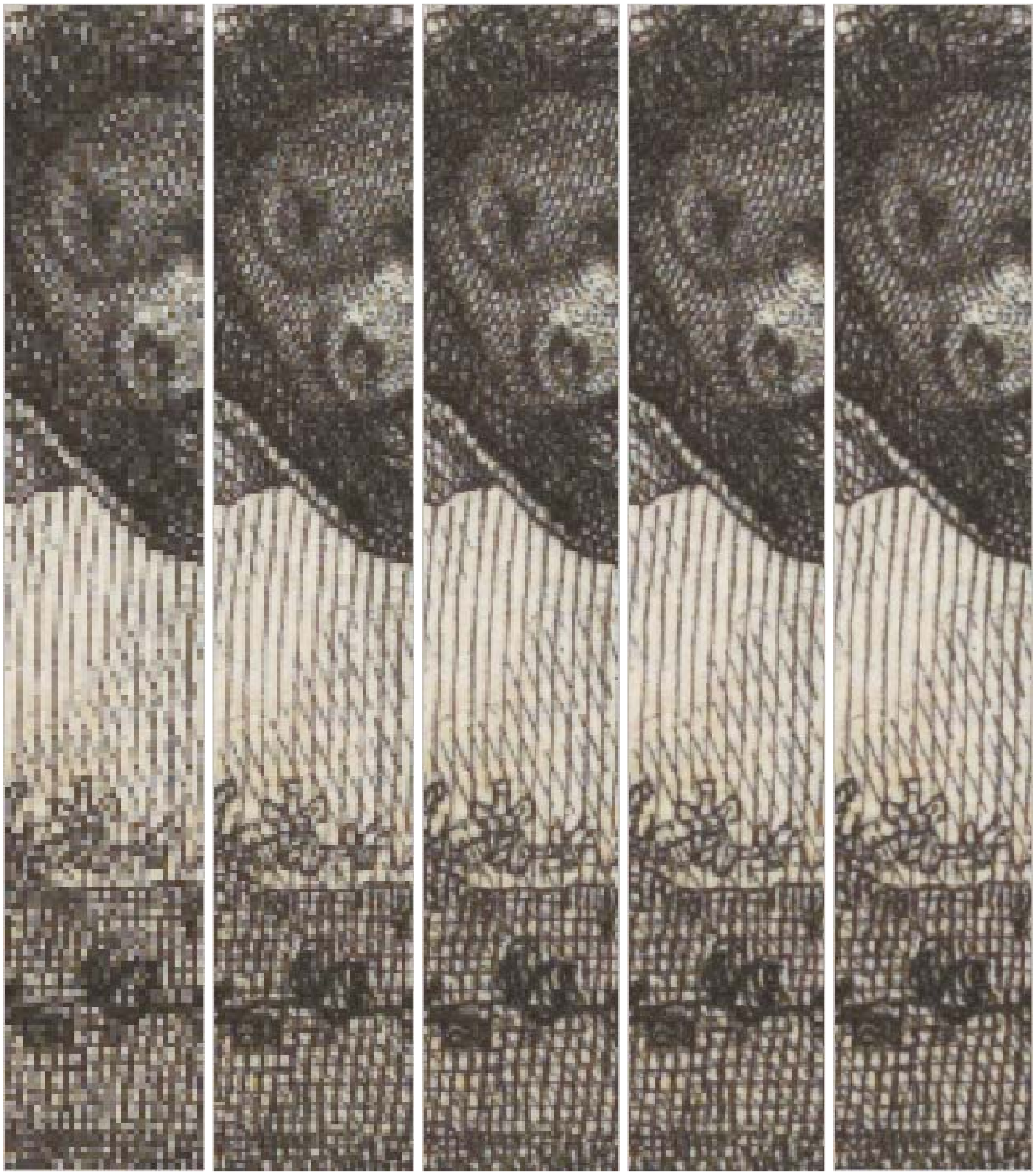
Example images (**download full set here**):



Above: Image increasing resolution from 200-ppi on the left to 700-ppi on the right.



Above: Image blown up 9x increasing resolution from 200-ppi on the left to 700-ppi on the right.



Above: Image blown up 50x increasing resolution from 200-ppi on the left to 700-ppi on the right.

[top]

Rare Books

Min Resolution	Min Color Space	Min Bit Depth
400	Color	24

Several factors influence the resolution required to digitize rare books. Less standardized fonts, or highly ornate and irregular fonts make the minute details in rare books potentially more interesting to scholars; while at the same time, making legibility more difficult. Colors, stains, holes and other markings may be important to the contextual information and should be represented accurately. Color images also help identify obstructions, like stains as opposed to holes, making the document more legible.

Most rare books will be captured adequately at 400-ppi in color. This should capture serifs and other embellishments for any elements larger than 1mm that have fine detail. Contextual evidence in the paper, inks, and illustrations may need to be captured at a higher resolution. The example has relatively large lettering measuring about 2.6mm on the small “e.”

For more information on digitizing rare books, see the **suggested reading for digitizing static media**.

Example images (**download full set here**):

Above: Image increasing resolution from 200-ppi on the left to 700-ppi on the right.

Above: Image blown up 9x increasing resolution from 200-ppi on the left to 700-ppi on the right.

Above: Image blown up 50x increasing resolution from 200-ppi on the left to 700-ppi on the right.

[top]

Manuscripts

Min Resolution	Min Color Space	Min Bit Depth	Notes
400	Color	24	Illegible or difficult to read scripts may require a higher resolution.

Manuscript materials may be written by hand and difficult to read. Additionally, there may be informational value in the inks, pen strokes, or even in the base media. Colors, stains, holes and other markings may be important to the contextual information and should be represented accurately.

Legibility and the most relevant physical information should be easily visible with 400-ppi color images. In the example, the handwriting is large with the small “e” measuring approximately 3mm. Even at this scale, faded ink is noticeably less legible at lower resolutions. In some cases, the resolution may need to be increased for legibility or when extreme magnification is necessary.

For more information on digitizing manuscripts, see the **suggested reading for digitizing static media**.

Example images (**download full set here**):

Above: Image increasing resolution from 200-ppi on the left to 700-ppi on the right.

Above: Image blown up 9x increasing resolution from 200-ppi on the left to 700-ppi on the right.

Above: Image blown up 50x increasing resolution from 200-ppi on the left to 700-ppi on the right.

[top]

Maps

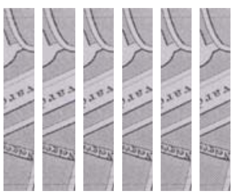
Min Resolution	Min Color Space	Min Bit Depth	Notes
300-600	Gray / Color	8 / 24	600 ppi will capture highly detailed information and good for reprinting and fine quality maps. Lower resolutions may be appropriate if detail is limited (large print etc...) or will not be printed.

Maps are quite diverse, and flexibility should be given when selecting a resolution. It is common that smaller maps have smaller details, but large maps can often have tiny elements. Digitizing large maps at high resolutions may create exceptionally large files that are difficult to handle without benefit unless the details are also small. Maps may also be created with a higher print resolution than other materials. Moiré patterning may present problems, which can be overcome in the digitization process by changing the angle of capture, software, resolution or a combination of all three.

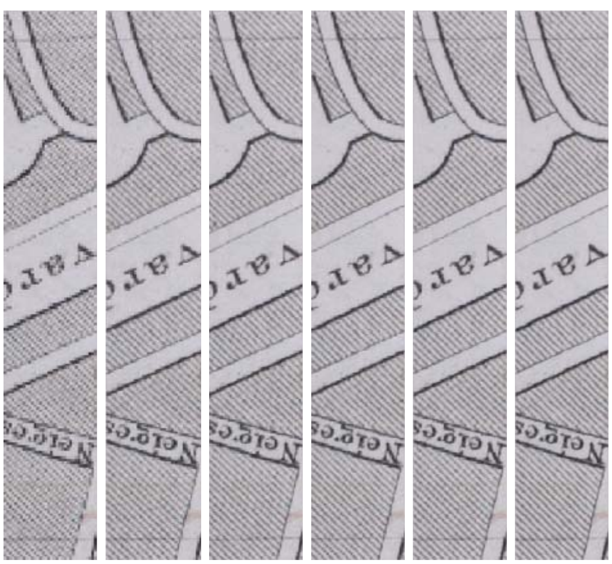
Maps with large details may be adequately digitized at 300-ppi, though maps with very small details may require 600-ppi resolution or higher. Large file sizes may be worth the extra detail clarity for some items. The example at the right comes from a relatively small map. The “e”s in “Notre Dame aux Neiges” are all approximately 0.5mm and the hash marks are approximately 0.25mm apart. A QI of 8 for a character that is 0.5mm requires 820ppi. The example used here is an example of a situation where increasing the resolution from the minimum has clear benefits.

For more information on digitizing maps, see the **suggested reading for digitizing static media**.

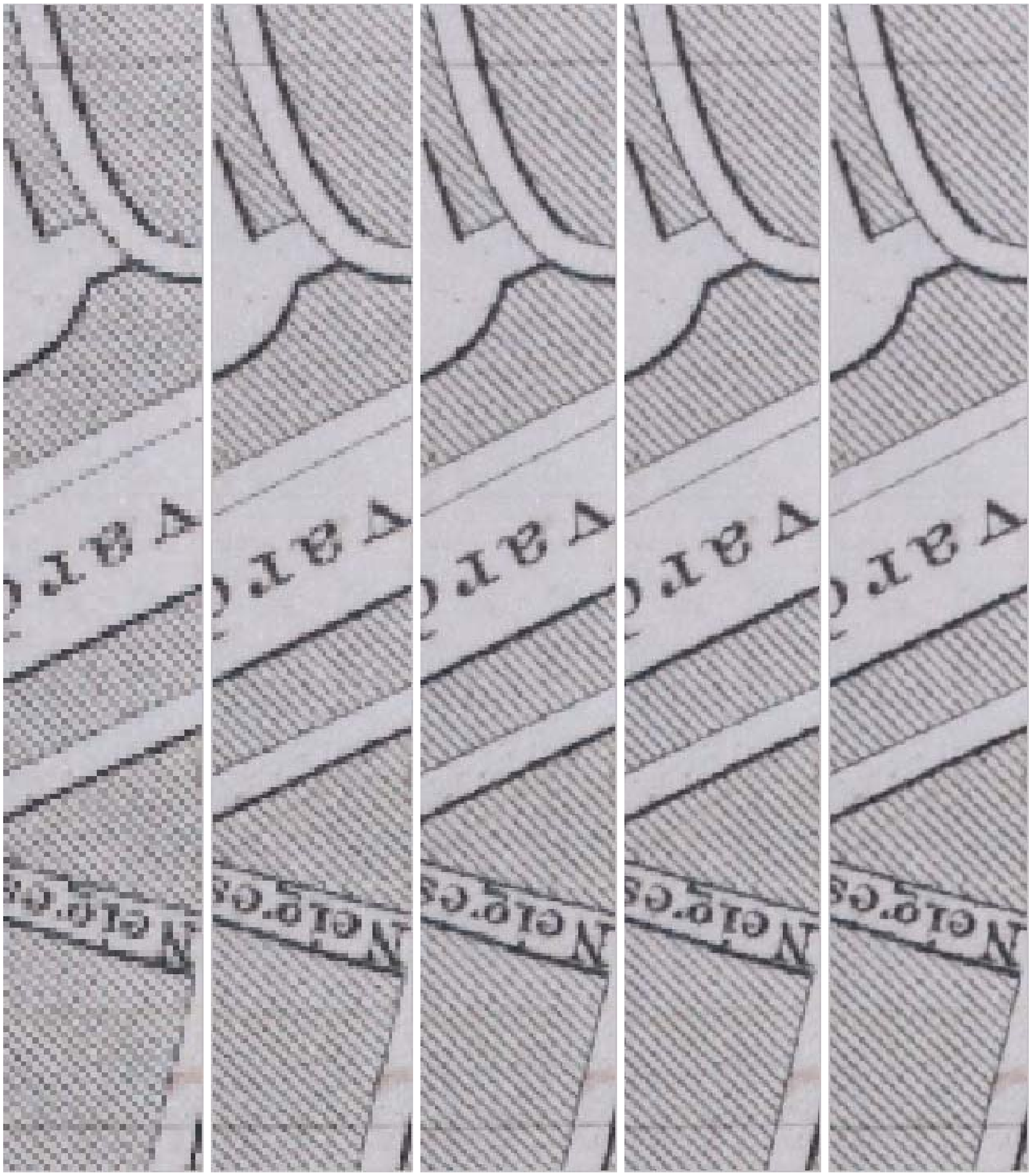
Example images (**download full set here**):

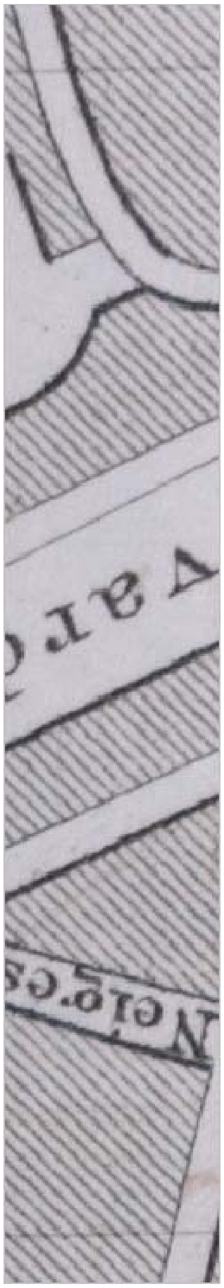


Above: Image increasing resolution from 200-ppi on the left to 700-ppi on the right.



Above: Image blown up 9x increasing resolution from 200-ppi on the left to 700-ppi on the right.





Above: Image blown up 50x increasing resolution from 200-ppi on the left to 700-ppi on the right.

[top]

Photographic Processes

	Min Resolution	Min Color Space	Min Bit Depth	Notes
Aerial Photographic Prints (<8"x10")	400	Gray	8	Use 24-bit color when appropriate. Resolution is based on the commonly used 4000 pixels along the long edge commonly used

	Min Resolution	Min Color Space	Min Bit Depth	Notes
				for photographic digitization. Use 4000 pixels whenever possible ultimately creating a higher resolution.
Aerial Photographic Prints (8"x10" - 11"x14")	400-600	Gray	8	Use 24-bit color when appropriate. Resolution is based on the commonly used 6000 pixels along the long edge commonly used for photographic digitization. Use 6000 pixels whenever possible ultimately creating a higher resolution.
Aerial Photographic Prints (>11"x14")	600	Gray	8	Use 24-bit color when appropriate. Resolution is based on the commonly used 8000 pixels along the long edge commonly used for photographic digitization. Use 8000 pixels whenever possible ultimately creating a higher resolution.
Aerial Photographic Film (70mm - 4"x5")	1200-2150	Gray	8	Use 24-bit color when appropriate such as infrared or false color UV. Resolution is based on the commonly used 6000 pixels along the long edge commonly used for photographic digitization. Use 6000 pixels whenever possible ultimately creating a higher resolution.
Aerial Photographic Film (4"x5" - 5"x7")	1200-1600	Gray	8	Use 24-bit color when appropriate such as infrared or false color UV. Resolution is based on the commonly used 8000 pixels along the long edge commonly used for photographic digitization. Use 8000 pixels whenever

	Min Resolution	Min Color Space	Min Bit Depth	Notes
				possible ultimately creating a higher resolution.
Aerial Photographic Film (>5"x7")	1450	Gray	8	Use 24-bit color when appropriate such as infrared or false color UV. Resolution is based on the commonly used 10000 pixels along the long edge commonly used for photographic digitization. Use 10000 pixels whenever possible ultimately creating a higher resolution.
Photographic Film (35mm - 4"x5")	800-2800	Gray / Color	8 / 24	Resolution is based on the commonly used 4000 pixels along the long edge commonly used for photographic digitization. Use 4000 pixels whenever possible ultimately creating a higher resolution.
Photographic Film (4"x5" - 8"x10")	800-1200	Gray / Color	8 / 24	Resolution is based on the commonly used 6000 pixels along the long edge commonly used for photographic digitization. Use 6000 pixels whenever possible ultimately creating a higher resolution.
Photographic Film (>8"x10)	800	Gray / Color	8 / 24	Resolution is based on the commonly used 8000 pixels along the long edge commonly used for photographic digitization. Use 8000 pixels whenever possible ultimately creating a higher resolution.
Photographic Prints (<8"x10")	400	Gray / Color	8 / 24	
Photographic Prints (8"x10" - 11"x14")	400-600	Gray / Color	8 / 24	Resolution is based on the commonly used 6000 pixels along the long edge

	Min Resolution	Min Color Space	Min Bit Depth	Notes
				commonly used for photographic digitization. Use 6000 pixels whenever possible ultimately creating a higher resolution.
Photographic Prints (>11"x14")	600	Gray / Color	8 / 24	Resolution is based on the commonly used 8000 pixels along the long edge commonly used for photographic digitization. Use 8000 pixels whenever possible ultimately creating a higher resolution.

Accurately reformatting historic photographs is among the most challenging of the static media types. Text based materials, and even most printing methods, have limits to the size of the smallest discrete elements. Photographs do too, the light sensitive graduals, but they may not be the most effective way to determine resolution. Prints and film are both representational media that can have subtleties in tones and colors. One should consider the intent for the digital object when selecting an adequate resolution. Digitizing for informational content is somewhat different than creating a surrogate for artifactual reasons where granules are clearly seen. Capturing the image such that the important elements are represented is usually adequate when one is concerned with the informational content, though it should be noted that photographs are commonly enlarged or magnified to view smaller elements clearly. The minimum recommendations given here are starting points. Increasing the resolution or bit depth may be required for several reasons. Additionally, it may be necessary to capture individual light sensitive granules when specific information on the original photographic process is important.

Digitization recommendations intended to only represent the image are commonly based on the size of the original. Because prints and film come in different sizes, and to keep images a reasonable file size, most film and prints are grouped in three sizes with 4,000, 6,000, and 8,000 pixels along the long edge for small, medium, and large items. Aerial film is an exception because the grain is much smaller and the artifacts that need to be represented are also smaller, digitized with 6,000, 8,000, and 10,000 pixels along the long edge. Aerial prints have similar grain as photographic prints, and can be captured similarly, though sometimes the discernible artifacts are smaller and may require higher resolution. There are situations where fine details are not captured adequately at the recommended minimum and resolution should be increased accordingly. True black and white images can be captured in grayscale, but color images are preferred for many photographic processes. Aerial photography is usually in black and white, but occasionally uses infrared or false UV film. These latter types should be imaged in color.

Many suggest that digitizing fine art and photographic objects require greater depth and should be imaged at 16-bit grayscale or 48-bit color. There is some evidence that capturing at this higher bit depth and reducing to 8-bit grayscale or 24-bit color provides better images than imaging directly to 8-bit grayscale or 24-bit color. It is frequently recommended to capture photographs in 16-bit grayscale or 48-bit color. There are good arguments for using these increased bit-depths for photographs, but the payoff is not obvious for all objects.

8-bit Grayscale bar

16-bit Grayscale bar

Capturing granular details of the photographic process is more difficult. This is not about the minimum requirements to adequately capture an object but rather closer to the maximum level an item should be digitized. Once the granules in a photograph are fully captured, there is essentially no more information that can be captured. Photographic prints alone constitute a wide array of surface emulsions from salt prints to albumen, cyanotypes, silver gelatin, printing out paper, carbon, Platinum/Palladium, Photogravure, Bromoil, C-prints and more. Photographs have been produced for over one hundred eighty years with an expansive variance in appearance. Accurately reformatting historic photographs is one of the most challenging digitization processes of Static Media types. The majority of B&W photographs populating photographic archives are silver gelatin prints stemming from the late 1890’s to present day. Color printing introduced dye layers creating additional considerations for reformatting. Investigations by the Library of Congress found that granular detail was lost when digitizing historic photographs at less than 1,200-ppi. Not all photographic prints require as high a resolution but one should try to avoid undersampling. Most historic photographic prints generically scanned result with unwanted artifacts due to uninformed workflows typified by flatbed scanning.

Photographic film scanning and post processing is technically challenging and analogous to printing in a darkroom that determines tonal values. High reformatting resolution, proper shadow and highlight values, color encoding and varied substrate issues are critical to proper workflows. Kodachrome is uniquely different than Ektachrome, color film from B&W,

glass plates and nitrate, Autochromes to Lantern Slides, all exhibit unique considerations.

Recent tests have shown that granules in black and white negatives from the first half of the 20th century are fully captured by 1,200-ppi but those in the second half of the century have smaller granules and need to be digitized at as high as 2,800-ppi before no further information can be captured. Some early color processes have much larger granules and captures higher than 750-ppi does not provide any more information.

Aerial Films are designed with extremely fine emulsion resolutions for recording precise subject details. Reformatting requires high resolution to preserve all subject information. The Library of Congress conducted a scientific resolution study by analyzing several aerial film types. The samplings resulted in determining the upper thresholds for reformatting resolutions. Information was discernible upwards of 2,200ppi or greater for the sample films.

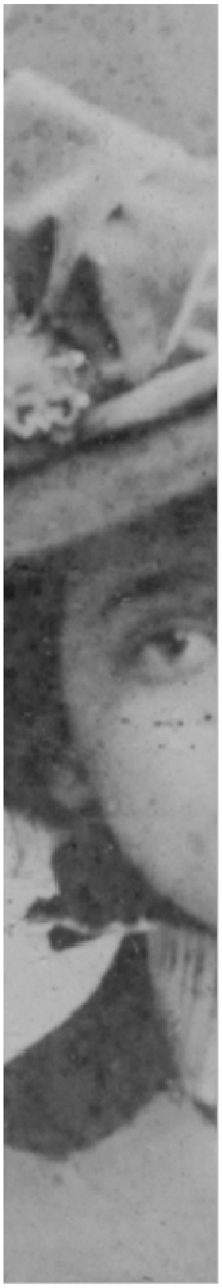
For more information on digitizing photographs, see the **suggested reading for digitizing static media**, or specifically for transmissive photographic media:

- Fleischhauer, Carl. “Information or Artifact: Digitizing Photographic Negatives and Transparencies, Part 1.” *The Signal*. Library of Congress, 18 Oct 2011. Web. 24 Apr. 2013. **http://blogs.loc.gov/digitalpreservation/2011/10/information-or-artifact-digitizing-photographic-negatives-and-transparencies-part-1/**.
- Fleischhauer, Carl. “Information or Artifact: Digitizing Photographic Negatives and Transparencies, Part 2.” *The Signal*. Library of Congress, 18 Oct 2011. Web. 24 Apr. 2013. **http://blogs.loc.gov/digitalpreservation/2011/10/information-or-artifact-digitizing-photographic-negatives-and-transparencies-part-2/**.
- Wheeler, Barry. “What Resolution Should I Use? Part 3.” *The Signal*. Library of Congress, 1 Mar 2013. Web. 24 Apr. 2013. **http://blogs.loc.gov/digitalpreservation/2012/12/what-resolution-should-i-use-part-1/**.
- Williams, Don, Michael Stelmach, and Steven Puglia. *Establishing Spatial Resolution Requirements for Digitizing Transmissive Content: A Use Case Approach*. Williamson: Image Science Associates, 2011. **http://www.imagescienceassociates.com/mm5/pubs/TransmissiveResolution_Williams.pdf**.

Example photographic print images (**download full set here**):

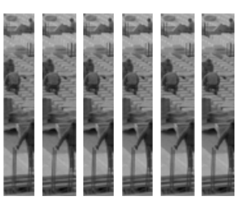
Above: Image increasing resolution from 200-ppi on the left to 700-ppi on the right.

Above: Image blown up 9x increasing resolution from 200-ppi on the left to 700-ppi on the right.



Above: Image blown up 50x increasing resolution from 200-ppi on the left to 700-ppi on the right.

Example photographic film images (**download full set here**):



Above: Image increasing resolution from 200-ppi on the left to 700-ppi on the right.

Above: Image blown up 9x increasing resolution from 200-ppi on the left to 700-ppi on the right.

Above: Image blown up 50x increasing resolution from 200-ppi on the left to 700-ppi on the right.

[top]

Posters/Broadsides/Oversize Documents

Min Resolution	Min Color Space	Min Bit Depth	Notes
300	Color	24	The off-set print resolution of posters and broadsides may present screening artifacts creating moiré patterns requiring adjustments. The off-set print resolution of posters and broadsides may present

Min Resolution	Min Color Space	Min Bit Depth	Notes
			screening artifacts creating moiré patterns requiring adjustments.

After finding equipment large enough to digitize them, the biggest hurdle with large documents is that the file sizes can become large and hard to manage. Most posters, broadsides, and oversize documents are meant to be viewed from a distance and therefore do not have smaller informational elements, though one may need to print from a digital file for exhibition, reproduction, or other purposes.

Most poster, broadsides, and oversize documents will be adequately digitized at 300-ppi in color or grayscale depending on whether the original has colors or shades that should be represented. This should allow for a quality print reproduction, of course some documents may require higher resolution. Moiré patterning may present problems, which can be overcome in the digitization process by changing the angle of capture, software, resolution or a combination of all three.

For more information on digitizing posters, broadsides, or oversize documents, see the **suggested reading for digitizing static media**.

[top]

Art on Paper

Min Resolution	Min Color Space	Min Bit Depth
400	Color	24

Art on paper is a broad category of materials covering many printing, drawing, and illustration techniques. One must be particularly careful selecting appropriate resolutions for the media and intended purpose of the digital product. Many researchers are interested in the production methods of these works so magnification is not uncommon.

Works of art on paper should not be imaged below 400-ppi, 24-bit color, but there are many instances where this will be inadequate and the resolution or bit depth should be increased. It is frequently recommended to capture fine art in 48-bit color. There are good arguments for using the increased bit-depth for fine art, but the payoff is not obvious for all objects. Moiré patterning may present problems, which can be overcome in the digitization process by changing the angle of capture, software, resolution or a combination of all three.

For more information on works of art on paper, see the **suggested reading for digitizing static media**.

[top]

Microforms

Min	Min	Min	Notes

Resolution	Color Space	Bit Depth	
300	Gray	8	Resolution should be calculated at 100% of the size of the original object. 300-ppi will suffice if the film has innately low resolution (e.g. some newspapers), but microfilm of non-oversize materials should be 400-ppi. Use 16-bit gray if the film is continuous tone or 24-bit color if the film is color.

Microforms are essentially photographic film with highly reduced images of textually based or manuscript materials. The film itself has a very fine grain allowing for extraordinary amount of detail in a small space. There can be great variation in the quality of microfilm depending on how it was filmed and the type of film used. In 1979 the first microfilm standard was published helping improve the quality overall.

The digitization resolution should be calculated by the size of the original document, not the film itself. Microforms created in accordance with preservation standards will state the reduction ratio on a frame at the beginning of the roll. Because digitizing from microfilm is not imaged directly from an original, there is some concern of ill-defined letters becoming less legible because of poor registration. Most microforms should be digitized at 400-ppi with 8-bit grayscale, which accounts for some imperfection in the image quality. Very poor quality film may be adequately digitized at 300-ppi as long as reduction in quality does not further degrade the image. Continuous tone film should be scanned at 16-bit grayscale and color film at 24-bit color.

For more information on digitizing microforms, see the **suggested reading for digitizing static media**.

[top]

Three-Dimensional Objects

Min Resolution	Min Color Space	Min Bit Depth
300	Color	24

The intended purpose of imaging a three-dimensional object is very different from the intent of many other types of digitization. Even with the rise of inexpensive 3D printers, one cannot currently make an adequate reproduction from a three-dimensional object, and so the intent is not to digitize so users have a surrogate to use, but rather to give the user general information about the object. Three-dimensional objects will most likely be reimaged at a later point. In libraries, three-dimensional objects are typically photographed at the native camera resolution of 300ppi. The size of the sensor and size of the capture area determine the achieved resolution. It is not uncommon that 3D objects receive several views (potentially presented as a rotating object). Lighting and camera angle of view are essential considerations and details of specific regions may be useful.

For more information on digitizing three-dimensional objects, see the **suggested reading for digitizing static media**.

[top]

Time-Based Media

The time element of time-based media makes it different from other analog materials. On a basic level, they are comprised of tiny elements that fit in a defined order and are each perceived for a defined period of time. The object is the whole that can only be experienced over a period of time. Similar to how a digital image is composed of small pixels, each frame of a video, or tone in a sound recording, is one small piece that requires all of the others to make a complete work. While digitizing time based media, one must be aware of the perceptual limitations, not only with each individual unit, but also the limits in how one perceives changes over time. Time-based media uses this to try and string individual elements into a single continuous work. Some time-based media, like moving images, can incorporate several layers such as visual and audio elements.

Digitizing time-based media has not been performed as regularly nor for as long as static media so there are fewer articles and publications discussing the issues and nuances. While there is still some interpretation when digitizing analog media, because it is machine based and because of the limitations of the original media, many of the debates of how fine a resolution is necessary, even in terms of a maximum, is more agreed upon than some of the static media types.

[top]

Audio

Min Resolution	Min Bit Depth
96000 kHz	24

In order to experience an audio recording, a speaker must create compression waves that will move small bones in the ear that will then be perceived as sound. Audio analog to digital converters now capture frequencies beyond the range of human perception, and dynamic range at the limits of the laws of physics. Further “improvements” are not likely to have any measurable or perceptible improvement.

The audio community has coalesced around digitizing analog audio at 96kHz with 24-bits per sample. There are arguments for digitizing some types of audio sources at lower quality, but for consistency and standardization, most institutions comply with this standard for all sources. There is doubt to whether analog audio, or any audio for that matter, would need to be captured with higher quality.

In audio digitization a bit is equal to 6dB of dynamic range. The more bits the more dynamic range you can capture. Also called signal to noise ratio, it is a measure of the range of signals from full scale (saturation) to the smallest signal that can be resolved. A 24 bit system can capture 144dB of dynamic range. At room temperature an ideal electrical circuit will have a theoretical dynamic range of about 130dB, or about 21.5 bits. Even though we cannot capture 144dBs we use 24 bits because computers work in bytes that are groupings of bits. There are 8 bits in a byte, and 3 bytes is 24 bits.

Digital audio should be migrated natively whenever possible. It is useless to convert it with different resolution and bit-depth from the original if a file migration is necessary.

For more information on digitizing audio:

- Bradley, Kevin, ed. "Guidelines on the Production and Preservation of Digital Audio Objects (web edition)." International Association of Sound and Audiovisual Archives, n.d. Web. 20 Feb 2013. **http://www.iasa-web.org/tc04/audio-preservation**.
- "Guidelines: Audio Digitization System Performance." *Federal Agencies Digitization Guidelines Initiative*. FADGI, 15 Mar 2011. Web. 8 Jun 2012. **http://www.digitizationguidelines.gov/guidelines/digitize-audioperf.html**.
- "Noise Calculator." *Sengpielaudio*. N.p., n.d. Web. 8 Jun 2012. **http://www.sengpielaudio.com/calculator-noise.htm**.
- Robin, Michael. "Digitizing Audio." *Broadcast Engineering*. May. (2006): n. page. Web. 20 Feb. 2013. **http://broadcastengineering.com/mag/digitizing-audio-0**.
- Schüller, Dietrich. *Safeguarding of the Audio Heritage: Ethics, Principles and Preservation Strategy*. Version 2. International Association of Sound and Audiovisual Archives, 2005. Web. **http://www.iasa-web.org/sites/default/files/downloads/publications/TC03_English.pdf**.

Analog calibration references are available from:

- *Audio Analog-to-Digital Converter Performance Specification and Test Method*. Federal Agencies Digitization Guidelines Initiative, 2012. Web. **http://www.digitizationguidelines.gov/audio-visual/documents/ADC_performGuide_20120820.pdf**.
- "MRL Calibration Tapes." *Magnetic Reference Laboratory*. N.p., 19 Feb 2010. Web. 8 Jun 2012. **http://home.comcast.net/~mrltapes/**.
- "78 rpm Calibration Disc Set." *Audio Engineering Society(AES)*. Audio Engineering Society, 2012. Web. 8 Jun 2012. **http://www.aes.org/publications/standards/calibration.cfm**.

[top]

Video

	Min Resolution	Min Color Space	Min Bit Depth	Notes
Analog NTSC Video	720 x 486		8	Use 10-bit whenever possible.
Digital Video Source Tape, where possible to access bits	Native		Native	
Digital Video Source Tape, where not possible to access bits	Decompressed 720 x 486		8 or 10	

	Min Resolution	Min Color Space	Min Bit Depth	Notes
Digital Video File	Native			Subject to file format obsolescence evaluation. If deemed obsolete, decompress to 10-bit native raster (horizontal x vertical pixel count).
Video Optical Disc	Native		Native	Reformat to ISO disc image to capture all video, all angles, all subtitle and multiple languages, and menus.

The electronic signal that is video is organized in ways that both facilitate digitization and thwart efforts to do so efficiently. Like film, analog video is divided into discrete elements called frames. The image portion of these frames is further organized into 486 separate horizontal lines. These discrete elements, frames and lines, lend themselves naturally to the discrete nature of digitizing otherwise analog content. Along the lines, however, there is high frequency information modulated (“encoded”) in the electronic signal. In the analog domain this information scans an electron beam across a field of multi-colored phosphors to produce an image on a cathode ray tube (CRT). Video digitization captures 720 pixel elements along each line. In essence, uncompressed video digitization is equivalent to producing 30 TIFFs each second, with a resolution of 720 x 486 and an additional audio track.

The resolution of each pixel falls within a fixed range. As bits are added to resolution, the fixed range is successively subdivided into finer and finer shades and colors. By comparison in audio a bit has a fixed range, 6dB, and the more bits you add the greater the dynamic range. If we have 2 bits of video resolution, we have only black and white values. As we move through 3, 4 and higher resolutions we have finer and finer gradations of gray as well as the extreme values of black and white. In some cases, 8 bits may suffice, but to avoid banding artifacts within the limits of human perception, 10 bits of resolution are required. Static image grayscale scanning works exactly the same way.

Lossy compression is always bad in digitizing analog video. Even “visually lossless” compression is a bad thing. Compression algorithms change over time. The compromises that one compression codec utilizes to fool the eye today may be exaggerated in the codec of tomorrow, yielding visible artifacts. These artifacts will be uncorrelated to the picture making them even more visible and annoying.

Most born-digital video is compressed at inception. As with all born-digital objects an institution must decide if they are going to support that file format and codec or not. If they are, store the digits as they are. If the file format and codec are not going to be supported within the institution's repository, the file should be decoded/decompressed to a similar frame size and bit depth as the original and captured as an uncompressed video to avoid generational loss.

For more information about digitizing video, please see “**Refining Conversion Contract**

Specifications: Determining Video Formats for Medium-term Storage," a white paper commissioned by the Library of Congress on specifications for digitizing video.

[top]

Moving Image Film

Film is the last area where there are large numbers of people who believe analog duplication, that is film to film, is the proper form of preservation. Preservation on film is expensive, time consuming, creates hazardous waste and is subject to the same challenges of all analog-to-analog duplication.

There is significant disagreement whether digitization will ever create an appropriate surrogate for film. One side argues it is only a matter of time until the resolution of digitization surpasses the amount of information that can be captured in photographic film. The other side argues digital projection will never match luminesce of light projected through film onto a screen. Digits may capture the information on film, but it will never reproduce the experience of projected film.

We may look back upon two recent events as catalysts for moving into a file-based solution for film preservation. Recently, Kodak filed for bankruptcy and an earthquake in Japan shut down the only factory where HDCAM tape is manufactured. The demise of Kodak removes a major player in the manufacturer and processing of film. The inability to purchase tape for $130,000 professional video machines hastened the migration toward file-based workflows. As data storage has become larger and cheaper, the economics of working on computers are fast becoming significantly less than purchasing, handling, editing and storing film.

While major motion picture studios can be expected to continue to do high end color separation, photo-chemical preservation of major films, and to care for the analog preservation work performed to date, the rapidly falling cost and increasing resolution of digitization will make it harder and harder for smaller institutions to justify film-to-film preservation.

The authors feel that there are currently too many unknowns to make a well informed recommendation on digitalizing moving image film at this time.

This topic is discussed in detail in the following document produced by the Academy of Motion Pictures Arts and Sciences:

- "Digital Dilemma 2." Academy of Motion Picture Arts and Sciences, 2012. Web. 8 Jun 2012. **http://origin-www.oscars.org/science-technology/council/projects /digitaldilemma2/download.php**.

[top]

Recommended Minimum Capture Summary

	Min Resolution	Min Color Space	Min Bit Depth	Notes

	Min Resolution	**Min Color Space**	**Min Bit Depth**	**Notes**
Books and Textual Based Materials Without Images (non-rare)	300	Gray	8	Capture in color (24 bit) whenever possible. 300-ppi will capture details larger than 1.4 mm at a QI of 8. The resolution may be adequately adjusted according to the largest detail to be represented.
Books and Textual Base Documents With Images (non-rare)	400	Gray	8	Capture in color (24 bit) whenever possible. Fine lines in etchings or other pictorial elements require more definition than text alone. 400-ppi will capture details larger than 1 mm at a QI of 8. The resolution may be adequately adjusted according to the largest detail to be represented.
Manuscripts	400	Color	24	Illegible or difficult to read scripts may require a higher resolution.
Microforms	300	Gray	8	Resolution should be calculated at 100% of the size of the original object. 300-ppi will suffice if the film has innately low resolution (e.g. some newspapers), but microfilm of non-oversize materials should be 400-ppi. Use 16-bit gray if the film is continuous tone or 24-bit color if the film is color.
Rare Books	400	Color	24	

	Min Resolution	Min Color Space	Min Bit Depth	Notes
3D Objects	300	Color	24	
Aerial Photographic Prints (<8"x10")	400	Gray	8	Use 24-bit color when appropriate. Resolution is based on the commonly used 4000 pixels along the long edge commonly used for photographic digitization. Use 4000 pixels whenever possible ultimately creating a higher resolution.
Aerial Photographic Pritns (8"x10" - 11"x14")	400-600	Gray	8	Use 24-bit color when appropriate. Resolution is based on the commonly used 6000 pixels along the long edge commonly used for photographic digitization. Use 6000 pixels whenever possible ultimately creating a higher resolution.
Aerial Photographic Prints (>11"x14")	600	Gray	8	Use 24-bit color when appropriate. Resolution is based on the commonly used 8000 pixels along the long edge commonly used for photographic digitization. Use 8000 pixels whenever possible ultimately creating a higher resolution.
Aerial Photographic Film (70mm - 4"x5")	1200-2150	Gray	8	Use 24-bit color when appropriate such as infrared or false color UV. Resolution is based on the commonly used 6000 pixels along the long edge commonly used

	Min Resolution	Min Color Space	Min Bit Depth	Notes
				for photographic digitization. Use 6000 pixels whenever possible ultimately creating a higher resolution.
Aerial Photographic Film (4"x5" - 5"x7")	1200-1600	Gray	8	Use 24-bit color when appropriate such as infrared or false color UV. Resolution is based on the commonly used 8000 pixels along the long edge commonly used for photographic digitization. Use 8000 pixels whenever possible ultimately creating a higher resolution.
Aerial Photographic Film (>5"x7")	1450	Gray	8	Use 24-bit color when appropriate such as infrared or false color UV. Resolution is based on the commonly used 10000 pixels along the long edge commonly used for photographic digitization. Use 10000 pixels whenever possible ultimately creating a higher resolution.
Art on Paper	400	Color	24	
Photographic Film (35mm-4"x5")	800-2800	Gray / Color	8 / 24	Resolution is based on the commonly used 4000 pixels along the long edge commonly used for photographic digitization. Use 4000 pixels whenever possible ultimately creating a higher resolution.

	Min Resolution	Min Color Space	Min Bit Depth	Notes
Photographic Film (4"x5" - 8"x10")	800-1200	Gray / Color	8 / 24	Resolution is based on the commonly used 6000 pixels along the long edge commonly used for photographic digitization. Use 6000 pixels whenever possible ultimately creating a higher resolution.
Photographic Film (>8"x10")	800	Gray / Color	8 / 24	Resolution is based on the commonly used 8000 pixels along the long edge commonly used for photographic digitization. Use 8000 pixels whenever possible ultimately creating a higher resolution.
Photographic Prints (<8"x10")	400	Gray / Color	8 / 24	
Photographic Prints (8"x10" - 11"x14")	400-600	Gray / Color	8 / 24	Resolution is based on the commonly used 6000 pixels along the long edge commonly used for photographic digitization. Use 6000 pixels whenever possible ultimately creating a higher resolution.
Photographic Prints (>11"x14")	600	Gray / Color	8 / 24	Resolution is based on the commonly used 8000 pixels along the long edge commonly used for photographic digitization. Use 8000 pixels whenever possible ultimately creating a higher resolution.
Oversize Documents	300	Gray / Color	8 / 24	

	Min Resolution	Min Color Space	Min Bit Depth	Notes
Maps	300-600	Gray / Color	8 / 24	600 ppi will capture highly detailed information and good for reprinting and fine quality maps. Lower resolutions may be appropriate if detail is limited (large print, etc...) or will not be printed.
Posters / Broadsides	300	Color	24	The off-set print resolution of posters and broadsides may present screening artifacts creating moiré patterns requiring adjustments.
Audio	96000		24	
Analog NTSC Video	720 x 486		8	Use 10-bit whenever possible.
Digital Video Source Tape, where possible to access bits	Native		Native	
Digital Video Source Tape, where not possible to access bits	Decompressed 720 x 486		8 or 10	
Digital Video File	Native			Subject to file format obsolescence evaluation. If deemed obsolete, decompress to 10-bit native raster (horizontal x vertical pixel count).
Video Optical Disc	Native		Native	Reformat to ISO disc image to capture all video, all angles, all subtitle and multiple languages, and menus.

[top]

Appendix I: File Naming Conventions for Digital Collections

Electronic files should be well organized and named in such a way that they are easily identifiable and accessible. The examples in this appendix center primarily on documents, but the guidelines below can easily apply or be adapted to all file formats. The guidelines are considered best practices, however, not all may be relevant to everyone or every situation. These do provide groundwork for designing a consistent and easy to use file-naming standard when creating digital objects.

1. Use only alpha-numeric characters for both files and folders.
    - Exceptions are dashes (-) and underscores (_). Periods (.)_should be used to separate the base filename from the file extension.
2. Do not use special characters.
    - Examples include, but are not limited to: / > < + = ' ^ | \ { } [ ] # , ; ? ! $ * & these characters are used by the operating system.
3. Use a valid file extension, which is usually has three and sometimes four characters (i.e. .tif .pdf .html, etc.).
4. Do not use spaces in file/folder names – use dashes or underscores instead.
5. Use leading zeros.
    - When file names need numbering, use zeros as placeholders. Example – a collection of 900 items should be numbered: abc001.tif, abc002.tif, etc.
    - This will facilitate proper sorting.
6. When dates are used, choose a standard format.
    - Example: yyyymmdd, 20110501 or 2001_05_01
7. Be brief! Not all systems are the same, or allow for lengthy file names. Long filenames may prohibit portability.
8. File names should contain necessary descriptive information, independent of their storage location.
    - Examples:
        - jones_diary_18900420_0001.tif
        - Not: 00001.tif
9. If versioning is needed, include a version numeral.
    - v01, v02, etc.
    - Suggestions: for the final version, include the word “final”, without a version number.

[top]

Appendix II: Metadata

Hundreds of metadata standards exist within the broader cultural heritage community. A fantastic overview of these often interrelated standards is available in the new resource, Seeing Standards, a Visualization of Metadata Universe[6]. Minimally, this should be basic descriptive and technical metadata collection sufficient to allow retrieval and management of the digital copies and to provide basic contextual information for the user. Additionally, the inclusion of preservation metadata through the use of PREMIS is advisable.

Below are the applicable technical and preservation metadata standards related to these recommendations.

General Standards

- *A Framework of Guidance for Building Good Digital Collections*. 3rd ed. Baltimore: National Information Standards Organization, 2007. Web. **http://www.niso.org/publications/rp/framework3.pdf**.[7]
- "Metadata Encoding and Transmission Standard." Library of Congress, Apr 2010. Web. 11 Jun 2012. **http://www.loc.gov/standards/mets/mets-schemadocs.html**.

Technical

Still Images:

- "Metadata for Images in XML Standard (MIX)." *NISO Metadata for Images in XML Schema*. Library of Congress, 18 May 2008. Web. 8 Jun 2012. **http://www.loc.gov/standards/mix/**.
- NISO Z39.87 2006 (R2011) *Data Dictionary - Technical Metadata for Digital Still Images*. Bethesda: National Information Standards Organization, 2006. Web. **http://www.niso.org/apps/group_public/download.php/6502/Data Dictionary - Technical Metadata for Digital Still Images.pdf**.

Audio:

- *AES Standard for Audio Metadata- audio object structures for preservation and restoration*. New York: Audio Engineering Society, 2011. Web. **http://www.aes.org/publications/standards/search.cfm?docID=8**.
- *AES Standard for Audio Metadata- core audio metadata*. New York: Audio Engineering Society, 2011. Web. **http://www.aes.org/publications/standards/search.cfm?docID=85**.
- "Audio Technical Metadata Extension Schema." Library of Congress, 02 May 2011. Web. 8 Jun 2012. **http://www.loc.gov/standards/amdvmd/audioMD.xsd**.
- "78 rpm Calibration Disc Set." *Audio Engineering Society (AES)*. Audio Engineering Society, 2012. Web. 8 Jun 2012. **http://www.aes.org/publications/standards/calibration.cfm**.

Video:

- "Schema Documentation for videoMD.xsd." Library of Congress, n.d. Web. 8 Jun 2012. **http://www.loc.gov/standards/amdvmd/htmldoc/videoMD.html**.

Preservation (PREMIS)

- *PREMIS Data Dictionary for Preservation Metadata*. version 2.2. Washington DC: Library of Congress, 2011. Web. **http://www.loc.gov/standards/premis/v2/premis-2-2.pdf**.
- "PREMIS: Preservation Metadata Maintenance Activity." Library of Congress, 1 Jun 2012. Web. 8 Jun 2012. **http://www.loc.gov/standards/premis/**.
- "Schemas for PREMIS." *PREMIS: preservation metadata maintenance activity*. Library of Congress, 16 May 2012. Web. 8 Jun 2012. **http://www.loc.gov/standards/premis/schemas.html**.

Structural

Still Image:

- *Best Practices for Structural Metadata*. New Haven: Yale University Library, 2008. Web. **http://www.library.yale.edu/dpip/bestpractices/BestPracticesForStructuralMetadata.pdf**.

Descriptive

General:

- *Descriptive Metadata Guidelines*. Mountain View: Research Libraries Group, 2005. Web. **http://www.oclc.org/research/activities/past/rlg/culturalmaterials/RLG_desc_metadata.pdf**.
- "Dublin Core Metadata Element Set, Version 1.1." *Dublin Core Metadata Initiative*. Dublin Core, 11 Oct 2010. Web. 11 Jun 2012. **http://dublincore.org/documents/dces/**.
- *Guidelines for the Creation of Digital Collections: best practices for descriptive metadata*. Champaign: Consortium of Academic and Research Libraries in Illinois, Web. **http://www.carli.illinois.edu/comms/board/Metadata_guidelines.pdf**.
- "Metadata Object Description Schema: MODS." Library of Congress, 18 Oct 2011. Web. 11 Jun 2012. **http://www.loc.gov/standards/mods/**.

Still Image:

- *DIG35 Specification: metadata for digital images*. Boynton Beach: Digital Imaging Group, 2001. Web. **http://www.i3a.org/wp-content/uploads/docs/dig35-v1.1.pdf**.

Audio/Visual:

- *Access Options for Embedding Metadata in WAVE Files and Plan the Audio Metadata File Header Tool Development Project*. New York: AudioVisual Preservation Solutions, 2009. Web. **http://www.digitizationguidelines.gov/audio-visual/documents/AVPS_Audio_Metadata_Overview_090612.pdf**.
- *Embedding Metadata in Digital Audio Files*. Washington DC: FADGI, 2012. Web. **http://www.digitizationguidelines.gov/audio-visual/documents/Embed_Guideline_20120423.pdf**.
- "Information Technology - MPEG-21 Multimedia Framework." Klagenfurt University, 03 Jan 2005. Web. 11 Jun 2012. **http://mpeg-21.itec.uni-klu.ac.at/cocoon/mpeg21/_mpeg21Parts.html**.

[top]

Appendix III: Storage

There are two kinds of preservationists: those who have lost data and those who will. Losing data isn't in and of itself a crisis. It should be little more than a short term inconvenience. Data can be lost due to four causes: an inability to access the information due to file format obsolescence, corruption of data, failure of the physical carrier, and accidental deletion. File format obsolescence is beyond the scope of this document. For more information on tools for checking a file's viability explore JHOVE, DROID, and PRONOM. The solution to both of the other two cases is to retrieve the data from a second copy. The first action, then of any preservation strategy is to have more than one copy. Ideally that additional copy (or those additional copies) should be geographically isolated, and be on a different storage technology. Examples of two different technologies would be hard disc drives and data tape. A system to manage the data for preservation consists of a method to detect errors due to corruption and errors due to loss, either to erasure or media failure. The next stage is to have a system that both self-monitors and when loss is discovered, automatically replaces the loss.

OAIS describes such a system, including the metadata for discovery and management of the content.

- *Open Archival Information System Reference Model: introductory guide*. Dublin: Online Computer Library Center, 2004. Web. **www.dpconline.org/docs/lavoie_OAIS.pdf**.
- *Reference Model for an Open Archival Information System (OAIS)*. PINK BOOK: draft recommended standard. Washington DC: Consultative Committee for Space Data Systems, 2009. Web. **http://public.ccsds.org/sites/cwe/rids/Lists/CCSDS 6500P11/Attachments/650x0p11.pdf**.

The most widely known implementation of the data integrity management portion of OAIS is LOCKSS. Multiple copies of data are stored in disparate locations, with each node monitoring its own health and reporting to the other nodes. When there is loss, the other nodes provide a replacement copy of the lost or corrupted file. If a node is making frequent requests for replacements it is deemed untrustworthy, administrators are notified automatically, and corrective action is taken. LOCKSS can be installed and configured in less than an hour.

- "LOCKSS: Lots of Copies Keep Stuff Safe." Stanford University, n.d. Web. 8 Jun 2012. **http://www.lockss.org/**.

Preservation systems apply “fixity” values to assure files have not changed, and thereby guarantee authenticity. MD5 and SHA-1 checksums are commonly used fixity algorithms. Any computer file, regardless of size, is fed into the algorithm. It generates a “hash” value. If a single bit, or multiple bits are changed within the file, a different hash value results. Due to a high “avalanche” or “cumulative error”, the values are very different due to even a small change; that is, it’s easy to tell the hash value has changed. It is theoretically possible for multiple bits to change and the same hash value to result. However, you are 50 million times more likely to be struck by lightning twice than for this to happen. It is possible for two files to have the same hash value, but the probability is very low. The MD5 algorithm has fewer combinations than SHA-1 but there are still 340,000,000,000,000,000,000,000,000,000,000,000,000 unique combinations. As a practical matter, the high improbability of accidental hash duplication, combined with the extremely low cost of implementation, makes these fixity algorithms highly useful for preservation.

As defined by the Preservation and Reformatting Section of the American Library Association, a key tenant of digital preservation is migration.

- "Definitions of Digital Preservation." Association for Library Collections and Technical Services, 24 Jun 2007. Web. 8 Jun 2012. **http://www.ala.org/alcts/resources/preserv /defdigpres0408**.

Within the window of obsolescence, typically approximately 5 years, data files are migrated to the next generation of storage technology. At the point of migration you confirm the readability of the media by the act of access the file, verify authenticity with checksums, evaluate the file format obsolescence status using file validation tools, perform authority control on file names and embedded metadata, update metadata and perform preservation file retention actions, update checksums as needed, and copy to new media. As computer processing speeds have gotten faster, only verifying and generating new checksums consume much time. Indeed most of the other actions would not be noticeably slower than a simple file copy.

If all you do is copy your master files to another storage medium, by having a second copy in storage at another location, you will have performed an important first step in digital preservation. This first step is more important than having a fully deployed OAIS environment. But it is only the first step.

By exploring the Weblinks and terms above you can find many resources to learn and implement these strategies.

[top]

Appendix IV: Institutional Guidelines

- **Bibliographical Center for Research** - Audio, Artwork, Graphic Art, Text Documents, Maps, Photographs, 3D Objects
  http://www.bcr.org/cdp/best/digital-imaging-bp.pdf
  http://www.bcr.org/cdp/best/digital-audio-bp.pdf

- **California Digital Library** - Aerial Photographs, Color Documents, Manuscripts, Photographs, Text Documents, 3D objects
  http://www.cdlib.org/services/dsc/tools/docs/cdl_gdi_v2.pdf

- **Columbia University** - Illustrations, Transparencies, Manuscripts, Maps, Microfilm, Photographs, Pictorial Materials, Printed Text, 3D Objects
  http://www.columbia.edu/acis/dl/imagespec.html

- **Cornell University** - Artworks, Graphic Art, Manuscripts, Maps, Microfilm, Photographs, Printed Text, Rare Books, Transparencies
  http://www.library.cornell.edu/imls/image%20deposit%20guidelines.pdf
  http://www.library.cornell.edu/preservation/tutorial/conversion/table3-1.html

- **Digital Library Federation** - Color Documents, Illustrations, Printed Text
  http://www.diglib.org/standards/bmarkfin.htm

- **Digital Library of Georgia** - Manuscripts, Photographs, Printed Text
  http://dlg.galileo.usg.edu/guide.html#04

- **Digital Library of the Caribbean** - Graphic Art, Manuscripts, Maps, Oversize Documents, Photographs, Printed Text
  http://ufdc.ufl.edu/design/aggregations/dloc/training/en/Section6.pdf

- **Federal Agencies Digitization Guidelines Initiative** - Artworks, Aerial Photographs, Color Documents, Film, Illustrations, Photographs, Printed Text, 3D Objects
  http://www.digitizationguidelines.gov/guidelines/FADGI_Still_Image-Tech_Guidelines_2010-08-24.pdf

- **Government Publications Office** - Film, Illustrations, Photographs, Printed Text

http://www.gpoaccess.gov/legacy/registry/DigitizationSpecification3.0.pdf

- **Historical Voices** - Audio
  http://www.historicalvoices.org/papers/audio_digitization.pdf

- **Indiana University** - Audio
  http://www.dlib.indiana.edu/projects/sounddirections/papersPresent/sd_bp_07.pdf

- **International Association of Sound and Audiovisual Archives** - Audio
  http://www.iasa-web.org/downloads/publications/TC03_English.pdf

- **Johns Hopkins University** - Audio, Illustrations, Manuscripts, Photographs, Printed Text, Video
  http://www.library.jhu.edu/collections/institutionalrepository/irpreservationpolicy.html#Standards

- **Library of Congress** - Audio, Graphic Art, Illustrations, Manuscripts, Maps, Microfilm, Music Scores, Photographs, Printed Text, Rare Books
  http://memory.loc.gov/ammem/about/standardsTable1.pdf
  http://www.loc.gov/rr/mopic/avprot/audioSOW.html
  http://memory.loc.gov/ammem/dli2/html/document.html#page
  http://www.loc.gov/ndnp/pdf/NDNP_200911TechNotes.pdf

- **Memory of the Netherlands Project** - Audio, Color Documents, Illustrations, Microfilm, Moving Images, Printed Text, Video
  http://www.kb.nl/coop/geheugen/extra/downloads/Richtlijnen%20en%20procedures%20%204.0%20english.pdf

- **Michigan State University** - Audio
  http://www.historicalvoices.org/papers/audio_digitization.pdf

- **National Anthropological Archives** - Images (General)
  http://voom.si.edu/anthro/imaging_standards.htm

- **National Archives and Records Administration** - Aerial Photographs, Color Documents, Manuscripts, Oversize Materials, Photographs, Printed Text, Pictorial Materials
  http://www.archives.gov/preservation/technical/guidelines.pdf

- **National Library of Australia** - Artworks, Color Documents, Illustrations, Manuscripts, Maps, Microfilm, Music Scores, Photographs, Printed Text, Transparencies
  http://www.nla.gov.au/digital/capture.html

- **New York State Archives** - Color Documents, Illustrations, Photographs, Printed Text
  http://www.archives.nysed.gov/a/records/mr_erecords_imgguides.pdf

- **New York University** - Audio, Video
  http://www.nyu.edu/its/humanities/ninchguide/VII/
  http://aic.stanford.edu/sg/emg/library/pdf/mcdonough/McDonough-EMG2004.pdf

- **New Zealand Archives** - Illustrations, Microfilm, Photographs, Printed Text
  http://continuum.archives.govt.nz/files/file/standards/s6/s6-app5.html
- **NISO Framework Working Group** - Audio, Illustrations, Manuscripts, Moving Image Film, Oversize Documents, Photographs, Printed Text, Video
  http://www.niso.org/publications/rp/framework3.pdf
- **Oxford University** - Color Documents, Manuscripts, Paintings, Photographs, Printed Text, Rare Books, Transparencies
  http://www.odl.ox.ac.uk/papers/odlprice.pdf
- **Penn State University** - Color Documents, Illustrations, Printed Text
  http://www.libraries.psu.edu/psul/digipres/bestpractices.html
- **Rutgers University** - Audio, Color Documents, Photographs, Printed Text, Video
  http://rucore.libraries.rutgers.edu/collab/ref/dos_avwg_video_obj_standard.pdf
  http://rucore.libraries.rutgers.edu/collab/ref/dos_repteam_digitization_of_analog_formats.pdf
  http://rucore.libraries.rutgers.edu/collab/ref/dos_avwg_audio_obj_standard.pdf
- **Smithsonian Institution Archives** - Images (General)
  http://siarchives.si.edu/records/electronic_records/records_erecords_digitization_images.html
- **Stanford University** - Audio, Video
  http://lib.stanford.edu/node/8544/moving-image-digitization
  http://lib.stanford.edu/node/8544/audio-digitization
- **State Library of Queensland** - Audio, Images (general), Video
  http://www.slq.qld.gov.au/__data/assets/pdf_file/0009/139815/SLQ-_DS2Capture_v2.05.pdf
- **University of California, Berkeley** - Imaging (general)
  http://www.lib.berkeley.edu/LSO/digital_imaging_best_practices
- **University of Illinois at Urbana-Champaign** - Artworks, Illustrations, Manuscripts, Maps, Microfilm, Photographs, Printed Text
  http://images.library.uiuc.edu/resources/digitalguidev3.pdf
- **University of Maryland Libraries** - Artworks, Audio, Graphic Arts, Illustrations, Manuscripts, Photographs, Printed Text, Rare Books, Transparencies
  http://www.lib.umd.edu/dcr/publications/best_practice.pdf
- **University of Michigan** - Illustrations, Photographs, Printed Text
  http://www.lib.umich.edu/files/UMichDigitizationSpecifications20070501.pdf
  http://www.lib.umich.edu/lit/dlps/dcs/UMichDigitizationSpecifications20070501.pdf
- **University of Southern California** - Audio, Microfilm, Photographs, Printed Text, Video
  http://www.usc.edu/libraries/collections/digitallibrary/documents/USCDL_Formats_and_Resolutions.pdf
  http://www.usc.edu/libraries/collections/digitallibrary/documents

/USCDL_Audiovisual_Digitization_Overview.pdf

- **University of Southern Mississippi Libraries Digital Program** - Images (General)
  http://www.lib.usm.edu/legacy/spcol/crda/guidelines/index.html

- **University of Virginia** - Images (General)
  http://guides.lib.virginia.edu/digitization-guidelines
  http://etext.lib.virginia.edu/services/helpsheets/scan/specscan.html

- **University of Wisconsin** - Audio
  http://uwdcc.library.wisc.edu/documents/AudioWorkflow.pdf

- **Yale University** - Audio, Graphic Arts, Maps, Photographs, Printed Text
  http://www.yale.edu/digitalcoffee/downloads/DigitalCoffee_SharedPractices_%5Bv1.0%5D.pdf

[top]

Notes

1. Pixels per inch. There is frequently confusion between PPI and DPI (dot per inch). A pixel is a representation on a computer screen; this pixel can be of any color. A dot references how something is printed. A printer has a limited amount of colors (usually three plus black) that need to be mixed in different ratios to create other colors. Since this document discusses capturing images and not printing them, it uses PPI. **[return to text]**
2. The University of Pennsylvania provided the images used in the examples. A special thanks to Chris Lippa for having the materials photographed to the specifications required to make the examples. **[return to text]**
3. Actual size will depend on the screen settings. The original size of each example is 1" high. If you adjust your adjust your monitor's pixel density to 96ppi, the examples should be at the intended sizes. Otherwise, you can measure the height of the example and square it to determine the magnification. For example, if the example measures 3" high, it has been magnified 9x. **[return to text]**
4. The embedded images are close approximations because of the limits of the presentation software. The original images are available for download for reference of exact differences. **[return to text]**
5. It may be interesting to note that standard student microscopes, with the three swivel lenses, usually have powers of 40x, 100x, and 400x. One can easily see many plant and animal cells at 100x magnification. An internet search will show many examples. **[return to text]**
6. Riley, Jenn. *Seeing Standards: A Visualization of the Metadata Universe*. 2009-2010. **http://www.dlib.indiana.edu/~jenlrile/metadatamap/**. **[return to text]**
7. A new version of this publication is being compiled. **[return to text]**

[top]

ALA American Library Association
50 E Huron St., Chicago IL 60611 | 1.800.545.2433