Proposed Digital Imaging Standards and Best Practices
Indiana Memory and LSTA Digitization Projects
February 8, 2007

This document provides information on the application of published standards and best practices for digital imaging to determine specifications for individual projects. It includes general principles based on best practices, minimum scanning guidelines for various material types, and provides links to other published standards and best practices at the end.

Indiana Memory and LSTA-funded projects must adhere to these guidelines.

GENERAL PRINCIPLES

Capture once, use many times

Digitization is expensive, time-consuming, and requires extensive handling of original materials. Any digitization project should thus focus on creating high-quality master images from which many derivative images can be created for specific uses (e.g., web delivery). The master image should capture all "important" information from the original material, which should be explicitly defined for each digitization project. The master image should also be flexible enough to allow derivatives to be created meeting a wide variety of current and future needs. Therefore, no image processing (such as sharpening) should be done to the master file. The best practices described in this document are designed to achieve this goal of flexibility.

Create a faithful reproduction of the original

Access images displayed via Indiana Memory should look as close as possible to the original material from which they were derived. Many cultural heritage materials are aged and may show signs of damage from years of wear and tear. For example, black and white photographs can yellow over time, or there may be breakage around the edge of the pages of a diary from the 1800's. It is tempting to eliminate those signs of aging by digitally erasing a tear in the paper, cropping items so as to hide damaged areas, or color-correcting images to get rid of evidence of sun damage or discoloration. It is important not to take these measures, as these "imperfections" are often of interest to artists, historians, archaeologists, archivists, and researchers for a variety of reasons. There are some red, orange, and fluorescent hues that will not look exactly like the original no matter what digital imaging software is used to achieve a color balance, but digital scanning technicians should do their best to ensure that digital surrogates are faithful to the look and condition of the original material.

Scan from earliest generation practical

As copies are made of analog materials, each generation loses some detail. From a photographic negative to a print to a copy negative, from a book to microfilm, there is generational loss of information. To capture the most information in a scanned image, always use the earliest generation of the original material that it is practical to use. In general, scanning from negatives rather than prints and scanning from original printed material rather than microfilm or photocopies is preferable. However, there are cases where practical considerations dictate using a second- or third-generation original as the source of a scanned image. A set of cracked or broken glass-plate negatives might benefit from professional printing, then scanning the prints. A large series of bound volumes that have been microfilmed would be considerably cheaper to scan from microfilm rather than to unbind the volumes for scanning or invest expensive face-up scanning equipment. In these cases, a determination must be made if images created from later-generation originals can still meet the flexibility goals of master images for the project.

Technical issues

When setting technical specifications for digitization projects, higher is not always better. There is no advantage to scanning at a resolution higher than what is needed to capture the amount of detail on the original. In fact, there is a large disadvantage to this practice in that this excess

resolution adds file size without adding detail to the digital image. The guidelines in this document are designed to help determine appropriate specifications and ensure files are as large as they need to be, but no larger.
A digitization program should employ some sort of color management solution to ensure scanners, monitors, and printers all represent image color accurately. Using “canned” International Color Consortium (ICC) profiles for each imaging device is a low-cost, somewhat effective mechanism, while using professional profiling software is a much more accurate but higher-cost solution.
Using digitization equipment appropriate to the materials being scanned is essential to an effective digitization project. Unfortunately, there are no one-size-fits-all digitization equipment solutions. For example, flatbed scanners are useful for unbound textual materials and photographic prints, while transparencies and negatives are much better imaged with dedicated film scanners. Never use a scanner at a resolution setting above its listed optical resolution (known as an interpolated resolution).

Quality control
A structured quality control program is essential to a good digitization project. An effective program might combine automated checking of objective criteria such as image resolution, file size, dimensions, and bit depth for all images with manual checking of subjective criteria such as color fidelity on a subset of scanned images.

Technical metadata
Recording adequate technical metadata about scanned images is essential for long-term maintenance of master files. The NISO draft standard Technical Metadata for Digital Still Images in its XML Schema form from the Library of Congress at <http://www.loc.gov/standards/mix/> offers guidance on what sorts of technical metadata are appropriate to record.

MINIMUM SCANNING GUIDELINES

In general, the specifications below conform to the guidelines put forth in the February 2005 edition of the Hudson Valley River Heritage Digital Imaging Best Practices document at <http://www.hudsonvalleyheritage.org/about/standards/bestpractice.htm>. The Western States Best Practices found at <http://www.cdpheritage.org/resource/scanning/documents/WSDIBP_v1.pdf> expound more on the general principles outlined above and contain a great deal of additional helpful information.

Originals as TEXT-based materials (Books, pamphlets, etc.)

	Master	Access	Thumbnail
File Format	TIFF	JPEG	JPEG
Bit Depth	1 bit bitonal 8 bit grayscale 24 bit color	1 bit bitonal 8 bit grayscale 24 bit color	1 bit bitonal 8 bit grayscale 24 bit color
Spatial Resolution	300 - 600 ppi (400 ppi and up for OCR purposes)	150 ppi	96 ppi
Spatial Dimensions	100% of original	600 pixels across the long dimension	150-200 pixels across the long dimension

Originals as PHOTOGRAPHS [i]

	Master	Access	Thumbnail
File Format	TIFF	JPEG	JPEG
Bit Depth	8 bit grayscale 24 bit color	8 bit grayscale 24 bit color	8 bit grayscale 24 bit color
Spatial Resolution	300-800 ppi, or 3000 to 5000 pixels across the long dimension	150 ppi	96 ppi
Spatial Dimensions	100% of original	600 pixels across the long dimension	150-200 pixels across the long dimension

Originals as MAPS

	Master	Access	Thumbnail
File Format	TIFF	JPEG	JPEG
Bit Depth	8 bit grayscale 24 bit color	8 bit grayscale 24 bit color	8 bit grayscale 24 bit color
Spatial Resolution	3000 pixels across the long dimension, or 300-400 ppi	150 ppi	96 ppi
Spatial Dimensions	100% of original	600 pixels across the long dimension	150-200 pixels across the long dimension

Originals as GRAPHIC MATERIALS (Broadsides, sheet music, etc.)

	Master	Access	Thumbnail
File Format	TIFF	JPEG	JPEG
Bit Depth	8 bit grayscale 24 bit color	8 bit grayscale 24 bit color	8 bit grayscale 24 bit color
Spatial Resolution	300-600 ppi, or 3000 pixels across the long dimension	150 ppi	96 ppi
Spatial Dimensions	100% of original	600 pixels across the long dimension	150-200 pixels across the long dimension

[i] Photographic prints are generally not scanned at a fixed resolution but instead at a fixed number of pixels across the long side, resulting in two differently-sized prints from one negative yielding similarly-sized digital files. The appropriate resolution is determined by dividing the desired number of pixels (e.g. 3000) by the number of inches of the long side of the photograph (e.g. 10" for an 8x10 photo). In this case 3000 / 10 = 300, so an 8x10" print should be scanned at 300ppi.

In general, color photographs should be scanned as 24-bit RGB color and black & white photographs in 8-bit grayscale. There are many cases, however, when black & white photographs would benefit from color scanning, for example, when they are sepia-toned or badly faded.

PUBLISHED STANDARDS AND BEST PRACTICES

- CPD Digital Audio Group. Digital Audio Best Practices. Version 2.0. November 2005.
  <http://www.cdpheritage.org/digital/audio/documents/CDPDABP_1-2.pdf>.
- Digital Library Federation Benchmark Working Group. Benchmark for Faithful Digital Reproductions of Monographs and Serials, Version 1. December 2002
  <http://www.diglib.org/standards/bmarkfin.htm>.
- Hudson River Valley Heritage. Digital Imaging Guidelines. February 2005.
  <http://www.hudsonvalleyheritage.org/about/standards/bestpractice.htm>.
- Kenney, Anne R., and Oya Y. Rieger. Moving Theory into Practice: Digital Imaging for Libraries and Archives. Mountain View, CA: Research Libraries Group, 2000.
- TEI Text Encoding in Libraries: Guidelines for Best Encoding Practices, Version 1.0. July 30, 1999
  <http://www.diglib.org/standards/tei.htm>.
- UIUC Digital Imaging & Media Technology Initiative. Image Quality Calculator. 2000
  <http://images.library.uiuc.edu/projects/calculator/>.
- Western States Digital Standards Group, Digital Imaging Working Group. Western States Digital Imaging Best Practices, Version 1.0. January 2003
  <http://www.cdpheritage.org/resource/scanning/documents/WSDIBP_v1.pdf>.