

GUIDELINES FOR THE CREATION OF DIGITAL COLLECTIONS

Digitization Best Practices for Images

This document sets forth guidelines for digitizing two-dimensional, non-textual materials for the CARLI Digital Collections. The issues described concern image quality, file formats, storage, and access.

This document was created by the CARLI [Digital Collections Users' Group](#) (DCUG). For questions about this document, please contact CARLI at support@carli.illinois.edu

Introduction

Images may include such items as photographs, maps, plans, blueprints, drawings, paintings, and other two-dimensional visual media.

In many instances, images will contain or have accompanying textual material. Due to this dual nature, the digitization of images is very similar to the digitization of text. For further information on text digitization, consult the "[Guideline for the Creation of Digital Collections: Digitization Best Practices for Text](#)" document maintained by DCUG.

The sections below provide guidance on the processes of creating digital images.

Image Collections

Although no universal standards for quality image capture exist and technical standards are constantly evolving, member institutions participating in the CARLI Digital Collections should adhere to the best practices adopted by recognized leading institutions when the collections' purpose would benefit from those actions.

Digital Images

A digital image is a two-dimensional array of small square regions known as pixels. Images typically fall into one of three categories: monochrome (bitonal), grayscale, and color. In the case of a monochrome image, the brightness of each pixel is represented by a numeric value. Black and white are the most common pairing of pixel values, but any two colors may be used. Grayscale images typically contain values in the range from 0 to 255, where 0 represents black, 255 represents white, and values in between represent shades or intensity of gray. A color image can be represented by a two-dimensional array of Red, Green and Blue triples, where 0 indicates that none of that primary color is present in that pixel and 255 indicates a maximum amount of that primary color.

Creating Images

At least one copy of a digital master or archival image file should be created for each object photographed or scanned. From that master file, at least two derivative files should be created:

- An access image (an image used for detailed on-screen viewing)
- A thumbnail image (for fast access during search, browse and retrieval)

A minimum of three types of images should be generated when an object is digitized:

Master Image	Access Image	Thumbnail Image
<ul style="list-style-type: none"> • Represents as closely as possible the information contained in the original • Uncompressed, or lossless compression • Unedited • Serves as long term source for derivative files and print reproductions • Can serve as surrogate for the original • High quality • Large file size • Stored in the TIFF file format, though JPEG2000 is gaining acceptance as a master file format 	<ul style="list-style-type: none"> • Used in place of master image for general web access • Generally fits within viewing area of average monitor • Reasonable file size for fast download time; does not require a fast network connection • Acceptable quality for general research • Compressed for speed of access • Usually stored in JPEG or JPEG2000 file format 	<ul style="list-style-type: none"> • A very small image usually presented with the descriptive record • Designed to display quickly online; allows user to determine whether they want to view access image • Usually stored in GIF or JPEG file formats • Not always suitable for images consisting primarily of text, musical scores, etc.; user cannot tell what content is at so small a scale

From the Western States Digital Standards Group, Digital Imaging Working Group, *Digital Imaging Best Practices*, available on the Lyrasis site (<http://www.lyrasis.org/>), January 2003 and Technical Guidelines for Digitizing Cultural Heritage Materials: Creation of Raster Image Master Files http://www.digitizationguidelines.gov/guidelines/FADGI_Still_Image-Tech_Guidelines_2010-08-24.pdf, August 2010

Master Images

Due to the stress of digitizing unique materials, a digital master should be generated for every object created. The digital master image represents as accurately as possible the visual information in the original object. This image’s primary function is to serve as a long-term archival record, as well as a source for derivative files and printed materials. Digital master files are measured in ppi (pixels per inch). Master files are most often saved to a designated server or other long-term storage device.

Master images should be scanned at an appropriate level of quality to avoid re-handling of any original materials. Scanned master images should not be edited for any specific output or use, and should be saved as large TIFF files with lossless or no compression.

	File Format	Pixel Array and Resolution	Bit Depth
Master Image	TIFF	4000-6000 pixels across the long dimension. Scan at a minimum of 300 dpi, more if necessary, to acquire an image conforming to 4000-6000 pixels across the long dimension.	1-bit bitonal mode, 8-bit grayscale, or 24-bit color

Values can vary outside of the ranges given depending on source material. Specifications taken from CDL Guidelines for Digital Images http://www.cdlib.org/services/dsc/tools/docs/cdl_gdi_v2.pdf, January 2011

Creating digital master files:

- Guidelines for file size and resolution of digital master files will vary by collection based on end user needs, sizes and types of original objects, software specifications, available file storage space, etc.
- Each library should develop specific scanning guidelines based on individual collection needs and requirements.
- Where possible, scanning guidelines for creation of digital master files should follow the specifications outlined in the CDL Guidelines for Digital Images: Guidelines for Digital Master Files: http://www.cdlib.org/services/dsc/tools/docs/cdl_gdi_v2.pdf
- CARLI member libraries using CONTENTdm **should not** upload full resolution TIFF files to the CARLI server. Archival image file storage is the responsibility of each contributing institution and must be managed locally. The CONTENTdm Project Client can automatically convert TIFF files into JPEG2000 or JPEG display images..

Derivative Images

Derivative files are used for editing and enhancement, conversion to different formats, and presentation or transmission over networks. For each master image, two derivative files are usually created: an access image (for more detailed onscreen viewing) and a thumbnail image (for searching and browsing). In the case of collections using CONTENTdm, the software can be configured to automatically generate access and thumbnail images from the master file.

Recommendations

	File Format	Pixel Array and Resolution	Bit Depth
Access Image	JPEG or JPEG2000	800-3000 pixels across the long dimension (72 – 300 ppi)	1-bit bitonal, 8-bit grayscale, or 24-bit color
Thumbnail Image	GIF or JPEG	150-200 pixels across the long dimension (72 ppi)	1-bit bitonal, 4 - 8 bit grayscale, or 8 – 24 bit color

Specifications taken from CDL Guidelines for Digital Images
http://www.cdlib.org/services/dsc/tools/docs/cdl_gdi_v2.pdf, January 2011

File Naming Conventions

Each digital object in a collection should be assigned a unique identifier. Unique identifiers should follow a consistent naming format to ensure ongoing identification and retrieval of digital files.

Guidelines for file names will vary by collection based on local needs and specifications. Each library should develop specific file naming conventions based on individual collection needs and local requirements.

Examples from the CARLI CONTENTdm collections include the following:

- ACWC0002
- bru004_03_nF
- Images_W-04.jpg
- Lor137_01.jpg

Monitor Calibration

Monitors used for image editing and color correction should be calibrated according to the following specifications:

- Set to 24 millions of colors
- Set monitor Gamma at 2.2
- Color temperature at 5000 degrees K

From Technical Guidelines for Digitizing Cultural Heritage Materials: Creation of Raster Image Master Files
http://www.digitizationguidelines.gov/guidelines/FADGI_Still_Image-Tech_Guidelines_2010-08-24.pdf, August 2010

Monitor calibration software can be selected and purchased by member libraries and will vary depending on local budgets, equipment and software specifications.

Technical Metadata

In the interest of preservation and reproduction, it is helpful to capture technical metadata in the creation of the digital image file. Digital cameras and scanners can automatically capture this information and embed it in the object file. NISO Standard Z39.87 (Data Dictionary - Technical Metadata for Digital Still Images) is widely accepted for use in the management of technical metadata. Among the attributes that can be described by the technical metadata are the following:

- file format
- file resolution (pixels per inch)
- dimensions (image dimension or size in inches or centimeters)
- bit-depth (e.g., 8-bit, 16-bit, 24-bit, etc.)
- color mode (e.g., RGB, CMYK, or grayscale)
- scanner or digital camera brand, name, and model number
- software used to manipulate or compress the image, including the software name and version.

From Best Practices for Technical Metadata:
http://www.library.illinois.edu/dcc/bestpractices/chapter_10_technicalmetadata.html

Collection of technical metadata will vary according to equipment used and local policy. Inclusion of technical metadata is not required for items in CARLI Digital Collections, but it can be included.