

Suggested Minimum Guidelines for Creating Quality Digital Objects

by Adam Northam
Digital Collections Librarian

The building of digital collections is a relatively new area; therefore there is not one set of definitive guidelines describing how to create digital objects. Collections often vary depending on hardware/software used, the format of the content within collection, the purpose and audience of the collection, and the resources of the institution/organization /individual building the collection. This document is intended to help achieve a certain level of consistency when building a digital collection. The standards discussed have been established over time by other groups involved in creating digital collections, and they are the standards to which Texas A&M University – Commerce wishes to adhere in creating our digital collections. We would encourage or partner organizations to follow these guidelines whenever possible, feasible, and/or appropriate.

Digitizing Text and Images

There are 3 types of files to consider in the digitization process:

1. **Master File** – Used for archival purposes; should remain unedited
2. **Access File** – Used for web display
3. **Thumbnail** – Small, low resolution image, often displayed with item record

If possible, one of each type should be created for each digital object, but issues such as cost, time, storage, and purpose of the digital object must all be considered when deciding whether or not master files are desired.

Master Files

If you plan to modify, re-use, or print high quality reproductions of an image after initial digitization, it is strongly recommended that you create a master file that will remain unaltered. Varying sizes and resolutions can then be derived from that one. Our master image files are typically saved in Tagged Image File Format (TIFF). TIFF is not proprietary, and lends itself to long term preservation. Most scanners typically allow files to be saved in TIFF format, as do some digital cameras. If a digital camera does not allow files to be downloaded and saved as TIFF, use the highest resolution setting possible. There are emerging file formats, such as JPEG 2000 and Motion JPEG 2000, that offer lossless compression, which means smaller file sizes with no discernable loss of quality. One current drawback of such formats, and the main reason we do not currently use them at our institution, is that they are not widely supported by most available hardware; therefore, they are often difficult and expensive to create.

Access Files

An access file is typically used for web access and display on a computer monitor. File sizes are considerably smaller than master files, which allows for faster transmission and

download. The typical file format for access image files is JPEG. If a JPEG image is saved and re-saved, the image quality will degrade with each generation.

Thumbnails

A thumbnail is a very small JPEG or GIF image that is often displayed alongside a bibliographic record if one exists. A thumbnail acts as sort of an index print that allows the viewer to quickly decide if they want to see the access image. Thumbnails are usually only useful when dealing with image files, not text.

About Scanners

Most scanners can operate in one of three modes depending on types of materials you are working with:

1. **Bitonal** – Black and white; Good for black and white, text-only documents
2. **Grayscale** – Good for black and white photographs
3. **Color** – Good for full-color image or text documents

Aside from choosing which mode to capture images in, it is also important to consider resolution, bit depth, and dimensions of the digitized image. Resolution is measured in dots per inch (dpi) or points per inch (ppi)—both terms are interchangeable. A higher ppi means a higher resolution.

Bit depth refers to the number of colors available to reproduce the digitized image. Greater bit depth means more colors are used.

- 1 bit (2^1) = 2 tones-Bitonal
- 2 bits (2^2) = 4 tones
- 3 bits (2^3) = 8 tones
- 4 bits (2^4) = 16 tones
- 8 bits (2^8) = 256 tones
- 16 bits (2^{16}) = 65,536 tones
- 24 bits (2^{24}) = 16.7 million tones-true color

Guidelines for Scanning Text

	Master File	Access File	Thumbnail
<i>File Format</i>	TIFF	JPEG	JPEG or GIF
<i>Bit Depth</i>	8 bit grayscale – text w/ B&W Pictures	8 bit grayscale – text w/ B&W Pictures	8 bit grayscale – text w/ B&W Pictures
<i>Resolution</i>	600 ppi	300 ppi	72 ppi
<i>Dimensions</i>	100% of Original Size	long side of image – about 600 pixels	long side of the image – 150-200 pixels

Guidelines for Scanning Images

	Master File	Access File	Thumbnail
<i>File Format</i>	TIFF	JPEG	JPEG or GIF
<i>Bit Depth</i>	8 bit grayscale – text w/ B&W Pictures	8 bit grayscale – text w/ B&W Pictures	8 bit grayscale – text w/ B&W Pictures
	24 bit color – color text	24 bit color – color text	24 bit color – color text
<i>Resolution</i>	long side of images – 3000-5000 pixels	300 ppi	72 ppi
<i>Dimensions</i>	100% of Original Size	long side of image – about 600 pixels	long side of the image – 150-200 pixels

It may be worthwhile to scan the back side (verso) of a photograph as a separate object as well, particularly if there is information there that does not exist anywhere else. The decision of whether or not to scan the verso will depend on the nature of a given collection, the needs of the collection creator, and those of the collection's intended audience.

Digitizing Maps

The important factor when digitizing maps is to make sure that the object is scanned at a high enough resolution to make the smallest detail distinguishable, which means that it may be necessary to scan maps at a higher resolution than other images. This depends on the size, clarity, and detail of the original map.

Guidelines for Scanning Maps

	Master File	Access File	Thumbnail
<i>File Format</i>	TIFF	JPEG	JPEG or GIF
<i>Bit Depth</i>	8 bit grayscale – text w/ B&W maps	8 bit grayscale – text w/ B&W maps	8 bit grayscale – text w/ B&W maps
	24 bit color – color maps	24 bit color – color maps	24 bit color – color maps
<i>Resolution</i>	long side of images – 3000-5000 pixels	300 ppi	72 ppi
<i>Dimensions</i>	100% of Original Size	long side of image – about 600 pixels	long side of the image – 150-200 pixels

Since the proper resolution for scanning maps varies greatly depending on the size of the original, it may be necessary to use an image quality calculator, which can be found on the Internet by typing **image quality calculator** into a search engine.

Copy Photography

Traditional flatbed scanners work very well for capturing images, documents, and other flat media of a certain size, but there may be times when it is not possible to capture an image of an object with a scanner due to irregular shape, large size, location etc. The largest flatbed scanners typically have a scanning surface of 12" X 17". Overhead scanners do provide some capability to scan larger or three-dimensional objects but they are often prohibitively expensive for some, particularly smaller organizations with limited resources. A digital camera can be a cost-effective way to capture images that cannot be scanned with a flatbed scanner.

When selecting a digital camera, it is a good idea to buy the best that your budget will allow. One of the most obvious specs to consider regarding digital cameras is Megapixels. Digital images are composed of tiny dots called pixels. One Megapixel equals 1 million pixels. The following chart gives a frame of reference for how the Megapixel count of a camera affects the image quality:

Pixel Dimensions Related to Print Size

Pixels	Megapixels	Max Print Size at 300dpi (Inches)
640 x 480	.3	1.6 x 2.1 (good for web, email, and PowerPoint only)
1024 x 768	.8	2.6 x 3.4
1280 x 960	1.2	3.2 x 4.3
*1600 x 1200	1.9 ~ 2	4 x 5.3
2048 x 1536	3.1	5 x 6.8
2272 x 1704	3.9 ~ 4	5.7 x 7.6
2304 x 1728	3.9 ~ 4	5.8 x 7.7
2560 x 1920	4.9 ~ 5	6.4 x 8.5
2592 x 1944	5.0	6.5 x 8.6
3072 x 2048	6.3	6.8 x 10.2

It is recommended that you take pictures at the highest available resolution on your camera. It will give the best picture quality. See your camera documentation regarding resolution settings. When transferring images from digital camera to computer, it is strongly recommended that you save the unaltered master images in TIFF format. Many photo management/editing software packages allow this to be done. It may be possible to save the images in RAW format, which is completely uncompressed, but this doesn't lend itself to sustained access because RAW formats are proprietary, and not widely supported.

Guidelines for Copy Photography

	Master File	Access File	Thumbnail
<i>File Format</i>	TIFF	JPEG	JPEG or GIF
<i>Bit Depth</i>	24 Bit Color	24 Bit Color	24 Bit Color
<i>Resolution</i>	Highest Possible	300 ppi	72 ppi
<i>Dimensions</i>	100% of Original Size	long side of image – about 600 pixels	long side of the image – 150-200 pixels

Copy photography is somewhat trickier than scanning due to inconsistencies with lighting/other environmental factors. We found that white lights placed on either side of the object being photographed produces satisfactory results.

Digitizing Video

Standards for digitizing video are less well defined as those for digitizing text and images. This is partially due to the rapidly evolving video formats that are being created. Creation and preservation of video files is one of the most expensive and difficult processes of digitization. Video files are often very large, and are therefore not easily stored unless significant resources are available. The current standards used for capturing and digitizing video are as follows:

- Digital camcorder records video on a Mini DV tape (Storage media may vary depending on specific camera)
- Mini DV tape is retained in archive
- Video is captured to an external hard drive as a **.AVI** (Audio/visual interleaved) file
- Lossless copies can be produced from this file – DVDs are burned
- The **.AVI** file is transferred to an external hard drive and taken to our Technology Services Department where it is backed up on a server – this digital master serves as a backup for the tape master.
- Currently, small clips are rendered in .mp4 format and used as access files in our online digital collection; Full video files may be placed online in the near future in **.MP4** format to serve as access files

If an item exists in an analog format, such as a VHS cassette, then the item is digitized and saved in MPEG-2 format. VHS cassettes are generally of lesser quality, therefore there is no significant advantage in saving them in AVI format.

Quality control is an essential part of video digitization. It is recommended that the person(s) in charge of quality control watch the majority of each DVD in order to ensure proper function. It is recommended that the DVDs be watched on a standalone DVD player and TV, rather than a computer drive. It is important to watch for audio/visual sync problems, dropped frames, audio quality issues etc. The DVD may or may not need to be re-burned based on the severity of the problems, and the judgment of Digital Collections and Special Collections staff.

Materials that are sent via FTP to our digital collection server should be access quality (JPEG for Images, .MP4 for video files). Video files may also be sent to us on DVD via mail.

Digitizing Audio

Sampling rate and bit depth are 2 factors that determine the overall quality of digitized audio. Sampling rate refers to the number of times per second the amplitude of a sound wave is measured. The standard sample rate of a consumer compact disc is 44.1 kilohertz (kHz), which means that 44,100 measurements occur in one second. A higher sample rate means a more accurate representation of the original sound. Sampling rate is related to the pitch of the recorded sound, and is measured in bits.

The consensus seems to be that the human ear cannot hear above the 20 kHz range.

Bit Depth refers to the range of numbers that represent the value of the amplitude measurement. Eight bit range is 0-255; 16 bit range 0-65,535; 24 bit range 0-16,777,215. Greater bit depth means a more precise amplitude measurement. Twenty-Four bit range should be used whenever possible for digital conversion.

Master audio files should be saved in WAV format with a sampling rate of no less than 44.1 kHz (some prefer 96kHz) and a bit depth of 24. Access files (which may be on Compact Disc or digital format) can be 44.1 kHz 16 bit files. MP3 files should be at least 128kbps.

Metadata

Metadata may be simply described as descriptive information about resources. A metadata record contains prescribed elements that describe a record in some way that gives it value to the user of the collection. There are several metadata standards currently in use, including MARC, METS, VRA Core, and Dublin Core. OCLC's CONTENTdm collection management system is the software we use to mount our collections on the Internet, and make the accessible to users. CONTENTdm supports both the simple and qualified versions of Dublin Core. A simple Dublin Core record is made up of 15 basic elements, whereas a qualified Dublin Core record contains 3 additional elements, plus several qualifiers for greater specificity. Dublin Core was chosen due to its flexibility and relative simplicity. It can easily describe a wide range of objects in various formats, and records can be easily created by non professional staff, if necessary.

Simple Dublin Core Elements

Title	Format	Creator
Identifier	Subject	Source
Description	Language	Publisher
Relation	Contributor	Coverage
Date	Rights	Type

Qualified Dublin Core contains 3 additional fields: Audience, Provenance, and Rights Holder. Each Dublin Core element is optional, and can be repeated if necessary.

When items are imported into CONTENTdm, the Dublin Core element names may be changed to offer a greater level of customization to records. It is important, however, to make sure that

each field is mapped to a Dublin Core Element to ensure that the structure of the metadata is maintained if it ever needs to be exported.

It is important that digital item records have as much relevant information as possible to make them findable and useful. Different collections may require different elements depending factors such as, the types of items in the collection, the audience, collection size etc. A core set of elements that appear on all records establishes consistency. The elements that must appear in all of our collections are: **Title, Subject, Description, Format, and Rights**. Other fields may be added as necessary.

It should be noted that this university hosts several collections for external institutions. We provide server space and support for these collections, but all content decisions related to those collections including item selection, metadata etc. are made by each respective institution.

References

Carignan, Y., Evander, J., Gueguen, G., Hanlon, A., et.al. (2007). *Best Practice Guidelines for Digital Collections at University of Maryland Libraries*

Hillman, D. (2007). *Using Dublin Core*. Retrieved June 10, 2009 from <http://dublincore.org/documents/usageguide/>.

Photoshare. *Digital Photography Tips-Best practices for Maximizing Quality*. Retrieved March 13, 2008 from <http://www.photoshare.org/phototips/digitalphoto.php>

Western States Digital Standards Group, Digital Imaging Working Group. *Western States Digital Imaging Best Practices, Version 1.0*. January 2003

Document Details

Original publication: 18 August 2008

Last update: 17 November 2009

Contact Information

Name

Adam Northam
Digital Collections Librarian

Phone Number

Office 903-468-8738
Fax 903-886-5723



Federal Depository
Library
Program participant

Texas A&M University –
Commerce
Libraries
PO Box 3011, Commerce,
Texas 75429 – 3011
(903) 886-5720



TexShare program
participant

To report a problem while using the Libraries' website please contact [Sean Anderson](#).

[Contact Us](#) | [Directory](#) | [Employment](#) | [Accessibility Policy](#) | [Privacy Policy](#) | [myLeo](#) | [Open Records](#)
[State of Texas](#) | [Texas Homeland Security](#) | [A&M System](#) | [Statewide Search](#) | [Risk & Misconduct Hotline](#)

