Florida Statewide Digital Action Plan: Guidelines for Creating Good Digital Content

By Liz Bishoff, Tom Clareson and the Florida Digital Action Plan Content Creation Working Group
November 2015

Table of Contents

Florida Statewide Digital Action Plan: Guidelines for Creating Good Digital Content

Background

Since 2013, the Florida Division of Library and Information Services and representatives of Florida's libraries, museums and archives have been developing a statewide plan and strategy for providing access to the wealth of digital materials in the state. Under the leadership of Florida librarians, archivists and museum professionals, working groups developed individual guidelines for the creation of digital content and metadata. These working groups researched national standards and best practices, investigated statewide programs in other states and reviewed existing Florida practices.

Guidelines for Creating Good Digital Content was informed by a statewide survey of Florida's archives, libraries and museums conducted in winter 2014. This document is designed to be used by Florida's archives, libraries and museums in the creation of digital content and is a foundation upon which local digital projects can be developed. It builds on the experience of those within Florida and across the country who have created digital collections. This document should be used in conjunction with other documents in the Florida Statewide Digital Action Plan series.

Getting Started on Creating Good Digital Content

This document is designed to guide planning and use of appropriate standards to create original or reformat existing digital objects. The need and desire to create digital copies of existing materials or to manage "born digital" files for research or historical purposes opens exciting possibilities. These guidelines will help libraries, archives, museums and other organizations (large and small) with making decisions for creating quality digital collections that will have lasting value to their owning institutions and others.

Like the Association for Library Collections and Technical Services' (ALCTS) *Minimum Digitization Capture Recommendations*,[1] this document was created as a guideline for producing sustainable digital items that will not need to be re-digitized in the future. 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. To ensure lasting value of the digital collection, institutions need to implement programs to ensure that multiple copies are made of their digital objects, that the copies are stored in multiple locations and that they participate in a digital preservation program to assure long-term access to the digital objects.

The awareness and use of these content creation guidelines and standards will provide institutions with a background and understanding of:

- Digitization standards required by funding organizations;
- Methods for digitizing content with the objective of producing a sustainable product that will not need to be re-digitized;
- The resources, budget, staffing and monitoring needs of in-house versus outsourced digitization projects;

---

[1]Association for Library Collections & Technical Services, *Minimum Digitization Capture Recommendations*, ala.org/alcts/resources/preserv/minimum-digitization-capture-recommendations.

- The importance of creating digital items accessible for collaboration and of sharing the digitized content; and
- Generalized minimum standards for cultural heritage organizations as a whole, though individual institutions or projects may wish to exceed these standards for certain collections.

The ALCTS *Minimum Digitization Capture Recommendations* state that:

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.[2]

Bibliographical Center for Research's (BCR) *Colorado Digital Program (CDP) Digital Imaging Best Practices, Version 2.0* (June 2008) suggests that "carefully thinking through the many components of a digitization project will go a long way towards ensuring a successful outcome."[3] A first step is to "consider how digitization fits into your institution's overall strategic plan, technology plan and project workflows."[4] The CDP recommends asking the following questions before beginning a digitization project:

- *What is your purpose?*
- *Does the project support the institution's mission?*
- *Who is your audience?*
- *Who owns the original objects and resulting files?*
- *What are the physical characteristics of the collection?*
- *What is your timeframe?*
- *How is the project being funded?*
- *Who will be responsible at different stages of the project?[5]*

Collaboration, Partnerships and Outsourcing

One of the earliest decisions to be made in planning a digitization program is whether to digitize the collection in-house, work in partnership with another organization, or outsource the digitization project to a third party.

A partnership, unlike an outsourced work for hire, requires contributions from both partners for a successful outcome. The contributions may be time, staffing, funding or arrangements for access. Successful partnerships are most likely to exist between institutions that have other cooperative arrangements or a history of shared activities. When deciding whether to work in partnership with another institution, consider the following partnership pros and cons.

Partnership pros:
- A potential partner organization with an existing digitization program may have similar collections and would welcome the ability to add access to the partner organization's digital collections.
- A partner organization may have facilities, equipment, staff or consultants that could be shared.

---

[2] Ibid.
[3] Bibliographical Center for Research's Colorado Digital Program, *Digital Imaging Best Practices, Version 2.0* (June 2008), 3.
[4] Ibid.
[5] Ibid.

- A partner organization may have an established venue for access that could give broader exposure to the digital collections.

Partnership cons:
- The potential partner organization may be fully committed to projects or grants, limiting the resources available for added projects.
- The partner organization may need to prioritize its own work ahead of partner projects.
- The ownership of, access to and storage of the digital masters and access copies need to be carefully agreed upon at the beginning of the project.

While organizations can generally implement a digitization program to reformat text and photographs/negatives, the reformatting of audio, video and large format materials may require a level of technical expertise, hardware and software that the local organization cannot easily implement. *BCR CDP Digital Imaging Best Practices* recommends that every organization carefully consider the pros and cons of outsourcing digitization projects versus digitizing collections in-house. For projects being considered for outsourcing, visiting the potential vendor's facilities is recommended to ensure that they have the space, staff and technical expertise to provide digital objects that meet the standards set for the project. The *BCR CDP Digital Imaging Best Practices* offers these pros and cons to aid in decision making:

In-house pros:
- *Development of digital content creation experience by "doing it" (project management, familiarity with technology, etc.)*
- *More control over the entire content creation process as well as handling security and storage of originals*
- *Requirements for quality, access and creation process can be adjusted as you go instead of defined up front*
- *Direct participation in development of collections that best suit your organization and users*

In-house cons:
- *Requires large initial and ongoing financial investment in equipment and staff*
- *Longer time needed to implement digitization process and technical infrastructure*
- *Limited production level*
- *Staffing expertise not always available*
- *Institution must accept costs for network downtime, equipment failure, staff training, etc.*
- *Need to enforce standards and best practices*

Outsourcing pros:
- *Pay for cost of content creation only, not equipment or staffing*
- *High production levels*
- *On site expertise*
- *Less risk*
- *Vendor absorbs costs of technology obsolescence, failure, downtime, etc.*

Outsourcing cons:
- *Organization has less control over content creation process, quality control*

- *Complex contractual process: specifications must be clearly defined up front, solutions to problems must be negotiated, communication must be open and problems must be accommodated*
- *Vendor may know more than client or may presume a level of understanding on part of library/museum/archives that they may not have*
- *Lack of standards with which to negotiate services and to measure quality against*
- *Originals must be transported, shipped and then also handled by vendor staff*
- *Possible inexperience of vendor with library/archival/museum/historical society[6]*

Documentation

Whether the digitization project is done in-house or outsourced, the CDP suggests that documentation of the choices made for your project can be a key factor in the long-term success of digitization efforts.

"Good documentation can offset the impact of staff [or vendor] turnover and allow future staff an ability to deal with digital collections created by their predecessors," says *BCR's CDP Digital Imaging Best Practices*. Among the items to document, they list:

- *Local guidelines and benchmarks for quality and resolution*
- *Resources that contributed to local practice guidelines*
- *Types of metadata captured*
- *File naming schemes*
- *Sustainability plans and procedures (storage, archiving, refreshing media, etc.)[7]*

Partners and collaborative efforts should also be documented.

Staffing

Many in-house digitization projects will utilize existing staff from other areas in the organization, including student assistants or volunteers. According to the CDP, it may benefit the project manager to look at transferable skills that project staff members already possess and to provide sufficient time for training. Creating high quality digital images, audio and video, including the required quality control, are labor-intensive activities.

By nature, digitization projects require a team approach – perhaps more than any other project – bringing together diverse sets of skills from different areas of the organization.

Sample project staff and their roles:

- *Project manager*
- *Digital librarian/Digital archivist*
- *Collection development librarian*
- *Collection curator*
- *Cataloger/Metadata specialist*
- *Digital photographer*

---

[6] Ibid., 5-6.
[7] Ibid., 3.

- *Digital audio/video technician*
- *Digitization technician*
- *Quality control technician*
- *Systems administrator/Network administrator*
- *User interface developer or designer[8]*

Equipment Selection

As part of a digitization project, you may need a variety of equipment, including a digital camera, a flatbed scanner, a slide scanner, digital video recorders and digital audio recorders. For more information on equipment selection criteria, please refer to one of the guides listed in the Selected Content Creation Resources section.

Format Conversion General Principles

When an institution embarks on a program to create digital content, Howard Besser recommends the following general principles. Specific minimum recommendations for various formats are in the tables below.

- *Digitize at the highest resolution appropriate to the nature of the source material.*
- *Digitize at an appropriate level of quality to avoid re-digitizing and re-handling of the originals in the future.*
- *Digitize an original or first generation (i.e., negative rather than print) of the source material to achieve the best quality image possible. In the case of art prints, the developed print is considered the original piece. Conservation concerns may prevent digitizing original negatives.*
- *Create and store a master image file that can be used to produce surrogate image files and serve a variety of current and future user needs.*
- *Use system components that are nonproprietary.*
- *Use image file formats and compression techniques that conform to standards within the cultural heritage community.*
- *Create backup copies of all files on servers and have an off-site backup strategy.*
- *Create meaningful metadata for image files or collections.*
- *Store digital files in an appropriate environment.*
- *Monitor data as necessary.*
- *Document a migration strategy for transferring data across generations of technology.*
- *Plan for future technological developments.[9]*

Digitization Specifications

Whether the organization digitizes their own materials or outsources the digitization to another organization, the same specifications should be followed to assure consistent quality. Similarly, if born-digital materials are being created, the same specifications should be followed by the creating

---

[8] Ibid., 4.
[9] Howard Besser, "Procedures and Practices for Scanning," *Canadian Heritage Information Network*, 1997, besser.tsoa.nyu.edu/howard/imaging.

organization. If the organization is accepting digital content from a donor, it is recommended that the collections be reviewed in light of these specifications.

The following content creation guidelines specify minimum digitization standards for major formats currently held by Florida's cultural heritage organizations. Organizations should remember that digital objects are used in different ways and that this variety of uses needs to be considered in determining the level of digitization required. While the minimum level might meet most needs, an adjustment to the higher specification might be necessary for specialized needs.

Text

When scanning text documents, spatial resolutions should be based on the size of text found in the document and resolutions should be adjusted accordingly. Documents with smaller printed text may require higher resolutions and bit depths than documents that use large typefaces. Projects planning on applying Optical Character Recognition (OCR) may wish to test pages at several resolutions to find the most satisfactory results. Images that produce the best results for OCR may not be pleasing to the eye and may require separate scans for OCR and human display.

Projects with large amounts of textual materials, particularly hard-to-read materials such as manuscripts, should consider providing transcriptions of the materials in addition to the digital image. Access to textual material can be further enhanced through Standard Generalized Markup Language/Extensible Markup Language (SGML/XML) schemes such as the Text Encoding Initiative (TEI). As re-keying text can be cost prohibitive, projects considering transcriptions should investigate including Optical Character Recognition (OCR) software in their toolkit.[10]

Text	Master	Access	Thumbnail
File Format	Tagged Image File Format (TIFF)	JPEG	JPEG
Bit Depth	1 bit bi-tonal 8-16 bit grayscale 48 bit color	8 bit grayscale 24 bit color	8 bit grayscale 24 bit color
Spatial Resolution	Adjust scan resolution to produce a minimum pixel measurement across the long dimension of 6,000 lines for 1 bit files and 4,000 lines for 8-16 bit files	150-200 pixels per inch (ppi)	144 ppi
Spatial Dimensions	4000-6000 pixels across the long dimension	600 pixels across the long dimension	150-200 pixels across the long dimension

---

[10] BCR's CDP, *Digital Imaging Best Practices*, 25.

November 2015

Photographs

Photographs can present many scanning challenges. Scan from the negative (or the earliest generation of the photograph) to yield a higher-quality image. Another consideration is whether to scan sepia-tone photographs as color or black and white images. Scan as color images to create a better image, although this will increase the file size. Yet another consideration with photographs is whether to scan the backs of photographs as separate image files if there is significant information on the back of the photo (which may be of interest to users) that may not be included elsewhere. If a scanned image of the verso of the photograph is available, the digital image may serve as a more successful surrogate for the original.

Photographs	Master	Access	Thumbnail
File Format	TIFF	JPEG	JPEG
Bit Depth	16 bit grayscale 48 bit color	8 bit grayscale 24 bit color	8 bit grayscale 24 bit color
Spatial Resolution	400-800 ppi	150-200 ppi	144 ppi
Spatial Dimensions	4000-8000 pixels across the long dimension, depending on size of original, excluding mounts and borders	600 pixels across the long dimension	150-200 pixels across the long dimension

Table from BCR's CDP Digital Imaging Best Practices

Graphic Materials

Graphics include the various techniques used to reproduce words and images from originals such as engraving, lithography, line art, graphs, diagrams, illustrations, technical drawings, and other visual representations. Nearly all graphics will be two dimensional, and should be scanned using the following guidelines.[11]

Graphic Materials	Master	Access	Thumbnail
File Format	TIFF	JPEG	JPEG
Bit Depth	16 bit grayscale 48 bit color	8 bit grayscale 24 bit color	8 bit grayscale 24 bit color
Spatial Resolution	600-800 ppi	150-200 ppi	144 ppi
Spatial Dimensions	6000-8000 pixels across the long dimension, excluding mounts and borders	600 pixels across the long dimension	150-200 pixels across the long dimension

Table from BCR's CDP Digital Imaging Best Practices

---

[11] Ibid., 27.

<u>Artwork / 3 Dimensional Objects</u>

For projects where the physical dimensions of the non-3D artwork match the equipment available, the following standards can be used. If scanning photographic copies of objects and artifacts, see recommended requirements in the appropriate photo and film charts.[12]

Artwork	Master	Access	Thumbnail
File Format	TIFF	JPEG	JPEG
Bit Depth	48 bit color	24 bit color	24 bit color
Spatial Resolution	Device Maximum	300 ppi	144 ppi
Spatial Dimensions	100% of original	600 pixels across the long dimension	150-200 pixels across the long dimension

Table from BCR's CDP Digital Imaging Best Practices

<u>Maps</u>

Scanning maps may involve items that vary widely in size, condition and amount of detail. Small maps may fit easily onto a flatbed scanner, but large plat maps may need to be scanned in sections, even on a large format scanner or captured by a camera. The size of the image can become a problem for storage, but also for viewing, serving over the Web or processing.[13]

Smaller maps (less than 36 inches on the longest dimension) should be digitized at 600 ppi, 48-bit color or 16-bit grayscale if possible. For larger maps, 300-400 ppi may be more practical. If it becomes necessary to digitize a map in sections and stitch the image together in Photoshop, keep both the original images of the sections as well as the combined image.[14] Stitching images together is very time consuming and can be difficult. Consider outsourcing the reformatting of large format maps.

---

[12] Ibid., 28
[13] Ibid.
[14] Ibid.

Maps	Master	Web	Thumbnail
File Format	TIFF	JPEG	JPEG
Bit Depth	16 bit grayscale 48 bit color	8 bit grayscale 24 bit color	8 bit grayscale 24 bit color
Spatial Resolution	600 ppi 300-400 ppi for larger maps	150-200 ppi	144 ppi
Spatial Dimensions	6000-8000 pixels across the long dimension	1078 pixels across the long dimension	150-200 pixels across the long dimension

Table from BCR's CDP Digital Imaging Best Practices

<u>Film</u>

For duplicates (negatives, slides, transparencies), match the original size. However, if original size is not known, the following recommendations are supplied: for a copy negative or transparency, scan at a resolution to achieve 4000 pixels across the long dimension. For duplicates, follow the scanning recommendations for the size that matches the actual physical dimensions of the duplicate.

Master scans of camera originals may be captured and saved in RGB, particularly those negatives that contain color information as a result of staining, degradation, or intentional color casts. Derivative files could later be reduced to grayscale in the scanning software or during post-processing editing. [15]

---

[15] Ibid., 28-29.

Film	Master	Access	Thumbnail
File Format	TIFF	JPEG	JPEG
Bit Depth	16 bit grayscale 48 bit color	8 bit grayscale 24 bit color	8 bit grayscale 24 bit color
Spatial Resolution	Resolution to be calculated from actual image format and/or dimensions – approx. 2800 ppi for 35mm originals, ranging to approx. 600 ppi for 8"x10" originals	150-200 ppi	144 ppi
Spatial Dimensions	4000-8000 pixels across long dimension of image area, depending on size of original and excluding mounts and borders	600 pixels across the long dimension	150-200 pixels across the long dimension

Table from BCR's CDP Digital Imaging Best Practices

Sound Recordings

Sound Recordings	Master	Access	Web Streaming
File Format	Broadcast WAVE (BWF)	Broadcast WAVE (BWF)	MP3 (MPEG-2 Audio Layer III)
Encoding	Uncompressed Pulse-code modulation (PCM)	Uncompressed Pulse-code modulation (PCM)	
Sampling Frequency	96 or 48 kHz, depending upon characteristics of source item	44.1 kHz	128 kbps (bitrate)
Configuration	Monophonic or stereo, depending upon characteristics of source item	Two channel (dual mono or stereo), according to Red Book CD standards	
Enhancements		Minimal cleanup (some removal of unrecorded segments, hisses, hums, artificial tones and clicks as needed; compression and gain adjustments as needed)	

File Naming Conventions

Systematic file naming is important for system compatibility, interoperability and to demonstrate ownership of the digital asset. It is critical that your file names are unique, and it is recommended that they follow an established convention to assure consistency and ease of use. File naming recommendations include:

- *Use lowercase letters of the Latin alphabet and the numerals 0 through 9.*
- *Avoid punctuation marks other than underscores and hyphens* [e.g. / > < + = ' ^ | \ { } [ ] # , ; ? ! $ * &].

- *Begin each file name with a two- to three-character acronym representing the institutional name followed by a second two- to three-character acronym representing the department or unit name (when applicable).*
- *Follow the institutional and departmental acronyms with an object ID. The object ID consists of any unique numbering scheme already in use to represent the object or, if no such number exists, a short description representing the item.*
- *File names should be limited to 31 characters, including the three character file extension.*
- *Use a single period as a separator between the file name and the three letter extension.*
- *Include a part designator after the object ID, when applicable.*

When selecting a file naming convention, think long-term. Select a system that will outlast staff involved in the current project. Consider the number of files your institution will ultimately be managing. Remember human error — if technicians will manually be assigning file names, how simple or easy will it be to make a mistake? Remember, file names do not take the place of metadata. Keep them simple and straightforward. [16]

Quality Control

The digital object creation process is not complete until the resulting objects are checked for quality and adherence to the standards selected for the project. Quality control should play a prominent role in scanning operations. Visual inspection together with a query of the file header should be completed by spot check. Spot check requires inspection of every image in thumbnail view and of no less than 10% of the images in full-image view. A variety of factors effecting on-screen and print legibility should be examined. BCR's CDP suggests the following visual inspection criteria:

- *Image is the correct size*
- *Image is the correct resolution*
- *File name is correct*
- *File format is correct*
- *Image is in correct bit depth and color mode (i.e., color image has been scaled as grayscale)*
- *No loss of detail in highlight or shadows*
- *No excessive noise especially in dark areas or shadows*
- *Even tonal values, no flare*
- *Correct focus*
- *Not pixelated*
- *Excessive dust spots or other objects*
- *No digital artifacts (such as very regular, straight lines across picture)*
- *Image not cropped*
- *Image not rotated or reversed*
- *Correct color balance[17]*

Storage and Backup

Storage and backup procedures are important to consider when planning and implementing digital projects to help ensure the sustainability of the digitized content. Digital files should be stored on

---

[16] Ibid., 29.
[17] Ibid., 32.

servers or networked storage. Copies of master files should not be stored on CDs, DVDs or other portable media. These should be used only as transport media, not as long-term storage or backup. RAID (Redundant Array of Independent Disks – a data storage virtualization technology) network drives can be used for medium-range storage and backup.

Backups guard against data loss and are an essential element of any storage plan. It is important to consider creating off-site backups in case of disaster. Best practice for storage and backup is three copies in two different media with geographic distribution. Backups should be done daily for ongoing digitization projects. Cloud storage offers an opportunity for realizing the geographic distribution and a second media type.

Digital Preservation

Digital preservation is one of the components in the life cycle of digitization, along with creation, storage, search and discovery. The goal of digital preservation is to assure long-term accessibility of born-digital and reformatted objects. Tools and software applications are being developed to support the range of activities necessary to a comprehensive digital preservation plan. Capturing the administrative, structural and technical data associated with each object means following industry-accepted best practices and open standards. This is critical to supporting digital preservation. Digital perseveration also relies on sustainable storage strategies.

Training

For staff adapting to new roles, training in the creation of good digital collections is essential. The following are examples of classes offered as in-person workshops and webinars that organizations beginning a digitization project should plan for staff to take at introductory and advanced levels:

- Introduction to Digital Imaging
- Introduction to Digital Project Management
- Introduction to Creating Digital Audio Content
- Introduction to Creating Digital Video Content
- Using Adobe Photoshop
- Safe Handling of Materials
- Introduction to PDF/PDF-A

Many national, state and regional organizations offer training and continuing education opportunities. Look at current classes and conferences from organizations such as:

- American Association of Museums (AAM)
- American Library Association (ALA), Association for Library Collections and Technical Services (ALCTS)
- American Association for State and Local History (AASLH)
- Archives & Museum Informatics
- Florida state and regional library, museum and archival organizations
- Institute of Museum and Library Services (IMLS)
- LYRASIS
- Museum Computer Network (MCN)
- Northeast Document Conservation Center (NEDCC)

- Society of American Archivists (SAA)

Selected Content Creation Resources

American Library Association, Association for Library Collections and Technical Serivces. *Minimum Digitization Capture Recommendations*. 2013.
ala.org/alcts/resources/preserv/minimum-digitization-capture-recommendations

Anderson, Seth*. Feet on the Ground: A Practical Approach to Cloud Storage – Nine Things to Consider When Assessing Cloud Storage*. 2014. avpreserve.com/wp-content/uploads/2014/02/AssessingCloudStorage.pdf

Bibliographical Center for Research's Colorado Digital Program. *Digital Imaging Best Practices, Version 2.0*. 2008.
lyrasis.org/LYRASIS%20Digital/Pages/Preservation%20Services/Resources%20and%20Publications/Digital%20Toolbox/Digital-Imaging.aspx

Bibliographical Center for Research's Colorado Digital Program. *Digital Audio Best Practices, Version 2.1*. 2006.
lyrasis.org/lyrasis%20digital/pages/preservation%20services/resources%20and%20publications/digital%20toolbox/best-practices.aspx

Federal Agencies Digitization Guidelines Initiative (FADGI) Digitization Guidelines.
digitizationguidelines.gov

Library of Congress. *Recommended Format Specifications 2014-2015.*
loc.gov/preservation/resources/rfs/textmus.html

National Information Standards Organization. *A Framework of Guidance for Building Good Digital Collections, 3rd edition*. 2007.
niso.org/publications/rp/framework3.pdf

Casey, Mike, and Bruce Gordon. *Sound Directions: Best Practices for Audio Preservation*. 2007.
dlib.indiana.edu/projects/sounddirections/papersPresent/sd_bp_07.pdf