



PROCEDURAL LEAFLET

Approved by State Records Commission and Local Government Records Commission

April 2014

DIGITIZATION FAQS: HELPFUL INFORMATION TO READ BEFORE STARTING SMALL TO MEDIUM COLLECTION DIGITIZATION PROJECTS

The ADAH Government Services Division frequently receives calls regarding digital imaging. This document reflects the most common questions received and the most common issues present when starting an imaging/digitizing program. This document is by no means a comprehensive list of all issues and considerations but should be viewed as a starting point for any digitization project/program.

WHY DIGITIZE?

Patrons of archives, libraries, museums, and government agencies are increasingly expecting to find collections and government documents available in digital format. From a user's perspective, the digital image is easier to use and quickly available. However, an organization needs to define for itself the specific goals of a digitization project before starting the project. Are you interested in:

- **Producing Surrogates for Access** – Use of surrogates enables the original records to be stored away from daily use yet still provides access to users. Digital surrogates are ideal alternatives to original materials that are often requested, historically valuable, and/or fragile.
- **Producing Replacement Formats for Preservation of Content** – Some materials are in a format that is or will become obsolete. For these types of materials, digital replacements ensure future access to the content.
- **Producing Images for Public Programming** – Occasions for digitization projects include exhibitions, anniversary celebrations, funding appeals, donation requirements, and efforts to build institutional support through demonstrating to the community the value of the institu-

tion's collections. All of these contribute to the locale's sense of community and shared history. Most of the time these projects are limited in scope and direction.

- **Gaining Intellectual Control and Managing Collections** – Digitizing collections creates the ability to manage material on an item level. However, this adds additional layers to the management and control work required by the collections, which usually translates into additional staffing. Digitization has no value if the images cannot be identified or located. That is why an accompanying index linked to the images, full-text-search capability, or another method of identification/linking using metadata is important.

Note: Digitization does not mean preservation, nor does it solve inventory, arrangement and description, management, or preservation problems.

WHAT TO DIGITIZE?

Once an institution has identified why it is digitizing materials, it needs to determine which materials will be digitized. Not all items deserve or require digitization. There must be strong reasons to justify the expenditure of limited funds to digitize material that is not unique, is available elsewhere, or is not owned by the organization. The organization must develop clear protocols for selection, including:

- **Availability of description information** – Scanning the material is the easy part of any digitization project. Adding the appropriate descriptive information for each imaged item is the most time-consuming and labor-intensive part of the project.
- **Condition and size** – Materials in poor condition that see regular use are good candidates for digitization; however, items in poor condition may be too fragile to withstand the digitization process. The material's size must also be considered, as some items may be too large or cumbersome to fit on a scanner or to be photographed.
- **Historical, cultural, or financial significance or uniqueness** – Materials that are the only records of a significant community event or have value on the collectible market may be good candidates for digitization.
- **Rights** – Any materials for which the institution is the clear owner of rights are good candidates for digitization. Does the institution own the copyright to the material to be digitized? Does it have the right to reformat and present the material on the web? If not, can the organization get permission to digitize? Does the organization need the rights?

HOW TO DIGITIZE?

Once an institution establishes digitization priorities, it should determine whether the work is to be performed in-house or by an outside vendor. The format of the material to be digitized may play a major role in this decision.

Outsourcing

Outside vendors are generally very efficient at digitizing materials of similar format and size. Most vendors, however, may not be familiar with archival standards, or how to handle rare and fragile objects. Remember that even when a project is outsourced, good project management remains crucial.

It is also important to make sure that the bid package specifically addresses an organization's expectations. Sending materials off with a mandate to digitize the files will not result in the organization receiving fully searchable files in an image management system. The organization will likely receive only image files for each of the items digitized, usually organized by folder. For example, if an institution sent off Will Books dated 1819, 1820, and 1821, to a vendor for digitization and only stated in the work order that they wanted digital images of the books created and returned to the institution – it may end up with the following:

- Volumes disbound to facilitate scanning
- Images in .pdf format
- Image files organized in three folders, one labeled 1819, one labeled 1820, and one labeled 1821. Each image name would correspond to a page name – page1.pdf, page2.pdf, etc. This wouldn't necessarily be a problem unless the expectation was to be able to immediately find the will record for John Smith.

If the organization wants to be able to search for digitized items, it should ensure that searchable metadata is included in the bid request. For textual materials, this will involve creation by the vendor of optical character recognition (OCR) files to accompany the image of the textual document or for existing index information to be "tied" to the image. For photographic items, maps, or other visual materials, this will involve the creation of metadata (identifying information) for each image or linking existing index information to the image. Many organizations already have indices or inventories of the items to be digitized and their existence reduces the costs for creating the required metadata.

Request the production of a test batch of digitized items as part of the bid package. This provides an idea about the quality of the images created, identifies the final products the organization will receive, and demonstrates the processes and procedures to be undertaken during the project.

In-house

For in-house digitization work, choose equipment and software best suited to the types of materials to be digitized. Textual and photographic materials equipment options include flatbed, digital-

camera, and planetary scanners. For large-volume business record applications, automated scanners may be useful:

- Flatbed scanners – choose one that will scan at least 11” x 14”; (11” x 17” would be better). A scanner that accommodates books and 35 mm slides and negatives will accomplish 80% of the needs of many local organizations. Organizations may want to scan photographs at a high pixels-per-inch (ppi) rate, so the range of ppi scanning options is also important. Some scanners that fit these parameters include:
 - Canon - CanoScan 9000F
 - Canon - CanoScan 5600F Flatbed Color Scanner
 - Epson - Perfection V750 Flatbed Scanner
 - HP - ScanJet N9120 Digital Flatbed/Sheetfed Scanner
 - Fijitsu – fi-6230Z Flatbed scanner

Check the reviews of scanners and other hardware at [IntelliReview](#) for the most current list of scanning equipment.

- Digital cameras - any digital single-lens reflex (DSLR) camera will create the desired images. Choose one that produces tif and jpeg raw files.

Scanning/Imaging Tips

- Determine the best image file format for the material – TIF, PDF, PNG, WAV, etc. – and any file formats your institution will use for long-term preservation of electronic files.
- Consider carefully the use of sheet-fed scanners for digitizing one-of-a-kind documents that cannot be replaced.
- Scanning software that accompanies the scanner provides some level of ability to manipulate the scan. Choose settings best suited for the material being digitized.
- Determine the best ppi for creation of the digital item. For example, if the original material is an historic map or unique photograph, the organization may want to create the master file at 600 ppi or higher to get the best capture of the content of the item. Black and white business documents may only require 200 or 300 ppi.
- Lighting can raise some big issues. Overhead lighting may cause glare, while objects and improper camera angles may cast shadows on the digitized images. Also ensure fingers, chopped edges, and other foreign objects are not captured and that the images are in focus. Use diligence and quality control to eliminate these problems.
- Cameras and scanners will occasionally default to the factory settings. Check the settings before imaging and throughout the project to ensure that, if the device defaults to the original settings, not much time is wasted by taking lower quality images.

- Purchase extra lenses in the event that the original lens breaks.
- If using a camera, mounting it will provide stability.
- Are you digitizing one-of-a-kind archival materials that you hope to never scan again? If so, at a minimum create a master image file and a use file. The master image file is usually the raw file and the preservation file. Create the use files from the master file. Consider also creating a copy master file – high quality to be used to create images for researchers as they request copies for their use. This will allow the master/raw file to be stored as it was when it was created.
- Occasionally images may lose bits and bytes (ones and zeros) when transferred. Be diligent and check the images to ensure they do not corrupt in the transfer process.
- Purchase software that permits manipulation of image files – Photoshop, GIMP, Fireworks, Inkscape, or Pixelmator. Any of these would allow the types of changes that may need to be made to the copy master image file (the file used by staff to create copies for researchers). Always document the manipulations performed on master files.
- Consider using low resolution files such as 72 ppi for presentation on the internet. A 72 ppi resolution is not high enough for the image to be used commercially, and the file loads quickly.
- Some institutions may want to watermark their material. Remember – 1) Anyone skilled in Photoshop or other image software can remove the watermarks inserted and 2) Watermark technology is usually proprietary and should not be placed on the master file.
- Audio materials – Digital Audio Tapes (DATs) should be a reformatting priority as they are unstable and the players are discontinued. The organization must have the best-quality digital converter it can afford to convert DATs. Some machines starting at \$1,000 include RME Fireface and Rosetta. The alternative is to have a good digital soundcard. There are many on the market that can get the job done. A general rule would be to replace the pc stock soundcard in a computer. Expect to spend at least \$150.00.
 - DAT (digital audio tape) Players: Sony RCM-500 and Tascam DA-30 (MK I or MK IIs)
 - Audio Digital Software: SoundForge, Adobe Audition, and WaveLab
 - Digital Soundcards: Digital Audio Labs and Creative Labs
 - Analogue Digital Converter: RME Fireface and Rosetta Apogee
 - Reel to Reel Players: Otari and Studer
 - Turntables: Technics-1200, Stanton, and Nu-mark

WE HAVE THE FILES – NOW WHAT?

Files Maintenance

Managing all of the versions of the files (preservation master file, copy master file, and use file) is important. Having several copies of files, stored in separate locations on external hard drives and not on CDs and DVDs, which are no longer considered archival storage media, is a good best practice to implement. Cloud storage is another option for storing copies; however, consider the vendor track record and security practices before making any decision to rely solely on the cloud for storage if you are storing archival files.

Staff will need to periodically check stored files to ensure that there has been no file degradation and make plans to migrate the files before degradation or technological obsolescence occur. For temporary records, staff will also need to ensure the records are destroyed when their retention periods expire.

File and Folder Names

Efficient management of electronic records begins with accurate file naming. If not managed, a computer assigns a unique name to each file when saved; but the name does not provide a context for the file, nor is it logical for ready human access. Images are frequently labeled with sequential numbers. These types of file names do not promote accessibility and ease of identification.

1. Establish and organization-wide folder management system
2. File names should include all necessary descriptive information, independent of where it is stored
3. Avoid using special characters or spaces in a file name.
4. Format dates consistently.

Metadata

Good metadata makes it possible to catalog and effectively present digital information to the public. Metadata typically describes how the image was digitized, its format, ownership and copyright information, and much more. One short descriptive scheme is the Simple Dublin Core, the fields for which are listed below. See dublincore.org/documents/dces/ for more information.

- | | |
|----------------|----------------|
| 1. Title | 10. Format |
| 2. Creator | 11. Identifier |
| 3. Subject | 12. Source |
| 4. Description | 13. Language |
| 5. Publisher | 14. Relation |
| 6. Contributor | 15. Coverage |
| 7. Date | 16. Rights |
| 8. Type | |
| 9. Format | |

Usually the image metadata will reside in the image/document management software. If the organization does not have an image management software product, images and metadata may be linked in Excel or Access. This data would need to be structured and uniform in nature so that each entry reflects the same type of information for each digital file. For example, if the word “street” is entered as St. in some entries, as STR in other entries, and as Street in other entries, any search for the term “Street” will only produce one-third of the possible candidates.

Presentation on the Web

Once the files are organized, digitized, and identified, the institution may decide to place them online. The organization may choose to place all or part of a collection online. The organization has the options of running its own website or contracting with a web hosting company.

Options for web presentation:

- Small collections
 - eHive - <http://ehive.com/> (pricing <http://ehive.com/pricing>)
 - Flickr - <http://www.flickr.com/>
- Medium to large collections management:
 - Omeka - <http://omeka.org/>
 - CollectiveAccess - Available free of charge under the GPL open-source license <http://collectiveaccess.org/about/overview>
 - Greenstone - <http://www.greenstone.org/>
 - CONTENTdm - www.contentdm.org

If an organization is not using a document or content management software product that facilitates web presentation, the other options are to use spreadsheets or databases to present the information on the web.

- For small collections an institution can place the Excel or PDF file on the web as a static web page.
- For large collections or collections that an institution wants to be searchable, some code script creation on the part of the institution’s IT staff or the ISP staff will be required.
- Because the script is going to be different for each entity, please contact the IT staff, the ISP, or the ADAH for specific guidance based on the organization’s web set up.

Questions to ask when working with a web hosting company:

- How reliable is the service?
- What kind of downtime should be expected?
- What is the backup strategy?
- What is the hardware and software used for the site?
- What kind of security do they provide for the site?
- How secure is the site?
- Is the service able to assure that no unauthorized users will be able to change or delete files?
- Is there a disk space limit? This will determine the quantity and size of files that can be stored on the site.

- What is the cost to increase the size?
- What kind of searching software is available on the site?
- Do they have the capability of referencing files from another storage site and displaying them from the organization's site? This would decrease the need for additional storage space on the host site.
- Does the web host provide any technical assistance?
- If the host has technical support, is there an additional cost for support? When is it available? What support services are provided?
- If there is metadata in an excel spreadsheet/database, will the organization have the capability to display the information with url links to the files?
- Does the web host charge more for additional traffic on the site?
- Will they charge for file transfers?
- Are there any limitations on bandwidth?

FINAL THOUGHTS

Digitization is more complicated than just scanning a few documents. As with all successful projects, a clear idea of goals, clear communication of the work to be done, and planning will serve the institution well. Don't scan just because everyone else is scanning.

For additional information please call 334-242-4452 or e-mail records@archives.alabama.gov.

BIBLIOGRAPHY

“Collection Management Systems: Museums and the Web 2011,” Institute for Dynamic Educational Advancement, <http://www.idea.org/blog/2011/04/07/collection-management-systems-museums-and-the-web-2011/>

Fulton, Wayne. “A Few Scanning Tips,” <http://www.scantips.com/basics09.html>.

Hirtle, Peter. “Copyright Term and the Public Domain in the United States,” Corenell University, <http://copyright.cornell.edu/resources/docs/copyright2014.pdf>.

Jones, Trevor. “An Introduction to Digital Projects for Libraries, Museums, and Archives,” Illinois Digitization Institute, <http://images.library.uiuc.edu/resources/introduction.htm>.

Smith, Abby. “Strategies for Building Digitized Collections,” Council on Library and Information Resources, <http://www.clir.org/pubs/reports/pub101/pub101.pdf>.

Last updated on April 23, 2014