LSC 652 Digitization Best Practice Report

Bridget Ryan

February 12, 2014

Introduction

This report begins with an explanation of what is required to digitize textual documents, including the equipment required. It then reviews two sets of best practices for developing digitization guidelines, one from Arizona State Libraries and the other for the Montana Memory Project. Then a list of what sections should be included in any digitization guidelines document will be provided. Examples of digitized collections follow the review as does a list of good reference sources. The report concludes with a list of definitions for important terms.

Process to Digitize Textual Documents

Once items have been selected, the process of digitization can begin. For any project, this involves first deciding whether the digitization will take place on-site or offsite. Many advantages and disadvantages exist for each option. On-site digitization is better if fragile or rare materials that should or cannot be moved are involved, or if the organization desires a lot of control over the conversion process. However, on-site digitization most likely requires a lot of up-front investment of monetary and staffing resources; equipment will probably need to be purchased and staff will probably need to be trained to digitize the resources. Institutions often choose to have the digitization done off-site to avoid this investment if the materials can survive the process of being moved. However, some time will still need to be spent on developing bids and monitoring the progress so outsourcing does not mean that no time will need to be spent on the digitization process by internal staff. Also, outsourcing is not a good option when small collections are involved because vendors often are not interested in taking on small projects unless there will be several of them from the institution. However, it can be more cost-effective for large projects than having internal staff time spent on digitizing a large amount of documents and can provide cost stability that is not possible when the digitization is done on-site (Cohen). It

also can be a good option for those institutions that do not require much control over the digitization of their documents. The choice of whether to do the digitization on-site or offsite is a very subjective yet important one that must be made before digitization can begin.

If the institution decides to digitize the materials on-site, then equipment will need to be gathered. For text documents and most other digitization projects, a flatbed scanner is recommended to convert the analog items into digital form. The reviewed project guidelines recommend using a scanner with a minimum resolution of 1200 dpi and 32 bit depth for digitization projects. Another way to digitize textual documents is to use a digital camera, although it is considerably more difficult to get a good digital copy of the analog text document with a camera setup than with a scanner. Also, a 35mm Digital Single Lens Reflex camera is the minimum required camera for good quality digitization; average point-and-shoot cameras do not offer good file size, lens quality, or studio flash synchronization (Montana, 2008). No matter which method is chosen, a computer will also be necessary to capture the files for storage. As of 2007, the ideal computer would have the highest processing speed that is affordable, 512MB RAM, a 40GB or larger hard drive, a CD-ROM drive and CD-RW and/or DVD writer. It would be attached to at least a seventeen- to twenty-one-inch monitor (Arizona, 2007). The software requirements for digitization are a powerful image editing software such as Adobe Photoshop and scanning software that is "descreen" capable. For long-term storage, CD-Rs or DVD-Rs were recommended by the guidelines reviewed since external hard drives are more susceptible to failure and magnetic media is not a good option for backup either. A lot of equipment is necessary to get a digitization project off the ground if the institution chooses to digitize on-site.

Once the equipment is gathered, the digitization process can begin. Depending on the type of equipment that has been chosen, the documents will either need to be scanned or

photographed. Once the files have been gathered, they will need to be organized. In most cases, an institution will develop a standard storage hierarchy and naming convention for files so that it is easy to maintain consistency (Arizona, 2007). Also, an access file will need to be created for each master file captured. After the files have been organized, they need to be described and cataloged so they can be found. Metadata is used to describe the documents, and the Dublin Core format is the most-used standard (Arizona, 2007). Once the documents have been described, this information must be stored so that the items can be searched in a catalog or in the library's interface. The next step in the process is presenting the digitized objects to potential users. This is done by developing or choosing a presentation system. When selecting or developing a system, usability must be considered; if a system is not usable then users will not use it. Ensuring usability involves having a consistent and useful organization of documents, as well as a usable interface design. Having a usable interface design requires error prevention and simple steps to use the system as well as good help and other documentation. As a system is being made available to the public, its longevity must be considered. A digital library is no good if it will not be available for viewing five years in the future. Many steps exist in the digitization process from capturing the files to presenting them to users, but they are all essential to the success of the library.

Review of the Guidelines Documents

I think that both the guidelines provided by the Arizona State Library and for the Montana Memory project do an adequate job of providing assistance for those wishing to get started on a digital library project. However, I feel that each of the documents is better in some ways than the other and also not better in other ways. If I had to choose a favorite of the two, I would say the Arizona guidelines are the best.

I think that the Arizona guidelines are a better starting place for institutions who are just getting their feet wet in the digital library world than the Montana guidelines. The Arizona guidelines are significantly shorter, at 27 pages, than the Montana guidelines, which are 45 pages long. If I were starting a project, the shorter document would seem more approachable. A lot of the Montana document's length comes from several pages of tables, which are not the same thing as paragraphs of text, but they still must be examined when reading the guidelines to gain information. The shorter length of the Arizona document does nothing to take away from its ability to provide good, understandable information; I think any novice in Arizona looking to begin a digital library project would find all of the information he or she needs to know in a way that can be understood.

Another reason why I favor the Arizona guidelines more than the Montana ones is they offer a lot of good resources for individuals to consult for additional information while working on their digital library project. They are primarily informational websites, however directions for accessing a few printed articles were provided. There was information on everything from Dublin Core rules and other cataloging and metadata help to examples of projects that have used the guidelines and other guidelines documents to consider for additional information. In all cases except the Dublin Core information, at least two resources, if not several more than that, were provided. The Montana guidelines do not offer any additional resources.

A third reason why I think the Arizona guidelines are slightly more novice-friendly than the Montana guidelines is the document begins with a list of names (with e-mail addresses and phone numbers) of people to contact in the various departments of the Arizona State Library. That way, if someone who is working on a project encounters a roadblock and does not know

how to proceed, he or she can contact one of these individuals for assistance. Individuals should never feel like they are in their effort alone.

Another thing that made the Arizona guidelines more approachable in my opinion is the chart that is provided with specific digital guidelines. It is one page, as compared to the eleven pages of the Montana document that are devoted to digitization guidelines. One page is a lot easier to deal with and more approachable than eleven when attempting to determine what one needs to do in a situation. Often when people are getting started on a new project simpler is better.

The Arizona guidelines may be more novice-friendly, but that does not mean that the Montana document is not good and should be ignored when one is considering guidelines documents. Sometimes, and for some people, being more thorough can be better. The eleven pages of tables that explain the specific requirements to follow when digitizing documents may seem overwhelming at first, but can actually be better because they provide a much more specific description of what is required for format and resolution for the files. With more thorough information, it is easier to know exactly what to do so you can be sure you are doing the right thing. The information provided by the one-page table in the Arizona document may be sufficient for the vast majority of cases, but will not satisfy those looking for very specific details for all file and types and all document types.

The fact that the Montana document is longer allows it to provide more information than the shorter Arizona guidelines document. The Arizona document may give enough information for novices to do an adequate job of creating a digital project for inclusion with other Arizona State Library digital initiatives, but the greater amount of detail could lead to higher quality resources. Also, an institution that has done other digital projects so is not new to the idea may

be able to deal with and expect more information about what the requirements are than would a novice institution.

The Montana document may provide these reasons for people to favor it over other documents, but there were a few things besides its lack of friendliness to novices that made it the least favored for me between it and the Arizona document. First, the Montana document began with a very cold, official-sounding rights statement, which is somewhat off-putting in my opinion. I think information about copyright is important, but in a digitization guidelines document it should not be the first thing addressed and with such cold language. The Arizona document discussed copyright, but used much more approachable language. Another thing that lessened my opinion of the Montana document is that one of its appendices had a heading but no content. Perhaps if these issues were resolved I could think more highly of the Montana document.

Overall, I feel that both the Arizona State Libraries and Montana Memory Project digitization guidelines documents are good documents. They cover all of the necessary information for an institution that wishes to create a digital project for inclusion with other resources in the state. However, they each have their appropriate applications and I do slightly favor the Arizona document for its novice-friendliness. Also, in my research about the process of digitization of textual documents I found a long discussion of machine-readable text versus page images. Neither of the two guidelines documents mentions that is a choice that has to be made when digitizing textual documents. I realize these documents are intended to be a guide for institutions wishing to begin a digitization project so they are not expected to include all information necessary to undergo a project; however, I think that the fact that this choice must

be made should be mentioned, and I would if I were creating a guidelines document. Otherwise, I do not see any issues with these guidelines documents.

Sample Outline for Digitization Guidelines Document

1) Introduction
2) Planning Your Project
    a) Things to Consider Before Deciding to Start Your Project
        i) Project Mission and Goals
        ii) Questions to Address
            (1) Who will be in charge?
            (2) What resources (staff, money, equipment) are available to complete the project?
            (3) What will be digitized?
            (4) How will the items be stored?
            (5) Copyright issues
            (6) Preservation
    b) Equipment You'll Need to Digitize Files
        i) Hardware and software requirements
3) Digitizing Your Materials
    a) File Specifications
        i) File type information and recommendations for all file types (master, access, and thumbnails) for the various item types that could be encountered (text, images, maps, etc.)
    b) Organizing and Storing Your Files
        i) Developing a Storage Hierarchy

ii) Choosing How to Store Your Files Physically

c) Describing and Cataloging Your Materials

i) Metadata Requirements and Guidelines

d) Presenting Your Materials to the User

i) Best Practices for a Good User Interface

4) Planning for the Future

a) How to keep the items in the digital library project available in the future

5) Appendices

a) Glossary of Important Terms

b) Recommended Resources

Examples of Digitized Collections

The following are examples of digital collections that have been developed by groups using the reviewed guidelines:

Arizona Digital Newspaper Program

This is a joint effort between the National Endowment for the Humanities and the Library of Congress in which Arizona State Libraries chose to participate to make some Arizona newspaper titles available anywhere in the world.

Link: http://adnp.azlibrary.gov/

Arizona Memory Project

This is a collection of digitized resources from various cultural institutions throughout the state of Arizona. It is mainly concerned with government information, but some of the information could be interesting to genealogists or others interested in the history of the area.

Link: http://azmemory.azlibrary.gov/

Montana Memory Project

This is a collection of digitized resources from the government and cultural institutions of the state of Montana. It offers a wide variety of items from the past and the present, and hopes to be a resource for all needs including education, business, pleasure, and lifelong learning.

Link: http://mmp.montanastatelibrary.org/

Reference Sources

- Copyright Information
  - Description: The American Library Association provides a very good resource of basic copyright information, which those involved with digital library initiatives would need.
  - Link: http://www.ala.org/advocacy/copyright
- D-Lib Magazine
  - Description: This online resource provides good quality up-to-date information about the latest developments in the digital library world.
  - Link: http://www.dlib.org/
- Digital Library Federation
  - Description: This organization, part of the Council on Library and Information Resources, can be a great one to look to for information related to digital libraries.
  - Link: http://www.diglib.org/
- Dublin Core Information
  - Description: This is the website created to host Dublin Core information, created by the Association for Information Science and Technology which is the creator of the Dublin Core Metadata Initiative
  - Link: http://dublincore.org/
- Library of Congress Subject Headings
  - Description: This website provides the Library of Congress Subject Headings that correspond to an entered search term, and can be very useful when developing metadata for items to be included in a collection.
  - Link: http://id.loc.gov/authorities/subjects.html

- List of Digital Library Projects
  - Description: This Wikipedia article provides a list of many digital library projects throughout the world.
  - Link: http://en.wikipedia.org/wiki/List_of_digital_library_projects
- NINCH Guide to Good Practice
  - Description: This website from New York University's National Initiative for a Networked Cultural Heritage gives a very good overview of the process involved with digitizing a collection of cultural heritage objects.
  - Link: http://www.nyu.edu/its/pubs/pdfs/NINCH_Guide_to_Good_Practice.pdf

Key Concepts and Terminology

The following are terms that are necessary for understanding the process of digitizing textual documents, but whose definitions might not be known:

- Access file: smaller file created from the original file for presentation purposes
- Bit depth: indicates how many unique colors are available in an image's color palate in terms of the number of 0s and 1s. This is more of a concern for images and less so for text documents, but is still important to understand.
- Compression: There are two types of compression, lossy and lossless
  - Lossy: reduces the storage space needed for a file while eliminating some information. When decompressed, the resulting file will be different from the original
  - Lossless: reduces the storage space necessary for a file while maintaining the integrity of the data
- Dots per inch (dpi): measure of resolution used for printed text or images
- JPEG (Joint Photographic Experts Group): file format commonly used for access files because they are smaller therefore quicker to download online than other formats such as TIFF.

- Master file: original file created when the document is digitized
- Metadata: data about data, used to describe text and other documents for inclusion in a digital library collection. It includes information about the object, intellectual rights, and digital representation data, among many other possibilities
- TIFF (tagged image file format): file storage format typically used for archival digital capture of objects.

References

Arizona State Library, Archives and Public Records. (2007). *Digital projects guidelines.*

Retrieved from:

http://www.azlibrary.gov/amp_resources/DigitalProjectGuidelinesv3.5.pdf.

Cohen, D. & Rosensweig, R. *Digitizing text: What do you want to provide?*.

Retrieved from: http://chnm.gmu.edu/digitalhistory/digitizing/2.php

Montana Digital Library and Archives. (2008). *Montana memory project guidelines.*

Retrieved from:

http://msl.mt.gov/Statewide_Projects/Montana_Memory_Project/Get_Started/MMPGuidelines.pdf