How to Begin

SELECTING ITEMS FOR DIGITIZATION

How does one choose the best materials to digitize? Below are the central elements that should be considered in selection, with some questions to help you assess the materials. These elements focus on areas that should form the framework for your selection criteria. Each section provides questions that should help you to think through defining your selection criteria.

Audience

- Who are the expected users? Who is the intended audience?
- Will the material be of interest to a large public?
- Will the original materials be appropriate for multiple levels of users or a specific audience?
- Will the project make materials available to a population that otherwise would be unable to use the collection, (e.g., disabled population, home-bound, or international users)?

Intellectual control

- Will digitization provide better indexing and better bibliographic control of the material?
- Will digital capture enhance use through a contextual presentation?
- Will the project raise the knowledge base of staff about the materials within the institution?

Preservation

- Will digitization aid in the preservation of deteriorating materials by diverting resources to their conservation or decreasing the wear and tear on originals by providing a digital surrogate?
- Will the materials hold up under the handling and processing required by digitization? Will special handling of the material to prevent damaging it be necessary? Will that special handling be costly? Do the materials require special technology considerations in order to digitize them without damaging them?

Value

- Does the project duplicate materials available at another repository or are they unique to your collection?
- Will the resulting digital collection have enduring value?
- Will the project make the content more broadly available?
- Will digitization give the collection "added value"?
- Will digitization improve legibility of originals?
- Will the project provide educational material that can be used in resource-based learning?

- Will the digitized collection have the potential to attract funding, either through external grants (i.e., it meets the criteria of funding agencies) or in terms of raising revenue (i.e., is it marketable)?
- Will the project be in keeping with policies at the institutional level?

Document Attributes
- Does the material lend itself to digitization?
- Can the informational content be adequately captured in digital form?
- Do the physical formats and condition of the material represent major impediments?
- Are intermediates, such as microfilm or slides, available and in good condition?
- How large and complex in terms of document variety is the collection?

Organization and Available Documentation
- Is the material in a coherent, logically structured order? Is it paginated or is the arrangement suggested by some other means? Is it complete?
- Is there adequate descriptive, navigational, or structural information about the material, such as bibliographic records or a detailed finding aid?

Documenting Your Selection Criteria
As part of the selection process, you should record the criteria that you are using to choose materials for digitization. This documentation process serves several purposes. First, it allows you to revisit the original materials to ensure you have consistently applied the selection criteria. Once the digitization project is underway, you may decide to change individual items selected. A set of well-documented selection criteria will guide any changes you make in your selection and remind you of the decisions made during this process. In addition, your documentation allows for more productive teamwork because all members of the team will follow the same protocol. Finally, documentation will provide a framework for the next digitization project, allowing for consistency across digitization projects.

Additional Resources on Selection
- North Carolina Exploring Cultural Heritage Online (ECHO) - Selection Guidelines: http://www.ncecho.org/guide/selection.htm
- Cornell University's "Moving Theory Into Practice: Digital Imaging Tutorial," Section 2 - Selection: http://www.library.cornell.edu/preservation/tutorial/contents.html

NAMING YOUR DIGITAL COLLECTION
When naming your collection of digital images, there are several criteria to consider. The name of your collection will serve as one of the primary means that patrons use to access your digital materials. When considering a name for your collection, **choose a name that is both distinctive and descriptive**. The name for your collection should be distinct from the names of other collections in the MMP, so as to avoid confusion between collections. The name should also be descriptive; that is, it should be explanatory enough to allow patrons to understand what topic or types of materials are covered within your collection of digital images. *A generic name can cause confusion for patrons*. In general, remember that the name of your digital collection should attract patrons and succinctly describe the contents.

COPYRIGHT

Federal law determines copyrights, trademark and patents, which protect original, creative works done by individuals and corporations. Copyright relates to the distribution, creation of derivatives, performances, display and exhibition and the reproduction of original works. Copyright applies the moment a creative work is published in any physical form. Works protected under copyright include literary work, artwork, multimedia works, music, photographs, and correspondence, in any format, among others. As appropriate, projects must be careful to obtain copyright permissions from repositories or copyright holders prior to distribution.

Before beginning a digitization project, establish which objects are in the public domain and which objects will require permission from the copyright holder. Items in the public domain may in fact drive the selection of digital images for a digitization project because the issues of copyright are already resolved. If a collection has been chosen for digitization and copyright is not yet cleared, locating the copyright holder and obtaining permission can be a lengthy and costly process. It is important to allot staff time and to document the procedures followed, along with any results in order to demonstrate due diligence.

There are few rules that apply to everything where copyright is concerned. However, the following can be used as a general guide when trying to determine copyright.

- If a published item (ie a book) was published before 1923, it is in the public domain.
- Unpublished items (ie a letter, diary, photograph) <u>with</u> a known author are covered by copyright until 70 years following the death of the author (as of 2008, items written by someone who passed away prior to 1938 are in the public domain).
- Unpublished items (ie a letter, diary, photograph) <u>without</u> a known author are covered by copyright until 120 years following the date of creation (as of 2008, items written in 1888 are in the public domain).

There are several online tools that can be helpful in determining if copyright may be an issue for the materials you plan to digitize. A few websites with information about copyright include:

- American Library Association's Library Copyright site:
  - http://librarycopyright.net/digitalslider/
- Cornell University – Copyright Term and the Public Domain in the United States
  - http://www.copyright.cornell.edu/public_domain/
- Library of Congress, U.S. Copyright Office:
  - http://www.copyright.gov/
- Stanford University's Copyright and Fair Use
  - http://fairuse.stanford.edu/

The Montana Memory Project adopted the following best practices to help inform contributing institutions about how to manage copyright concerns and to document due diligence for orphaned works and unpublished materials:

- The Society of American Archivists Best Practices for Orphan Works: http://www.archivists.org/standards/OWBP-V4.pdf;

- Well-intentioned practice for putting digitized collections of unpublished materials online: http://www.oclc.org/research/activities/rights/practice.pdf.

All participants in the MMP agree to respect copyright and individual privacy rights. An assessment of copyright and privacy issues is part of the planning process for any project submitted to the MMP. Materials should not be digitized without an understanding that all created digital images will be within the legal restrictions of copyright and privacy laws.

If you have questions concerning copyright and privacy within the MMP, please contact the Montana Historical Society at mmp@mt.gov.

FILE FORMATS RECOMMENDED BY THE MMP

The MMP recommends the creation of both a Master and an Access version of each digital file to be placed on the MMP. The Access, or Derivative, file will be the file made available through the CONTENTdm software.

Master Files

Digital imaging projects should create a high-quality master or archival image and then derive any additional versions in smaller sizes or alternative formats for a variety of uses. There are compelling preservation, access and economic reasons for creating an archival-quality digital master image: it provides an information-rich, research-quality surrogate. A high-quality master image will make the investment in the image capture process worthwhile. Since user expectations and technologies change over time, a digital master should be rich enough to accommodate future needs and applications. The master image should be the highest quality you can afford; it should not be edited or processed for any specific output; and it should be uncompressed.

Master digital images should be stored in a file format that supports the fidelity and long term preservation of the image. The master image file format should be:
- Nonproprietary / open source
- Uncompressed

The recommended format most frequently used for master digital images is the Tagged Image File Format (TIFF). Another format currently being considered for master digital images is the Joint Photographic Experts Group File Interchange Format (JPEG 2000).

Access or Derivative Files

Derivative files are created from the master file and are used for general access. Derivative files typically include an access image which is sized to fit within the screen of an average monitor or other delivery mechanism and/or a thumbnail image. With the proper image editing software, it is not necessary to subject source materials to multiple scans.

File formats using lossy compression are commonly used when creating derivative files. Derivatives are also generally optimized for computer monitor viewing so that visual details may be viewed as clearly as possible.

MMP recommends the use of the JPEG file format for Access files for all types of materials and the JPEG or PDF file format for Access files for text documents. The creation of a thumbnail file is not necessary, as CONTENTdm creates the thumbnail file automatically.

LONG-TERM STORAGE OF DIGITAL OBJECTS

Longevity of a digital medium depends on many factors - the type of media (CD, DVD, tape, etc.), how often and the way in which the media is handled, and how the media is stored. It is important to keep in mind that even with proper maintenance and great luck, no digital format can be considered permanent. The very best result that cultural institutions can hope to accomplish is long-term sustainability of digital material through good preservation planning and vigilant management. The storage media is an essential part of that process.

There are two types of digital storage media - portable and non-portable. Each has advantages and disadvantages for long-term storage.

Portable Media

- CD (CD-R and CD-RW)
  - CD-R (Compact Disk Recordable) is a format that requires a CD-ROM drive to read and to write. The CD-R format is an inexpensive way to store digital object masters. These disks are susceptible to scratches, to fingerprints and to extremes in temperature and light. They should be handled and stored with great care. If writing on the disk, only a water-based felt-tip pen should be used. An alcohol-based felt-tip pen can migrate through the protective layer and possibly affect the integrity of the data.
- DVD (DVD-R and DVD-RW)
  - DVD technology (Digital Video Disk or Digital Versatile Disk) is a format that requires a DVD-ROM drive to read. A DVD-R disk will hold approximately 4.7 gigabytes. Gold standard DVDs have been developed to meet archival standards.
- DAT Tape, DLT Tape, ZIP® and JAZ® drives
  - Tape, ZIP® and JAZ® drives are all magnetic media, and magnetic media is NOT recommended for long-term storage. Tape is, however, an excellent intermediate medium, particularly for transport of data and for backup.

  ### Tips for Improving the lifespan of CD's and DVD's
  - Always:
    - Store media in controlled archival environment
    - Store media in a jewel case or protective sleeve when not in use
    - If using sleeves, use those that are of low-lint and acid-free archival quality
    - Wear gloves when handling the master disks
  - Avoid:
    - Damage to the upper and lower surfaces and edges of the disk

- - o Scratching and contact with surfaces that might result in grease deposits (e.g. human hands)
    - o Exposing disks to direct sunlight
  - **Never:**
    - o Attach or fix anything to the surface of a disk
    - o Write on any part of the disk other than the plastic area of the spindle

Tips for improving the lifespan of DLT's
- **Always:**
  - o Keep tape in its protective case when not in use
  - o Move tapes in their cases
  - o Store the tapes in appropriate archival environment
  - o Store the tapes vertically
- **Avoid:**
  - o Placing the tapes near magnetic fields
  - o Moving the tapes about
  - o Exposing disks to direct sunlight
- **Never:**
  - o Stack the tapes horizontally
  - o Put adhesive labels on the cartridge
  - o Touch the surface of the tape
  - o Put a tape that has been dropped in a drive without first visually inspecting it to make certain that the tape has not been dislodged or moved

(The above information, modified from tables in the *NINCH Guide* (available at http://www.nyu.edu/its/humanities/ninchguide/XIV/), can assist in making sure digital storage media lasts as long as possible.)

Non-portable Media
- Network Servers (drives):
  - o If a server is required for your project, it should be purchased to be adequate for the first two years of the project.
- Hard Drives:
  - o It is recommended that institutions purchase the largest hard drive they can afford. If it is possible to purchase two hard drives, this will provide a more flexible storage system. If managers of digital projects use hard drives for image storage, they should defragment them on a regular basis to maintain optimum performance. Hard drives are not recommended for long-term storage.

Recommended storage standards:
- Master file storage:
  - o Minimum recommendation: Gold CR-R
  - o Best practice recommendation: Redundant Hard Disk storage and/or Hard Disk with Tape Backup

Resources for Digital Image Storage and Preservation:

- North Carolina ECHO Project: http://www.ncecho.org/guide/preservation.htm
- Cornell University, Digital Preservation Tutorial:
  http://www.icpsr.umich.edu/dpm/
- NEDCC Digital Preservation Leaflet:
  http://www.nedcc.org/resources/leaflets/6Reformatting/05DigitalPreservation.php