

Guidelines for Digital Imaging Projects

Digital Imaging and Media Technology Initiative

University Library

University of Illinois at Urbana-Champaign

December 6, 2001

I met a traveler from an antique land
Who said: Two vast and trunkless legs of stone
Stand in the desert . . . Near them, on the sand,
Half sunk, a shattered visage lies, whose frown,
And wrinkled lip, and sneer of cold command
Tell that its sculptor well those passions read
Which yet survive, stamped on these lifeless things,
The hand that mocked them, and the heart that fed.

And on the pedestal these words appear:
"My name is Ozymandias, king of kings:
Look on my works, ye Mighty, and despair!"

Nothing beside remains. Round the decay
Of that colossal wreck, boundless and bare
The lone and level sands stretch far away.

Ozymandias, *Percy Bysshe Shelley (1792-1822)*

Introduction

Without standards, the likelihood is slim that digital images will endure as long as the works of Ozymandias. The following guidelines are designed to assist the Library and other campus units in planning and completing projects involving digital images. The guidelines suggest minimum standards for digital images and metadata, which will allow the Library to provide the UIUC community with the broadest possible access to its collections. Adherence to these standards will make it possible for the Library to connect digital projects with other campus units and outside institutions. The guidelines have been written for those who are planning or beginning digital projects, and are by no means comprehensive. Library faculty and staff are encouraged to exceed these standards whenever possible, depending on the needs and purposes of specific projects. These guidelines are drawn from published standards and widely accepted best practices, and references to further information are provided throughout this document.

Standards Addressed in these Guidelines

In addition to broad information about the digitization process, this document also includes specific guidelines addressing the following areas:

- I. Getting Started with Digital Imaging Projects

- II. Minimum Quality Requirements for Digital Images
- III. Metadata Standards
- IV. Long-term Maintenance of Digital Files

I. Getting Started With Digital Imaging Projects

1. Available Resources

No two digital imaging plans are exactly alike, but there are similarities across projects that can reduce costs and human effort. Library faculty and staff are encouraged to discuss their project with the Digital Imaging and Media Technology Initiative (DIMTI). DIMTI provides free consultation regarding digital projects, and possesses a wide array of equipment that can be used for Library projects. In addition, DIMTI's website contains numerous links to information on all aspects of digitization and digitization training. If you have questions regarding the digitization process please feel free to contact us.

Resources:

- The DIMTI website (<http://images.library.uiuc.edu>)
- A comprehensive introduction to digitization is the *Handbook for Digital Projects: A Management Tool for Preservation and Access*, Maxine K. Sitts, editor. Northeast Document Conservation Center, 2000. (Also available online at <http://www.nedcc.org>)
- Another excellent book is *Moving Theory to Practice: Digital Imaging for Libraries and Archives*, Anne R. Kenney and Oya K. Rieger. Research Libraries Group, 2000.

2. Why Digitize?

The main reasons to digitize are to enhance access and improve preservation. By digitizing their collections, libraries can make information accessible that was previously only available to a select group of researchers. Digital projects allow users to search collections rapidly and comprehensively from anywhere at anytime.

Resources:

- "Why Digitize?" by Abby Smith, published by the Council on Library and Information Resources, February 1999. (<http://www.clir.org/pubs/reports/pub80-smith/pub80.html#preservation>)

3. Preservation

Digitization can also help preserve precious materials. Making high-quality digital images available electronically may reduce wear and tear on fragile items. This does not mean, however, that digital copies should be seen as a replacement for the original artifact. Even after digitization, original documents and artifacts must be cared for, as digital files are not permanent and must be periodically transferred to new formats. Please

see *Section V: Long-term Maintenance of Digital Files* for information on preserving digital data.

4. The Downside of Digitization

Despite everything that digitization can accomplish, there are also some very good reasons to stay out of the digital realm. First, not every collection is worth digitizing. The idea of entire libraries or museums being completely online is a long way off, and many experts believe that it will never happen. Successful digital projects are the result of the careful evaluation of collections and the digitization of only those items that will provide the greatest benefit to the user.

5. Planning for Digitization

The success of digital projects hinges not on expensive technology, but rather on sound project planning. Perhaps because digitization is relatively new, institutions too often concentrate on technology before deciding on a project's goals. **Technology should never drive digital projects.** User-based goals should be determined first, and only then should the appropriate technology be selected in order to meet a project's objectives.

It is best to ask a series of questions before starting a digital project. What will be gained by digitizing? Could the same ends be reached with a book, exhibit, pamphlet, presentation or video? How will the project fit into the institution's goals? What benefits will be realized? How will it be determined if the project was successful?

6. Setting Goals

Good goal setting is important for any new initiative, and digitization is no exception. Simply having a goal that states "we want to make our materials more accessible on the web" is not specific enough. Consider the following: Who will access this collection? What will they be looking for? How will they use it? How many people will use it? How will it be advertised? How will increased use of this material benefit the user and the institution?

Contacting current and potential users is an excellent way to determine the answers to these questions. Consider sending out a survey to the project's intended audience in order to learn how they are currently using the material, and how they might use it differently if it were digitized. It is also helpful to contact other institutions that have digitized similar collections and learn from their successes and failures.

A pilot digitization project should start with a manageable collection. Focusing on items with consistent or standard formats (photographs of all one size or type, documents from one collection, etc.) provides the best chance of success.

Resources:

- Colorado Digitization Project's Questions to Ask (<http://coloradodigital.coalition.org/question.html>)

7. Copyright

Copyright is a complex issue that strongly limits the selection of materials for digitization. Many cultural heritage institutions have chosen to avoid the complexities of copyright law by only digitizing materials that have passed into the public domain and are no longer covered by copyright restrictions. This practice, however, means that many collections are not digitized comprehensively. Consulting information on what is covered by copyright is an essential step in the selection process. Library digitization projects must consider the following questions before beginning to digitize:

- *Does the Library own rights to the material to be digitized?* (remember that ownership of an item does not necessarily mean the Library also holds the copyright. Examining the donor agreement can help determine if copyright was transferred with the item).
- *Is the material to be digitized in the public domain?* (Consult the websites listed in the “resources” section for further information).
- *If the material is not in the public domain and the Library does not hold copyright, will the copyright holder give permission to digitize the material?*
- *Would the prospective digitization of the material be eligible for a “fair use” defense against infringement claims?* (Consult the websites listed in the “resources” section for further information).

If the answer to the first two questions is “yes”, then copyright is unlikely to be an issue with the project. If someone else holds copyright, permission often must be obtained before the material can be placed on the web. The office of Instructional Copying and Copyright Clearance on the UIUC campus (333-5425) can assist in obtaining copyright clearances.

Resources:

- “When Works Pass into the Public Domain: Copyright Term for Archivists” by Peter Hirtle (<http://cidc.library.cornell.edu/copyright/>)
- Copyright Guidelines by Lolly Gasaway of the University of North Carolina at Chapel Hill, (<http://www.unc.edu/%7Euncng/public-d.htm>)
- For a more detailed discussion of copyright restrictions, take the superb *Copyright Crash Course* from the University of Texas at (<http://www.utsystem.edu/ogc/intellectualproperty/cprtindx.htm>)
- For a guide to decisions on “fair use” see the Copyright Management Center’s “Fair Use Checklist” at (<http://www.iupui.edu/~copyinfo/fucheckintro.html>)

8. Evaluation

Evaluation is an oft-neglected aspect of digitization projects. Project evaluations should move beyond easily quantifiable figures and attempt to determine a program’s impact on the user. Many digital projects are judged solely by the number of items they digitize, but this is actually one of the least useful measures of a project’s success. Digitizing 500,

1,000 or even 100,000 images means nothing if they are of low quality, hard to locate in a database, or not interesting to the public. Surveying users to learn how they are utilizing digital materials provides a more effective evaluation tool. At the bare minimum, all Library projects should be formally evaluated based on whether they have met their goals.

III. Minimum Image Quality Requirements for Library Digital Imaging Projects

1. Image Types

There are generally three types of digital imaging files created for Library projects:

- Master files (sometimes called archival files) are the source files for all other digital files and ensure the long-term usability of the digital information. Master files should be saved in TIFF file format, and should not be compressed, altered or resized. Master files should be stored on a stable medium and should remain in a controlled environment. Please see *Section V: Long-term Maintenance of Digital Files* for more information.
- Access files are for day-to-day use. They may be placed on web pages or used for printing. Access files are expected to be of good quality, but they may be modified to meet the needs at hand.
- Thumbnail files are very small files for use in databases or web pages.

Resources:

- Digital Library Foundation (<http://www.diglib.org/standards/presreformatsum.htm>)
- Illinois State Library's Digital Resource Guide (http://www.cyberdriveillinois.com/library/digital/di_res.htm)
- Howard Besser's "Procedures and Practices for Scanning" for the Canadian Heritage Information Network (CHIN) (<http://sunsite.Berkeley.EDU/Imaging/Databases/Scanning>)

2. Image Quality Requirements

The imaging quality requirements for Library digital imaging projects have been placed in table format for ease of use. These requirements should be viewed as the **minimum** necessary to create quality digital images, and Library units are encouraged to exceed these basic requirements when it meets the needs of their digital project. These standards have been drawn from the following sources:

- Cornell University Library's "Report of the Digital Preservation Policy Working Group on Establishing a Central Depository for Preserving Digital Image Collections", Version 1.0, March 2001. (<http://www.library.cornell.edu/imls/image%20deposit%20guidelines.pdf>)

DRAFT VERSION #3

- Library of Congress, “Requests Proposals for Digital Images of Pictorial Materials”, 1997. (<http://memory.loc.gov/ammem/prpsal9/coverpag.html>)
- Library of Congress “Requests Proposals For Digital Images from Original Documents Text Conversion and SGML-Encoding National Digital Library Program”, 1996. (<http://lcweb2.loc.gov:8081/ammem/prpsal/coverpag.html>)
- Colorado Digitization Project, “General Guidelines for Scanning”, 1999. (<http://coloradodigital.coalliance.org/scanning.html#Standards>)
- California State Library Scanning Standards, 1999. (<http://www.library.ca.gov/assets/images/scandocrev1122.PDF>)

Image Type	Printed Text	Damaged Printed Text	Handwritten Manuscripts	Maps, Drawings
Master	<p>Scan Type: Bitonal</p> <p>Resolution: 600 DPI</p> <p>Format: Uncompressed TIFF</p>	<p>Scan Type: 8-bit grayscale or 24-bit color</p> <p>Resolution: 400 DPI</p> <p>Format: Uncompressed TIFF</p>	<p>Scan Type: 8-bit grayscale or 24-bit color</p> <p>Resolution: 600 DPI</p> <p>Format: Uncompressed TIFF</p>	<p>Scan Type: 8 bit grayscale or 24-bit color</p> <p>Resolution: 200-400 DPI</p> <p>Format: Uncompressed TIFF</p>
Access	<p>Type: 8-bit grayscale/24-bit color</p> <p>Format: JPEG Compression: Medium</p> <p>Spatial Resolution: Resize to 1024 x 768 pixels</p>	<p>Type: 8-bit grayscale/24-bit color</p> <p>Format: JPEG Compression: Medium</p> <p>Spatial Resolution: Resize to 1024 x 768 pixels</p>	<p>Type: 8-bit grayscale/24-bit color</p> <p>Format: JPEG Compression: Medium</p> <p>Spatial Resolution: Resize to 1024 x 768 pixels</p>	<p>Type: 8-bit grayscale/24-bit color</p> <p>Format: JPEG Compression: Medium</p> <p>Spatial Resolution: Resize to 1200 pixels across the long dimension (large maps) size to 640 x 480 pixels (small maps)</p>
Thumbnail	<p>4-bit grayscale/8-bit color</p> <p>Format: GIF</p> <p>Spatial Resolution: Resize to 150-200 pixels across the long dimension 72 DPI</p>	<p>4-bit grayscale/8-bit color</p> <p>Format: GIF</p> <p>Spatial Resolution: Resize to 150-200 pixels across the long dimension 72 DPI</p>	<p>4-bit grayscale/8-bit color</p> <p>Format: GIF (or JPEG)</p> <p>Spatial Resolution: Resize to 150-200 pixels across the long dimension 72 DPI</p>	<p>4-bit grayscale/8-bit color</p> <p>Format: GIF (or JPEG)</p> <p>Spatial Resolution: Resize to 150-200 pixels across the long dimension 72 DPI</p>

Image Type	Black and White Photographs	Color Photographs	Works of Art on Paper	Microfilm
Master	Scan Type: 8-bit grayscale Resolution: 600 DPI Format: Uncompressed TIFF	Scan Type: 24 bit color Resolution: 600 DPI Format: Uncompressed TIFF	Scan Type: 8-bit Grayscale/ 24-bit color Resolution: 400 DPI Format: Uncompressed TIFF	Scan Type: Bitonal Resolution: 600 DPI (blown back to original size) Format: Uncompressed TIFF
Access	Type: 8-bit grayscale Format: JPEG Compression: Medium Spatial Resolution: Resize to 1024 x 768 Pixels	Type: 24-bit color Format: JPEG Compression: Medium Spatial Resolution: Resize to 1024 x 768 Pixels	Type: 8-bit grayscale/24-bit color Format: JPEG Compression: Medium Spatial Resolution: Resize to 1024 x 768 Pixels	Type: 8-bit grayscale/24-bit color Format: JPEG Compression: Medium Spatial Resolution: Resize to 1024 x 768 pixels
Thumbnail	4-bit grayscale Format: GIF (or JPEG) Spatial Resolution: Resize to 150 - 200 pixels across the long dimension 72 DPI	8-bit color Format: GIF (or JPEG) Spatial Resolution: Resize to 150 - 200 pixels across the long dimension 72 DPI	4-bit grayscale/8-bit color Format: GIF (or JPEG) Spatial Resolution: Resize to 150 - 200 pixels across the long dimension 72 DPI	4-bit grayscale/8-bit color Format: GIF Spatial Resolution: Resize to 150-200 pixels across the long dimension 72 DPI

III Metadata:

1. Introduction

The creation of robust metadata is of paramount importance for the success of any digital project. Metadata is typically defined as “data about data.” Good metadata makes it possible to catalog and present digital information effectively to the public. Metadata typically describes how the image was digitized, its format, and copyright restrictions. A wide variety of metadata schemes currently exist, but as of yet no single metadata standard has gained worldwide acceptance. Deciding which metadata standard to use should be determined before materials are digitized.

Resources:

- The California State Library has posted an excellent and novice-friendly introduction to metadata standards at: (<http://www.library.ca.gov/assets/acrobat/metadocfinalrev.PDF>)
- IFLA’s List of Metadata Resources: (<http://www.ifla.org/II/metadata.htm>)

2. Metadata Recommendations

- The Library recognizes that no single metadata scheme will meet the needs of every digital project, and does not require the use of a single metadata standard for digital images. Metadata for digital imaging projects, however, must conform to one of the following accepted standards:
 - Text Encoding Initiative (TEI) (<http://www.tei-c.org/>)
 - Encoded Archival Description (EAD) (<http://www.loc.gov/ead/>)
 - Dublin Core (<http://dublincore.org/>)
- Collection-level MARC records must be created for all monographs and serials. The Library follows the Digital Library Foundation's Draft Benchmark for Digital Reproductions of Printed Books and Serial Publications, which is available at: <http://www.diglib.org/standards/draftbmark.htm#appendix>
- Images that are cataloged at the item level must include the fifteen Dublin Core metadata elements, and a greater level of description is recommended. The Library follows the metadata standards outlined in the Making of America II Testbed Project. An explanation of these metadata fields is available at: <http://www.clir.org/pubs/reports/pub87/tables.html>
- The Library requires that generated metadata for all digital images enable:
 - Resource Discovery
 - Resource Presentation and Navigation
 - Rights Management
 - Preservation

V. Long-term Maintenance of Digital Image Files

2. Introduction

The long-term maintenance and preservation of digital files is a major concern for institutions worldwide. Some experts have estimated that it could cost as much as 10% of the price of digitization **per year** to preserve a digital image. In order to maintain the ability to display, retrieve and use digital collections, digital files must be cared for, periodically refreshed, and migrated to new formats. Digital preservation is a rapidly changing and complex field, but there are some simple practices that will help ensure the long-term stability of digital data. An important component in the preservation of digital files is the establishment of a central digital repository. The Library is currently working to develop repository standards. Cornell and Harvard University's Digital repository standards are available at:

- <http://www.library.cornell.edu/imls/image%20deposit%20guidelines.pdf>

- http://hul.harvard.edu/ldi/html/storage_access.html

2. Key Preservation Principles

- Select and create digital collections of long-lasting value.
- Store digital media with care.
- Develop preservation strategies and implement them systematically.
- Consider a hybrid approach where hard copy print or microfilm files are created in addition to digital files.

3. Storage Requirements for Digital Files

- Store master images on high quality, industry standard digital tape, magnetic disks, or CD-R. (The gold colored CD-R are of much better quality than the silver ones).
- CD-Rs should only be labeled with archival quality adhesive labels and not with markers.
- Create backups of the master files and store them off-site in a secure location. The Library recommends that three copies be kept of each file.
- Store media in a controlled environment. The accepted ranges for temperature are from 62-68 degrees (65 degrees is optimum) and relative humidity should be within a 35-45% range with 40% optimum. Consistency of temperature and humidity is more important than absolute levels. The storage area should not be in an attic, basement or other area prone to sudden changes in temperature or humidity.
- Store all media away from strong magnetic fields.
- Maintain a clean, dust-free environment.
- Minimize the handling of master images and the media on which they are stored.
- Digital files maintained for an extended period should be refreshed periodically by placing them onto new media. Follow the recommendations of the manufacturer.
- Develop a system to periodically determine the readability of digital files. Follow a verification procedure such as checksum or MD5 to ensure the integrity of the data after the refreshing process is complete.