

Report of the Digital Preservation Policy Working Group
on
Establishing a Central Depository for Preserving
Digital Image Collections

PART 1: Responsibilities of Transferee

Version 1.0, March 2001

Anne R. Kenney and Oya Y. Rieger, Co-Chairs

David Block
Erla Heyns
Peter Hirtle
George Kozak
Joy Paulson
Tom Turner

Cornell University Library

Central Deposit Guidelines for Digital Image Collections

INTRODUCTION	3
REQUIREMENTS FOR DEPOSIT	5
1. SELECTION AND CONTENT CONSIDERATIONS.....	5
A. <i>Scope</i>	5
B. <i>Content/Functional Criteria</i>	6
C. <i>Priorities for Deposit</i>	6
2. LEGAL CONSIDERATIONS	7
3. TECHNICAL REQUIREMENTS FOR CONVERSION.....	7
A. <i>Source Material for Digitization</i>	8
Master vs. Derivative Files	9
B. <i>Technical Considerations for Image Files</i>	9
Table 1: Digital Master Image Files— Recommended Imaging Requirements	11
Table 2: Digital Master Image Files— Minimal Imaging Requirements	13
C. <i>Quality Control</i>	15
QC Recommendations	15
4. PRE-DEPOSITORY STORAGE AND MAINTENANCE REQUIREMENTS.....	16
5. METADATA REQUIREMENTS.....	17
A. <i>Descriptive Metadata</i>	17
B. <i>Structural Metadata</i>	21
Table 3: List of Structural Metadata Elements for Digital Image Collections.....	22
C. <i>Preservation Metadata</i>	23
Digital Image Collections Inventory	24
Table 4: Digital Image Collections Inventory: Data Fields	25
Technical Metadata.....	26
Authenticity	27
ROLE AND RESPONSIBILITIES OF A CENTRAL DEPOSITORY.....	28
NEXT STEPS.....	29
A. NEEDS ASSESSMENT SURVEY.....	29
B. SCHEDULE AND PROCEDURES FOR UPDATING THE DEPOSIT GUIDELINES	30
C. PUBLICITY AND TRAINING.....	30
APPENDIX 1: IMAGE QUALITY ASSESSMENT	31
APPENDIX 2: DESCRIPTIVE METADATA EXAMPLE.....	35
APPENDIX 3: SAMPLE ENTRY FOR DIGITAL IMAGE COLLECTIONS INVENTORY	37

INTRODUCTION

This document presents recommendations from the Digital Preservation Policy Working Group, which was charged with developing a prospective strategy for managing Cornell's digital image assets over time. The Working Group itself represented a logical extension of the planning efforts undertaken as part of a project funded by the IMLS to develop a digital preservation solution for Cornell's retrospective digital image collections created over the past decade. The Working Group membership included individuals from across the library system who have the requisite expertise and/or responsibility for selecting, managing, and serving digital image collections. Anne Kenney and Oya Rieger co-chaired the Working Group, and the membership included: David Block and Erla Heyns (selection and content considerations), Joy Paulson (image conversion), Peter Hirtle (legal and technical requirements), George Kozak (technical requirements), and Tom Turner (metadata).

This report begins with two important recommendations to the Library Management Team. The first one is to establish centralized responsibility for ensuring continuing access to digital image collections over time. *The Working Group recommends that this responsibility take the form of a central depository to be administratively located within the Digital Library and Information Technology (D-LIT) infrastructure.* It is the Working Group's strong belief that centralized responsibility will facilitate the long-term use of digital resources in the most cost-effective manner.

Although preservation is the driving force behind this recommendation, in the digital realm, preservation is met by the ability to continue to provide reliable and trusted access. Current forms of access to materials may be managed remotely and/or locally. *The Working Group recommends that the staff of the central depository collaborate with the transferee to provide both access to and security of files entrusted to their care.* The transferee may continue to exercise a principal role in supporting access to the content or assume a consulting role in the event of any significant change in functionality or service. In some cases, such as with text-based images, both preservation and access can be served by placing the master image files at the heart of the delivery system.

With these two recommendations in mind, the Working Group has outlined a plan for establishing a central depository. This report represents the first installment in that plan. It defines requirements for those wishing to transfer materials to the central depository and outlines the role and responsibilities that a central depository might have. The Working Group decided not to outline these latter responsibilities in detail until the Library Management Team approves the concept of a central depository in principal. The Working Group also acknowledges that a fulsome description of the depository should be based on an assessment of the Cornell University Library's readiness to support such a responsibility. This assessment was beyond the scope of the Working Group's charge. The report concludes with a recommendation to conduct a three-month feasibility study, the details of which are outlined in the "Next Steps" section (page 29).

Following this feasibility study, the second installment of the plan can be undertaken, which will provide a detailed description of the Central Depository.

REQUIREMENTS FOR DEPOSIT

This section outlines the responsibilities of a *transferee* in preparing an image collection that will be centrally deposited. Long term management of digital collections requires a substantial commitment of institutional resources. The guidelines aim for a level of homogeneity, with the assumption that the maintenance of collections that are created based on consistent techniques is more practical and cost-effective in the long run.

Digital image collections can be of two types: those created initially via a digital process, and those that represent digital surrogates for analog source documents. Currently the vast majority of digital image files managed by Cornell units come from this latter category. To be eligible for inclusion in the central repository, and in order for them to be managed in an effective and efficient manner, image files must meet some minimal requirements. Therefore, guidance and recommendations for the selection, creation, and storage of digital image files and related documentation are included in this section. Although these recommendations may prove useful to projects of a short-term nature, they are not to be viewed as binding on such projects, nor are they designed to inhibit individual initiatives involving the use of digital image technology. The guidelines will be used only to determine whether to invest in the long-term care of digital collections by accepting them for central deposit.

1. Selection and Content Considerations

This section of the report focuses on the scope and nature of the digitized content that will be selected for the central depository.

A. *Scope*

Digitized materials, like their analog counterparts, should primarily support teaching and scholarship at Cornell University and secondarily the needs of other communities. Selection of digital image materials is based, in significant part, on their content. It is also essential that the functionality of the digitized items be considered as part of the selection decision, i.e., the extent to which the items' particular digital forms affect or extend their utility for specific scholarly or educational purposes. Additional evaluation elements, enumerated in subsequent sections of these guidelines, imply that a resource meets the standards of authority and relevance that characterize other segments of the library's holdings.¹

¹ Half a decade of reflection, deliberation and implementation has expanded selection criteria to include analysis of what digital collections imply for technical processing, physical access, interpretation and preservation. While these considerations await formal codification, their essence appears in scattered documents: Report of the Committee on Electronic Resources (1996) <http://www.library.cornell.edu/DLWG/CERREPOR.htm#Rec>, Report of the Full Text Working Group, www.mannlib.cornell.edu/frat/ftwg-rpt.pdf (1998) and the Criteria for Comparative Assessment of Networked Resources, an instrument for assessing electronic databases prior to initial selection or relicensing <http://www.english.cornell.edu/cul/a2i/drc/ccaform.html> [1999] among them.

B. Content/Functional Criteria

Digital image materials selected for central deposit should reflect the same content criteria used for evaluating print collections. They should:

- Support the curriculum, including emerging distributed learning initiatives
- Ensure standard source availability—the identification of “core” publications
- Facilitate faculty and student research
- Maintain national collection strengths
- Honor inter-institutional and other commitments

These considerations are codified in the collection development policy of the Library:

Cornell University Library, Collection Development Policy Statements
www.library.cornell.edu/colldev/cdhome1.html

Cornell Primary Collecting Responsibilities
www.library.cornell.edu/colldev/cpcr.pdf

C. Priorities for Deposit

At some point, CUL may have to make difficult choices about which materials to accept for deposit based on economic or managerial considerations. Technical and legal issues will most likely be the determining factors in selection but content decisions may also come into play. While not restricting selection to materials that meet one or more of the following criteria, preference for central deposit could be given to digital image materials that:

- Represent complete and credible versions of digitized resources
- Represent thematic or format-based aggregates, rather than idiosyncratic works
- Help create a comprehensive collection
- Enhance access to collections by making them easier to browse, search, and use
- Increase use of collections, by bringing little-known materials to light or by widening potential readership
- Are accessible campus-wide
- Help preserve, protect, store remotely, or replace materials by providing reliable surrogates for consultation
- Enable new kinds of research, not possible in the analog form
- Do not duplicate resources available to the Cornell community via other arrangements, e.g., digital resources managed at individual libraries within the system or other divisions within the University, or made accessible via other institutions, consortia, inter-institutional agreements, or existing licenses

2. Legal Considerations

The ready accessibility of digital documents distinguishes them from their print counterparts, and complicates their selection. The digital equivalent of practices that were common with print material may place the University at unacceptable risk merely because they are more visible and may have greater consequences. Selection for central deposit must begin with an understanding of the current uncertainty in the application of copyright to digital resources. Copyright law, still written on and, most cogently about paper, is in flux as it extends to digitized materials.

The uncertainty in how copyright law will be applied to digital materials leads to an additional selection criterion: Cornell's legal right to store digital copies of material can be justified. There are several possible justifications for central management of a digital image resource:

- The digital image materials are in the public domain.
- The copyright holder has granted permission for network distribution and use.
- An assessment of the risk involved in digitizing a collection reveals that it is unlikely Cornell could be found at fault, either because the copyright owners cannot be found or the use is presumed to be fair.
- The digital resources are copies of unpublished works found in a different library.
- The digital copy has been made to replace a damaged, deteriorating, lost, or stolen copy of a work that cannot be obtained at a fair price, and the digital copies of which will not be made available outside of the library.

Evidence of the copyright status and documentation of efforts to obtain permission to make copyrighted digital resources available—including a signed copyright waiver from the copyright holder or written documentation that details a good faith effort to secure such permission—are required as part of the deposit process. Under special circumstances, digitized material that is copyright-protected, but which will fall into public domain within a short time frame, may also be considered for deposit. A chart summarizing the terms of protection for published and unpublished materials is available at <http://cidc.library.cornell.edu/copyright/>. Note that copyright may cover software use as well as digital content. In addition to copyright, privacy and donor restrictions must be considered. It is the depositing unit's responsibility to ensure that these rights are not breached by the digitization and use of such materials.

3. Technical Requirements for Conversion

There are compelling preservation, access, and economic reasons for creating rich digital master image files that reflect all significant informational content contained in the original source materials. These files have the best chance of remaining useful and cost effective over time for a number of reasons. Preservation is one of the main arguments for rich digital masters. Digital files can be created to reduce use of or in some cases

replace a deteriorating or vulnerable original, provided the digital surrogate offers an accurate and trusted representation. Preservation of the digital files themselves is also served when digital images are captured consistently, the capture methods are well documented, and widely supported file formats are used. And it makes good economic sense to produce sufficiently high-level images to avoid the expense of reconverting when technology requires or can use a richer digital file. This point is particularly compelling since the expense of identifying, preparing, inspecting, and indexing digital information far exceeds scanning costs. The master file can also be used to create derivatives files, which meet a variety of current and future users' needs. The quality, utility, and expense of derivatives for publication, image display, and computer processing are directly affected by the quality of the initial scan.

A. Source Material for Digitization

Digital image files can be created to serve as surrogates for a range of document types. Source material can include, but may not be limited to, the following:

- Printed text—distinct edge-based representations that are cleanly produced, with no tonal variation, such as a book containing text and simple line graphics
- Book illustrations—representing the range of relief, intaglio, and planographic illustration processes reproduced in books produced in the 19th and 20th century, including halftones, etchings, and engravings
- Rare or damaged printed text—items that convey intrinsic information beyond the printed text or those in which the text may be obscured by surface dirt, stains, or other damage
- Manuscripts—soft, edge-based representations that are produced by hand or machine, but do not exhibit the distinct edges typical of machine processes, such as a letter or line drawing
- Maps, architectural drawings—oversized materials that contain fine details, line drawings, and text, either hand or machine-produced
- Graphics—original relief, intaglio, and planographic illustrations
- Works of art on paper—hand produced artwork, including water colors, charcoal sketches, pencil drawings, tempura, and oil painting
- Photographic prints—reflection print formats, including cartes de visite, cabinet cards, 3.5" x 5", 4" x 5", postcard, 5" x 7", and 8" x 10"
- Photographic transparencies and negatives—negatives and positive transparencies produced on film or glass, including 35 mm, lantern slides, 4" x 5", 5" x 7", and 8" x 10"
- Microformats, including 16 mm, 35mm, 70 mm microfilm and 105 mm microfiche

Scanning from Originals vs. Intermediates

As a general rule of thumb, scanning from the original will ensure the highest quality image file. When multiple copies of an item exist, scan from the best copy available, whenever possible. The use of an intermediate, such as a slide, transparency, microfilm,

or photocopy, will introduce another step in the imaging process, increasing the complexity of the workflow, and lowering the quality of the resulting image. It may also affect the accuracy of subsequent image processing, such as the use of Optical Character Recognition (OCR) programs. Quality and processing applications will be particularly compromised if the intermediate itself is poorly produced, damaged, or in deteriorated condition. If an intermediate is used, ensure that it has been prepared according to established standards and is in good condition, free of scratches, dust, light damage, and distortions.

Master vs. Derivative Files

These guidelines cover technical requirements in the creation of master images only. Derivative files that are currently being used for access (e.g., thumbnails, access images, printing files) may also be deposited but they will be maintained only as long as they are being used in the access system. There is no guarantee that they will be maintained for any set length of time, as derivative requirements may change rapidly based on developments in file formats, compression techniques, display technologies, and on-the-fly generation capabilities. *Projects considering the development of static derivatives should consult with Depository staff about current practice in creating access versions.* For instance, derivatives of text-based image files currently include: 10 dpi 3-bit GIF thumbnails, 100 dpi and 75 dpi 3-bit GIFs created on the fly, and text files produced via OCR. PDFs of individual pages are created for printing purposes from the master images (e.g., 600-1bit TIFF images).

B. Technical Considerations for Image Files

The following technical considerations are covered in the Recommended Imaging Requirements. Additional technical considerations are discussed under Quality Control.

- Resolution—the spatial frequency at which a digital image is sampled, often stated as dots per inch (dpi), pixels per inch (ppi), or pixel dimensions.
- Bit depth—determined by the number of bits used to define each pixel. Digital images may be produced in black and white (bitonal, or 1 bit images), in grayscale, or in color. All source documents containing intentional color or where discoloration provides important evidence of age and use should be imaged in color.
- Enhancement/image processing—any process applied to the raw scan to improve quality or legibility. Generally accepted enhancements included reduction of greater than 8-bit/channel linear data to 8-bit non linear data,; contrast stretching; minimal adjustments for color and tone; descreening/rescreening of halftones and other graphic content to reduce/minimize moiré.
- File format—consists of both the bits that comprise the image and the header information on how to read and interpret the file. Currently there is no clear archival format to recommend, although open, widely supported file formats are

recommended, with preference given to TIFF files 5.0 and 6.0. The goal is to limit the number of file formats that need to be managed by the Depository. The future of TIFF and other formats, however, is uncertain, and the depository staff will need to monitor format development. A table presenting information on some of the more common image formats in use today is available at <http://www.library.cornell.edu/preservation/tutorial/presentation/table7-1.html>.

- Compression—a process used to mathematically reduce or abbreviate the string of binary code in an uncompressed image. Compression techniques can be either lossless (no information discarded in the process) or lossy (where the least significant information is averaged or discarded). There is a clear preference for uncompressed files or for compressed files using lossless compression. The goal is to limit the number of compression processes that need to be managed long term. A table listing attributes of common compression formats is available at <http://www.library.cornell.edu/preservation/tutorial/presentation/table7-3.html>.

Two levels of technical imaging requirements are presented in the charts below. The first represents Recommended Requirements that will promote the long-term viability of digital image collections, and are detailed in Chart 1. *Depositors are strongly urged to meet or exceed these recommendations, especially in the development of prospective digital imaging projects. The Working Group proposes that the Recommended Guidelines become Required Guidelines for projects that begin after 2001.*

Under certain circumstances, digital image materials will also be accepted for deposit that meet only the Minimal Requirements (see Chart 2). It should be understood that such files may place a heavier burden on the central depository staff and these files may need to be reformatted to be managed effectively. *Further, these files may have a relatively short-term life expectancy (e.g., less than 10 years), and may be subject to de-accessioning should the expense associated with maintaining them outweigh the value of preserving and making them accessible.* Only under extraordinary circumstances (e.g., last copy of an item) will digital images files be accepted that do not meet the Minimal Requirements presented in Chart 2.

Table 1: Digital Master Image Files— Recommended Imaging Requirements

Document Type	Resolution	Bit Depth	Enhancements Allowed	File Format	Compression
Printed Text ²	600 dpi	bitonal	Sharpening, descreening, cropping, deskewing, and despeckling	TIFF 5 & 6	Lossless compression (e.g., ITU-G4)
Rare/damaged printed text	400 dpi	8-gray or 24-color	Contrast stretching Minimal adjustments for tone and color	TIFF 5 & 6	Uncompressed or lossless compression (e.g., LZW)
Book Illustrations	400 dpi ³ 600 dpi with enhancement	8-gray or 24-color ----- bitonal	Contrast stretching Minimal adjustments for tone and color ----- Descreen/rescreen, sharpen	TIFF 5 & 6	Uncompressed or lossless compression (e.g., ITU-G4, LZW)
Manuscripts	300-500 dpi	8-gray or 24-color, if color present in the original	Contrast stretching Minimal adjustments for tone and color	TIFF 5 & 6	Uncompressed or lossless compression (e.g., LZW)
Maps & other oversized items	300-400 dpi	8-gray or 24-color	Contrast stretching Minimal adjustments for tone and color	TIFF 5 & 6	Uncompressed or lossless compression (e.g., LZW)
Graphic Art	400-600 dpi	8-bit/ channel internal reduction	Contrast stretching Minimal adjustments for tone and color	TIFF 5 & 6	Uncompressed or lossless compression (e.g., LZW)
Photographic Prints	400 dpi	8-bit/ channel internal reduction	Contrast stretching Minimal adjustments for tone and color	TIFF 5 & 6	Uncompressed or lossless compression (e.g., LZW)

² Although 600 dpi 1-bit is a defacto standard for printed text, a comparable or richer text file may be produced in grayscale at 400 dpi.

³ Random or irregular halftones and those produces in color may be imaged at lower resolution, e.g., 300 because there is a lower incidence of moiré. It is recommended that high quality book illustrations, such as aquatints, collotypes, and engravings, especially those produced as separate plates, be retained for their artifactual value.

Works of art on paper	400 dpi	8-bit/ channel internal reduction	Contrast stretching Minimal adjustments for tone and color	TIFF 5 & 6	Uncompressed or lossless compression (e.g., LZW)
Transparencies	4000-5000 on long end or 400 dpi on output > 8" x 10"	8-bit/ channel internal reduction	Contrast stretching Minimal adjustments for tone and color	TIFF 5 & 6	uncompressed or lossless compression; (e.g., LZW)
Microfilm	600 dpi blown back to original size ---- 300-400 dpi blown back to original size	Bitonal ---- 8-bit gray	Sharpening, descreening; cropping deskewing, and despeckling	TIFF 5 & 6	Uncompressed or lossless compression (e.g., ITU-G4, LZW)

Table 2: Digital Master Image Files— Minimal Imaging Requirements

Document Type	Resolution	Bit Depth	Enhancements Allowed	File Format	Compression
Printed Text	300-400 dpi	bitonal	Sharpening, descreening; cropping, deskewing, and despeckling	TIFF 4, 5, & 6 JFIF/ JPEG	lossless or visually lossless compression; e.g., modest JPEG ($\geq 10:1$)
	----- 200 dpi	----- 8-gray	----- Minimal adjustments for tone		
Rare/damaged printed text	300 dpi	8-gray or 24-color	Contrast stretching	TIFF 4, 5 & 6 JFIF/ JPEG	lossless or visually lossless compression; e.g., modest JPEG ($\geq 10:1$)
	----- 600 dpi	----- bitonal	----- Minimal adjustments for tone and color Sharpening, descreening; cropping deskewing, and despeckling		
Book Illustrations	300 dpi	8-gray or 24-color	Contrast stretching adjustments for tone and color	TIFF 4, 5 & 6 JFIF/ JPEG KPCD	lossless or visually lossless compression; e.g., modest JPEG ($\geq 10:1$)
	----- 600 dpi, with enhancement	----- bitonal	----- Sharpening, descreening; cropping deskewing, and despeckling		
Manuscripts	200 dpi	8-gray or 24-color, if color present in original	Contrast stretching adjustments for tone and color	TIFF 4, 5 & 6 JFIF/ JPEG	lossless or visually lossless compression; e.g., modest JPEG ($\geq 10:1$)
	---- 400-600 dpi, with enhance-ment	---- 1 bit	----		
Maps & other oversized items	200 dpi	8-gray or 24-color	Contrast stretching adjustments for tone and color sharpening	TIFF 4, 5 & 6 JFIF/ JPEG	lossless or visually lossless compression; e.g., modest JPEG ($\geq 10:1$)
Graphic Art	300 dpi	8-gray or 24-color	Contrast stretching Documented color correction, sharpening	TIFF 4, 5 & 6 JFIF/ JPEG KPCD	lossless or visually lossless compression; e.g., modest JPEG ($\geq 10:1$), Image Pac
Photographic Prints	300 dpi	8-gray or 24-color	Contrast stretching Documented color correction, sharpening	TIFF 4, 5 & 6 JFIF/ JPEG KPCD	lossless or visually lossless compression; e.g., modest JPEG ($\geq 10:1$), Image Pac
Works of art on paper	300 dpi	8-gray or 24-color	Contrast stretching Documented color correction	TIFF 4, 5, & 6	lossless or visually lossless compression;

			sharpening	JFIF/ JPEG KPCD	e.g., modest JPEG ($\geq 10:1$), Image Pac
Transparencies	3000 pixels on long end	8-gray or 24-color	Contrast stretching Documented color correction sharpening	TIFF 4, 5 & 6 JFIF/ JPEG KPCD	lossless or visually lossless compression; e.g., modest JPEG ($\geq 10:1$), Image Pac
Microfilm	300-400 dpi ----- 200 dpi	bitonal ----- 8-gray	Sharpening, descreening; cropping deskewing, and despeckling ----- Minimal adjustments for tone	TIFF 4, 5 & 6 JFIF/ JPEG	lossless or visually lossless compression; e.g., modest JPEG ($\geq 10:1$)

C. Quality Control

Quality control (QC) is an integral component of creating digital content that will retain value and utility over time. QC encompasses procedures and techniques to verify the quality, accuracy, and consistency of digital products. The Digital Imaging Tutorial (<http://www.library.cornell.edu/preservation/tutorial/>) outlines the main points of a quality control program. A fully developed strategy for establishing such a program is presented in Oya Y. Rieger, "Establishing a Quality Control Program," in *Moving Theory into Practice: Digital Imaging for Libraries and Archives*, pp. 61-83 (Research Libraries Group, 2000). Depositors may request a copy of this chapter from the Department of Preservation by writing to preserve@cornell.edu.

QC Recommendations

1. *Scope of Inspection:* Inspect the quality of the digital image files, the accuracy and consistency of metadata, and the integrity of the storage media for delivery to the central depository.
2. *Extent of Inspection:* Establish a sampling frequency for each aspect of the QC program. Recommended frequency is 100% of all image files and accompanying metadata; minimal requirement is 10% sampling of each image/metadata batch.
3. *Type of Inspection:*
 - a. *Image Files*

The key factors in image quality assessment are resolution, color and tone, and overall appearance. QC can be conducted by visual inspection of images on-screen or via printouts, although it is important to note that quality assessment, especially for tone and color may be highly subjective and changeable according to the viewing environment and the characteristics of monitors and printers. The viewing environment and all links in the imaging system (including the scanner, monitor, and printer) should be carefully controlled. Image quality can also be judged through the use of technical targets (resolution, tone, and color) and increasingly through software. Appendix 1 includes a list of questions to ask in assessing resolution, detail, tone, and color appearance.
 - b. *Metadata*

Metadata has a central role in processing, managing, accessing, and preserving digital image collections. Because of the crucial role it plays in the life cycle of image collections, metadata review should be an integral part of a quality control program. Metadata QC can be automatic or manual or a combination of the two. QC should verify the following: data integrity, form and validity, accuracy of derived data, correctness of data, accuracy and completeness of components, and dynamic metadata. Richard Marisa

describes these aspects and makes recommendations for metadata QC in a sidebar to the chapter referenced above.

4. Pre-Depository Storage and Maintenance Requirements

This section deals with storage and maintenance of digital resources prior to deposit. To ensure long-term viability, transferees shall provide a secure and reliable storage environment for the digital files. Good storage practice plays a key role in preventive preservation, which is a crucial strategy to control and reduce resource requirements associated with preservation. The following list highlights the key requirements for proper storage and maintenance of digital collections that may ultimately be transferred to the central depository.⁴

1. *Media:*

- Store master files on high quality, industry standard digital tape, magnetic disks, CD-R, or other contemporary media approved by the Depository staff.
- Check media periodically for readability depending on the manufacturer's recommendations.

2. *Backups:*

- Create backups of the master files and store off-site in a secure location.

3. *Recording/Reading Devices:*

- Monitor the recording and access devices, such as tape drives, and make sure that they are of good quality and well-maintained (note that problems with the access devices e.g., head/media crashes are one of the most common causes of damage to magnetic storage media).

4. *Storage:*

- Store media in a controlled environment. The accepted ranges for temperature are 62°-68° (65° optimum) and for humidity are 35%-45% (40% optimum).
- Establish consistent levels within the acceptable range (this is more important than attempting to maintain the optimum temperature and humidity.)
- Store media away from strong magnetic fields.
- Maintain a clean operating environment.
- Minimize the handling and use of magnetic storage media to reduce wear.
- If media is stored off-line, store it vertically in appropriate containers.

⁴ The resources used in developing this section included:

Maggie Jones and Neil Beagrie, *Preservation Management of Digital Materials Workbook*, 2000, www.jisc.ac.uk/dner/preservation/workbook/

Neil Beagrie and Daniel Greenstein, *A Strategic Policy Framework for Creating and Preserving Digital Collections*, Arts and Humanities Data Service Executive, 1998, ahds.ac.uk/manage/framework.htm.

J. Van Bogart, *Magnetic Tape Storage and Handling*, CLIR, 1995, www.clir.org/pubs/reports/pub54.html

5. *Security:*
 - Control access to storage facilities and processing areas. Store media in a separate, preferably lockable area.
 - Employ appropriate security systems and procedures to protect the authenticity of the collections and ensure no deliberate or inadvertent changes take place.

6. *Refreshing:*
 - Digital files maintained for an extended period prior to deposit should be refreshed to new media regularly, taking into consideration the recommendations of the media supplier for certain environmental conditions and following the trends for more efficient storage technologies (e.g., refreshing may be necessary when new storage systems are purchased).
 - Follow a verification procedure such as checksum or MD5 to ensure the authenticity and integrity of the files after media refreshing.

7. *Documentation:*
 - Document actions taken during refreshing or other maintenance operations that may affect the integrity of files.

5. Metadata Requirements

A. Descriptive Metadata

Descriptive metadata, loosely defined as information used by the delivery system for resource discovery and to identify resources, are an integral part of any digital access system and are essential to the long-term maintenance of digital files. The Cornell University Library currently uses the online catalog as its database of record for descriptive metadata. This will continue to be the case as the library begins its work of providing long-term access to digital files. However, depending on the nature of the digital material, additional descriptive metadata records may be needed to support fully this endeavor.

Title-level MARC records should be created for all monographs and serials that are added to the digital depository. As has been observed repeatedly with aggregations of electronic journals and e-books, the presence of title-level MARC records in the online catalog will increase the use of these resources. Furthermore, records for these materials will be shared in the national utilities and may be used by other institutions to provide access to resources housed at Cornell. In addition to increasing use, MARC records can be used to form the basis of descriptive metadata used in an electronic delivery system - either through links to a system or through the duplication of a portion of the MARC data in the digital system. Although particular procedures and standards have changed over time, all MARC records created should follow current cataloging standards and should meet at least the minimal requirements defined by current, local procedures. Appendix 2 includes an example of descriptive metadata. Cornell's policies for cataloging digital materials are embodied in:

Cataloging Procedures for Networked Electronic Resources
<http://www.library.cornell.edu/voyager/Bibs/ECat/e-catTOC.html>

These procedures provide guidelines for handling basic issues as well as using separate and multiple-version records, treating multiple electronic versions, and other complex situations. As noted in section 2.1 of the *Cataloging Procedures for Networked Electronic Resources*:

This document provides local usage guidelines for cataloging networked electronic resources in Voyager and in the Library Gateway. It does not cover everything one needs to know to process these items. For a complete list of field definitions, appropriate tags, and national standards, consult USMARC and CONSER documentation. For more general instructions on the cataloging of networked e-resources, see Nancy Olson's *Cataloging Internet Resources: A Manual and Practical Guide*, 2nd ed.

This advice should also be followed when approaching the descriptive cataloging of digital image materials. Although local practice takes precedence, catalogers may also consider the following publication:

Library of Congress, *Draft Interim Guidelines for Cataloging Electronic Resources*, http://lcweb.loc.gov/catdir/cpso/elec_res.html

In addition to following the standards listed in these procedures, cataloging staff should be certain to take the following into account:

- Particular attention should be paid to the use of the 899 "Local series code" field. A unique code is to be created for each digital collection and recorded in the 899 field of each record that reflects a title in that collection. This code will collate each collection of MARC records in the online catalog to assist in record maintenance, data migration, and other functions. Instructions for assigning 899 codes are available in *Cataloging Procedures for Networked Electronic Resources*. Establishing an 899 code should be done in cooperation with staff from the Technical Services Support Unit.
- Records should use appropriate 007 values according to MARC 21 <http://lcweb.loc.gov/marc/bibliographic/ecbd007s.html>. In 2000, many 007 values were enhanced to capture needed data concerning preservation and reformatting issues and to record image production values. Adding this information to previously cataloged items may not be practical but is desirable if possible. In addition, all prospective cataloging should take advantage of this useful field.
- Special attention should be given to LCRI 1.11A "Facsimiles, photocopies, and other reproductions" (revised Summer, 2000). This LCRI provides assistance in handling non-microfilm reproductions of print materials (including digital reproductions) and suggests methods for recording information concerning the digitization process. Cataloging staff should also follow current, local procedures for handling such

reproductions. In particular, be sure to include either a 533 "Reproduction Note" in the bibliographic record or an 843 "Reproduction Note" in the holdings record. This information is essential in tracking the agency responsible for the creation of the reproduction.

- Notes on image file format are routinely added to the MARC record. This information is likely also to be contained in administrative metadata that will be part of a digital collection management system. When adding this information, be aware that if materials are migrated, this information will become outdated and may need to be changed in the MARC record.
- Cornell University Library has decided that persistent, stable URLs should be used in most MARC records for electronic resources, such as digital image files, and has chosen OCLC's PURL server as the method for creating such persistent identifiers. URLs that appear in the record may need to have PURLs created for them. The procedures for creating PURLs (including information on resources that do not require PURLs) are defined in Section 6 of the *Cataloging Procedures for Networked Electronic Resources*:
<http://www.library.cornell.edu/voyager/Bibs/ECat/e-cat6.html>.
- The Technical Services Support Unit handles the creation and maintenance of PURLs and should be consulted about setting up any needed PURLs. PURLs for collections of digital resources utilize the local series code listed in the 899 as well as the online catalog bibliographic record ID. The creation of PURLs for these materials is important. The PURL will reside in the online catalog and will also be part of all records added to the bibliographic utilities. This is particularly important for long-term access since image file locations are likely to change over long periods of time.

As cataloging issues arise, these documents will change and other documents will emerge to standardize the treatment of electronic resources. If you are unfamiliar with the treatment of these materials, be sure to check with other local catalogers about the most current methods for approaching the creation of MARC records for this type of material.

Unlike monographs and serials, many materials (such as archival collections, pamphlets, manuscripts, letters, photographs and other visual resources) often do not have title-level MARC records created for them. In many cases, there is a one-to-many relationship between the MARC record and the material itself. These materials present a unique challenge to the long-term accessibility of descriptive metadata. Although title-level records must be created for monographs and serials, they are not mandatory for other materials but do represent an ideal level of access. These materials must have at least collection-level MARC records. These records should follow the standards outlined above as closely as possible given the differences inherent in the material. More detailed descriptive metadata will be needed for the creation of adequate access systems for these materials. The record structure of this metadata is likely to be unique and idiosyncratic. The same is true of article-level (or chapter-level) metadata that might be recorded for electronic serial (or monograph) collections. This situation will present difficulties for the long-term maintenance of this data. As soon as it is practical and possible, idiosyncratic data structures should be migrated to more universal standards. For instance, the use of Dublin Core may be preferred as a long-term storage structure for

bibliographic data over a homegrown structure. However, this will only be practical when access systems are created to utilize metadata stored and structured in such a way.

When archival materials are digitized, that process should follow the creation of a collection finding aid, or the recording of equivalent information, and should be guided by that finding aid, which will contain the most comprehensive descriptive metadata about the collection. Archival finding aids should be expressible with the Encoded Archival Description DTD. Although the finding aid will contain information about the collection's physical organization or location, it should ideally describe the intellectual arrangement of the collection. The finding aid should be viewed as the principal access and navigation mechanism into the images. Collection-level MARC records should link to the finding aid when possible. To help track the relationship among the digital images, the EAD finding aid or equivalent, and the collection-level MARC record, the EAD finding aid or equivalent should include the MARC record bibliographic record ID and/or the collection number. Since library management systems will change over time, the collection number is likely to remain a stable identifier.

The Rare and Manuscript Collections has begun to define an "RMC EAD template," which can be found at <http://cidc.library.cornell.edu/xml/template/>. As this document evolves, it will define minimum field/element requirements, as well as a set of local content standards. In addition, it should reflect wider archival community standards concerning the interoperable use of EAD.

Following the criteria described above is necessary not simply for information retrieval and access but also for the maintenance of descriptive metadata. Since MARC is a rich, descriptive metadata standard, it is possible to crosswalk portions of this data to less-complex, short-term formats used by access systems. Unlike metadata structures like the Dublin Core, MARC has a clear process for supporting and maintaining both the form and content of records, can be shared easily among institutions, and is widely recognized, supported, and documented. Non-MARC descriptive metadata used in a delivery system should make reference to the MARC record(s) associated with the material, preferably by citing the online catalog record ID(s). This will enable CUL to update delivery system metadata to capture changes in names, titles, and subjects as well as new standards, such as Unicode. In addition, as common, non-MARC metadata standards develop, MARC will be used to populate records that follow those standards when possible to limit the amount of customized data migration that will need to take place.

If CUL widely adopts a metadata structure that can be used in place of MARC, materials submitted to the depository may need records in that structure as well as or instead of MARC records. In addition, as metadata formats and standards emerge, CUL may migrate all MARC data into other metadata structures for long-term preservation of descriptive metadata.

B. Structural Metadata

Structural metadata provides essential information needed to guarantee adequate long-term access to digital image files. Although less important for single-image objects (such as photographs, artwork, posters, or maps), structural metadata is invaluable for access to objects comprised of more than one image or for single images that are related to other objects (such as a series of letters or memos meant to be read in tandem). Unfortunately the way that structural metadata is stored will be dependent on the system used to provide access to the image material. However, each collection's structural metadata should be captured and stored in a uniform manner. This is needed to create stable and consistent access systems as well as to preserve both metadata and data files.

As part of the IMLS grant "Preserving Cornell's Digital Image Collections: Implementing an Archival Strategy" (1999-2000), the project team defined a series of structural elements that were identified as mandatory (M), mandatory if applicable (MA), and optional (O) for digital image collections at Cornell. Table 3 was generated by examining various preservation metadata proposals for descriptive and administrative metadata to facilitate preservation decisions.

Table 3: List of Structural Metadata Elements for Digital Image Collections

M = Mandatory; MA = Mandatory if applicable; O = Optional

All Materials

Relationship to Other Resources (MA)
Metadata Locations (M)
Start image/page (M)
End image/page (M)

Manuscripts

Title page (MA)
Colophon (O)
Caption (O)
Heading (O)
Leaves (O)
Illustrative matter (O)
List of illustrations (O)
Table of contents (MA)
List of tables (O)
Page numbers (M)
Blank page (M)
Marginalia (O)
Front matter (MA)
Back matter (MA)

Monographs

Title page (M)
Copyright page (M)
Table of contents (M)
List of illustrations (O)
List of tables (O)
Beginning segments (e.g., forward, preface, acknowledgements) (O)
End segments (e.g., epilogue, afterword, conclusion, etc.) (O)
Chapters/parts (O)
Notes (O)
Bibliography (O)
Index (M)
Colophon (O)
Errata (O)
Page numbers (M)
Blank page (M)

Pamphlets

Title page or cover (M)
Copyright page (MA)
Table of contents (MA)
List of illustrations (O)
List of tables (O)
Beginning segments (e.g., forward, preface, acknowledgements) (O)
End segments (e.g., epilogue, afterword, conclusion, etc.) (O)
Chapters/parts (O)
Notes (O)

- Bibliography (O)
- Index (MA)
- Colophon (O)
- Page numbers (M)
- Blank page (M)

Serials

Entire Publication

- Volume (M)
- Issue (M)
- Supplements (M)
- Table of contents (M)
- Index (at issue and volume level) (M)
- Corrections and retractions (O)
- Serial front matter (M)
- Serial part (O)
- Serial section (O)
- Name index (if separate from other index) (O)
- Subject index (if separate from other index) (O)
- Errata (O)
- Page numbers (M)
- Blank page (M)

Articles

- Article title (O)
- Author (O)
- Abstract (O)
- Date (O)
- Tables/figures (O)
- Errata (O)
- Page numbers (M)
- Blank page (M)

C. Preservation Metadata

Preservation metadata encompasses a range of information that is required for the short- and long-term management of digital image files. This category of metadata includes both micro information that describes the technical specifications of an image collection as well as administrative information that will support future preservation decision making and action. This section outlines the preservation metadata requirements in two categories. The first section, Digital Image Collections Inventory, outlines an information system that aims to collect and maintain high-level administrative metadata. The second category, Technical Metadata, attempts to collect technical information at a micro-level.

Digital Image Collections Inventory

The goal of the Digital Image Collections Inventory database is to provide profiling information on CUL's digital image collections. It aims to include general information that will support preservation administration and decision-making. This inventory approach tries to address the difficulty associated with gathering base-level information about the library's individual image collections. Every project team is required to complete this questionnaire during the project implementation phase as the collection of elementary information has proven to be very difficult to recreate after the fact. Table 4 lists the recommended data elements for such an inventory database. Currently, this type of information is not required of staff members who are involved in different stages of digital imaging projects. There is no formal recording or sharing obligation. Although this questionnaire does not address the need for detailed administrative metadata (and how it is recorded and maintained), it presents an easily attainable and effective approach for short-term management of administrative metadata until we have more sophisticated systems and standards in place. The technical metadata standard that is currently being developed by ANSI/NISO (<http://www.niso.org/commitau.html>) will complement this approach by providing a framework to record in-depth information on the technical specification of individual collections.

The short-term plan to implement this inventory is through the development of a Web survey (with an automated recording system). However, the long-term recommendation for the collection and manipulation of such data is to create a DTD for XML (or an XML schema) implementation. Project coordinators and participants can easily collect the information elements requested in the inventory, with an estimated time involvement of one hour. However, some of the required information is dynamic and cumulative (e.g., refreshing and migration history) and therefore would require ongoing updates. After the completion of the key sections by project staff during the implementation phase, the central depository staff may need to update different parts of the inventory throughout the life cycle of a particular collection. The suggested frequency for the revision of the dynamic fields is one year.

Table 4: Digital Image Collections Inventory: Data Fields

<i>Project description</i>
project title
year the collection was created
project leaders/coordinators, team members
project partners
sources of funding
reason for the project
<i>Source type and characteristics</i>
document type (e.g., printed text, book illustrations, color photographic prints, manuscripts, etc.)
physical dimensions (category: regular, oversize - if possible exact size, or "size varies" statement with min and max measurements -, size varies, but no greater than 8.5 x 11 or some such)
scanned from original or film intermediate
subject matter
<i>Collection size</i>
total file size of the collection including image and metadata files, programs, scripts, etc. (estimated or actual)
number of images
<i>Storage media</i>
type and location
<i>Scanning information</i>
resolution
bit depth
color space or CLUT information for color documents
file format and version
compression technique, version, and ratio
scanner used
vendor vs. in-house scanning
<i>Processing information</i>
any image enhancements on the master copy? E.g., how were halftones handled? Any special treatment?
derivatives created (access, processing; such as scaled/reformatted copies for Web delivery, OCR'ed images, etc.)
<i>Metadata</i>
file header (if possible tags used)
what kind of descriptive metadata – where and how recorded? (e.g., MARC, Dublin, PURL, etc.)
what kind of structural metadata – where and how recorded? (SGML, XML, structuring tags, external metadata, etc.)
what kind of technical metadata – where and how recorded?
special collections – finding aid information
<i>Access mechanisms</i>
online/offline
Web address

<i>System/interface design and characteristics</i>
system specifications (e.g., based on Hunter, OpenText, etc.)
known system requirements
key interface features (forms and style sheets, use of JavaScripts, etc.)
<i>Refreshing/migration history</i>
<i>Rights management & Authenticity</i>
document the process of clearing copyright issues
license information
display and transmission restrictions, right holders
any security/authenticity measures (e.g., watermark)
chain of custody

Appendix 3 includes a sample entry for the Save America's Treasures collection to demonstrate the use of this inventory.

Technical Metadata

One of the key requirements of a preservation policy is to have a framework for collecting and recording technical metadata to safeguard information that may be essential in monitoring and rescuing files in the face of changing technologies. Continued viability of the CUL digital image collections heavily depends on the availability of information on technical characteristics of collections, technological dependencies, change history, and rights management. Technical metadata serves several purposes. In a managerial context, it supports image quality assessment, image enhancement and processing, and facilitates work-flow management. Although there is limited evidence at this point, technical metadata is also seen as an important source for long-term collection management. The Cornell University Library intends to adopt the technical metadata standard that is currently being developed by NISO. Standardization of technical metadata will facilitate a systematic approach in recording and managing technical image information. The charge of the Technical Metadata for Digital Still Images Standards Committee (<http://www.niso.org/commitau.html>) is to review and revise the Data Dictionary for Technical Metadata for Digital Still Images (Working Draft, 1.0, July 2000). The proposed data dictionary presents a comprehensive list of elements required to describe the technical features of image files. The data fields are organized in four groups: basic image parameters, image creation, image performance assessment, and change history. The standard in development is not fully addressing the implementation question so it will become a CUL-based decision whether to develop and adopt a DTD for technical metadata to support an XML implementation.

Authenticity

Master digital images deposited in the Cornell Digital Library should represent accurate, complete, and trusted versions of the original source materials. Second, they should have documentation demonstrating an unbroken chain of custody since their creation. Third, master digital files deposited in the Cornell Digital Library should have documentation demonstrating how the files have been protected from un-documented or unintentional change, such as tampering. This documentation should cover the technical procedures followed to ensure that the files retain quality, integrity, authenticity, and reliability after creation. If more than one version of the digital image materials exists, the version that has been most closely monitored and safeguarded will be the preferred version for deposit. Once deposited, the Depository staff shall continue to document the chain of custody and assume responsibility for safeguarding the authenticity of the digital collections.

ROLE AND RESPONSIBILITIES OF A CENTRAL DEPOSITORY

This report began with a strong recommendation to the Library Management Team to establish centralized responsibility for ensuring continuing access to digital image collections over time. This responsibility should take the form of a Central Depository that is administratively located within the Digital Library and Information Technology (D-LIT) infrastructure. The Central Depository's role will be to facilitate the long-term management and use of digital resources in the most cost-effective manner, based on the distinctive characteristics of the Cornell University Library system.

A detailed description of the responsibilities of the Central Depository will constitute the second part of this document, and will be prepared after the proposed feasibility study (see below). This forthcoming section will focus on the responsibilities of the Central Depository, and will cover the following key issues:

- Guidelines on transmission methods, documentation, and media for deposit
- Acquisition procedures and protocols to: verify the arrival, completeness, validation, and readability of deposited material; reformat or copy materials to new media; and provide for inventory control
- Maintenance/updating of depository guidelines
- Respective responsibilities and rights of the transferee and the depository staff, including on-going interactions
- Outreach to potential transferees
- Relationship to Digital Library staff (access and availability, derivative creation), collection development (selection), Technical Services (documentation, cataloging, and creation of PURLs), Access Services (technical reference and user support), and Preservation (policy development)
- On-going system maintenance practices, including
 - Storage, backup, and redundancy procedures
 - Data security, integrity, and auditing requirements
 - Collection monitoring
 - Media refreshing
 - Disaster recovery plans
- Upgrading/modification procedures (e.g., replacement images, metadata updates, alternative access versions)
- Rights management (including authorized use of software and content for preservation purposes, creation of rights clearance forms)
- Outsourcing, contractual arrangements, and collaboration with other units/institutions (offsite storage, redundancy requirements)
- Access policies (in conjunction with Digital Library staff)
- Preservation alternatives, including risk assessment associated with various strategies
- Technology monitoring
- Documentation, including maintenance of the image collections inventory
- Collection review; deaccession and disposition guidelines
- Resource requirements, including cost assessments

NEXT STEPS

A. Needs Assessment Survey

The guidelines presented in this document represent an important first step in developing an institutional digital preservation strategy. A necessary next step is to address the managerial processes of the depository, including the day-to-day and long-term technical and financial management of data. To be able to accurately describe the role of the central depository and its processes, the Working Group recommends that the library conduct a preservation readiness feasibility study. Such a study will be instrumental in identifying the central depository requirements for image collections and also understanding the CUL's existing technical and administrative infrastructure necessary for the attainment of these objectives. The goal of the feasibility study is to explore the following issues:

- 1) Explore the short- and long-term role of a central depository for digital collections
 - Identify the role and working relationship of the key players
 - Evaluate the existing organizational structure and staffing patterns to assess the readiness of CUL to fulfill preservation requirements.
 - Identify resource requirements including staffing (level and skills), equipment and space needs, financial planning, CUL-CIT collaboration, etc.⁵
- 2) Evaluate the relevance and applicability of various institutional policy frameworks, preservation architectures and procedures (e.g., assessment of OAIS and Encompass for long-term maintenance purposes) for CUL
- 3) Assess the effectiveness of current maintenance practices for long-term preservation
- 4) Evaluate the extensibility of the policies developed for digital image collections in meeting the needs of CUL's other digital collections and initiatives (e.g., Euclid, Harvest, and Prism)
 - Identify the divergence and convergence of preservation policies and frameworks for different digital formats (e.g., images, HTML files, numeric files, etc.)
- 5) Consider the use of a rating system to identify the permanence level of CUL's electronic resources (e.g., see the National Library of Medicine's permanence rating model for electronic resources, www.arl.org/newsltr/212/nlm.html)

⁵ After a thorough analysis of the existing cost studies for digital preservation, the group decided that there were no existing models that could be readily implemented for our purposes. The main challenge is that most of the existing studies are not itemized and do not indicate what is included in the cost estimates. In addition, most of them present preservation costs on a per gigabyte basis, focusing mostly on storage costs. There are considerable economies of scale in large archives, so calculating costs based on a gigabyte unit may not be accurate as a system continues to grow. Expenses for digital preservation start accumulating soon after selection for digitization and continue as long as access to the digital collection must be ensured.

B. Schedule and Procedures for Updating the Deposit Guidelines

The members of the Digital Preservation Policy Working Group will reconvene once a year to review the document and to identify sections that need to be updated. The co-chairs of the committee will continue to organize the updating process by convening these annual meetings and also following up after the meetings to reflect the recommendations to this document.

During the first year of implementation of the depository, the Digital Preservation Policy Working Group will continue in an advisory role in the development of the depository guidelines.

C. Publicity and Training

After this document is approved by the Library Management Team, it will be distributed to the library staff for comments and questions. In addition, there will be a number of orientation and training sessions to familiarize the library staff with the requirements articulated in this document. The Digital Imaging and Preservation Unit of the Department of Preservation and Conservation Department offers several educational materials and can accommodate Cornell staff in their week-long digital imaging workshops (<http://www.library.cornell.edu/preservation/workshop/>). The following online tutorial prepared by the department will be useful in introducing the staff to basic concepts related to creating and managing digital image collections:

Moving Theory into Practice: Digital Imaging Tutorial

<http://www.library.cornell.edu/preservation/tutorial/>

Appendix 1: Image Quality Assessment

This section is reproduced from:

Oya Y. Rieger, "Establishing a Quality Control Program," in *Moving Theory into Practice: Digital Imaging for Libraries and Archives*, pp. 61-83 (Research Libraries Group, 2000).

A. Questions to Ask in Evaluating Resolution and Detail

Compare the digital images (or their printouts) to the original documents (or to the intermediates):

Text/line Art Documents

1. Is the stroke adequately reproduced?
2. Is the significant detail adequately reproduced?
3. Is the smallest text readable?
4. Are individual line widths (thick, medium, and thin) rendered faithfully?
5. Are serifs and fine detail rendered faithfully?
6. Are adjacent letters as separate as they should be?
7. Are the open regions of lowercase characters retained (i.e., not filled in)?
8. Are the edges of individual letters or shapes as smooth or well defined (not ragged) as the original?⁶
9. Is there good contrast or differentiation between the text and the background?
10. Is there even illumination across the image (i.e., is the image washed out or too dark)?
11. Is there a gray cast or streaking in the background?
12. Is the document fully reproduced?

Continuous-Tone and Halftone Documents

(The first three questions apply only to continuous-tone documents.)

1. Is the stroke adequately reproduced?
2. Is the significant detail adequately reproduced?
3. Is fine detail in the darkest and lightest portions retained?
4. Are there even gradations across the image (e.g., no banding, streaking, newton rings, or graininess)?⁷

⁶ Edge raggedness relates to the smoothness or straightness of edges along lines at very close inspection. Pay special attention to curved and diagonal lines on characters and line graphics.

⁷ Streaking and graininess are typical film attributes that might be evident in the displayed image. Banding (varying lightness and darkness) may be attributable to improper lighting. Newton rings (circular impressions) can be introduced during the scanning of transparencies.

5. Is the image free of a moiré effect?⁸
6. Is the significant informational content adequately reproduced?
7. Is the document fully reproduced?
8. Is the image too light or too dark?

B. Questions to Ask in Evaluating Color and Tone Appearance

Compare the image to the original document, an intermediate, or color/grayscale bars.

Grayscale Images

1. Evaluate tone appearance in the highlights (lighter sections), midtones, and shadows (darker sections). Are the details in these different sections captured without any loss?
2. Is the image too light or dark overall?

These questions are based on the grayscale targets used in photography and scanning:

3. How many grayscale bars can you count on your grayscale image?
4. If your grayscale target is numbered, at what numbers do you cease to discern distinct shades of white, gray, and black?
5. Is there an overall color shift on the grayscale target?
6. Use the information option of your image viewing software to read the color (RGB) values presented at different grayscale steps. How do they compare to the reference values provided by the grayscale bar? What is the difference between the smallest and largest values for each color channel for individual color patches (variance indicates color imbalance)?
7. Display a histogram of your grayscale bar image.⁹ Are all digital levels from 0 to 255 used? Do you observe any clipping?

Color Images

1. Do you observe a color shift in the overall image or an obvious shift to a certain color?
2. Study the red, green, blue, and yellow colors. Do any show a color shift? Is it minimal or obvious?
3. Evaluate colors in the highlights, midtones, and shadows, especially red, green, blue, and yellow. Do any show a color shift? Is it minimal or obvious?
4. Is the image light or dark overall?

⁸ Moiré patterns are most noticeable in the lighter regions of an image and in areas of “low activity” (e.g., in the sky portion of a landscape halftone). In portions with busy content (high activity), moiré is often hidden. Evaluate halftones onscreen only at 1:1 (100%); any other view might introduce halftone patterns not native to the image file.

⁹ Rely on your image viewing software's user guide to find out how to create and evaluate histograms.

These questions are based on the grayscale and color targets used in photography and scanning:

5. How many grayscale bars can you count on your grayscale image?
6. If your grayscale target is numbered, at what numbers do you cease to discern distinct shades of white, gray, and black?
7. Do you notice an overall color shift on the grayscale or color target? If so, does it fall within your tolerance range?
8. As shown in figure 4, use the Window/Show Info option of the Adobe Photoshop software to read the color (RGB) values presented at different grayscale steps. How do they compare to the reference values on the grayscale bar? What is the difference between the smallest and largest values for each color channel for individual color patches?
9. Use the hue, saturation, and brightness adjustments of your image viewing software to evaluate the individual colors of the color bar. Comparing the colors on the color target to the original color target, is there a color shift to a certain color? Is it minimal or obvious?
10. Even if the color bar evaluation is satisfactory, compare different sections of the document to the image: is the color satisfactory?

C. Overall Evaluation

The final overall evaluation of an image should combine all the individual factors that contribute to its quality, such as capture system performance, resolution, dynamic range, and color accuracy. Your subjective evaluation should confirm your objective conclusion that an image is satisfactory. Remember that it is *impossible* to generate an image that will fully replicate the look and feel of a document. Ask these questions to evaluate overall image quality:

Does the image convey all the significant information included in the original document (e.g., translucency in a watercolor painting, overlay in an oil painting, quality and texture of paper, etc.)? If not, how much does this affect your satisfaction with the image?

Compared to the original document, is the image:

- unacceptable?
- adequate but diminished?
- comparable?
- improved?

Will the user be satisfied with the image as a document surrogate, or will the image serve just as a basic access tool?

Even if the image passes your subjective and objective inspection based on the grayscale and color targets:

- Is the image's overall dynamic range adequate?

- Are you satisfied with the general color appearance of the image?

Appendix 2: Descriptive Metadata Example

MOA Multiple-Version OPAC View

Author/Creator: United States. Naval War Records Office.

Title: Official records of the Union and Confederate Navies in the War of the Rebellion.

Published: Washington, Govt. Print. Off.,

Description: 30 v. : ill., maps (part fold.) ports. ; 23 cm. index.

Electronic Access: <http://resolver.library.cornell.edu/moap/anu4547>

Subjects:

United States. Navy--History--Civil War, 1861-1865.

Confederate States of America. Navy--History.

United States--History--Civil War, 1861-1865--Naval operations--Confederate States.

Other Names: United States. Office of Naval Records and Library.

Series:

Office memoranda (United States. Naval War Records Office)

Office memoranda (United States. Naval War Records Office).

Notes:

Issued in the congressional series as House documents.

Series 1, v.1-27; Series 2, v.1-3.

Indexes: Ser. 1, v.1-13. 1 v. (Issued as the Office's Office Memoranda) (E591.U575)

Ser. 1, v.1-ser. 2, v.3. 1 v.

Location: *Networked Resource

Call Number: ONLINE

Status: Available

Volumes : Ser.1-2

Indexes: Index

Reproduction Note:

Computer file. Ser.1-2. Ithaca, N.Y. : Cornell University Library, 1995. [27,515] image files.

Notes:

Files for the images of individual pages are encoded in Aldus/Microsoft TIFF Version 5.0 using facsimile-compatible CCITT Group 4 compression.

Location: Olin Library

Call Number: E591 .U58

Copy Number: 2

Status: Available

Indexes: v.1

Location: Olin Library

Call Number: E591 .U58 1894a

Status: Available

Volumes : Ser.1-2

Indexes: Index

Reproduction Note: Reproduction from digital master. Ser. 1-2. Ithaca, N.Y. : Cornell University Library, 1995. 23 cm.

MOA multiple-version MARC record

000 01515cam 2200337 a 450

001 2797023

005 19990908120000.0

008 870716m18941922dcuabc f001 0 eng d

010 __ |a 06035188
 035 __ |a (NIC)notisANU4547
 040 __ |a NIC |c NIC
 043 __ |a n-us---
 110 1_ |a United States. |b Naval War Records Office.
 245 10 |a Official records of the Union and Confederate Navies in the War of the Rebellion.
 260 __ |a Washington, |b Govt. Print. Off., |d 1894-1922.
 300 __ |a 30 v. : |b ill., maps (part fold.) ports. ; |c 23 cm. |e index.
 490 1_ |a U. S. Naval War Records Office. Office memoranda
 500 __ |a Issued in the congressional series as House documents.
 500 __ |a Series 1, v.1-27; Series 2, v.1-3.
 500 __ |a Indexes: Ser. 1, v.1-13. 1 v. (Issued as the Office's Office Memoranda) (E591.U575)
 500 __ |a Ser. 1, v.1-ser. 2, v.3. 1 v.
 610 10 |a United States. |b Navy |x History |y Civil War, 1861-1865.
 610 10 |a Confederate States of America. |b Navy |x History.
 651 _0 |a United States |x History |y Civil War, 1861-1865 |x Naval operations |z Confederate States.
 710 1_ |a United States. |b Office of Naval Records and Library.
 830 _0 |a Office memoranda (United States. Naval War Records Office)
 830 _0 |a Office memoranda (United States. Naval War Records Office).
 899 _0 |a MOAProj
 856 40 |u <http://resolver.library.cornell.edu/moap/anu4547> |x <http://moa.cit.cornell.edu/MOA/MOA-JOURNALS2/OFRE.html>

Appendix 3: Sample Entry for Digital Image Collections Inventory

Save America's Treasures

Cornell University Library, Preservation and Conservation Department
[12/18/00 - Robert S. Glase]

Project description

<i>Project title:</i>
<i>Save America's Treasures</i>
<i>Year the collection was created:</i>
September 2000
<i>Project leaders/coordinators, team members:</i>
Barbara Berger Eden, John Dean, Elaine Engst, Preservation-Conservation staff
<i>Project partners:</i>
CUL Preservation-Conservation, CUL Division of Rare and Manuscripts Collections
<i>Sources of funding:</i>
Save America's Treasures Grant; Total: \$662,000 Direct Funds: \$331,000; Cost share \$331,000
<i>Reason for the project:</i>
Digitization of a historically significant anti-slavery collection

Source type and characteristics

<i>Document type (e.g., black and white text-based material, heavily illustrated text, color photographs, special collections materials, slides, etc.):</i>
Black and White, some illustrations/engravings
<i>Physical dimensions (category: regular, oversize - if possible exact size, or "size varies" statement with min and max measurements -, size varies, but no greater than 8.5 x 11 or some such):</i>
Average 4 x 7
<i>Scanned from original or film intermediate:</i>
Original
<i>Subject matter:</i>
Abolition of slavery in United States, Conditions of slaves in United States (early nineteenth century)

Collection size

<i>Total file size of the collection (estimated or actual):</i>
Project is approximately 20% complete: 4861.0 MB

<i>Number of images:</i>
Project is approximately 20% complete: 42,032
The total has been estimated at around 700,000 images.

Storage media

<i>Type and location:</i>
None - direct to library server

Scanning information

<i>Resolution:</i>
600 dpi bitonal
400 dpi grayscale
<i>Bit depth:</i>
600 dpi bitonal-1 bit
400 dpi grayscale-8 bit
<i>Color space or CLUT information for color documents:</i>
N/A
<i>File format and version:</i>
Tiff 6.0
<i>Compression technique, version, and ratio:</i>
N/A
<i>Scanner used:</i>
Xerox Doculmage 620s
<i>Vendor vs. in-house scanning:</i>
In-house

Processing information

<i>Any image enhancements on the master copy? (E.g., how were halftones handled? Any special treatment?):</i>
400 dpi grayscale used as needed for halftones
<i>Derivatives created (access, processing; such as scaled/reformatted copies for Web delivery, OCR'ed images, etc.):</i>
None to date (note: to be created by RMC)

Metadata

<i>File header (if possible tags used):</i>
N/A
<i>What kind of descriptive metadata – where and how recorded? (e.g., MARC, Dublin, PURL, etc.):</i>
MARC-To be created by RMC
<i>What kind of structural metadata – where and how recorded? (SGML, XML, structuring tags,</i>

<i>external metadata, etc.):</i>
Image tags to be created
<i>What kind of technical metadata (where and how recorded?):</i>
File size, image count. Recorded in Filemaker Pro.
<i>Special collections – finding aid information:</i>
N/A

Access mechanisms

<i>Online/offline:</i>
Online
<i>Web address:</i>
None to date

System/interface design and characteristics

<i>System specifications (e.g., based on Hunter, OpenText, etc.):</i>
MOA specifications (note: to be implemented by David Ruddy.)
<i>Known system requirements:</i>
Web interface through CUL digital library
<i>Key interface features (forms and style sheets, use of JavaScripts, etc.):</i>
None

Refreshing/migration history

Refresh
None to date
<i>Migrate</i>
None to date

Rights management & Authenticity

<i>Document the process of clearing copyright issues:</i>
To be resolved by RMC
<i>License information:</i>
None
<i>Display and transmission restrictions, right holders:</i>
None
<i>Any security/authenticity measures (e.g., watermark):</i>
None
<i>Chain of custody:</i>
RMC

