



Moving Theory into Practice *Digital Imaging Tutorial*

Questions?

Table of Contents

Using This Tutorial

Printing This Tutorial

Preface

- 1. Basic Terminology**
- 2. Selection**
- 3. Conversion**
- 4. Quality Control**
- 5. Metadata**
- 6. Technical Infrastructure**
 - A. Digitization Chain**
 - B. Image Creation**
 - C. File Management**
 - D. Delivery**
- 7. Presentation**
- 8. Digital Preservation**
- 9. Management**
- 10. Continuing Education**

View in:

← **Spanish**

← **French**

© 2000-2003 Cornell University Library/ Research Department



Moving Theory into Practice *Digital Imaging Tutorial*

Questions?

Use this form to send your questions and comments about the tutorial.

Name:

Complete Email Address:

Question/Comment:


Back Next
← Contents

View in:
← Spanish
← French

© 2000-2003 Cornell University Library/Research Department



Moving Theory into Practice *Digital Imaging Tutorial*

PREFACE	5. METADATA	7. PRESENTATION
1. BASIC TERMINOLOGY	definition	introduction
digital images	types and functions	formats/compression
resolution	creation	web browsers
pixel dimensions	additional reading	network
bit depth		scaling
dynamic range		monitors
file size	6. TECHNICAL INFRASTRUCTURE	image quality
compression	A. DIGITIZATION CHAIN	guidelines
file formats	introduction	additional reading
additional reading	components	8. DIGITAL PRESERVATION
2. SELECTION	system integration	definition
introduction	B. IMAGE CREATION	challenges
legal restrictions	introduction	technical strategies
other criteria	how scanners work	organizational strategies
selection policies	scanner types	additional reading
additional reading	image processing	9. MANAGEMENT
3. CONVERSION	C. FILE MANAGEMENT	introduction
introduction	introduction	project life cycle
scanning factors	keeping track	in-house vs. outsource
rich digital master	image databases	in-house facility
benchmarking	storage	project budgets
text	storage types	communication
stroke	storage needs	project monitoring
continuous-tone	D. DELIVERY	looking beyond
halftone	introduction	additional reading
proposed method	networks	10. CONTINUING EDUCATION
guidelines	concerns	introductory information
additional reading	speed	web-based journals
4. QUALITY CONTROL	trends	mailing lists
definition	monitors	
developing a program	evaluation	
assessing quality	image quality	
additional reading	printers	
	technologies	
	evaluation	

© 2000-2003 Cornell University Library/Research
Department



Moving Theory into Practice *Digital Imaging Tutorial*

Preface

This tutorial offers base-level information on the use of digital imaging to convert and make accessible cultural heritage materials. It also introduces some concepts advocated by Cornell University Library, in particular the value of benchmarking requirements before undertaking a digital initiative. You will find here up-to-date technical information, formulas, and reality checks, designed to test your level of understanding.

The tutorial can stand on its own, but it is intended to be used in tandem with another product, [*Moving Theory into Practice: Digital Imaging for Libraries and Archives*](#), by Anne R. Kenney and Oya Y. Rieger (RLG, 2000). This publication picks up where the tutorial leaves off and advocates an integrated approach to digital imaging programs, from selection to access to preservation and management. Over 50 international experts contributed to the intellectual content of this book.

You will note that at certain points within this [National Endowment for the Humanities](#) funded tutorial, we invite reader comments and suggestions. In particular, we are aware that the presentation is US-centric, and with your help we seek to augment that perspective to provide a broader international focus. We look forward to [hearing from you!](#)

← **Contents**

View in:

← **Spanish**

← **French**

© Cornell University Library/Research Department, 2000-2003

Prepared by:

Anne R. Kenney, Assistant University Librarian
Oya Y. Rieger, Coordinator of Distributed Learning
Richard Entlich, Digital Projects Librarian

Technical support by:

Carla DeMello, Design Coordinator, IRIS
Valerie Jacoski, Web Developer, IRIS
Greg McClellan, Digital Projects Librarian
David DeMello, Consultant

Spanish translation prepared by:

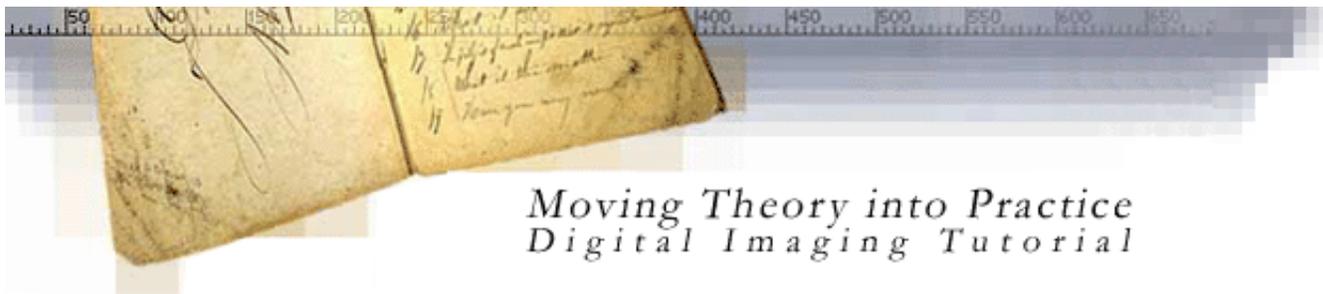
Global Listing

Spanish translation consultant:

Amparo R. DeTorres, Editor APOYO Newsletter

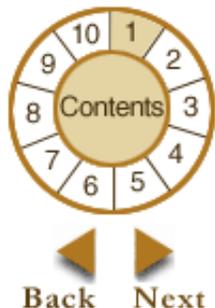
Support for this tutorial comes from the National Endowment for the Humanities. The Spanish translation was funded by the [Council on Library and Information Resources](#). Support for the French translation was received from the Food and Agricultural Organization of the United Nations. No part of this tutorial may be reproduced or transcribed in any form excepting for personal research use without prior written permission of Cornell University Library/Research Department. Requests for reproduction should be directed to the [Research Department](#). All URLs and internal links valid as of February 2003. Last revised on February 20, 2003.

© 2000-2003 Cornell University Library/ Research Department



Moving Theory into Practice

Digital Imaging Tutorial



DIGITAL IMAGES are electronic snapshots taken of a scene or scanned from documents, such as photographs, manuscripts, printed texts, and artwork. The digital image is sampled and mapped as a grid of dots or picture elements (pixels). Each pixel is assigned a tonal value (black, white, shades of gray or color), which is represented in binary code (zeros and ones). The binary digits ("bits") for each pixel are stored in a sequence by a computer and often reduced to a mathematical representation (compressed). The bits are then interpreted and read by the computer to produce an analog version for display or printing.

1. Basic Terminology

Key Concepts

digital images

- resolution
- pixel dimensions
- bit depth
- dynamic range
- file size
- compression
- file formats

additional reading

1	1	1	1	1	1	1	1	1	1
1	0	0	0	1	1	0	0	0	1
1	1	0	1	1	1	1	0	1	1
1	1	0	1	1	1	1	0	1	1
1	1	0	1	1	1	1	0	1	1
1	1	0	0	0	0	0	0	1	1
1	1	0	1	1	1	1	0	1	1
1	1	0	1	1	1	1	0	1	1
1	1	0	1	1	1	1	0	1	1
1	0	0	0	1	1	0	0	0	1
1	1	1	1	1	1	1	1	1	1

Pixel Values: As shown in this bitonal image, each pixel is assigned a tonal value, in this example 0 for black and 1 for white.

© 2000-2003 Cornell University Library/Research Department



View in:

- ← Spanish
- ← French



Moving Theory into Practice

Digital Imaging Tutorial



Back Next

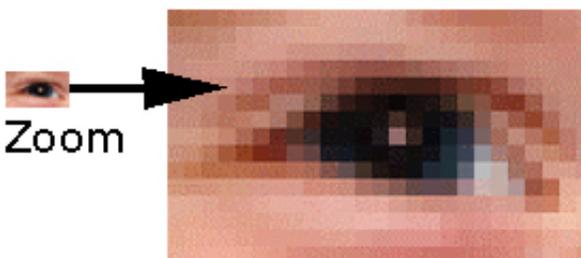
RESOLUTION is the ability to distinguish fine spatial detail. The spatial frequency at which a digital image is sampled (the sampling frequency) is often a good indicator of resolution. This is why dots-per-inch (dpi) or pixels-per-inch (ppi) are common and synonymous terms used to express resolution for digital images. Generally, but within limits, increasing the sampling frequency also helps to increase resolution.

1. Basic Terminology

Key Concepts

- digital images
- resolution**
- pixel dimensions
- bit depth
- dynamic range
- file size
- compression
- file formats

additional reading



Pixels: Individual pixels can be seen by zooming in an image.

© 2000-2003 Cornell University Library/Research Department



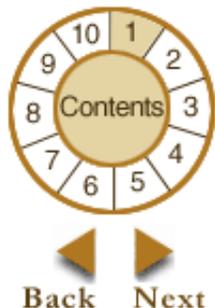
View in:

- Spanish
- French



Moving Theory into Practice

Digital Imaging Tutorial



PIXEL DIMENSIONS are the horizontal and vertical measurements of an image expressed in pixels. The pixel dimensions may be determined by multiplying both the width and the height by the dpi. A digital camera will also have pixel dimensions, expressed as the number of pixels horizontally and vertically that define its resolution (e.g., 2,048 by 3,072). Calculate the dpi achieved by dividing a document's dimension into the corresponding pixel dimension against which it is aligned.

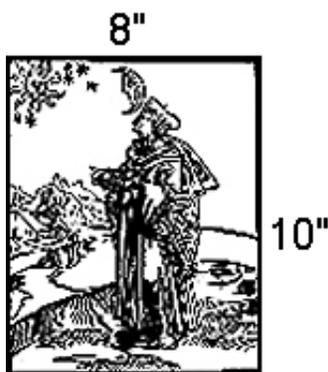
Example:

1. Basic Terminology

Key Concepts

- digital images
- resolution
- pixel dimensions**
- bit depth
- dynamic range
- file size
- compression
- file formats

additional reading



An 8" x 10" document that is scanned at 300 dpi has the pixel dimensions of 2,400 pixels (8" x 300 dpi) by 3,000 pixels (10" x 300 dpi).

Reality Check

What are the pixel dimensions of a 5x7-inch photograph scanned at 400 dpi?

Answer (check one):

2,000 x 2,800 pixels

1,300 x 1,800 pixels

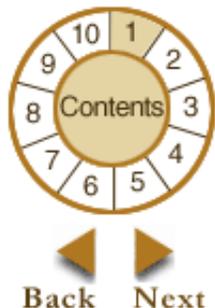


Reality Check

If an 8.5x11-inch page is scanned and has pixel dimensions of 2,550 x 3,300, what is the dpi?

dpi

© 2000-2003 Cornell University Library/Research Department



Moving Theory into Practice Digital Imaging Tutorial

BIT DEPTH is determined by the number of bits used to define each pixel. The greater the bit depth, the greater the number of tones (grayscale or color) that can be represented. Digital images may be produced in black and white (bitonal), grayscale, or color.

A *bitonal image* is represented by pixels consisting of 1 bit each, which can represent two tones (typically black and white), using the values 0 for black and 1 for white or vice versa.

A *grayscale image* is composed of pixels represented by multiple bits of information, typically ranging from 2 to 8 bits or more.

Example: In a 2-bit image, there are four possible combinations: 00, 01, 10, and 11. If "00" represents black, and "11" represents white, then "01" equals dark gray and "10" equals light gray. The bit depth is two, but the number of tones that can be represented is 2^2 or 4. At 8 bits, 256 (2^8) different tones can be assigned to each pixel.

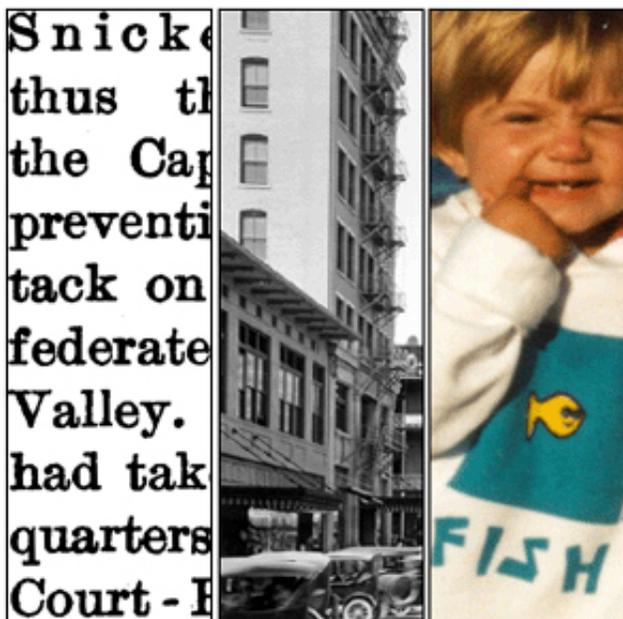
A *color image* is typically represented by a bit depth ranging from 8 to 24 or higher. With a 24-bit image, the bits are often divided into three groupings: 8 for red, 8 for green, and 8 for blue. Combinations of those bits are used to represent other colors. A 24-bit image offers 16.7 million (2^{24}) color values. Increasingly scanners are capturing 10 bits or more per color channel and often outputting 8 bits to compensate for "noise" in the scanner and to present an image that more closely mimics human perception.

1. Basic Terminology

Key Concepts

digital images
resolution
pixel dimensions
bit depth
dynamic range
file size
compression
file formats

additional reading



Bit Depth: Left to right - 1-bit bitonal, 8-bit grayscale, and 24-bit color images.

Binary calculations for the number of tones represented by common bit depths:

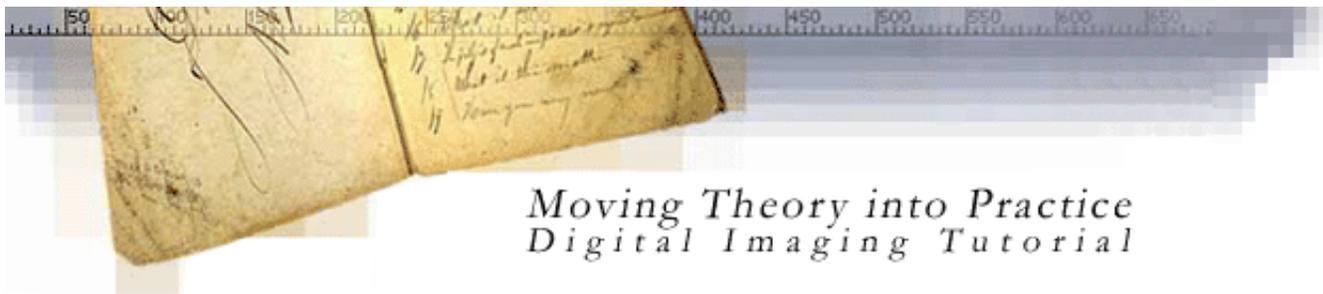
- 1 bit (2^1) = 2 tones
- 2 bits (2^2) = 4 tones
- 3 bits (2^3) = 8 tones
- 4 bits (2^4) = 16 tones
- 8 bits (2^8) = 256 tones
- 16 bits (2^{16}) = 65,536 tones
- 24 bits (2^{24}) = 16.7 million tones



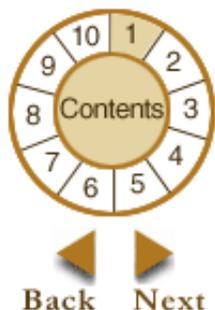
View in:



© 2000-2003 Cornell University Library/Research Department



Moving Theory into Practice Digital Imaging Tutorial



DYNAMIC RANGE is the range of tonal difference between the lightest light and darkest dark of an image. The higher the dynamic range, the more potential shades can be represented, although the dynamic range does not automatically correlate to the number of tones reproduced. For instance, high-contrast microfilm exhibits a broad dynamic range, but renders few tones. Dynamic range also describes a digital system's ability to reproduce tonal information. This capability is most important for continuous-tone documents that exhibit smoothly varying tones, and for photographs it may be the single most important aspect of image quality.

1. Basic Terminology

Key Concepts

- digital images
- resolution
- pixel dimensions
- bit depth
- dynamic range**
- file size
- compression
- file formats

additional reading

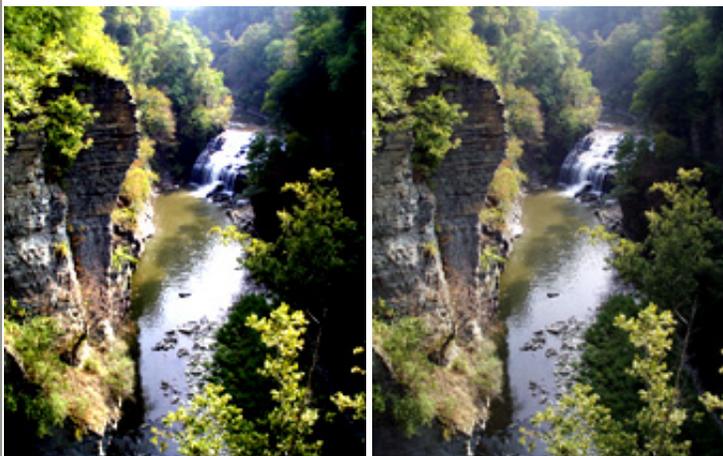


Dynamic Range: The image on top has a broader dynamic range, but a limited number of tones represented. The lower image has a narrower dynamic range, but a greater number of tones represented. Note the lack of detail in shadows and highlights in the top frame. Courtesy of Don Brown.

◀ ▶
Back Next
← Contents

View in:
← Spanish
← French

Reality Check



Which of these images has the more limited tonal representation?

Answer (check one):

The image on the left

The image on the right

© 2000-2003 Cornell University Library/Research Department



Moving Theory into Practice

Digital Imaging Tutorial



Back Next

1. Basic Terminology

Key Concepts

digital images
 resolution
 pixel dimensions
 bit depth
 dynamic range
file size
 compression
 file formats

additional reading

FILE SIZE is calculated by multiplying the surface area of a document (height x width) to be scanned by the bit depth and the dpi². Because image file size is represented in bytes, which are made up of 8 bits, divide this figure by 8.

Formula 1 for File Size

$$\text{File Size} = (\text{height} \times \text{width} \times \text{bit depth} \times \text{dpi}^2) / 8$$

If the pixel dimensions are given, multiply them by each other and the bit depth to determine the number of bits in an image file. For instance, if a 24-bit image is captured with a digital camera with pixel dimensions of 2,048 x 3,072, then the file size equals (2048 x 3072 x 24)/8, or 18,874,368 bytes.

Formula 2 for File Size

$$\text{File Size} = (\text{pixel dimensions} \times \text{bit depth}) / 8$$

File size naming convention: Because digital images often result in very large files, the number of bytes is usually represented in increments of 2¹⁰ (1,024) or more:

- 1 Kilobyte (KB) = 1,024 bytes
- 1 Megabyte (MB) = 1,024 KB
- 1 Gigabyte (GB) = 1,024 MB
- 1 Terabyte (TB) = 1,024 GB

Reality Check

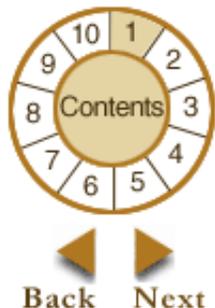
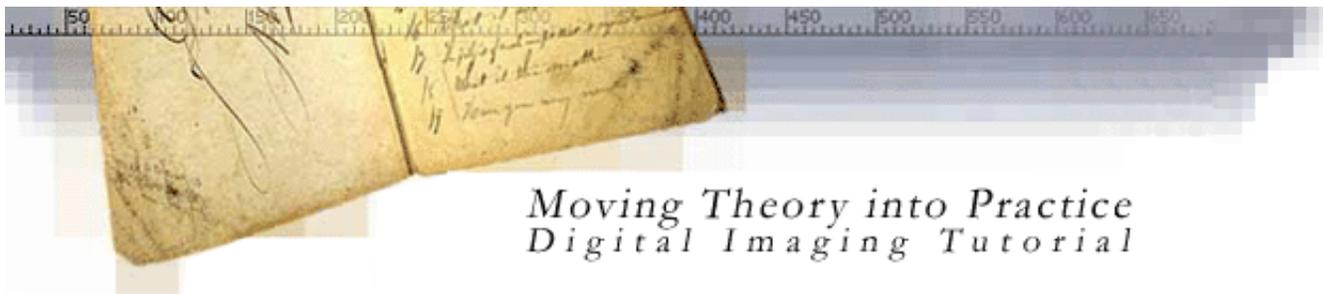
What is the file size for a [US letter-size page](#) captured bitonally at 100 dpi?

bytes

Back Next
 Contents

View in:

Spanish
 French



Moving Theory into Practice Digital Imaging Tutorial

COMPRESSION is used to reduce image file size for storage, processing, and transmission. The file size for digital images can be quite large, taxing the computing and networking capabilities of many systems. All compression techniques abbreviate the string of binary code in an uncompressed image to a form of mathematical shorthand, based on complex algorithms. There are standard and proprietary compression techniques available. In general it is better to utilize a standard and broadly supported one than a proprietary one that may offer more efficient compression and/or better quality, but which may not lend itself to long-term use or digital preservation strategies. There is [considerable debate](#) in the library and archival community over the use of compression in master image files.

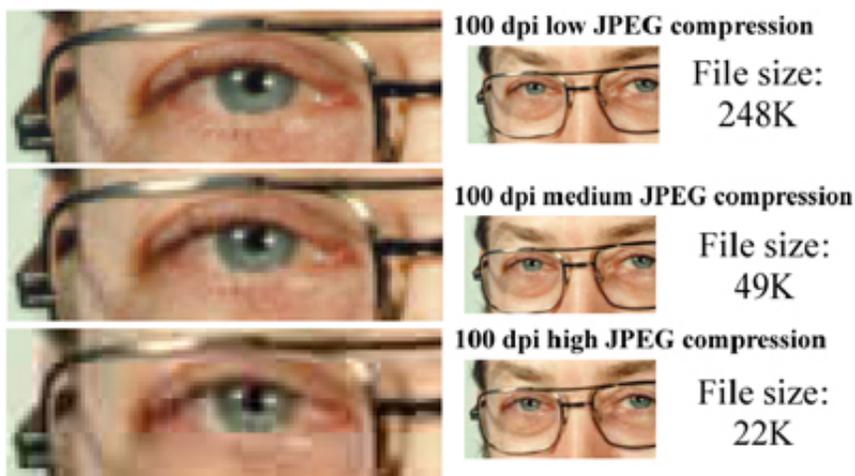
1. Basic Terminology

Key Concepts

- digital images
- resolution
- pixel dimensions
- bit depth
- dynamic range
- file size
- compression**
- file formats

additional reading

Compression schemes can be further characterized as either lossless or lossy. Lossless schemes, such as ITU-T.6, abbreviate the binary code without discarding any information, so that when the image is "decompressed" it is bit for bit identical to the original. Lossy schemes, such as JPEG, utilize a means for averaging or discarding the least significant information, based on an understanding of visual perception. However, it may be extremely difficult to detect the effects of lossy compression, and the image may be considered "visually lossless." Lossless compression is most often used with bitonal scanning of textual material. Lossy compression is typically used with tonal images, and in particular continuous tone images where merely abbreviating the information will not result in any appreciable file savings.



Lossy Compression: Note the effects of JPEG lossy compression on the zoomed image (left). In the bottom image, artifacts are visible in the form of 8 x 8 pixel squares, and fine details such as eyelashes have disappeared.

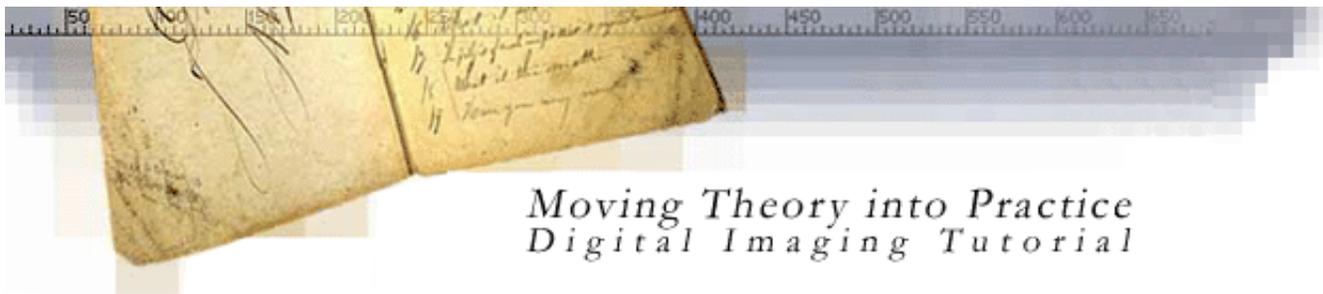
Emerging compression schemes offer the capability of providing multi-resolution images from a single file, providing flexibility in the delivery and presentation of images to end users.

To review a table summarizing important attributes for common compression techniques, [click here](#).

 
Back **Next**
 **Contents**

© 2000-2003 Cornell University Library/Research Department

View in:
 **Spanish**
 **French**



Moving Theory into Practice Digital Imaging Tutorial



Back Next

FILE FORMATS consist of both the bits that comprise the image and header information on how to read and interpret the file. File formats vary in terms of resolution, bit-depth, color capabilities, and support for compression and metadata.

To review a table summarizing important attributes for eight common image formats in use today, [click here](#).

1. Basic Terminology

Key Concepts

- digital images
- resolution
- pixel dimensions
- bit depth
- dynamic range
- file size
- compression
- file formats**

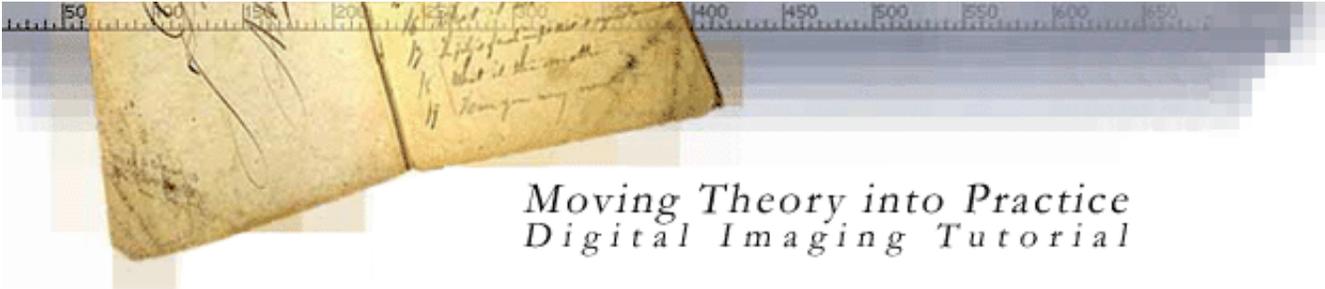
additional reading

© 2000-2003 Cornell University Library/Research Department



View in:

- Spanish
- French



Moving Theory into Practice Digital Imaging Tutorial



Back Next

1. Basic Terminology

Key Concepts

digital images
resolution
pixel dimensions
bit depth
dynamic range
file size
compression
file formats

additional reading

ADDITIONAL READING

Glossaries of Digital Imaging Terms:

Glossaries, PADI: Preserving Access to Digital Information,
<http://www.nla.gov.au/padi/format/gloss.html>

"Glossary" in Digital Toolbox (Colorado Digitization Project),
<http://coloradodigital.coalliance.org/glossary.html>

Anne R. Kenney and Oya Y. Rieger, *Moving Theory into Practice: Digital Imaging for Libraries and Archives*, Mountain View, CA : Research Libraries Group, 2000. <http://www.rlg.org/preserv/mtip2000.html>

Franziska Frey, *File Formats for Digital Masters, Guide 5 to Quality in Visual Resource Imaging*, <http://www.rlg.org/visguides/visguide5.html>

RLG DigiNews contains various features on file formats and compression techniques. Use the browse option to find articles, highlighted Web sites, and other information, <http://www.rlg.org/preserv/diginews/browse.html>.

Technical Advisory Service for Images, *New Digital Image File Formats*,
<http://www.tasi.ac.uk/advice/creating/newfile.html>

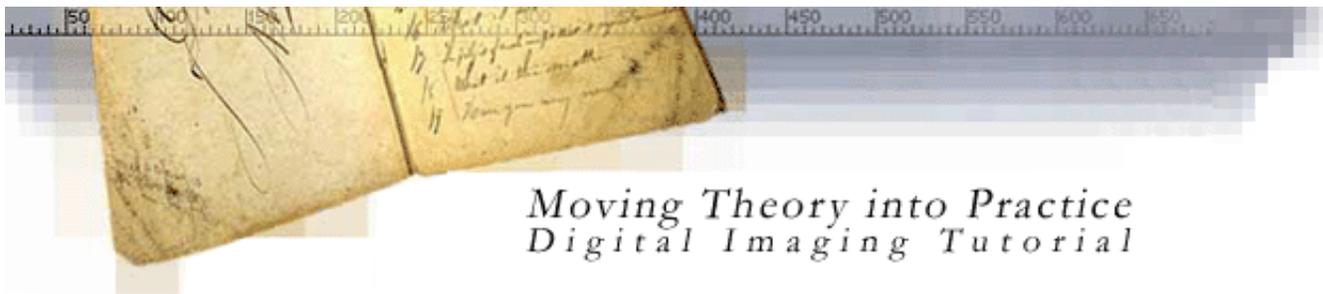
© 2000-2003 Cornell University Library/Research Department



Back Next
← Contents

View in:

← Spanish
← French



Moving Theory into Practice Digital Imaging Tutorial

2. Selection

Key Concepts

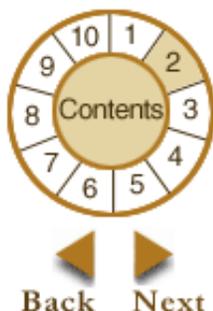
introduction

legal restrictions

other criteria

selection policies

additional reading



INTRODUCTION

Libraries and archives initiate imaging programs to meet real or perceived needs. The utility of digital images is most likely ensured when the needs of users are clearly defined, the attributes of the documents are known, and the technical infrastructure to support conversion, management, and delivery of content is appropriate to the needs of the project.

LEGAL RESTRICTIONS

Begin your selection process by considering legal restrictions. Is the material restricted because of privacy, content, or donor concerns? Is it copyright protected? If so, do you have the right to create and disseminate digital reproductions? Laura N. Gasaway, Professor of Law and Director of the Law Library at University of North Carolina at Chapel Hill, maintains an updated [chart](#) summarizing the terms of protection for published and unpublished works. Peter Hirtle of the Cornell Institute for Digital Collections has developed a [chart](#) specifically geared to archival and manuscript curators. Additional information on copyright in the digital world is available from the [Copyright Management Center](#) at Indiana University-Purdue University Indianapolis, and from the [Copyright Crash Course](#) at the University of Texas.

For copyright laws pertaining to the UK, TASI provides the "[Copyright FAQ](#)" co-developed with the Arts and Humanities Data Service.

The Canadian Heritage Information Network (CHIN) offers several publications via subscription or sale on [managing intellectual property](#).

Note: we'd like to include good sources on copyright for other countries; if you know of any, please [drop us a line](#).

Reality Check

My institution is interested in digitizing and making network accessible brittle books published in the United States from 1880-1920. Do we have the legal right to do so? Answer (check one):

Yes No

 
Back Next
 Contents

View in:
 Spanish
 French

© 2000-2003 Cornell University Library/Research Department



Moving Theory into Practice Digital Imaging Tutorial

2. Selection



Back Next
Key Concepts

introduction
legal restrictions
other criteria
selection policies

additional reading

OTHER SELECTION CRITERIA

The following issues should also be considered in choosing materials for digital conversion. Under each category, pose and answer a range of questions such as the ones suggested in order to highlight their effect on selection.

Document Attributes

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

Preservation Considerations

Would the material be put at risk in the digitization process? Would digital surrogates reduce use of the originals, thereby offering them protection from handling? Is the digital reproduction seen as a means to replace the originals?

Organization and Available Documentation

Is the material in a coherent, logically structured order? Is it paginated or is the arrangement suggested by some other means? Is it complete? Is there adequate descriptive, navigational, or structural information about the material, such as bibliographic records or a detailed finding aid? (see also [Metadata](#))

Intended Uses

What kinds, level, and frequency of use are envisioned? Is there a clear understanding of user requirements? Can digitization support these uses? Will access to the material be significantly enhanced by digitization? Can your institution support a range of uses, e.g., printing, browsing, detailed review? Are there issues around security or access that must be taken into account (e.g., access restricted to certain people or use under certain conditions?)

Digital Collection Building

Is there added incentive to digitize material based on the availability of complementary digital resources (including data and metadata?) Is there an opportunity for multi-institutional cooperation? For building thematic coherence or "critical mass?"

Duplication of Effort

Has the material already been digitized by another trusted source? If so, do the digital files possess sufficient quality, documentation, and functionality to serve your purposes? What conditions govern access and use of those files?

Institutional Capabilities

Does your institution have the requisite technical infrastructure to manage, deliver, and maintain digitized materials? Do your principal users have adequate computing and connectivity to make effective use of these

materials? See [Technical Infrastructure](#) for specific information on technical components to consider in such an evaluation.

Finances

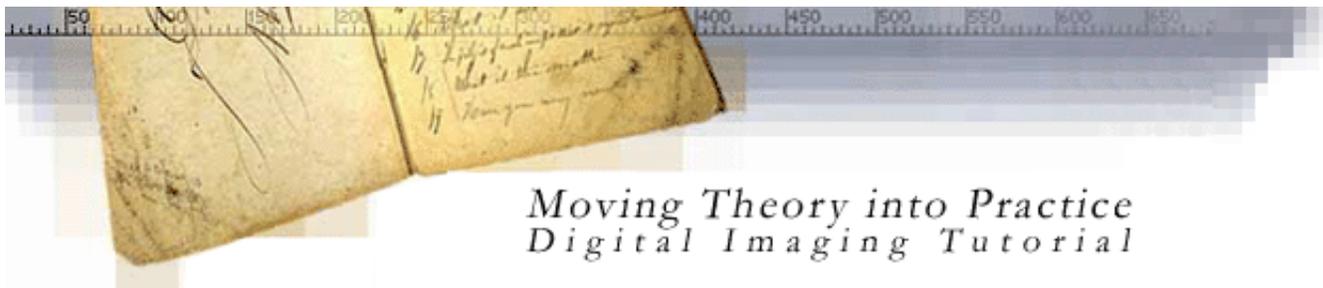
Can you determine the total cost of image acquisition (selection, preparation, capture, indexing, and quality control)? Is this cost justified based on real or perceived benefits accruing from digitization? Are there funds to support this effort? Is there institutional commitment to the on-going management and preservation of these files? See [Digital Preservation](#) and [Management](#) sections for more information.



View in:

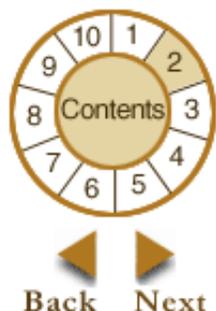


© 2000-2003 Cornell University Library/Research Department



Moving Theory into Practice Digital Imaging Tutorial

2. Selection



SELECTION POLICIES

Some institutions have developed selection policies or matrixes designed to assist staff in selection for digitization. The following may be of assistance to you in designing your own policies and procedures:

- Library of Congress, "[Selection Criteria for Preservation Digital Reformatting](#)"
- Columbia University, "[Selection Criteria for Digital Imaging Projects](#)"
- University of California, "[Selection Criteria for Digitization](#)"
- Harvard University, "[Selection for Digitization: a Decision-Making Matrix](#)"
- National Agricultural Library, "[Selection Criteria and Guidelines](#)"
- Oxford University, "[Decision Matrices and Workflows](#)" (Appendix B)
- [National Library of Australia Digitisation Policy, 2000-2004.](#)

Key Concepts

introduction

legal restrictions

other criteria

selection policies

additional reading

© 2000-2003 Cornell University Library/Research Department



View in:

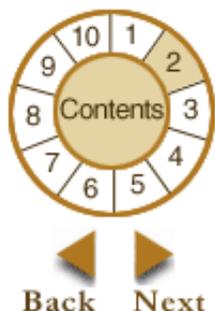
← Spanish

← French



Moving Theory into Practice Digital Imaging Tutorial

2. Selection



ADDITIONAL READING

Paula DeStefano, "Selection for Digital Conversion," in *Moving Theory into Practice: Digital Imaging for Libraries and Archives*, Mountain View, CA : Research Libraries Group, 2000; pp. 11-23.
<http://www.rlg.org/preserv/mtip2000.html>

Dan Hazen, Jeffrey Horrell, and Jan Merrill-Oldham, *Selecting Research Collections for Digitization*, <http://www.clir.org/pubs/reports/hazen/pub74.html>

Janet Gertz, "Selection Guidelines for Preservation," *Joint RLG and NPO Preservation Conference*, <http://www.rlg.org/preserv/joint/gertz.html>

Paul Ayris, "Guidance for Selecting Materials for Digitisation," *Joint RLG and NPO Preservation Conference*, <http://www.rlg.org/preserv/joint/ayris.html>

Angelica Menne-Haritz and Nils Brubach, "The Intrinsic Value of Archive and Library Material. List of Criteria for Imaging and Textual Conversion for Preservation,"
<http://www.uni-marburg.de/archivschule/intrinsengl.html>

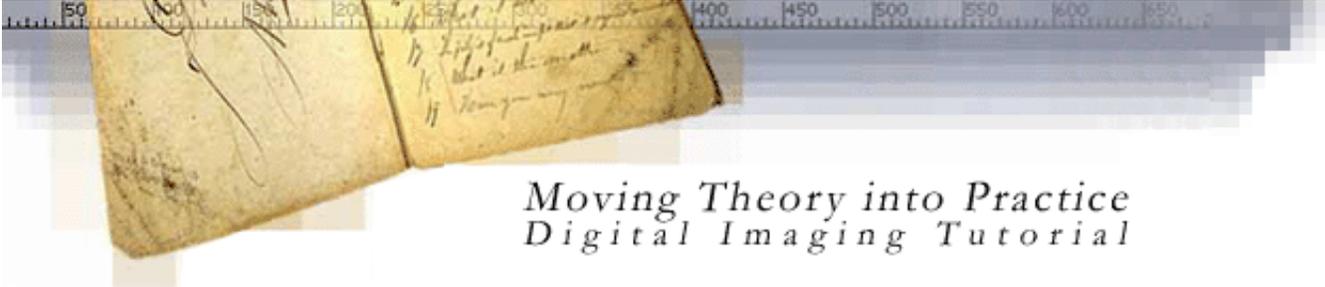
Key Concepts

- introduction
- legal restrictions
- other criteria
- selection policies

additional reading

© 2000-2003 Cornell University Library/Research Department





Moving Theory into Practice

Digital Imaging Tutorial

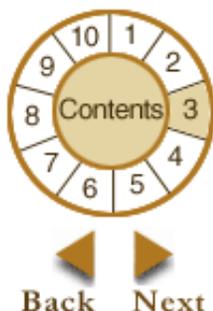
3. Conversion

Key Concepts

introduction

scanning factors
 rich digital master
 benchmarking
 text
 stroke
 continuous-tone halftone
 proposed method
 guidelines

additional reading

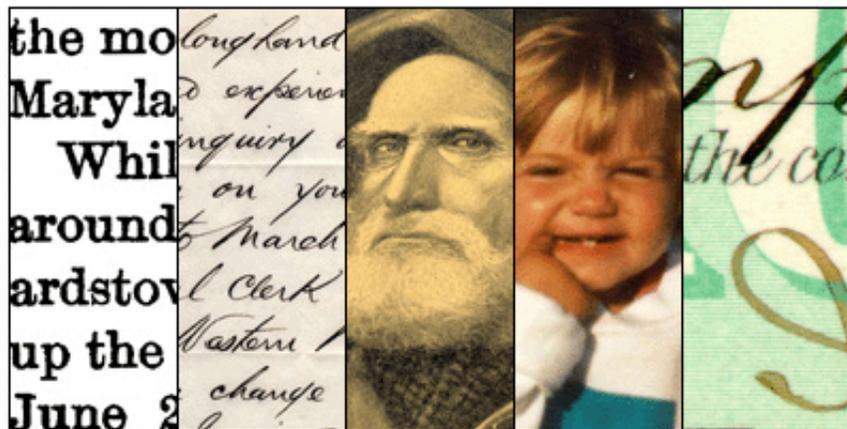


INTRODUCTION

Digital image capture must take into consideration the technical processes involved in converting from analog to digital representation as well as the attributes of the source documents themselves: physical size and presentation, level of detail, tonal range, and presence of color. Documents may also be characterized by the production process used to create them, including manual, machine, photographic, and more recently, electronic means. Further, all paper- and film-based documents will fall into one of the following five categories that will affect their digital recording.

Document Types

- *Printed Text/Simple Line Art*—distinct edge-based representation, with no tonal variation, such as a book containing text and simple line graphics
- *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
- *Halftones*—reproduction of graphic or photographic materials represented by a grid of variably sized, regularly spaced pattern of dots or lines, often placed at an angle. Includes some graphic art as well, e.g., engravings
- *Continuous Tone*—items such as photographs, watercolors, and some finely inscribed line art that exhibit smoothly or subtly varying tones
- *Mixed*—documents containing two or more of the categories listed above, such as illustrated books



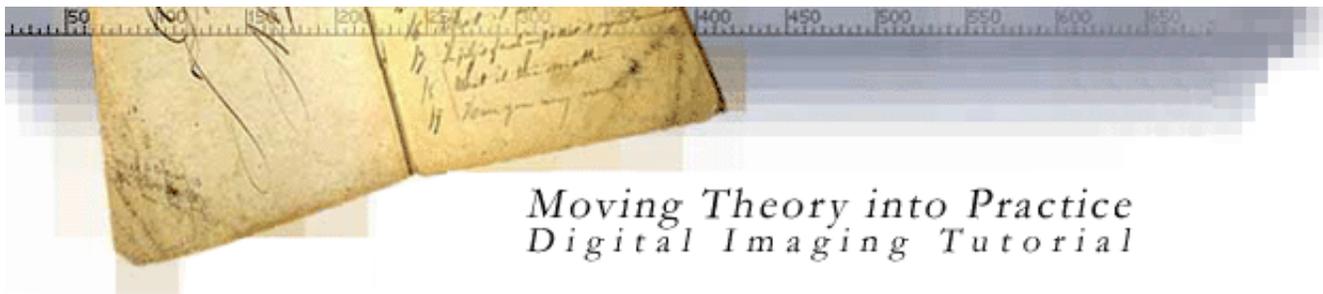
Back Next
← Contents

View in:

← Spanish
← French

Document Types: Left to right - printed text, manuscript, halftone, continuous tone, and mixed.

© 2000-2003 Cornell University Library/Research Department



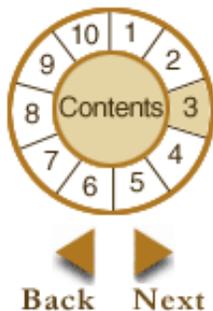
Moving Theory into Practice Digital Imaging Tutorial

3. Conversion

Key Concepts

- introduction
- scanning factors**
- rich digital master
- benchmarking
- text
- stroke
- continuous-tone halftone
- proposed method
- guidelines

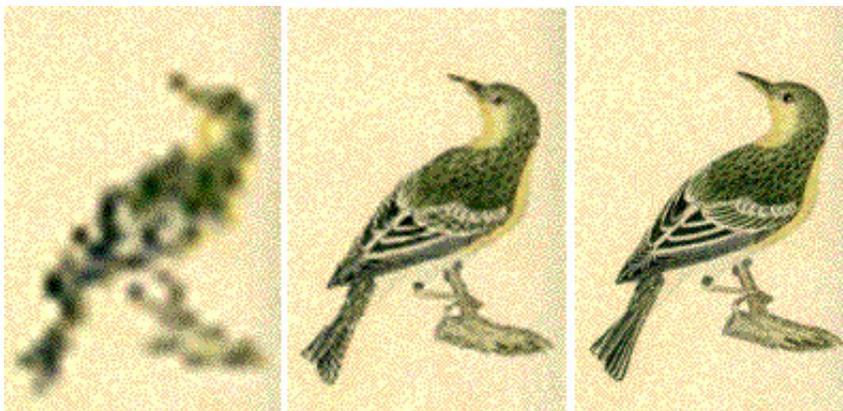
additional reading



SCANNING FACTORS AFFECTING IMAGE QUALITY

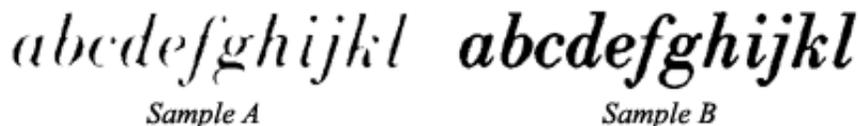
Resolution/threshold

Increasing resolution enables the capture of finer detail. At some point, however, added resolution will not result in an appreciable gain in image quality, only larger file size. The key is to determine the resolution necessary to capture all significant detail present in the source document.



Effects of Resolution on Image Quality: As the resolution increases, the gain in image quality levels off.

The *threshold* setting in bitonal scanning defines the point on a scale, ranging from 0 (black) to 255 (white), at which the gray values captured will be converted to black or white pixels. Note the effect of varying the threshold on typescript scanned at the same resolution on the same scanner.



Effects of Threshold on Resolution: Sample A has a lower threshold (60)

than Sample B (100).

Reality Check

Which sample has more gray values assigned to black?

Sample A

Sample B

Bit Depth

Increasing the bit depth, or number of bits used to represent each pixel, enables the capture of more gray shades or color tones. [Dynamic range](#) is the term used to express the full range of tonal variations from lightest light to darkest dark. A scanner's capability to capture dynamic range is governed by the bit depth used and output as well as system performance. Increasing the bit depth will affect resolution requirements, file size, and the compression method used.



Bit Depth: When a 24-bit JPEG image (left) is reduced to an 8-bit GIF image (right), the color reduction can result in quantization artifacts, evident in the appearance of visible tonal steps on the top left corner of the GIF image.

Enhancement

Enhancement processes improve scanning quality but their use raises concerns about fidelity and authenticity. Many institutions argue against enhancing master images, limiting it to access files only. Typical enhancement features in scanner software or image editing tools include descreening, despeckling, deskewing, sharpening, use of custom filters, and bit-depth adjustment. Here are several examples of image enhancement processes.

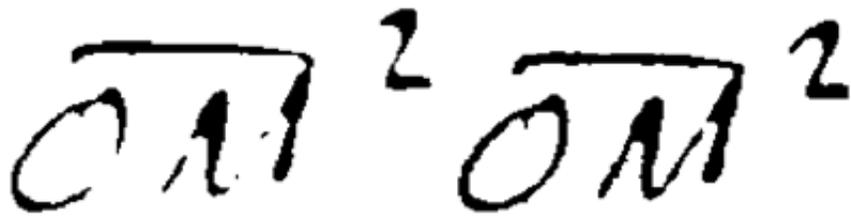


Image Enhancement: Letters scanned at the same resolution and threshold setting, but a sharpening filter has been applied to the one on the right.

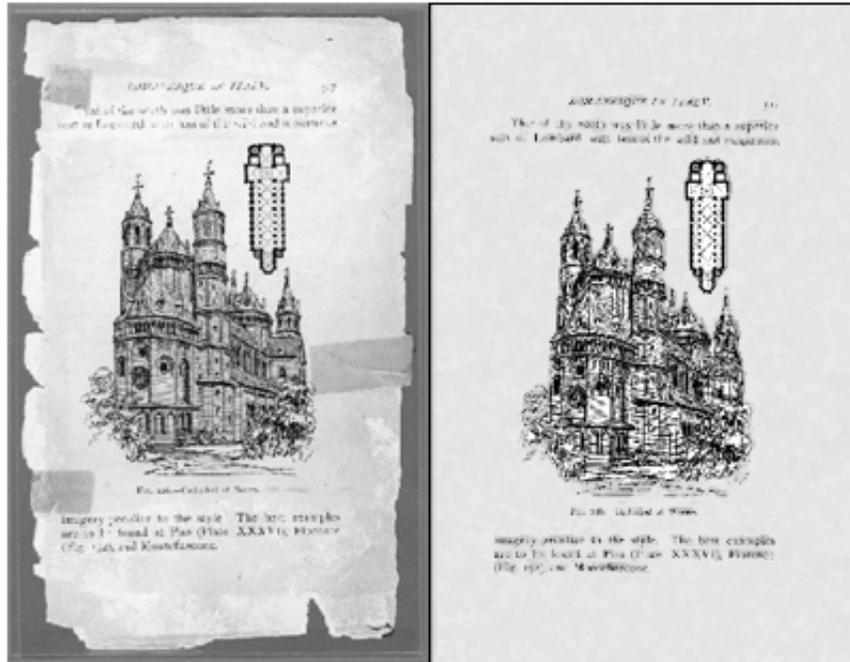


Image Enhancement: The left image was altered (right) at the pixel level using an image editing program.

Color

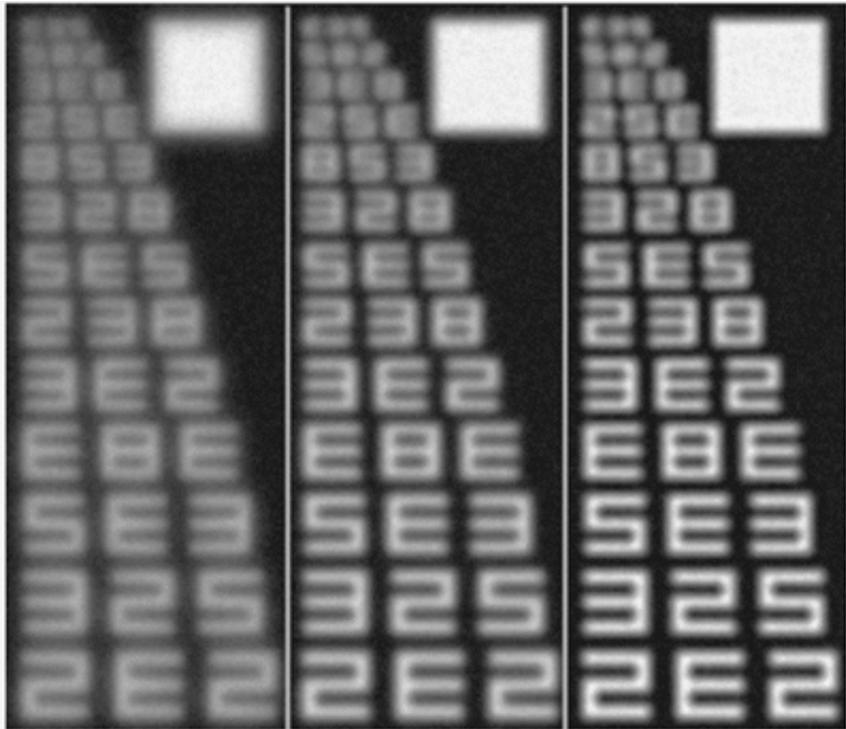
Capturing and conveying color appearance is arguably the most difficult aspect of digital imaging. Good color reproduction depends on a number of variables, such as the level of illumination at the time of capture, the bit depth captured and output, the capabilities of the scanning system, and mathematical representation of color information as the image moves across the [digitization chain](#) and from one color space to another.



Color Shift: Image with an overall red cast (left) and original colors (right).

System Performance

The equipment used and its performance over time will affect image quality. Different systems with the same stated capabilities (e.g., dpi, bit depth, and dynamic range) may produce dramatically different results. System performance is measured via tests that check for resolution, tone reproduction, color rendering, noise, and artifacts. (See [Quality Control](#).)



System Performance: Note the difference in image quality of the alphanumeric characters scanned on three different systems at the same resolution and bit depth.

File Format

The file format for master images should support the resolution, bit-depth, color information, and metadata you need. For example, there is little sense in creating a full color image, only to save it in a format that cannot support more than 8 bits (e.g., GIF). The format should also handle being stored uncompressed or compressed using either lossless and lossy techniques. It should be open and well-documented, widely supported, and cross-platform compatible. Although there is interest in other formats, such as PNG, SPIFF, and Flashpix, most cultural institutions rely on TIFF to store their master images. For access, derivative images in other formats may be created.

For a table listing attributes of common image formats, click on [Table: Commonly Used Image File Formats](#).

Compression

Lossy compression can have a pronounced impact on image quality, especially if the level of compression is high. In general, the richer the file, the more efficient and sustainable the compression. For instance, a bitonal scan of a page at 600 dpi is 4 times larger than a 300 dpi version, but often only twice as large in its compressed state. The more complex the image, the poorer the level of compression that can be obtained in a lossless or visually lossless state. With photographs, lossless compression schemes often

provide around a 2:1 file size ratio; with lossy compression above 10 or 20:1, the effect may be obvious.

For a table listing attributes of common compression processes, click on [Table: Commonly Used Compression Processes](#)

◀ ▶
Back Next
← Contents

View in:
← Spanish
← French



Effects of Lossy Compression on Text: Close-up comparison of a section from a map saved in lossless GIF (left) and lossy JPEG (right).

Operator Judgement and Care

The skill and care of a scanning operator may affect image quality as much as the inherent capabilities of the system. We have noted the effect of threshold in bitonal scanning; operator judgement can minimize line drop out or fill-in. When digital cameras are used, the lighting becomes a concern, and the skills of the camera operator will come into play. A quality control program must be instituted to verify consistency of output.

© 2000-2003 Cornell University Library/Research Department

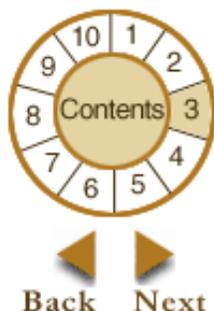


Moving Theory into Practice Digital Imaging Tutorial

3. Conversion

Key Concepts

introduction
 scanning factors
rich digital master
 benchmarking
 text
 stroke
 continuous-tone halftone
 proposed method
 guidelines
 additional reading



THE CASE FOR CREATING A RICH DIGITAL MASTER

There are compelling preservation, access, and economic reasons for creating a rich digital master image file (sometimes referred to as an archival image) in which all significant information contained in the source document is represented.

Preservation

Creating a rich digital master can contribute to preservation in at least three ways:

1. *Protecting vulnerable originals.* The image surrogate must be rich enough to reduce or eliminate the user's need to view the original.
2. *Replacing originals.* Under certain circumstances, digital images can be created to replace originals or used to produce paper copies or [Computer Output Microfilm](#). The digital replacement must satisfy all research, legal, and fiscal requirements.
3. *Preserving digital files.* It is easier to preserve digital files when they are captured consistently and well documented. The expense of doing so is more justifiable if the files offer continuing value and functionality.

Access

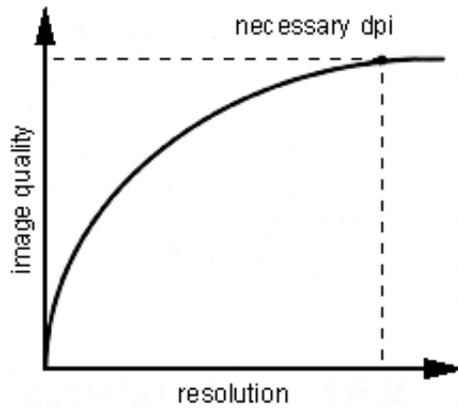
A digital master should be capable of supporting a range of users' needs through the creation of derivatives for printing, display, and image processing. The richer the digital master, the better the derivatives in terms of quality and processibility. User expectations will likely be more demanding over time--the digital master should be rich enough to accommodate future applications. Rich masters will support the development of cultural heritage resources that are comparable and interoperable across disciplines, users, and institutions.

Cost

Creating a high quality digital image may cost more initially, but will be less expensive than creating a lower quality image that fails to meet long-term requirements and results in the need to re-scan. Labor costs associated with identifying, preparing, inspecting, indexing, and managing digital information far exceed the costs of the scan itself.

The key to image quality is not to capture at the highest resolution or bit depth possible, but to match the conversion process to the informational content of

the original, and to scan at that level--no more, no less. In doing so, one creates a master file that can be used over time. Long-term value should be defined by the intellectual content and utility of the image file, not limited by technical decisions made at the point of conversion.



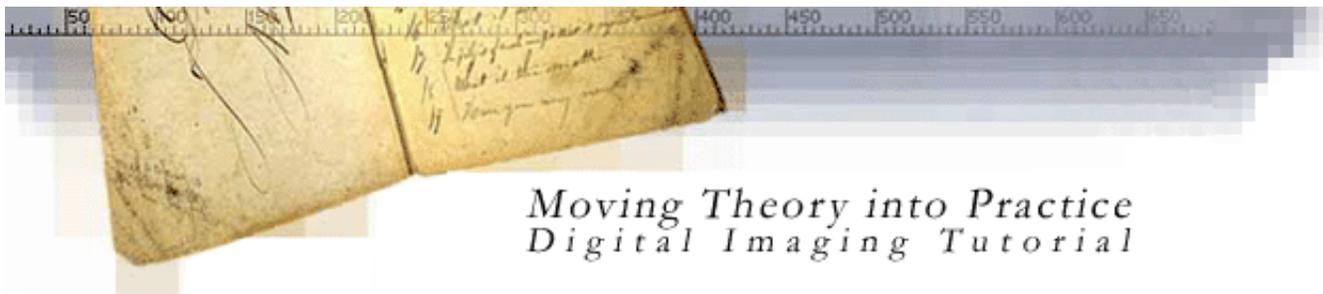
◀ ▶
Back Next
← Contents

View in:

← Spanish
← French

No More, No Less: As resolution increases, image quality will level off.

© 2000-2003 Cornell University Library/Research Department

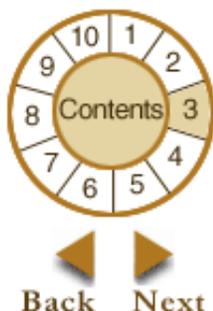


Moving Theory into Practice Digital Imaging Tutorial

3. Conversion

Key Concepts

- introduction
- scanning factors
- rich digital master
- benchmarking text**
- stroke
- continuous-tone halftone
- proposed method
- guidelines
- additional reading



BENCHMARKING FOR DIGITAL CAPTURE

Cornell advocates a methodology for determining conversion requirements that is based on the following:

- Assessing document attributes (detail, tone, color)
- Defining the needs of current and future users
- Objectively characterizing relevant variables (e.g., size of detail, desired quality, resolving power of system)
- Correlating variables to one another via formulas
- Confirming results through testing and evaluation

BENCHMARKING RESOLUTION REQUIREMENTS FOR PRINTED TEXT

Cornell adopted and refined a digital Quality Index (QI) formula for printed text that was developed by the C10 Standards Committee of AIIM. (An explanation of this approach is found in: [Tutorial: Determining Resolution Requirements for Reproducing Text-based Material](#)). This formula was based on translating the Quality Index method developed for preservation microfilming standards to the digital world. The QI formula for scanning text relates quality (QI) to character size (h) in mm and resolution (dpi). As in the preservation microfilming standard, the digital QI formula forecasts levels of image quality: barely legible (3.0), marginal (3.6), good (5.0), and excellent (8.0).

Table: Metric/English Conversion

1 mm = .039 inches
1 inch = 25.4 mm

The formula for bitonal scanning provides a generous over sampling to compensate for misregistration and reduced quality due to thresholding information to black and white pixels.

Bitonal QI Formula for Printed Text

$$QI = (dpi \times .039h)/3$$

$$h = 3QI/.039dpi$$

$$dpi = 3QI/.039h$$

Note: if the measurement of h is expressed in inches, omit the .039.

abcde

200 dpi

abcde

300 dpi

abcde

600 dpi

Resolution Requirements For Printed Text: Comparison of letters scanned at different resolutions.

Some printed text will require grayscale or color scanning for the following reasons:

- Pages are badly stained
- Paper has darkened to the extent that it is difficult to threshold the information to pure black and white pixels
- Pages contain complex graphics or important contextual information (e.g., embossments, annotations)
- Pages contain color information (e.g., different colored inks)



Scanning Text: Compare bitonal (left) and grayscale (right) scanning of a stained text page.

Because tonal images subtly "gray out" pixels that are only partially on a stroke, a separate formula was developed for grayscale/color scanning of printed text:

Grayscale/Color QI Formula for Printed Text

$$QI = (dpi \times .039h)/2$$

$$h = 2QI/.039dpi$$

$$dpi = 2QI/.039h$$

Note: if the measurement of h is expressed in inches, omit the .039.

Example: The Case of the Brittle Book

Cornell used benchmarking to determine conversion requirements for brittle books containing text and simple graphics, such as line art, charts, diagrams, and the like. Although some of the books contained darkened pages, in most cases the contrast between text and background was sufficient for capturing text in bitonal mode. We determined resolution requirements by assessing the level of detail and by defining our quality needs.

Printed text offers a fixed metric for detail: the height of the smallest significant lowercase letter. In a review of commercial typescripts commonly used from 1850-1950, Cornell discovered that virtually no publishers used fonts shorter than 1 mm in height. We were interested in creating paper replacements for the deteriorating originals, so our quality requirement was high--we wanted excellent rendering of the fonts, including full representation of the serifs and other attributes.



View in:



Once we had determined the size of the detail and the desired quality, our next step was to equate those requirements to the necessary resolution. Using the bitonal QI formula, and a fixed detail metric of 1mm, Cornell predicted that textual information could be captured with excellent quality at a resolution of 600 dpi. An extensive onscreen and print examination of digital facsimiles for a range of typescripts used during the brittle book period confirmed these benchmarks. Although many of the books did not contain such small text, to avoid an item-by-item review, all books are scanned at 600 dpi.

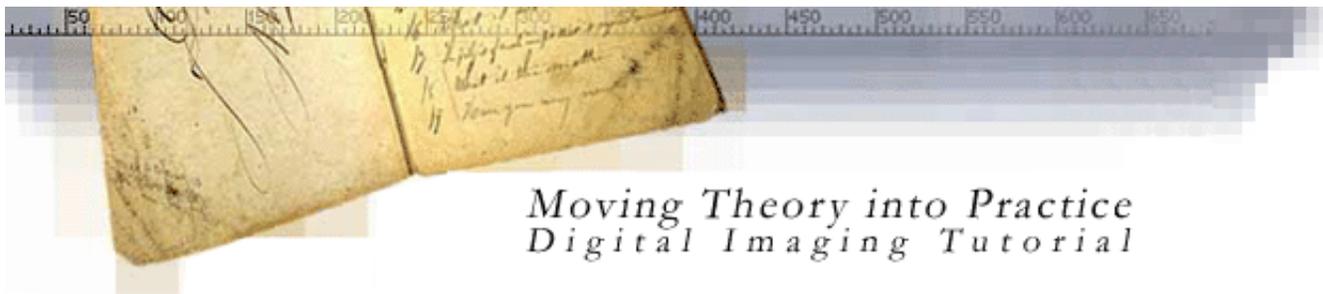
Reality Check

Calculate the bitonal scanning resolution required to achieve excellent quality (QI = 8) for a 3 mm high character. (Round to nearest whole number.)

dpi

When using a 400 dpi bitonal scanner, what would be the size of the smallest character that you could capture with medium quality (QI=5)? (Round your answer to the nearest hundredth of a millimeter.)

mm



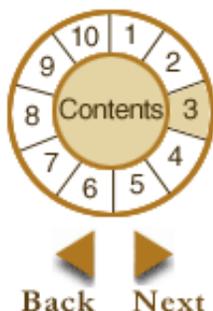
Moving Theory into Practice Digital Imaging Tutorial

3. Conversion

Key Concepts

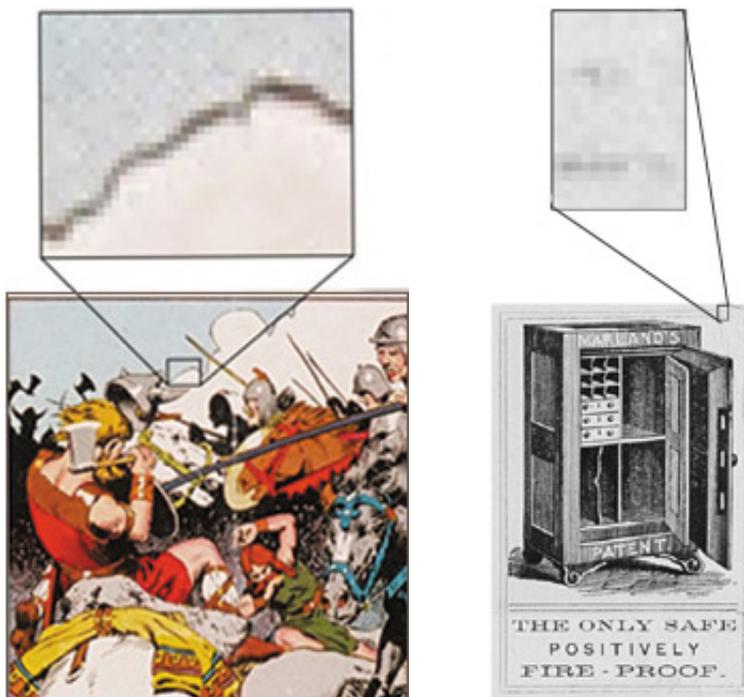
- introduction
- scanning factors
- rich digital master
- benchmarking**
- text
- stroke**
- continuous-tone halftone
- proposed method
- guidelines

additional reading



BENCHMARKING RESOLUTION REQUIREMENTS BASED ON STROKE WIDTH

The QI method was designed for printed text where character height represents the measure of detail. Manuscripts and other non-textual material representing distinct edge-based graphics, such as maps, sketches, and engravings, offer no equivalent fixed metric. For many such documents, a better representation of detail would be the width of the finest line, stroke, or marking that must be captured in the digital surrogate. To fully represent such a detail, at least 2 pixels should cover it. For example, an original with a stroke measuring 1/100 inch must be scanned at 200 dpi or greater to fully resolve its finest feature. For bitonal scanning, this requirement would be higher (say 3 pixels/feature) due to the potential for sampling errors and the thresholding to black and white pixels. A feature can often be detected at lower resolutions, on the order of 1 pixel/feature, but quality judgements come into play.



Stroke: Adequately rendered cloud outline (left) and inadequately rendered border line (right).

Cornell has developed the following correlation of perceived image quality to pixel coverage:

Table: Quality Index for Stroke Rendering

QI	Quality Assessment
2	excellent
1.5	good
1	questionable, confirm quality onscreen
<1	poor to unacceptable

Grayscale/Color QI Formula for Stroke

$$\text{dpi} = \text{QI} / .039w$$

This formula correlates QI with dpi and stroke width (w) measured in mm. QI in this case is based on the quality assessment above, which correlates to the number of pixels covering the stroke (e.g., 2 = excellent). *Note:* if the measurement of w is expressed in inches, omit the .039.

For bitonal scanning, the formula is adjusted to compensate for feature drop out in the thresholding process:

Bitonal QI Formula for Stroke

$$\text{dpi} = 1.5\text{QI} / .039w$$

Many items falling into this category exhibit features beyond simple edge-based representation, and resolution will not be the sole determinant of image quality. For example, a number of institutions have recommended scanning all manuscripts in grayscale or color.

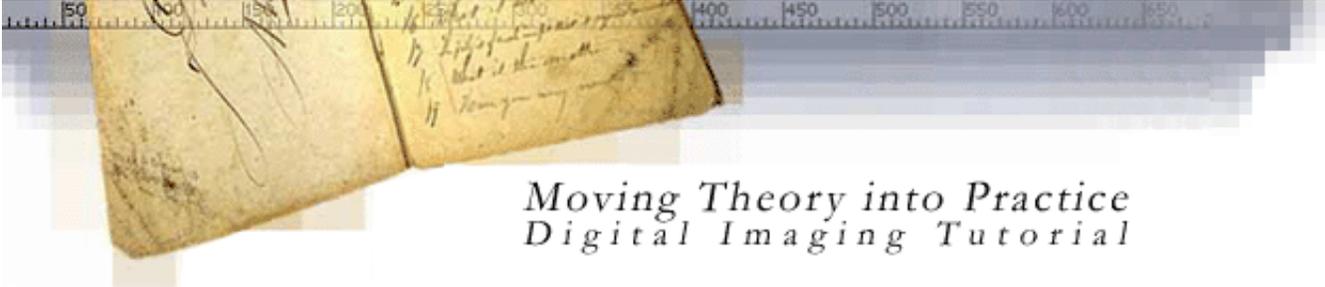
Reality Check

What's the minimum resolution I need to scan a manuscript page in grayscale to obtain good rendering of the finest stroke, which measures .1 mm?

dpi



View in:
← Spanish
← French



Moving Theory into Practice

Digital Imaging Tutorial

3. Conversion

Key Concepts

introduction

scanning factors

rich digital master

benchmarking

text

stroke

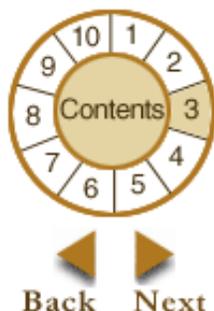
continuous-tone

halftone

proposed method

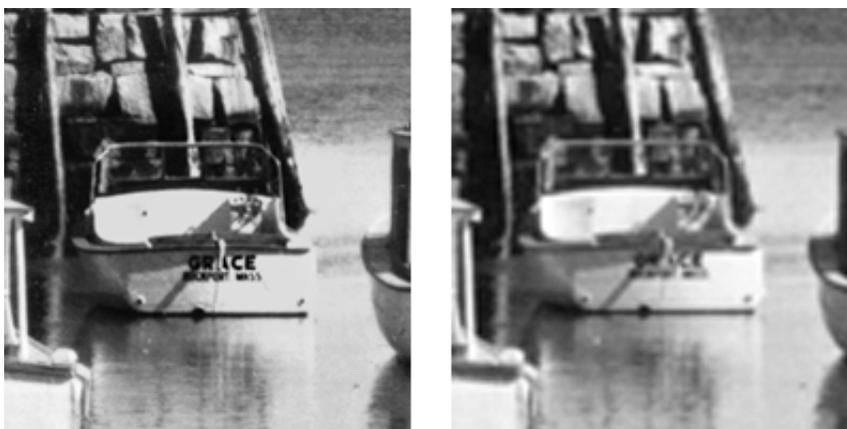
guidelines

additional reading



BENCHMARKING RESOLUTION REQUIREMENTS FOR CONTINUOUS TONE DOCUMENTS

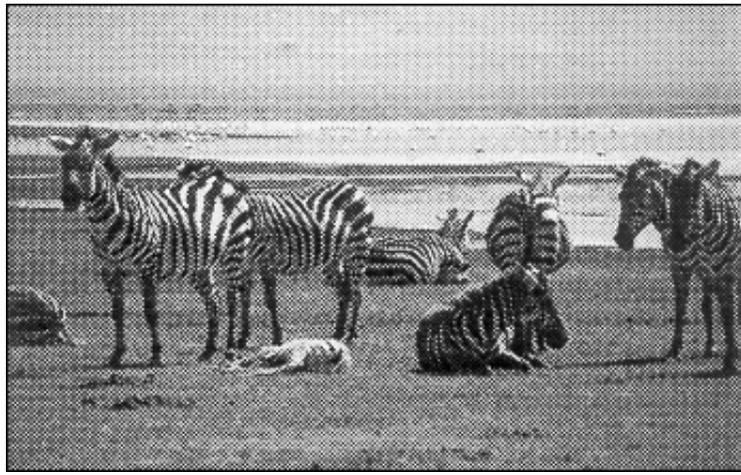
Resolution requirements for photographs and other continuous tone documents are difficult to determine because there is no obvious fixed metric for measuring detail. Detail may be defined as relatively small-scale parts of a document, but this assessment may be highly subjective. We might agree that street signs visible under magnification in a cityscape should be rendered clearly, but what about individual hairs or pores in a portrait? At the granular level, photographic media are characterized by random clusters of irregular size and shape, which can be practically meaningless or difficult to distinguish from background noise. Many institutions have avoided the issue of determining detail by basing their resolution requirements on the quality that can be obtained from prints generated at a certain size (e.g., 8 x 10-inch) from a certain film format (e.g., 35 mm, 4 x 5-inch). The important thing to remember about continuous tone documents is that tone and color reproduction is as important, if not more so, than resolution in determining image quality. See [Guides to Quality in Visual Resource Imaging](#).



Effect of Resolution on Continuous Tone Documents: The name of the boat (Grace) is legible in the left image, which was scanned at a higher resolution.

BENCHMARKING RESOLUTION REQUIREMENTS FOR HALFTONES

Halftones are particularly difficult to capture digitally, as the screen of the halftone and the grid of the digital image often conflict, resulting in distorted images with moiré (e.g., wavy patterns). Although a number of scanners have developed special halftoning capabilities, one of the more consistent ways to scan is in grayscale at a resolution that is four times the screen ruling of the halftone. This screen ruling can be determined using a [halftone screen finder](#), available from graphic arts supply houses. For high-end materials, such as fine art reproductions, this requirement will result in high resolutions (on the order of 700-800 dpi). For most halftones, 400 dpi, 8-bit capture is probably sufficient. Cornell did not discern any noticeable moiré when scanning a range of 19th- and early 20th-century halftones at that resolution. Lower resolutions can be used when special treatment scanning is employed. The [Library of Congress](#) has identified four distinct approaches to imaging halftone documents. See also the Cornell and Picture Elements study on [imaging book illustrations](#) for discussion of halftone treatments.



Effect of Resolution on Halftone Documents: The top image was scanned at 150 dpi, a resolution that clashed with the halftone's screen ruling of 85 lpi. The bottom image was scanned at 400 dpi and scaled for comparison purposes. Click on bottom image to view halftone grid pattern.



 Back Next

 Contents

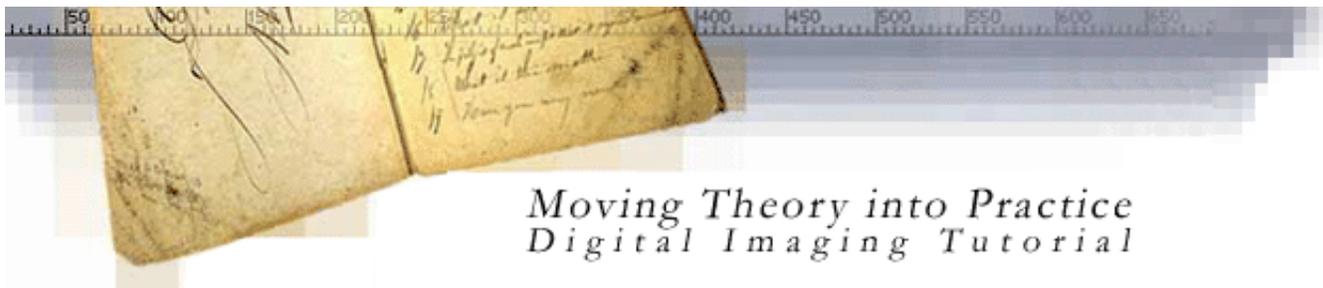
View in:



 Spanish

 French

© 2000-2003 Cornell University Library/Research Department



Moving Theory into Practice Digital Imaging Tutorial

3. Conversion

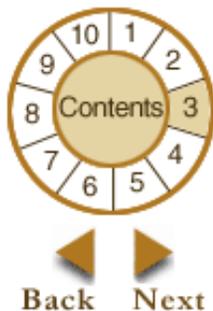
Key Concepts

- introduction
- scanning factors
- rich digital master
- benchmarking
- text
- stroke
- continuous-tone halftone

proposed method

guidelines

additional reading



PROPOSED METHODOLOGY FOR DETERMINING CONVERSION REQUIREMENTS

In *Moving Theory into Practice: Digital Imaging for Libraries and Archives*, a methodology is proposed for determining conversion requirements for a range of cultural heritage resources, including printed text, manuscripts, works of art on paper, and photographs. This methodology is based on the following steps:

- Document assessment & objective characterization
- Translation to digital equivalencies
- Assignment of tolerance values for pass/fail
- System calibration and performance testing
- Image evaluation via visual inspection and software analysis
- Recording of technical documentation

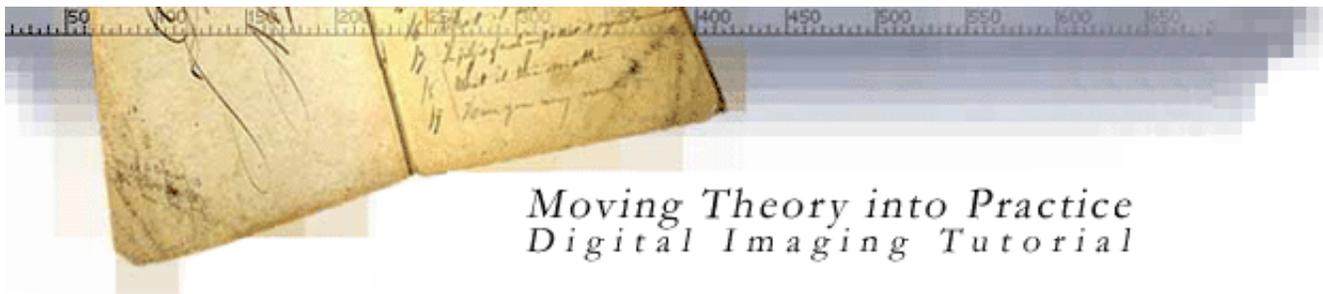


To see what some institutions recommend for conversion requirements, click on [REPRESENTATIVE INSTITUTIONAL REQUIREMENTS FOR CONVERSION](#).

View in:

- ← Spanish
- ← French

© 2000-2003 Cornell University Library/Research Department



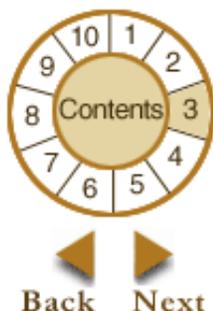
Moving Theory into Practice Digital Imaging Tutorial

3. Conversion

Key Concepts

- introduction
- scanning factors
- rich digital master
- benchmarking
- text
- stroke
- continuous-tone halftone
- proposed method
- guidelines

additional reading



ADDITIONAL READING

Anne R. Kenney, "Digital Benchmarking for Conversion and Access," *Moving Theory into Practice: Digital Imaging for Libraries and Archives*, Mountain View, CA : Research Libraries Group, 2000; pp. 24-60.

<http://www.rlg.org/preserv/mtip2000.html>

Guides to Quality in Visual Resource Imaging (July 2000), especially guides 2-4. <http://www.rlg.org/visguides>

Anne R. Kenney and Oya Y. Rieger, *Using Kodak Photo CD Technology for Preservation and Access*, 1998.

<http://www.library.cornell.edu/preservation/publications.html>

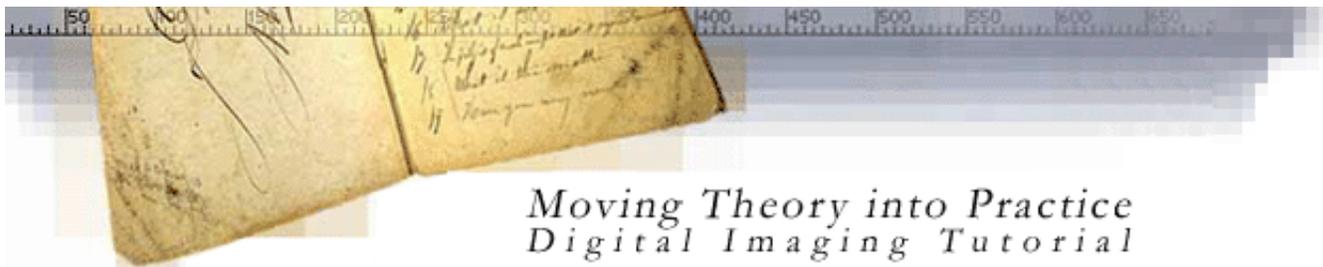
Anne R. Kenney and Stephen Chapman, *Tutorial: Digital Resolution Requirements for Replacing Text-based Material: Methods for Benchmarking Image Quality*, 1996. <http://www.clir.org/pubs/abstract/pub53.html>



View in:

- ← Spanish
- ← French

© 2000-2003 Cornell University Library/Research Department



Moving Theory into Practice Digital Imaging Tutorial



Back Next

4. Quality Control

Key Concepts

definition

developing a program
assessing quality

DEFINITION

Quality control (QC) is an integral component of a digital imaging initiative to ensure that quality expectations have been met. It encompasses procedures and techniques to verify the quality, accuracy, and consistency of digital products. Quality control strategies can be implemented at different levels:

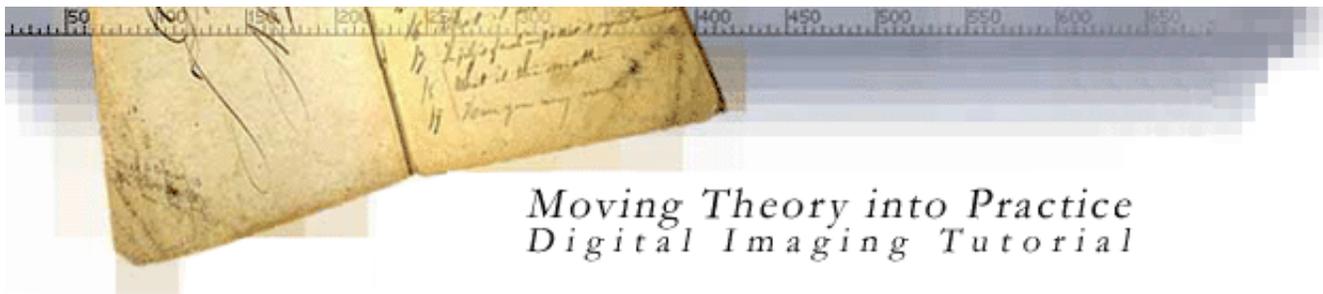
- Initial Evaluation
A subset of documents (to be converted in-house or by a service provider) is used to verify the appropriateness of technical decisions made during benchmarking. This evaluation occurs prior to implementing the project.
- Ongoing Evaluation
The same quality assurance process used to confirm benchmarking decisions can be scaled and extended to the whole collection to ensure quality throughout the digital imaging initiative.

© 2000-2003 Cornell University Library/Research Department



View in:

- ← Spanish
- ← French



Moving Theory into Practice

Digital Imaging Tutorial



Back Next

4. Quality Control

Key Concepts

definition
developing a program
assessing quality

additional reading

DEVELOPING A QC PROGRAM

The following steps outline the main points of a quality control program. A fully developed strategy for establishing such a program is presented in [*Moving Theory into Practice: Digital Imaging for Libraries and Archives*](#).

1. Identify Your Products

The first step is to clearly identify the products to be evaluated. These might include master and derivative images, printouts, image databases, and accompanying metadata, including converted text and marked-up files.

2. Develop a Consistent Approach

To measure quality and judge whether the products are satisfactory, clearly define baseline characteristics for "acceptable" and "unacceptable" digital products.

Example: Defining Image Quality Parameters for Different Project Goals

If the goal is *faithful representation*, quality assessment will be based on how well the image conveys the appearance of the original document (detail, color, tone, paper texture, etc.).



Faithful Representation: The color image (left) represents the essence of the original more fully than the grayscale image (right).

If the goal is *removing a color cast* introduced during the photographic process, quality will be judged against the original scene or document (rendering intent), rather than the photograph at hand.



Removing Color Cast: The color shift caused by photography (left) was detected and removed during quality inspection (right).

3. Determine a Reference Point

What are you judging the images against? Answering this question is not always straightforward. For example, if conversion is based on an intermediate, the digital image is two "generations" away from the original. It has been copied to film (first generation), which is then scanned (second generation). What should be the reference point in assessing such an image, the original document or the transparency? Will the master or derivative (or both) be the focus of image quality inspection?

4. Define the Scope and Methods

Determine the scope of your quality review. Will you inspect all the images, or only on a sampled subset (e.g., 20%)?

Describe your methodology and define how quality judgements will be made. For example, will you visually evaluate the images at 100% (1:1) magnification onscreen and compare them to the original documents? Or, will your evaluation be based only on a subjective assessment of images onscreen, without reference to the originals?

Example: Because Cornell University Library replaces brittle volumes with digital reprints, image quality evaluation is based on the printouts created from the digital images. A 100% inspection is conducted, comparing each printout to the corresponding original page.

5. Control the QC Environment

The impact of image-display conditions on perceived quality is often underestimated. Given an improper environment, even a high-quality image may come across as unsatisfactory. For example, a 24-bit color image might look heavily "posterized" when viewed using an improperly configured computer that cannot provide a full palette of colors. More information on controlling the viewing environment is provided in [Using Kodak Photo CD Technology for Preservation and Access](#).



QC Environment: Image quality evaluation conducted in a controlled environment. Courtesy of [William Blake Archive](#).

Factors affecting on-screen image quality

Hardware Configuration

It is difficult to prescribe the ideal hardware configuration. The rule of thumb is to assemble a system that supports your requirements for speed, memory, storage, and display quality. What kinds of images are being created? How many? To serve what purposes? What level of on-screen review is needed? You will need a fast and reliable computer with ample processing power and memory to be able to retrieve and manipulate the large files you are creating, especially when creating color images. See also: [Technical Infrastructure: Image Creation](#).

Image Retrieval Software

Use retrieval software appropriate to your images. For example, if you are evaluating images created and stored in Kodak ImagePac format, retrieve them using one of the viewing freeware and shareware products available on the Web that support the format and color space. Cornell used Adobe Photoshop with the Kodak Photo CD Acquire Module plug-in to ensure correct mapping of Photo CD colors. More information is provided in [Using Kodak Photo CD Technology for Preservation and Access](#).

Viewing Conditions

Control your viewing environment. Understand that the monitor and the source document require distinct viewing conditions. The original is best viewed in a bright surrounding, and the monitor works best in a low-light environment. However, a low-light environment does not equate to a dark room. Viewed in the dark, an on-screen image would appear deficient in contrast.

Human Characteristics

Image quality assessment requires visual sophistication, especially for subjective evaluations. Ideally, the same person should evaluate all images, using the same equipment and settings. In particular, staff need training to communicate color appearance information effectively.

Some color vision deficiencies are linked to a defective, recessive gene on

the X chromosome. Since females have two X chromosomes and males have one, the chance of color-deficient vision is 1 in 250 females, but 1 in 12 for males. Even among expert viewers, differences in judgments due to normal variation in the human eye are not uncommon. A [color vision test](#) can be used to evaluate a viewer's vision.

Monitor Calibration

Images may appear different on different monitors. Calibration is the process of adjusting monitor color-conversion settings to a standard, so that the image displays the same on a variety of monitors. The ideal method is to use monitor-calibration hardware and accompanying software. However, if you do not have access to these resources, use your application program's calibration tools. For example, Adobe Photoshop includes a basic monitor-calibration tool, which can be used to eliminate color cast and standardize the display of images.

Color Management

One of the main challenges in digitizing color documents is to maintain color appearance and consistency across the digitization chain, including scanning, displaying, and printing. Accurately reproducing colors is difficult because input and output devices treat colors differently. The goal of color management system (CMS) software is to ensure that the colors of the original match as precisely as possible the digital reproduction on-screen or printed out.

6. Evaluate System Performance

Whether conversion takes place in-house or is outsourced, system performance should be evaluated to ensure consistency throughout the conversion process. Among characteristics to evaluate are resolution, linearity, flare, scanner noise, color reproduction, and various artifacts. Several publications noted at the end of this section provide further information on system calibration.

7. Codify Your Inspection Procedures

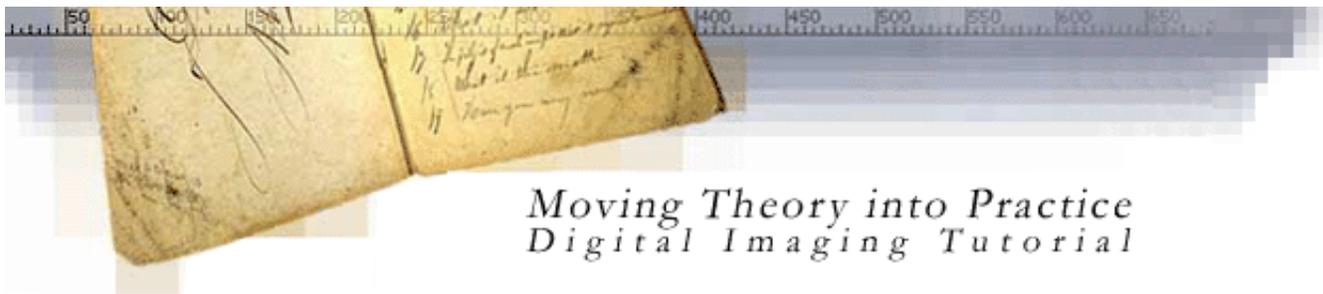
Quality control data has long-term value, from supporting different stages of quality inspection to facilitating future manipulation and migration. For in-house components of QC, we recommend detailing the inspection procedures in a short manual (or in a series of workforms) to be used in training and to facilitate workflow. Issues that need to be addressed include: QC procedures; staff involved and skills needed; instruments, hardware, and software needs; rejecting and replacing unacceptable products. An example of this approach is demonstrated by the Library of Congress in its [Internal Training Guide](#).



Reality Check

You retrieve an image from a CD that just arrived from the image production unit. The image is not crisp and looks darker than what you had expected. What is the first thing to do to determine the cause of the problem?

- (A) Confirm that your viewing environment is correct.
- (B) Use the adjustment tools (sharpening, color correction) available in your image viewing software to bring the image in line with expectations.
- (C) Call the image production unit to find out what happened.



Moving Theory into Practice Digital Imaging Tutorial



Back Next

4. Quality Control

Key Concepts

definition
developing a program
assessing quality

additional reading

ASSESSING IMAGE QUALITY

The key factors in image quality assessment are resolution, color and tone, and overall appearance. For further discussion of image quality metrics, see an *RLG DigiNews* [technical feature](#) by Don Williams.

Resolution

Resolution is the key factor in determining image quality for textual materials and other distinct, edge-based representations. For graphical material, especially continuous tone images, bit-depth, color representation, and dynamic range combine with resolution to determine the quality. Resolution attributes to inspect are legibility, completeness, darkness, contrast, sharpness, and uniformity. Measuring and evaluating stroke and detail are useful in assessing image quality. See the [RLG Model RFP](#) for examples of defining quality expectations and suggested QC procedures associated with these requirements. Follow emerging quality metrics for assessing resolution (see additional readings).



100 dpi

300 dpi

600 dpi

Effects of Resolution on Image Quality: Compare the quality of these three bitonal images of a wood engraving captured at different resolutions.

Color and Tone

For color, grayscale, and some monochrome images, color and tone reproduction are significant indicators of quality, complementing the "detail" provided by resolution. The goal behind assessing color and tone appearance is to determine the extent to which a digital image conveys the same appearance as the color and tone ranges of the original document (or intermediate used). Tone and color assessment may be highly subjective and changeable according to the viewing environment and the characteristics of monitors and printers.

The following Web site provides good information on viewing and evaluating color, but requires users to register before viewing the online guide.
X-Rite. 1998. [Color Guide and Glossary: Communication, Measurement, and Control for Digital Imaging and Graphic Arts.](#)

Below are illustrations that demonstrate the effects of color and tone on image quality.



Dynamic Range: Compare the left image to the right image, which has limited dynamic range; note the detail in shadows and highlights.



Bit Depth: When a 24-bit image (left) is reduced to an 8-bit one (right), the color reduction may result in quantization artifacts.



Brightness and Contrast: Compare the left image to the right one with high brightness and contrast settings.

Overall Evaluation

Image quality is cumulative, affected by a range of individual factors--capture system performance, resolution, dynamic range, and color accuracy. The final evaluation should be made on the overall image, appreciating all the individual factors that contribute to quality.

© 2000-2003 Cornell University Library/Research Department



Moving Theory into Practice Digital Imaging Tutorial



Back Next

4. Quality Control

Key Concepts

definition
developing a program
assessing quality

additional reading

ADDITIONAL READING

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

<http://www.rlg.org/preserv/mtip2000.html>

Don D'Amato, *Imaging Systems: The Range of Factors Affecting Image Quality, Guide 3 to Quality in Visual Resource Imaging*,

<http://www.rlg.org/visguides/visguide3.html>

Michael Ester, *Digital Image Collections: Issues and Practice*. The Commission on Preservation and Access, 1996.

<http://www.clir.org/pubs/abstract/pub67.html>

Franziska S. Frey and James M. Reilly, *Digital Imaging for Photographic Collections: Foundations for Technical Standards* (Rochester, NY: Image Permanence Institute, Rochester Institute of Technology, 1999)

http://www.rit.edu/~661www1/sub_pages/page3a.htm#7

Franziska Frey, *Measuring Quality of Digital Masters, Guide 4 to Quality in Visual Resource Imaging*, <http://www.rlg.org/visguides/visguide4.html>

Anne R. Kenney and Oya Y. Rieger, *Using Kodak Photo CD Technology for Preservation Access*. Ithaca, NY: Cornell University Library, Department of Preservation and Conservation, 1998,

<http://www.library.cornell.edu/preservation/kodak/cover.htm>

National Digital Library Program, Library of Congress, "Quality Review of Image Documents, Internal Training Guide," April 1999,

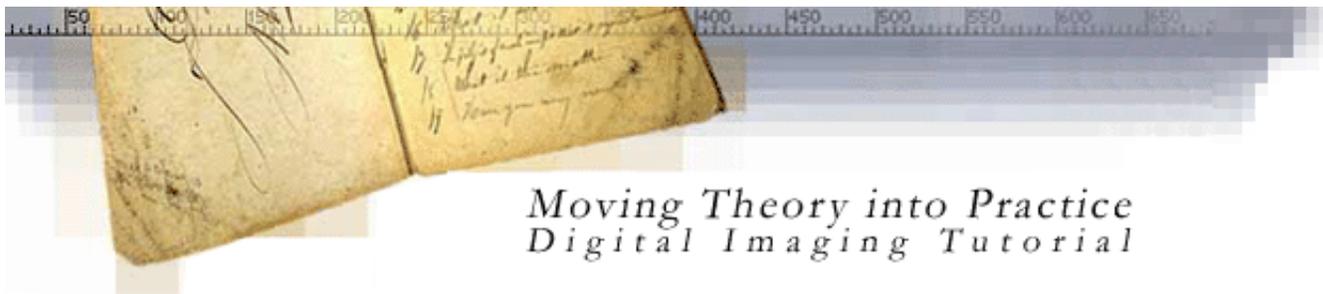
<http://memory.loc.gov/ammem/techdocs/qintro.htm>

© 2000-2003 Cornell University Library/Research Department

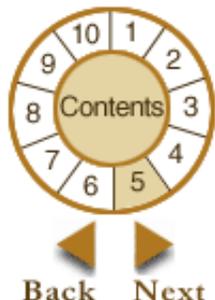


View in:

Spanish
French



Moving Theory into Practice Digital Imaging Tutorial



DEFINITION

Metadata describes various attributes of information objects and gives them meaning, context, and organization. Descriptive metadata theory and practice is a familiar area for many as its roots are embedded in the cataloging of print publications. In the digital realm, additional categories of metadata have emerged to support navigation and file management.

METADATA TYPES AND THEIR FUNCTIONS

For practical purposes, the types and functions of metadata can be classified into three broad categories: descriptive, structural, and administrative. These categories do not always have well-defined boundaries and often exhibit a significant level of overlap. For example, administrative metadata may include a wide range of information that would be considered descriptive and structural metadata.

5. Metadata

Key Concepts

definition

types and functions

creation

additional reading

To view a table summarizing the goals, elements, and sample implementations of the three categories of metadata, click on

[Table: Metadata Types](#)



View in:

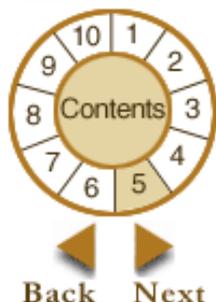
- ← Spanish
- ← French

© 2000-2003 Cornell University Library/Research Department



Moving Theory into Practice

Digital Imaging Tutorial



METADATA CREATION

Metadata creation and implementation are resource-intensive processes. Balance costs and benefits in developing a metadata strategy, taking into consideration the needs of current and future users and collection managers. Identify metadata requirements at the onset of an imaging initiative. These requirements should be tightly linked to functions that must be supported (e.g., rights management, resource discovery, and long-term care).

Consider the following issues:

- Although some metadata elements are static (e.g., date of creation, scanning resolution), certain fields (e.g., migration information) may continue to evolve and require continuous updating and maintenance.
- The creation and management of metadata is accomplished through manual (creating a [Dublin Core](#) record) and automated (generating a keyword index from OCR'ed text) techniques. Similarly, metadata quality control will be based on a mix of manual (evaluating the quality of subject access categories and keywords) and automated (using an SGML parser to validate tags) processes.
- Metadata can be internal (file naming, directory structuring, file headers, OCR, SGML) or external (external indexes and databases). The key factor in decision making is evaluating whether the location supports functionality and resource management. For example, TIFF file headers are instrumental in recording metadata internally; however, this metadata is usually lost when the TIFF files are converted to other file formats, such as JPEG or GIF.

There are several standards in development to facilitate interoperability among different metadata schemes. The [Resource Description Framework](#) (RDF) is an XML-based application to provide a flexible architecture for managing diverse metadata in the networked environment. The goal of the Digital Imaging Group's Metadata For Digital Images ([DIG 35](#)) initiative is to define a standard set of metadata that will improve interoperability between devices, services, and software, thus making it easier to process, organize, print, and exchange digital images. The [MPEG-7](#) (Moving Picture Experts Group) initiative targets audio-visual content description and aims to standardize a set of description schemes and descriptors, a language to specify description schemes, and a scheme for coding the description. The Interoperability of Data in E-Commerce Systems ([<indecs>](#)) project is an international collaboration to develop a metadata framework that supports network commerce of intellectual property.

5. Metadata

Key Concepts

definition

types and functions

creation

additional reading

Example

What kinds of metadata will be created for a journal collection that is converted as 600 dpi, 1-bit TIFF 6.0 images? The following metadata tasks might be undertaken. Each is identified by its principle metadata type (S = structural, D=descriptive, A=administrative). *Note:* The [RLG Model RFP](#) provides an example of metadata requirements for a text imaging project.

- Assign file names and directory structures to the image files and the associated metadata files. (S)
- Create or update MARC records (Fields 100, 110, 245, 260, 440, 650, etc.). (D)
- Create Dublin Core records. (D)
- Use [MARC Field 007](#) to record digital preservation and reformatting information. (A)
- Use appropriate TIFF 6.0 file headers to record technical information, e.g., ImageWidth, ImageLength, Compression, StripOffsets, RowsPerStrip, StripByteCounts, Xresolution, Yresolution, Resolution Unit, BitsPerSample. (A)
- Assign persistent, globally-unique, and location-independent file names ([PURL](#) or [Handle](#)). (D)
- Use appropriate TIFF 6.0 file headers for image description (Field 270) to record descriptive elements essential for identifying the file (e.g., project ID, institution, collection, year of publication, title, author, image sequence number). (D)
- Create a database to store and manage bibliographic information from the cumulative journal indexes to enable structured vocabulary search (e.g., journal volume, issue, title, author, beginning and ending page number). (D, S)
- Use [TEI Lite SGML](#) encoding to map the basic structural elements of the journals, such as volume, issue, title, author name, beginning and ending pages for each article, to facilitate online searching and browsing. (S)
- OCR images to provide free-text key word access. (D)
- Create HTML tags with Dublin Core information to facilitate resource discovery. (D)
- Register the Web site with relevant subject directories, specialized subject portals, and gateways to increase coverage by Web search engines. (D)

Example 2

What kinds of metadata will be collected and recorded for a collection of photographs?

In addition to many of the elements suggested above, consider whether to-

- Enhance an existing finding aid, and SGML-encode it using the EAD ([Encoded Archival Description](#)) Document Type Definition to create a map of the collection for searching and presentation. This will facilitate interoperability with other EAD-encoded finding aids (D, S, A)

Reality Check

Which of the following metadata would be important for preservation reasons? Select all correct answers.

Unique identifiers

Structuring tags

Physical description of source document

Scanner profile

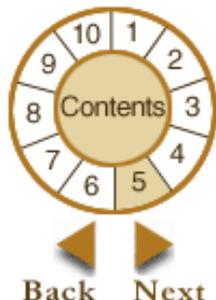
© 2000-2003 Cornell University Library/Research Department

 
Back Next
 Contents

View in:
 Spanish
 French



Moving Theory into Practice Digital Imaging Tutorial



5. Metadata

Key Concepts

definition
types and functions
creation

additional reading

ADDITIONAL READING

Carl Lagoze and Sandra Payette, "Metadata: Principles, Practices, and Challenges," in *Moving Theory into Practice: Digital Imaging for Libraries and Archives*, Mountain View, CA : Research Libraries Group, 2000; pp. 84-100
<http://www.rlg.org/preserv/mtip2000.html>

Fourth DELOS Workshop: Image Indexing and Retrieval. August 28-30, 1997: San Miniato. <http://www.ercim.org/publication/ws-proceedings/DELOS4/index.html>

NISO/CLIR/RLG. Technical Metadata Elements for Images Workshop, August 18-18, 1999: Washington, DC.
http://www.niso.org/news/events_workshops/image.html

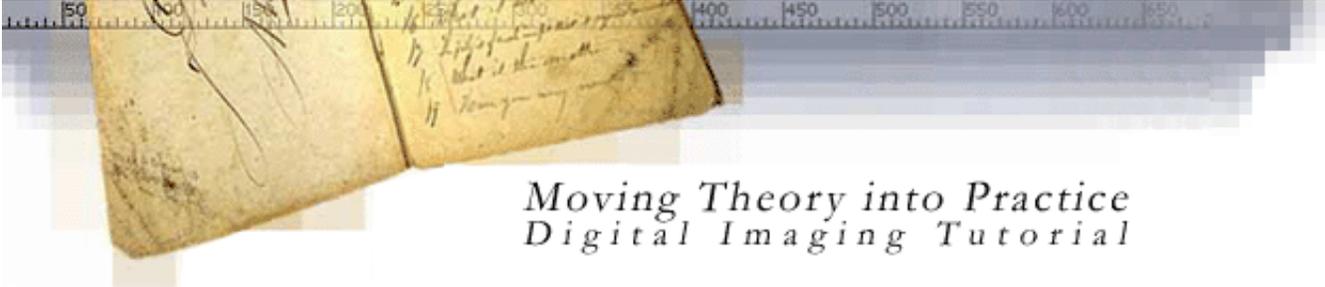
Getty Standards Program, "Introduction to Metadata: Pathways to Digital Information," Version 2.0, at
<http://www.getty.edu/research/institute/standards/intrometadata/index.html>

© 2000-2003 Cornell University Library/Research Department

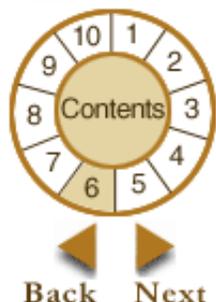


View in:

← Spanish
← French



Moving Theory into Practice Digital Imaging Tutorial



6A. Technical Infrastructure: DIGITIZATION CHAIN

Key Concepts

introduction

components

system integration

INTRODUCTION

Technical infrastructure refers loosely to the components that make digital imaging possible. The entire process is sometimes called the digitization chain, suggesting a series of logically ordered steps. In actual practice, the digitization chain can have side branches, loops, and recurring steps, but for simplicity's sake, we present it here as if it were linear.

The Digitization Chain

The technology necessary to navigate from one end of the digitization chain to the other consists mainly of hardware, software, and networks. These are the focus of this section. A truly comprehensive view of technical infrastructure also includes protocols and standards, policies and procedures (for workflow, maintenance, security, upgrades, etc.), and the skill levels and job responsibilities of an organization's staff.

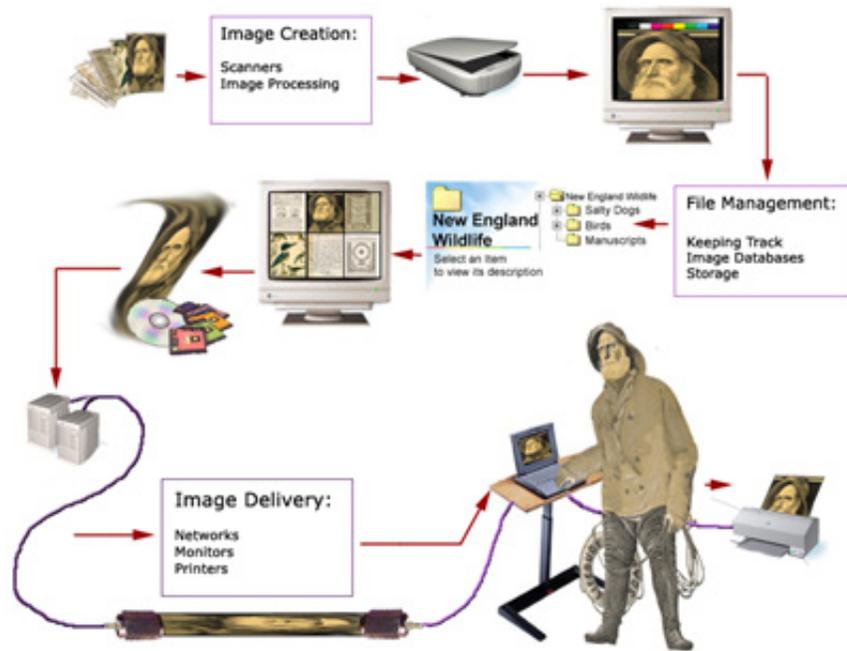
However, even the nuts and bolts of the technical infrastructure cannot be evaluated in complete isolation. Related actions and considerations that will affect decisions about the technical infrastructure include:

- Determining quality requirements based on document attributes ([Benchmarking](#))
- Assessing institutional strengths and weaknesses, timetable, and budget ([Management](#))
- Understanding user needs ([Presentation](#))
- Assessing long-term plans ([Digital Preservation](#))

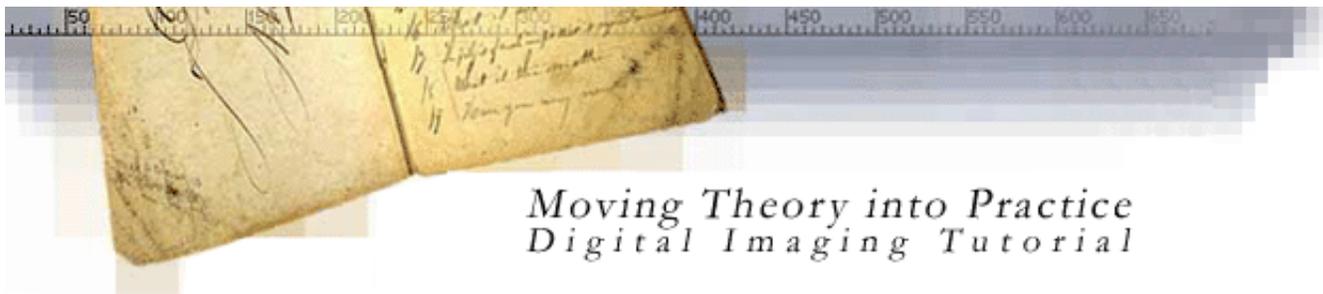
Technical infrastructure decisions require careful planning because digital imaging technology changes rapidly. The best way to minimize the impact of depreciation and obsolescence is through careful evaluation, and the avoidance of unique, proprietary solutions. If equipment choices are well-matched to intended uses and expected outcomes and synched to realistic timetables, return on investment will be maximized.

Digitization Chain: Click on this image to view a close-up version.

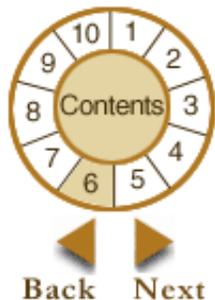
Back Next
Contents
View in:
Spanish
French



© 2000-2003 Cornell University Library/Research Department



Moving Theory into Practice Digital Imaging Tutorial



6A. Technical Infrastructure: DIGITIZATION CHAIN

Key Concepts

- introduction
- components**
- system integration

THREE MAJOR COMPONENTS

For the purposes of this tutorial, the digitization chain and the technical infrastructure that supports it is divided into three major components: creation, management, and delivery.

Image creation deals with the initial capture or conversion of a document or object into digital form, typically with a scanner or digital camera. There may then be one or more file or image processing steps applied to the initial image, which may alter, add, or extract data. Broad classes of processing include image editing (scaling, compression, sharpening, etc.) and metadata creation.

File management refers to the organization, storage, and maintenance of images and related metadata.

Image delivery incorporates the process of getting images to the user and encompasses networks, display devices, and printers. Issues associated with creating derivative images are covered in [Presentation](#).

Computers and their network interconnections are integral components of the digitization chain. Each link in the chain involves one or more computers and their various components (RAM, CPU, internal bus, expansion cards, peripheral support, storage devices, and networking support). Depending on the specific computing demands of each component, configuration requirements will change. Therefore we will revisit computer needs each step of the way.

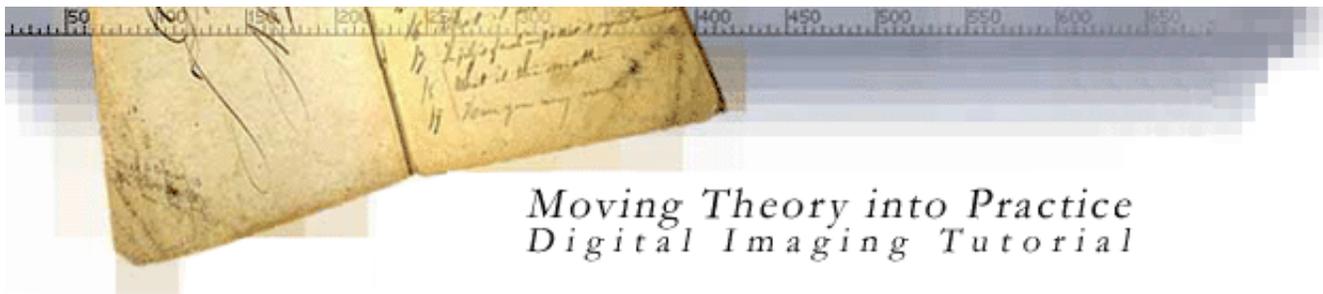
As we review each step, consider whether you'll conduct it yourself or rely on a vendor. (See [Management](#), for more information on the advantages and disadvantages of outsourcing). Those steps performed in-house require the most attention, though outsourcing doesn't reduce the need for a well thought out [quality control](#) program. However, in order to successfully evaluate and negotiate for contracted services, and to clearly communicate to vendors exactly what is expected, develop a baseline understanding of the concepts and procedures involved.



View in:

- ← Spanish
- ← French

© 2000-2003 Cornell University Library/Research Department



Moving Theory into Practice Digital Imaging Tutorial



Back Next

6A. Technical Infrastructure: DIGITIZATION CHAIN

Key Concepts

introduction
components
system integration

SYSTEM INTEGRATION: CONNECTING THE CHAIN

Keep a few overarching policy recommendations and caveats in mind as we discuss the technical infrastructure:

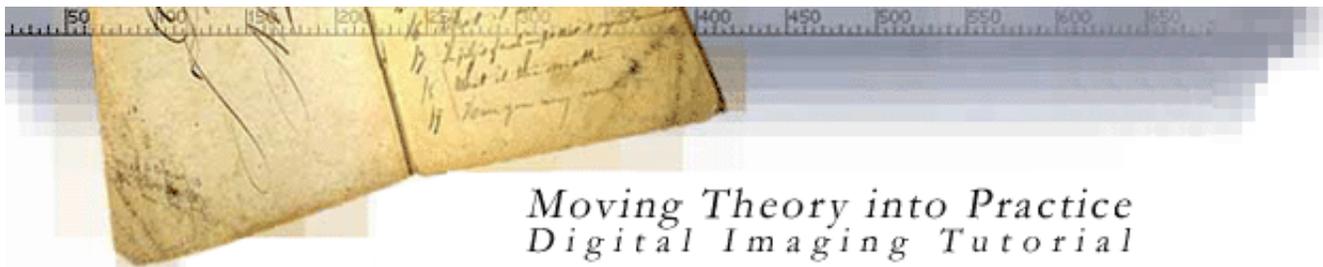
- 1) Consider using a systems integrator who can guarantee that all components interoperate without difficulty. If you decide to do all component selection yourself, keep the number of devices to a minimum.
- 2) Choose products that adhere to standards and have wide market acceptance and strong vendor support.
- 3) Despite all your best efforts, some things will go wrong, so be prepared for headaches. Claims to the contrary, plug 'n play doesn't always work. Digital imaging components must sometimes be adapted for library/archives use in creative ways.
- 4) Don't skimp—you'll pay more in the long run. If you're serious about making a commitment to digital imaging, buy quality and budget for upgrades and replacements at regular intervals. Waiting until you're stuck with obsolete, unsupported equipment or file formats can lead to time-wasting and expensive problems.
- 5) Involve technical staff early and often in planning discussions. As much as we'd like to think of it as linear, the digitization chain is really a complex shape that folds back on itself in several places. Technical staff can help identify the weak links resulting from the interdependencies of various steps in the process.



View in:

← Spanish
← French

© 2000-2003 Cornell University Library/Research Department



Moving Theory into Practice Digital Imaging Tutorial



Back Next

6B. Technical Infrastructure: IMAGE CREATION

Key Concepts

introduction

how scanners work
scanner types
image processing

INTRODUCTION

A dazzling array of devices that start the digitization chain now beckon the prospective digital imaging initiative. *Note:* We use the term scanner to refer to all image capture devices, including digital cameras.

Ask some key questions about any scanner you might consider.

- Is this scanner compatible with my documents? Can it handle the range of sizes, document types (single leaf, bound volume), media (reflective, transparent), and the condition of the originals? For additional details on matching a scanner to a particular set of document specifications, see Appendix A "Assessing Document Attributes and Scanning Requirements" of [The RLG Worksheet for Estimating Digital Reformatting Costs](#) and Don Williams, [Selecting a Scanner](#).
- Can this scanner produce the requisite quality to meet my needs? It is always possible to derive a lower quality image from a higher quality one, but no amount of digital magic can accurately restore detail that was never captured to begin with. Factors to consider include optical (as opposed to interpolated) resolution, bit depth, dynamic range, and signal-to-noise ratio.
- Can this scanner support my production schedule and conversion budget? (Pay attention to throughput claims—often a major factor in scanner cost.) What are its document handling capabilities? Its duty cycle, MTBF (Mean Time Between Failure), and lifetime capacity? What kind of maintenance contracts are available (on-site, 24-hour replacement, depot service)?

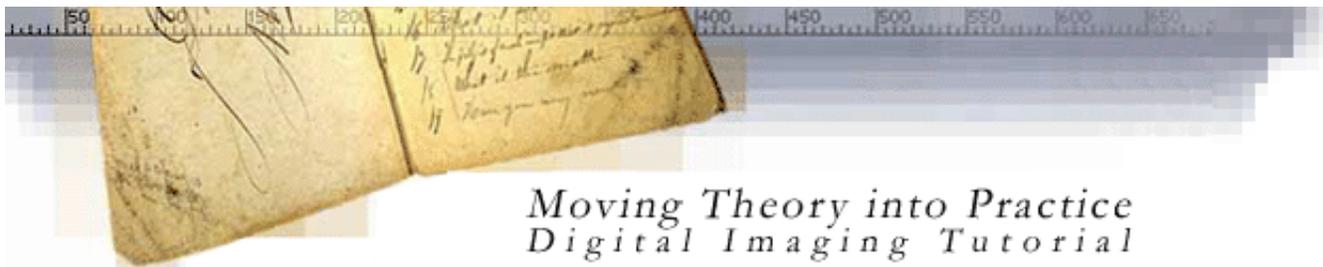
Scanner specifications can be difficult to interpret and often lack standardization, making meaningful comparisons impossible. The RLG/DLF guide, [Selecting a Scanner](#) examines scanner specifications related to image quality and can help the reader see past the marketing hype that is commonplace in the industry.

As you read through the details of available scanners, keep in mind that most scanners were designed for large markets such as the business and graphic arts segments. Few were designed to accommodate the specific needs of libraries and archives. Your goal will be to find one that best fits your needs with the fewest compromises.

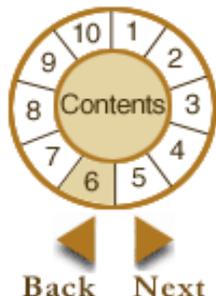


View in:

Spanish
French



Moving Theory into Practice Digital Imaging Tutorial



HOW SCANNERS WORK

Scanners operate by shining light at the object or document being digitized and directing the reflected light (usually through a series of mirrors and lenses) onto a photosensitive element. In most scanners, the sensing medium is an electronic, light-sensing integrated circuit known as a charged coupled device (CCD). Light-sensitive photosites arrayed along the CCD convert levels of brightness into electronic signals that are then processed into a digital image.

6B. Technical Infrastructure: IMAGE CREATION

Key Concepts

- introduction
- how scanners work**
- scanner types
- image processing

CCD is by far the most common light-sensing technology used in modern scanners. Two other technologies, CIS (Contact Image Sensor), and PMT (photomultiplier tube) are found in the low and high ends of the scanner market, respectively. CIS is a newer technology that allows scanners to be smaller and lighter, but sacrifices dynamic range, depth-of-field, and resolution. PMT-based drum scanners produce very high-quality images, but have limited application in library and archives scanning for reasons we'll discuss shortly.

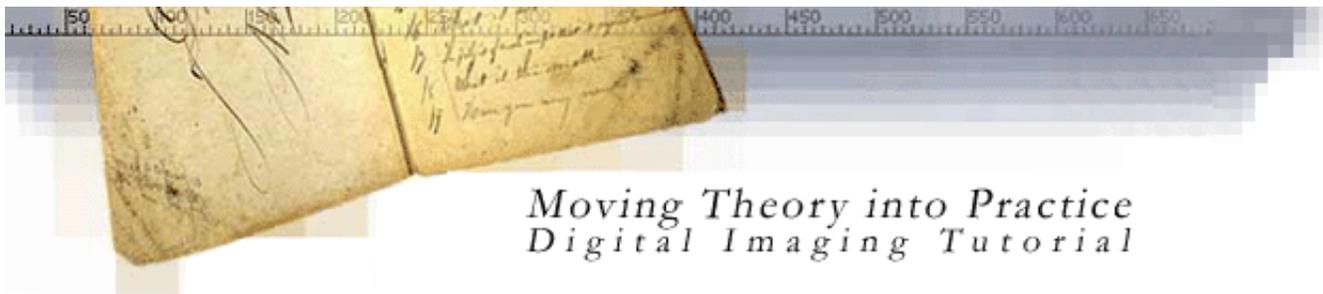
Another sensing technology, CMOS (Complementary Metal Oxide Semiconductor), appears primarily in low-end, hand-held digital cameras where its low cost, low power consumption and easier component integration permits smaller, less expensive designs. Traditionally, high-end and professional digital cameras employ CCD sensors, despite their expense and the complexity of their design, because they exhibit much superior noise characteristics. Although some innovative designs that render low-noise CMOS-based images are emerging, CCD still dominates the high end of the market. Click [here](#) for more details on scanner operation. Further technical details on digital cameras can be found [here](#).

© 2000-2003 Cornell University Library/Research Department

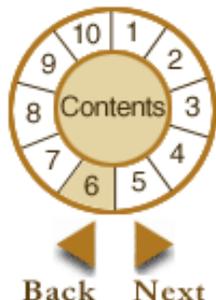


View in:

- ← Spanish
- ← French



Moving Theory into Practice Digital Imaging Tutorial



6B. Technical Infrastructure: IMAGE CREATION

Key Concepts

- introduction
- how scanners work
- scanner types**
- image processing

SCANNER TYPES

Flatbeds

Flatbed scanners are the best-known and largest selling scanner type, and with good reason. They're versatile, easy to operate, and widely available. Their popularity for Web publishing has opened up a huge market, pushing prices for entry level units below \$100. At the other end, professional units for the color graphics market now rival drum scanners in quality. All use the same basic technology, in which a light sensor (generally a CCD) and a light source, both mounted on a moving arm, sweep past the stationary document on a glass platen. Automatic document handlers (ADH) are available on some models, and can increase throughput and lessen operator fatigue for sets of uniform documents in reasonably good condition. A specialized variant of the flatbed scanner is the overhead book scanner, in which the scanner's light source, sensor array and optics are moved to an overhead arm assembly under which a bound volume can be placed face up for scanning.



Flatbed Scanner



Overhead Scanner

Sheetfeed Scanners

Sheetfeed scanners use the same basic technology as flatbeds, but maximize throughput, usually at the expense of quality. Generally designed for high-volume business environments, they typically scan in black and white or gray scale at relatively low resolutions. Documents are expected to be of uniform size and sturdy enough to endure fairly rough handling, although the transport mechanisms on some newer models reduces the stress. Using roller, belt, drum, or vacuum transport, the light sensor and light source remain stationary while the document is moved past. An important subclass of sheetfeed scanners are upright models specifically designed for oversize documents such as maps and architectural drawings.



Sheetfed Scanner

Drum Scanners

Drum scanners produce the highest resolution, highest quality scans of any scanner type, but at a price. Besides their expense, drum scanners are slow, not suitable for brittle documents and require a high level of operator skill. Thus they are typically found in service bureaus that cater to the color pre-press market.



Drum Scanner

Microfilm Scanners

Microfilm scanners are highly specialized devices for digitizing roll film, fiche, and aperture cards. Getting good, consistent quality from a microfilm scanner can be difficult because they can be operationally complex, film quality and condition may vary, and because they offer minimal enhancement capability. Only a few companies make microfilm scanners, and the lack of competition contributes to the high cost of these devices. Specifications for some microfilm scanners are available at

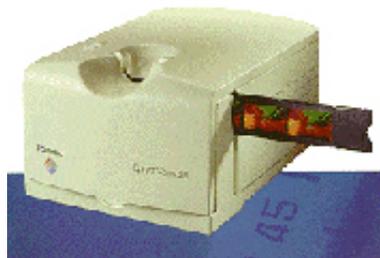
<http://www.rlg.org/preserv/diginews/diginews5-3.html#faq>.



Microfilm Scanner

Slide Scanners

Slide scanners are used to digitize existing slide libraries as well as photo intermediates of 3-dimensional objects and documents that are not well-suited for direct scanning, though more and more such objects will be captured directly by digital camera. The use of transparent media generally delivers an image with good dynamic range, but depending on the size of the original, the resolution may be insufficient for some needs. Throughput can be slow.



Slide Scanners

Digital Cameras

Digital cameras combine a scanner with camera optics to form a versatile tool that can produce superior quality images. Though slower and more difficult to use than flatbed scanners, digital cameras are adaptable to a wide array of documents and objects. Most fragile materials can be safely captured, though the need to provide external lighting means that light damage may be a concern. Digital camera technology continues to improve, helped along by the growing consumer market.



Digital Camera

To compare attributes of various capture devices, click on [Table: Comparison of Scanners](#)

Computer Considerations

A computer used as a scanning workstation must avoid becoming a bottleneck in the production process. Here are some characteristics to seek in a scanning workstation:

- Adequate RAM—512 MB recommended. More if the machine will also be used for image processing.
- A fast CPU—minimum 1.8 Ghz Pentium IV (or compatible) or 800 Mhz G4.
- Fast, capacious mass storage—enough space for at least temporary needs (40-60 GB), even if files are ultimately moved to other storage devices. (Methods for estimating storage needs are covered in [File Management](#)).
- Peripheral bus—Most low- and mid-range scanners now come with USB ports, commonly available on both Wintel and Mac computers. First generation USB (v.1.0/1.1) is quite slow and not suitable for large-scale production work. USB 2.0 is (theoretically) 40 times faster but is only just becoming widely available on Wintel machines in 2002, and scanners that support it are not yet common. Scanners offering Firewire connections (about the same speed as USB 2.0) are fairly widely available and Firewire is standard on Macintoshes, though it may have to be added to some Wintel machines. High-end scanners, including both high-speed monochrome and color scanners and low-speed (but very high quality) color scanners tend to offer only SCSI connectivity. SCSI has fallen out of favor on desktop systems, but can be installed on most systems with the addition of a peripheral card.
- High-bandwidth networking (10/100/1000 Base-T) to allow fast access to and transfer of scanned files.
- Platform/operating system—Most scanners offering USB connectivity work equally well on Wintel and Macintosh computers, though some manufacturers do not supply software drivers for Macs (third party products can sometimes solve this problem). Some scanners are platform specific, with high-end color graphics scanners more likely to support only Macintoshes, and high-speed production scanners more likely to support only Wintel machines. Be sure to check specifications to insure the scanner you want is compatible with your existing infrastructure.


Back Next
← Contents

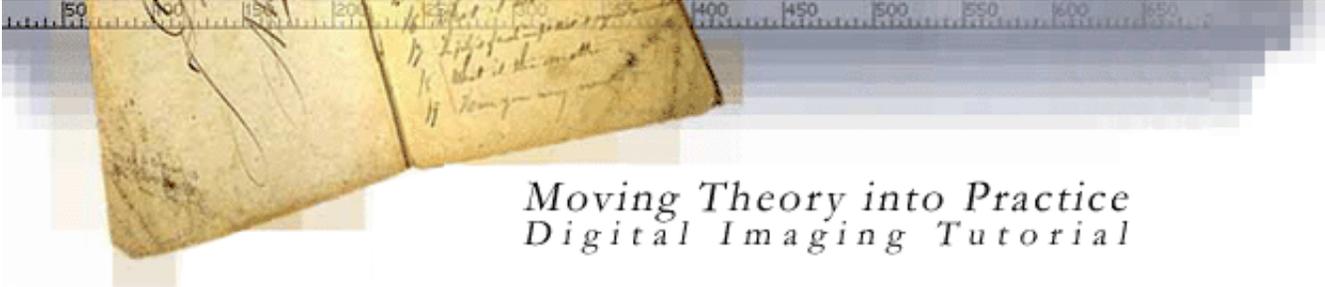
View in:
← Spanish
← French

Reality Check

Which scanner(s) can be used to image a 3-dimensional object?

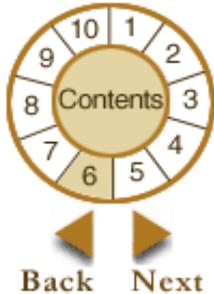
- Flatbed
- Sheetfeed
- Drum
- Slide/Film
- Microfilm
- Digital Camera

© 2000-2003 Cornell University Library/Research Department



Moving Theory into Practice

Digital Imaging Tutorial



FILE/IMAGE PROCESSING

A variety of processing steps follow scanning. Such procedures may occur at any point in the digitization chain, from immediately after scanning to just prior to delivery to end-users. These may be customized modifications that affect only certain files, or mass, automated processing of all files (batch processing). They may be one-time operations or done repeatedly on an as-needed basis.

Examples of file/image processing operations:

- Editing, touch-up, enhancement—this includes steps such as descreening, despeckling, deskewing, sharpening, use of custom filters, and bit-depth adjustment. In some cases, the scanning software performs these steps. In others, separate image-editing tools (e.g., Adobe Photoshop, Corel Photo-Paint, ImageMagick) are utilized.
- Compression—sometimes carried out by dedicated scanner firmware or dedicated hardware in the computer. Compression can also be a software-only operation though dedicated hardware is faster and should be considered when creating very large files or very large numbers of files.
- File format conversion—the original scan may not be in a format suitable for all intended uses, thus requiring conversion. See [Presentation](#).
- Scaling—it's likely that scans captured at high resolution will not be suitable for on-screen display. Scaling (that is, resolution reduction through bit disposal) is often necessary in order to create images for Web delivery. See [Presentation](#).
- OCR (optical character recognition)—conversion of scanned text to machine-readable text that can be searched or indexed.
- [Metadata](#) creation—addition of text that helps describe, track, organize, or maintain an image.

Computer Considerations

In some cases, image processing can be accommodated in the scanning workstation, especially if each image is checked as it's created. In the case of "on-the-fly" operations such as image scaling done just prior to delivery, image processing usually takes place on the image server.

Other operations may call for a separate computer. Image editing, especially for uncompressed 24-bit color images, requires large amounts of RAM and video memory. To work most efficiently, image editors require RAM several times the uncompressed size of the file being edited. A large, high-resolution monitor is also needed.

Image processing steps that may be carried out on every file (e.g. OCR, format conversion, deskewing) can be extremely CPU intensive. Batch processing requires a fast processor, lots of RAM, fast storage subsystems, and rapid and efficient routing of data within the system. These characteristics

6B. Technical Infrastructure: IMAGE CREATION

Key Concepts

[introduction](#)

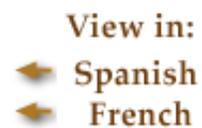
[how scanners work](#)

[scanner types](#)

[image processing](#)

are more often found on multi-user systems. In particular, Unix systems, with their inherent batch processing capabilities, are well-suited for these kinds of tasks, though computers running Linux or Windows 2000 Professional or XP Professional may also be suitable.

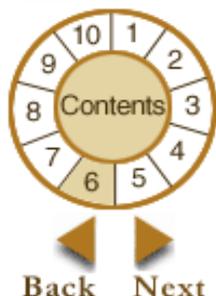
© 2000-2003 Cornell University Library/Research Department





Moving Theory into Practice

Digital Imaging Tutorial



INTRODUCTION

File management consists of a set of interrelated steps designed to ensure that files can be readily identified, organized, accessed, and maintained. Since there are strong connections between various aspects of file management, plan ahead to avoid making decisions that limit options later on. It is especially important to keep lines of communication open between technical staff and project staff during the planning stage.

6C. Technical Infrastructure: FILE MANAGEMENT

Key Concepts

introduction
 keeping track
 image databases
 storage
 storage types
 storage needs

File management steps examined here include:

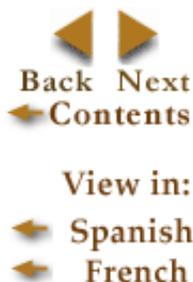
- *Keeping track* (basic file system considerations). Another aspect of keeping track is covered in [Metadata](#).
- *Image databases* and other image management solutions (special software for organizing image files)
- *Storage* (devices and media)
- *Maintenance* (backup, migration, preservation, and security) is addressed in [Digital Preservation](#)

KEEPING TRACK

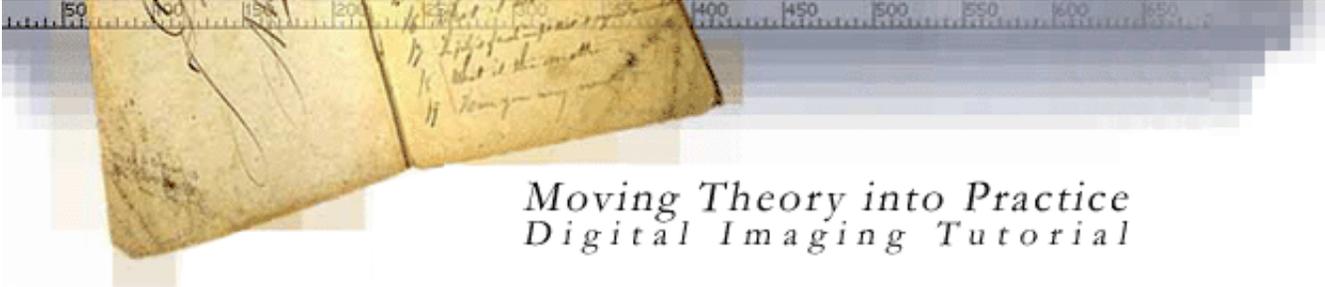
Default file and directory naming schemes are rarely optimal for a specific collection. Sound decisions about files and directories can help minimize chaos, especially for very large collections. To some degree, the nature of the material being scanned will suggest organizing principles. Serials are often divided into volumes and issues, monographs have page numbers, manuscript or photograph collections have folder or accession numbers, etc. In most cases, some aspect of these physical organizing principles can be translated into file system organization.

Follow some basic file system recommendations:

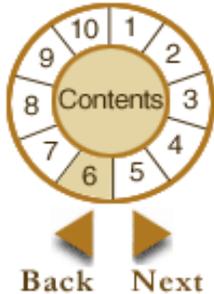
- Use a file naming scheme that is compatible with whatever operating systems and storage media you plan to use
- Use standard file extensions for different file types
- Don't overload directories with too many files
- Rely on storage management software to manage large collections across multiple physical disk drives
- Allow for generous amounts of collection growth



© 2000-2003 Cornell University Library/Research Department



Moving Theory into Practice Digital Imaging Tutorial



6C. Technical Infrastructure: FILE MANAGEMENT

Key Concepts

introduction
keeping track
image databases
storage
storage types
storage needs

IMAGE DATABASES

Many early digital initiatives relied heavily on custom programming for managing large collections of image files. Routines for batch processing, organizing, and delivering files were written using high-level scripting languages, such as Perl and Tcl. Today there are many off-the-shelf products that can dramatically simplify the process of managing a large collection of image files. However, even the simplest system requires some degree of customization. Larger collections and those with complex metadata require more sophisticated tools, which in turn require a higher degree of staff maintenance and oversight. Thus, programming experience is still a desirable skill for staff who manage image databases.

Image databases vary significantly in ease-of-use and level of functionality. They keep track of your files, provide search and retrieval functions, supply an access interface, monitor level and type of usage, and provide some security by controlling who gets access to what. No one tool is likely to meet all your needs, and even the most carefully chosen set of tools needs to be regularly re-evaluated to determine if it's still the best choice.

General criteria for evaluating image databases include the following:

- Purpose for which the digital file collection was created
- Size and growth rate of the image collection
- Complexity and volatility of accompanying metadata
- Expected level of demand and performance
- Existing technical infrastructure, including availability of skilled systems staff
- Expense

Basic categories of image database systems

A thorough assessment of image management systems, including pros and cons of each type and example applications, are discussed by Peter Hirtle in [*Moving Theory into Practice: Digital Imaging for Libraries and Archives*](#). He suggests the following major categories:

Common desktop databases are fairly low cost and simple to use, but limited in size and functionality.

Client-server database applications are more costly and more sophisticated than desktop databases, but are correspondingly more difficult to use and maintain.

Specialized image management systems can offer a complete off-the-shelf solution with pre-defined data structures, but are more expensive and less flexible in terms of customizability and compatibility.

More **library systems** are becoming image-enabled. Those that offer good linkage between existing catalog records and digital images, but suffer

from lack of standardization and a preference for item-level linking. Library systems staff may not be prepared to take on the additional burden of managing large image collections. However, this is an area of intensive development. More library systems are now accommodating image databases. A detailed look at some of the products is available in [Digital Object Library Products](#).

Computer Considerations

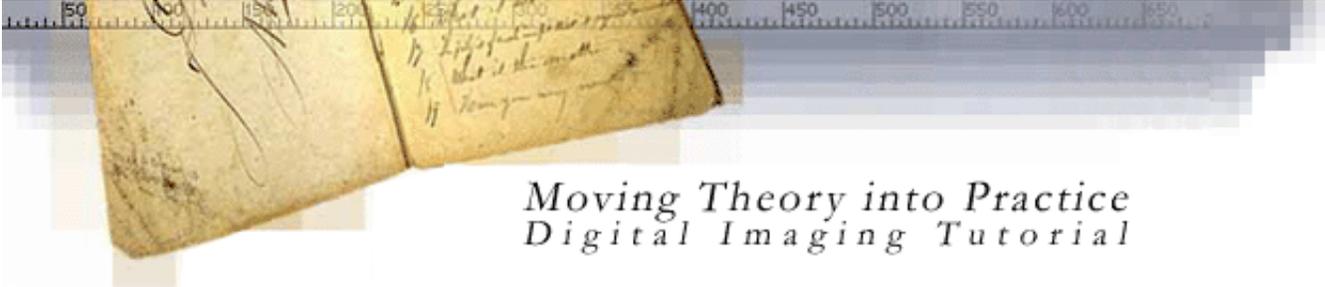
Desktop databases, by definition, are designed to run on desktop systems under MacOS or Windows. However, even a small collection may be overwhelmed on a desktop system if too many users attempt to access it simultaneously. Most larger database applications are designed to run in multi-user environments such as Unix, Linux, or Windows NT/2000, which run on machines offering fast processors, lots of RAM, fast i/o and peripheral buses, and fast storage devices.



View in:

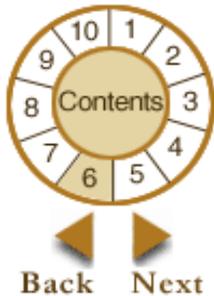


© 2000-2003 Cornell University Library/Research Department



Moving Theory into Practice

Digital Imaging Tutorial



6C. Technical Infrastructure: FILE MANAGEMENT

Key Concepts

introduction
 keeping track
 image databases
storage
 storage types
 storage needs

STORAGE

Typically, the component of the technical infrastructure that gets the most attention is the capture device, because it interacts directly with the tangible object being digitized and has the greatest influence on the quality and fidelity of the resulting image. Much less thought goes to the storage medium on which the captured bits will reside. This is unfortunate, since poor choices in storage technology can be detrimental to every step of digitization and can lead to production slowdowns, inefficient delivery, needlessly high short- and long-term costs, and the corruption and loss of data.

The reluctance to focus on storage technology is understandable. Storage devices perform a routine and utilitarian function within the digitization chain and are easy to take for granted. Additionally, mass storage is one of the most competitive and rapidly advancing computing technologies. As a result, it can be quite daunting even for the technically savvy to keep up with the ever-changing storage landscape, let alone understand some of its more complex aspects.

Except for relatively small installations, decisions about storage technology will probably be made in close consultation with systems staff. For that consultation relationship to be an effective partnership, knowledge of the basic terminology and concepts lays the foundation for asking the right questions.

General criteria for evaluation include:

- Speed (read/write, data transfer)
- Capacity
- Reliability (stability, redundancy)
- Standardization
- Cost
- Fitness to task

Rapid changes in storage technology have altered the impact of these criteria on digitization planning. In the early 1990s, storage was expensive, slow and of relatively limited capacity. Projects creating multiple gigabytes of image files experimented with various new (and often proprietary) optical disk technologies in order to find affordable means to safeguard their new digital treasures, often sacrificing speed and reliability in the process.

Today, the spinning magnetic disk drive is the undisputed king of storage. For all but the most ambitious projects, the production phase of digitization will be well-served by everyday, inexpensive parallel ATA drives, now commonly available in capacities of 120 GB per drive and interface transfer speeds up to 133 MB/second. Neither speed nor capacity is likely to create performance bottlenecks.

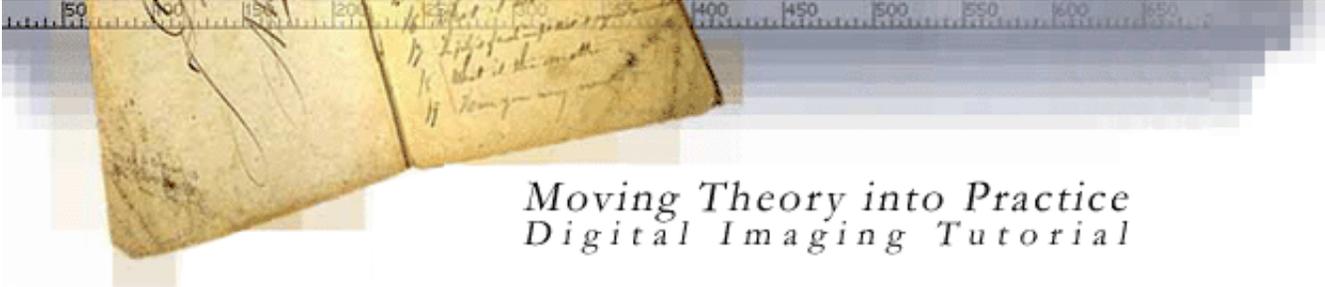
Today, the storage challenge is more likely to arise in the delivery stage, from efforts to consolidate disparate digital collections into a large digital library,

sometimes containing terabytes of data (a terabyte is 1000 gigabytes). Efficient management, delivery and maintenance of such collections is not a trivial task, and the premium pricing of large storage arrays with high reliability, excellent performance and integrated backup facilities can still strain budgets. Smaller collections that are in great demand may also require higher performance storage systems.

Within the range of available storage technologies, it is generally safest to choose one that is at or near its peak of popularity and acceptance. Technologies too close to the leading edge may never achieve widespread support from manufacturers or users, leaving early adopters with orphaned, unsupported hardware or media. Technologies too close to the trailing edge may suffer from declining product support and diminished upgrade paths. Also, don't buy substantially more storage than you think you'll need within the next couple of years. Under-utilized storage is not cost effective given the rapidly declining price and relatively short life expectancy. Most storage systems today are designed to accommodate incremental growth.

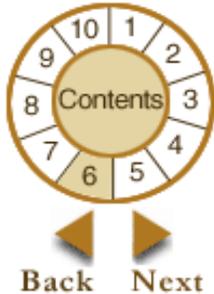
© 2000-2003 Cornell University Library/Research Department





Moving Theory into Practice

Digital Imaging Tutorial



6C. Technical Infrastructure: FILE MANAGEMENT

Key Concepts

introduction
 keeping track
 image databases
 storage
storage types
 storage needs

BASIC TYPES OF MASS STORAGE

Mass storage technologies can be classified in several ways. The underlying storage system (magnetic, optical or magneto-optical), the drive type (fixed or removable), the media material (tape, rigid platter, flexible platter), and the hardware interface (ATA, ATAPI, SCSI, USB, Firewire/IEEE 1394, Fibre Channel) jointly define the characteristics of each technology.

Storage systems are also distinguished as either direct attached storage or network attached storage. Direct attached storage includes standard desktop drives that are either installed within a computer case or cabled directly to it. Network attached storage generally encompasses storage that is accessible to multiple computers and may be either connected to a server and accessed via special file system protocols (e.g. Network File System or Common Internet File System), or part of a storage system that functions independently of any particular server (e.g. a SAN—Storage Area Network).

Storage hierarchies refer to the allocation of files to different kinds of storage depending on the frequency of use. When magnetic disk storage was very expensive, it was common to place the highest usage files on magnetic disk (online access), less frequently used files on less expensive (and slower) optical media (near line storage) and very infrequently accessed files on magnetic tape (offline storage). Due to the fact that magnetic disk storage has declined in price at a much more rapid pace than optical storage, the incentive to establish such hierarchies has lessened.

A table characterizing available technologies based on speed, capacity, and cost may be viewed by clicking on [Table: Comparison of Storage Media](#)

Trends in Mass Storage

Since the hard disk drive was invented in 1952, regular and rapid technological advancement has led to astonishing improvements in capacity, speed, reliability and price/performance ratio. The driving force in these improvements has been the unrelenting increase in the amount of data that can be stored in the same area (known as "areal density"). Unit cost for basic hard disk storage dropped by approximately a factor of 100 from 1997 to 2002. With predictions that the cost per unit of storage will continue to decline at a steep pace and that drive capacity will continue to increase, there is little likelihood that even the largest and fastest growing digital image collections will face capacity or affordability problems for mass storage. Other forms of mass storage, such as optical disk and magnetic tape systems, are also seeing improvements in price and performance, but at a lesser rate than seen with magnetic disk.

The downside of such rapid technological change is equally rapid obsolescence. The need to replace storage systems at short intervals

(perhaps every 3-5 years) cancels out some of the cost benefits. Maintenance budgets for digital imaging systems should anticipate these needs.

Another downside is the confusing proliferation of new technologies. This is particularly true in two areas. One is hardware interfaces for magnetic disks. In order to take advantage of the increasing storage density (and consequent increases in the speed of data retrieval) new hardware interfaces must be developed that can keep up with the drives. Otherwise, there would be no advantage to the faster drives.

The result has been intense competition to increase the data flow rate the interfaces can handle, with each interfaces' stakeholders attempting to one-up the others and win a larger share of the market for high-performance applications. Examples include the move from USB 1.1 to 2.0, the regular introduction of new SCSI standards, and the impending shifts from IEEE 1394a to 1394b and parallel ATA to serial ATA. The new versions offer superior performance, but may cause problems such as incompatibilities (with earlier version devices and the computer system itself), lack of operating system support, and delayed availability of device drivers.

The other area where technology proliferation has caused confusion and headaches for users is formats for compact disk media. This is especially true for the high-density DVD formats, where at least five different formats compete, including three different rewritable formats (DVD-RAM, DVD+RW and DVD-RW). The lack of standardization leads to incompatibilities amongst drives and media and makes it risky for users to settle on any one technology. For more information on this topic, see the [DVD FAQ](#).

A good discussion of many of the trends discussed above can be found [here](#).

Reliability Considerations

Storage reliability takes on many different meanings at different points along the digitization chain. During capture, the concern centers on accurate recording of the bits and the maintenance of fidelity as the files go through various processing steps before arriving in a permanent storage archive. Once ready for delivery, short-term concern shifts to maintaining high availability of important files by minimizing storage system down time, and recovering rapidly from failures. In the long-term, reliability is focused on replacing storage systems before hardware and/or media fails, loses integrity or becomes obsolete.

Overall, the reliability of storage systems has been steadily improving. Almost all storage technologies now have some form of error correction built-in. As storage has become faster and higher in capacity, the extra time and redundant storage necessary to implement error correction has become less of a burden to implement. More and more disk drives have features such as S.M.A.R.T (Self-Monitoring, Analysis, and Reporting Technology) that allow a drive to constantly monitor its own performance and send out an alert if something is starting to go wrong (for example, if the drive's rotational speed is changing, perhaps indicating that a motor or bearing problem is developing).

Larger storage arrays are available with a variety of reliability features. RAID (Redundant Array of Independent or Inexpensive Disks) allows several performance and reliability related configuration options, such as data mirroring, so there is complete redundancy. Some systems can be configured with "hot spares" and "automatic failover" so in the event of a complete drive failure, the contents will automatically be reconstructed on a powered up spare, which then takes its place, all without human intervention. Others permit "hot swapping" of drives, so that replacements can be installed without powering down the entire storage system. As hard drive storage comes down

further in price, it becomes less of a luxury to have empty drives spinning solely for the purpose of taking over in the event of a failure.

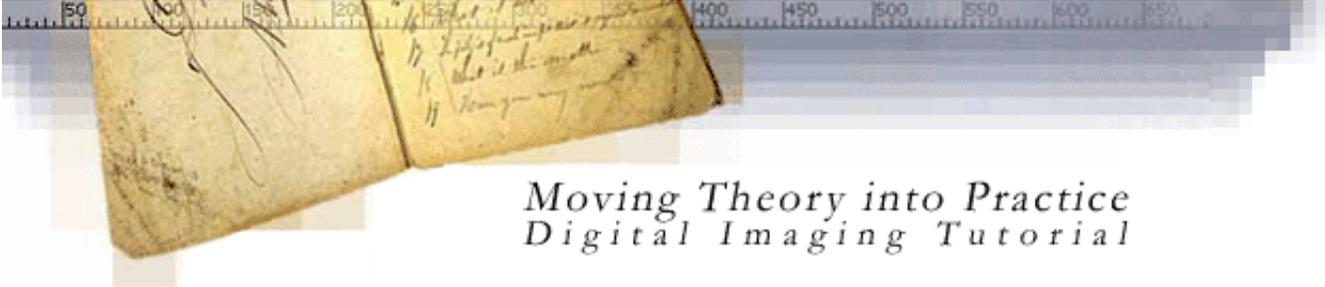
Unfortunately, these impressive features cannot be solely relied upon for protection of data. No technology is fail-safe, and entire storage installations can be destroyed by unpredictable events such as fires, floods, and earthquakes. For this reason, it is generally recommended that all unique data (especially master image files and all associated metadata) be stored on at least two kinds of media, in different physical locations. Often, the choice for secondary storage is removable media such as optical disks or magnetic tapes.

Most removable media can have reasonable life spans (claims vary from 10-100 years), though many of these figures are based on accelerated aging tests, not actual experience. However, improper storage conditions (e.g. high temperature and humidity) can dramatically lower media longevity. Some hard disk drive manufacturers are now claiming MTBFs (Mean Time Between Failure—a statistical measure of the likelihood of drive failure) of 100 years or more. How much attention should you pay to these numbers?

Given that all technologies are subject to failure, and new technologies are being introduced at ever-shrinking intervals, it is possible to get too caught up in concerns over the lifespan of digital storage media. Removable media drives are subject to rapid obsolescence (many formats have come and gone without ever achieving broad market acceptance). As discussed in [Digital Preservation](#), long term survival requires a comprehensive plan that includes attention to media lifespan, storage environment, handling procedures, error detection, backup, disaster response, and monitoring for hardware, media and format obsolescence.

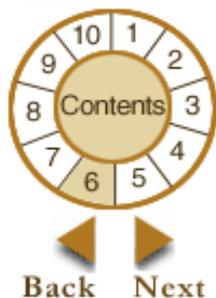
© 2000-2003 Cornell University Library/Research Department





Moving Theory into Practice

Digital Imaging Tutorial



6C. Technical Infrastructure: FILE MANAGEMENT

Key Concepts

[introduction](#)
[keeping track](#)
[image databases](#)
[storage](#)
[storage types](#)
[storage needs](#)

DETERMINING STORAGE NEEDS

Formula to compute storage needs

Basic storage capacity requirements can be estimated by simple calculation:

Total storage needed = # of image files x average file size x 1.25

Example: A Collection of 3,000 text images, each averaging 75KB, would require about 225MB of storage. However, many other factors can increase storage needs. OCR text for the same pages might run 3KB per page, or about 1/25th the space required for the corresponding image file. The number and size of derivative files, as well as whether they're permanently stored or created on the fly could add further to storage requirements. In addition, all storage technologies involve a certain amount of wasted space. The precise amount depends on factors such as the storage technology used, total capacity, partition size, and average file size. Some experimentation may be necessary to determine the approximate percentage of wasted space, but it needs to be taken into account in estimating storage needs. The formula above factors in a generous overage to cover such concerns.

Cost of storage can be approximated as follows:

Total storage cost formula

Total storage cost = total storage needed x cost per unit of storage

This will provide a rough estimate, since it includes only basic drive and media costs. Other costs related to storage could include racks and enclosures, backup power supply, cables, cards, storage management software, etc. . Expect unit storage costs for large systems that include redundancy, high reliability and very high performance to be substantially higher than for routine desktop storage. Check with your systems staff for a more complete picture.

Reality Check

A collection of 10,000 4 x 5-inch transparencies is scanned at 400 dpi, 24-bit color, and then losslessly compressed at a 1.3:1 ratio. Calculate the cost of hard disk storage (at \$2/GB) needed for this collection.

US dollars

Choosing a particular technology can be confusing. For example, consider magnetic disk, where there are many options-ATA (also called EIDE and UDMA), SCSI (wide/narrow, Ultra II/III/160/320, LVD, etc.), Firewire (IEEE-1394), USB, Fibre Channel, etc. The number of choices is growing, with higher performance versions of most of these technologies in the works.

For small collections, both during image capture and delivery, desktop ATA, USB and Firewire storage may be all that's required. The current implementation of ATA (now being called parallel ATA to distinguish it from its successor) has topped out at 133 MB/sec transfer rate and will gradually be replaced by serial ATA, which starts at 150 MB/sec and goes up from there. USB 2.0 and Firewire (IEEE 1394a) both run at about 50 MB/sec, though IEEE 1394b is expected to double that performance.

SCSI is an older storage technology that has, through a continuing series of upgrades, maintained a performance lead over most other technologies. SCSI used to be the choice for high-performance (and high-cost) desktop storage, but while still available, it is less and less common in desktop systems. However, SCSI is still very popular for high-performance networked disk arrays. It is also one of the most important technologies for NAS and SAN installations.

NAS (networked attached storage) can provide large quantities (terabytes) of hard disk storage in a storage appliance that attaches to existing, traditional network servers. NAS is fairly simple to set up and maintain and is usually quite reliable. NAS does suffer from some limitations on expandability and can become difficult to manage in large numbers. NAS is usually based on SCSI drives, though some use ATA.

SAN (storage area network) is primarily for very large installations that require maximum performance and flexibility. SANs allow better integration and sharing of backup facilities, and help to keep traffic between storage devices (e.g. for backup) off of Ethernet networks. However, SANs can be quite complex to establish and often require outside help in order to install the required infrastructure and avoid interoperability problems. SANs operate over a Fibre Channel infrastructure (not Ethernet), using either SCSI or Fibre Channel drives.

The various removable media technologies (both disk and tape) can be considered mostly secondary storage technologies. That is, they are well suited for backup, off-site storage, and storage of material that doesn't need to be immediately accessible. Also, if scanning is outsourced, many vendors return image files on some form of removable media. Despite its low density, CD-R is now a low-cost and widely accepted standard. However, at 650 MB capacity, it may not be suitable for large collections and/or very large files. DVD-R, at up to 9.4 GB/disk for double-sided media is a possible alternative and some manufacturers are predicting 100 year life expectancy. However, if experience with CD-Rs is any indication, media quality can vary substantially

amongst manufacturers, and even from batch to batch. It is also unclear how long any of the DVD formats will remain usable, since higher-capacity, next-generation DVD formats are already on the horizon, and backward compatibility questions are unanswered. Committing to new removable media formats for archival storage remains something of a risky business, and all media should be regarded as temporary .

Computer Considerations

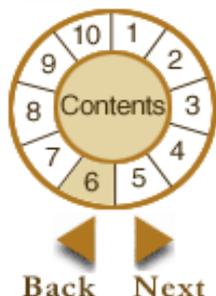
The main consideration will be the level of support provided for the chosen peripheral bus (i.e., SCSI, Firewire) and the computer's ability to keep up with its peripherals. Peripheral bus speeds now routinely exceed those of the computer's internal bus, meaning that some bottlenecks are unavoidable, but attempts should be made to minimize these, otherwise the performance advantage of high-speed storage is lost. Advanced storage architectures such as RAID or Fibre Channel are mostly supported on multi-user platforms such as Windows NT/2000 or Unix/Linux. SCSI is an option on many systems, but won't necessarily come with the base configuration. Make sure that the operating system and system BIOS support the size disk array you need and that there is sufficient space for needed expansion cards.

© 2000-2003 Cornell University Library/Research Department





Moving Theory into Practice Digital Imaging Tutorial



INTRODUCTION

Delivery encompasses the processes of getting digital images and auxiliary files to your users. The most important components are networks and display devices (mainly monitors and printers). It is at this stage in the chain where knowing your users becomes at least as important as knowing your documents.

Unless your digital images are strictly for in-house use, some delivery components are beyond your control. For example, if most users are connected to the Internet with 56Kbps modems, a collection of beautiful 24-bit color images, averaging 500KB in size and taking over two minutes each to download, will frustrate users.

Successful delivery to a mixed audience of in-house and off-site users requires careful advance planning. If resources allow, the best approach is to offer multiple versions of images, taking advantage of greater capacity where it exists, but also supporting low bandwidth connections with lower quality images. Beware of the "lowest common denominator" approach, which may seem egalitarian, but ultimately deprives high-end users of the potential value of your images.

Decisions about file formats, compression ratios, and scaling all will have an impact on delivery. The [Presentation](#) section covers these issues. New and emerging file formats offer multi-resolution capability, providing an alternative to creating multiple versions of the same image.

© 2000-2003 Cornell University Library/Research Department

6D. Technical Infrastructure: DELIVERY

Key Concepts

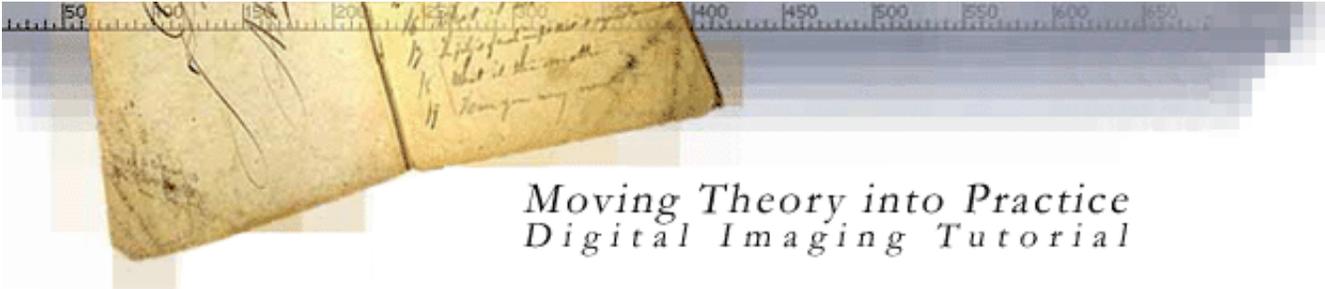
introduction

- networks
- concerns
- speed
- trends
- monitors
- evaluation
- image quality
- printers
- technologies
- evaluation

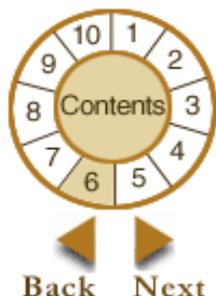


View in:

- ← Spanish
- ← French



Moving Theory into Practice Digital Imaging Tutorial



6D. Technical Infrastructure: DELIVERY

Key Concepts

introduction
networks
 concerns
 speed
 trends
 monitors
 evaluation
 image quality
 printers
 technologies
 evaluation

NETWORKS

Networks are probably the least visible portion of the technical infrastructure. Network cards lie hidden within computers; network hardware is tucked away in machine rooms or communications "closets"; and cable is buried underground, in walls, and/or runs overhead. But nothing can bring a digital imaging initiative to a halt faster than a network that is undersized, too slow, or unreliable. We have already mentioned the need for fast, reliable networks in shuttling files around during creation and file management. A heavily used digital image collection will place even greater demands on your network.

Networking infrastructure decisions are usually made at the institutional level. Large institutions anticipate growth in overall networking need and are prepared to handle significant volumes of network traffic. Small institutions could find that a digital imaging initiative makes demands on an existing network that have implications for the entire organization. Even limiting certain high-intensity network use to traditional low-traffic times may interfere with other activities. A discussion with network administrators about the anticipated network demands should come early in the planning stage.

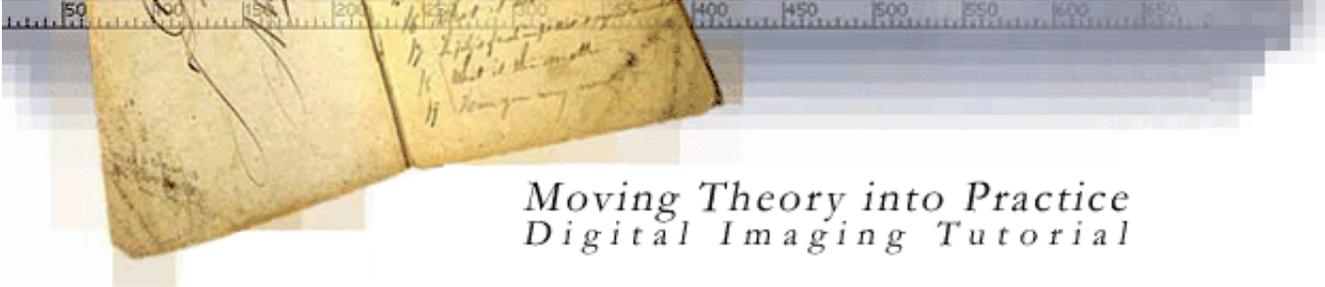
An organization that has used its network primarily for email and some Web surfing may find that its connection to the Internet is completely inadequate for serving up large volumes of digital images. Most Internet connections are asymmetrical, allowing more data to pass downstream (from the Internet) than upstream (towards the Internet). An Internet connection that allows large volumes of data to pass upstream can be quite expensive. Again, network administrators and your network provider need to be consulted if you anticipate significant demand in network traffic.

© 2000-2003 Cornell University Library/Research Department



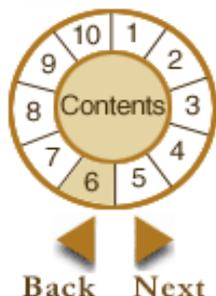
View in:

← Spanish
 ← French



Moving Theory into Practice

Digital Imaging Tutorial



6D. Technical Infrastructure: DELIVERY

Key Concepts

introduction
networks
concerns
 speed
 trends
 monitors
 evaluation
 image quality
 printers
 technologies
 evaluation

NETWORKS: KEY CONCERNS

Compatibility may still be a problem at some institutions. Although the Internet's communications protocol TCP/IP (Transmission Control Protocol/Internet Protocol) has become ubiquitous, legacy protocols are still in use at some institutions. Check with network administrators to be sure your plans are compatible with the existing network.

Reliability is another concern. Network outages reduce productivity and frustrate users. Although the reliability of network hardware has improved in recent years, failures still occur. Some older networks have "grown like topsy" and are a patchwork of different technologies, cabling, and hardware. Fragmented responsibility for network administration also undermines reliability.

Internet Security is an escalating concern (see, for example, [CERT statistics](#), for a summary of the trends). Image servers are subject to security breaches, potentially jeopardizing access by legitimate users or leaving data vulnerable to malicious deletion or modification. System and network administrators may propose remedies such as firewalls, special monitoring software, or requiring authentication of all users. Some security measures can be onerous, either because they require specially skilled personnel to maintain or because they restrict access to your materials more than you would like. Institutional policy may constrain what choices you have, but become familiar with the options.

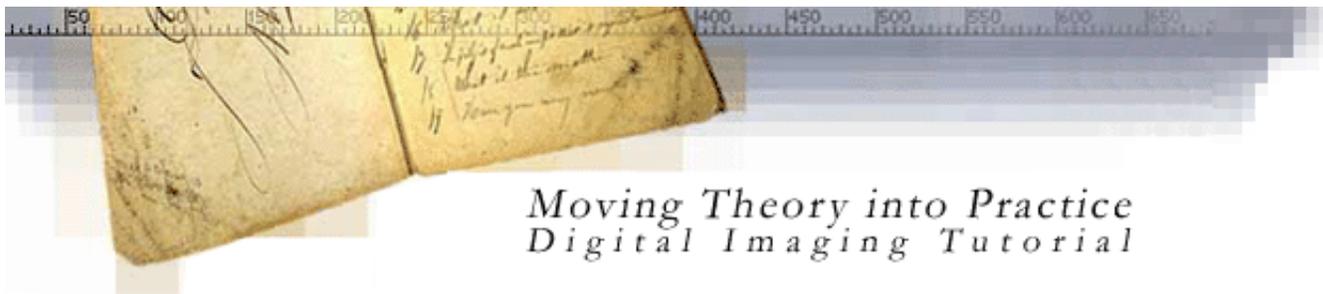
Expense may or may not be a significant issue. At some institutions, you will simply plug your equipment into existing networks and be on your way. But if your undertaking requires a major network upgrade or a new type of connection to the Internet, expense can be substantial. Testing is key, and it may be necessary to scale back your plans to something that the existing infrastructure accommodates.

© 2000-2003 Cornell University Library/Research Department

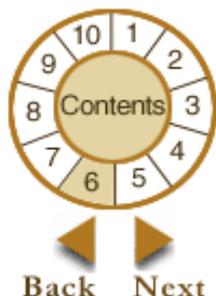


View in:

← Spanish
 ← French



Moving Theory into Practice Digital Imaging Tutorial



6D. Technical Infrastructure: DELIVERY

Key Concepts

- introduction
- networks**
- concerns
- speed**
- trends
- monitors
- evaluation
- image quality
- printers
- technologies
- evaluation

NETWORKS: SPEED

Speed and capacity issues are determined by a multitude of factors. Some are within your control, some are not. As with so many other performance issues, avoidance of bottlenecks is an important objective. Network transmission is governed by the slowest link. Factors affecting network delivery include:

- Carrying capacity (bandwidth) of the local area network
- Bandwidth of the institution's Internet connection
- Speed and capacity of the network server
- Read speed and data transfer rate of storage devices
- Image file size
- User demand at any particular time
- Amount of competing network traffic (at all network levels)
- Speed of any "on-the-fly" processing steps
- Time required for authentication and other security checks
- Capabilities of the end user's computer, including:
 - CPU speed
 - RAM/disk caching
 - Video subsystem performance
 - Speed of Internet connection

There are a variety of network technologies that might be encountered between an image server and the ultimate recipient. The following table presents some of the more important ones, in declining order by speed, in MB/second.

Table: Network Data Transfer Rates

Network Type	Speed in MB/sec
OC-192	1250
OC-48 (Abilene backbone)	300
1000BaseT Ethernet	125
vBNS (NSF/MCI backbone)	77.8
FDDI	12.5
100BaseT Ethernet	12.5
DS-3 (T-3)	5.6
10BaseT Ethernet	1.25
Cable modem (downstream)	.2-5
ADSL (downstream)	.19 -1
DS-1 (T-1)	.19
ISDN (home use)	.018
v.90 modem	.007

The fastest of these networks are used only for major Internet backbones. The next tier are local area networks, while the slowest are consumer services. The speeds given are theoretical maximums, which are rarely, if ever, encountered in real installations. Note that the fastest network is more than 175,000 times faster than the slowest.

Once one knows the transmission speed of a network it is possible to compute the approximate time it will take a file of any particular size to make its way across. Use this formula:

Formula on Transmission Speed

t (time in seconds) = number of megabytes in file ÷ (transmission speed (in MB/sec) x .8)

Example:

A 1 MB file can theoretically make it across a 10BaseT Ethernet network in $1 / (1.25 \times .8) = 1$ second. The .8 takes into account that 80% of rated speed is about the best one can expect to realistically encounter. Since most networks share bandwidth amongst users, the more traffic they handle, the lower the overall transmission speed. When saturated, performance can fall dramatically.

Reality Check

Using the formula for transmission speed and the table of network data transfer rates shown above, calculate the least amount of time it will take a 1 MB file to be accessed via 100BaseT Ethernet and v.90 modem (round to the nearest tenth second).

seconds (Ethernet speed)

seconds (v.90 modem speed)

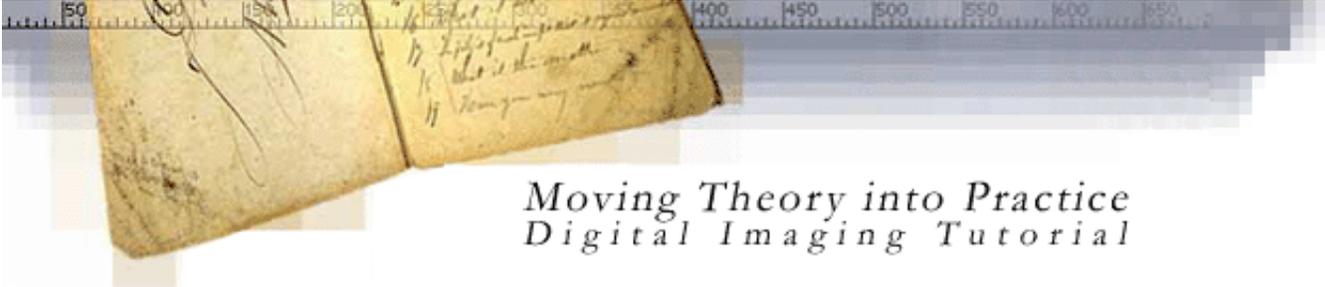
How much faster is 100BaseT Ethernet?

times faster



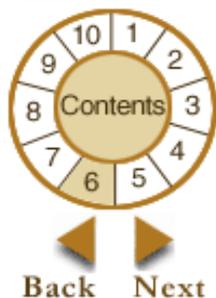
View in:





Moving Theory into Practice

Digital Imaging Tutorial



6D. Technical Infrastructure: DELIVERY

Key Concepts

introduction

networks

concerns

speed

trends

monitors

evaluation

image quality

printers

technologies

evaluation

NETWORKS: TRENDS

Efforts are continuing to increase network throughput. One strategy is to improve speeds on existing networks through new forms of compression or presentation. However, the need to reduce file size to speed delivery may be a limited-term concern as broad bandwidth information pipelines and wireless high speed data transfer capabilities are developed in the next 5-10 years to support research, electronic commerce, and entertainment. The increasing deployment of cable modem and DSL services to residences will ease bandwidth concerns at the user's end. The potential of digital television, in particular High Definition Television (HDTV), to provide new and different kinds of information to a broad range of users—including access to digitized cultural resources—is tantalizing. Current US Federal Communications Commission (FCC) rules require all analog broadcasts to be phased out by the end of 2006. Beginning with Internet 2, the U.S. government is funding efforts to build the Next Generation Internet (NGI) to link research labs and universities to high-speed networks that are 100 to 1,000 times faster than the current Internet. Designed to handle high volumes of information, the NGI will make access to digital image files easy and high quality audio and moving image transfer practical.

Computer Considerations

Most of the requirements for a network server have already been touched upon. Such machines are very resource hungry, especially if heavily used. Keeping a network server optimally tuned requires a skilled systems administrator. Perhaps the best advice is not to skimp on personnel for managing networks and servers.

Reality Check

User feedback on your Web site indicates a large number of complaints about how long it takes to view images. What first step(s) should be taken to respond?

Install more or larger servers

Upgrade to a higher bandwidth Internet connection

Reduce the resolution or bit-depth of your images

Gather more information from the complainants

Moving Theory into Practice Digital Imaging Tutorial



Back Next

6D. Technical Infrastructure: DELIVERY

Key Concepts

introduction

networks

concerns

speed

trends

monitors

evaluation

image quality

printers

technologies

evaluation

MONITORS

If storage devices are among the most rapidly evolving technologies, monitors are among the slowest. Though the price/performance ratio of monitors has greatly improved, even the most technically advanced products still require significant compromises.

A monitor will serve as the user's window into your digital image collection. As in the case of networks, sometimes that monitor is under your control, sometimes not. When it is, the opportunity is there to minimize the compromises inherent in current monitor technology. Beyond choosing a quality product, characteristics such as resolution setting, calibration, external lighting, and even how often the screen is dusted can all affect perceived image quality.

When the user is off-site, you can provide recommended settings, but the image confronting the user may still be a far cry from your expectations. Delivering images off-site may require accommodation. If most users have 640 x 480 size displays, images sized for comfortable display on a 1280 x 1024 display will not have the intended impact.

Table: Common Desktop Settings (PCs)

VGA	640 x 480
SVGA	800 x 600
XGA	1024 x 768
SXGA	1280 x 1024
UXGA	1600 x 1200

However, not all monitor deficiencies can be corrected by buying the right product or adjusting your display settings. Limitations related to color fidelity, image completeness, and dimensional fidelity have to be addressed during file processing. These issues are considered in [Presentation](#).

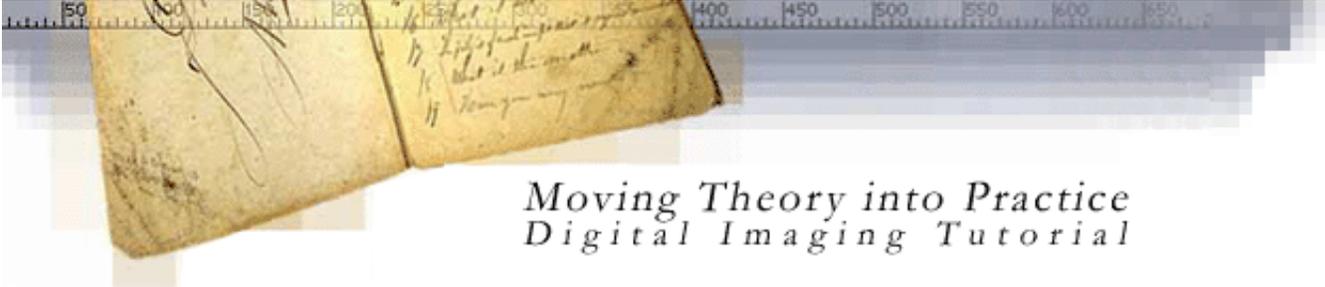
© 2000-2003 Cornell University Library/Research Department



View in:

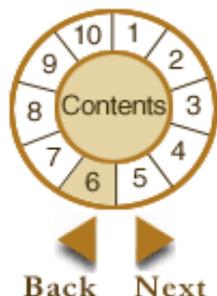
← Spanish

← French



Moving Theory into Practice

Digital Imaging Tutorial



6D. Technical Infrastructure: DELIVERY

Key Concepts

introduction

networks

concerns

speed

trends

monitors

evaluation

image quality

printers

technologies

evaluation

MONITORS: EVALUATION

The following factors should be integral to the evaluation process:

- Image quality
- Size
- Ease of use and sophistication of adjustment and calibration controls
- Fitness to task
- Cost

The market for monitors consists mainly of two very different technologies: CRT (Cathode Ray Tube)-based devices and LCD (Liquid Crystal Display)-based devices. CRTs are built around what in electronics terms is an ancient technology, but still dominate the market, especially for graphics intensive work. However, major improvements in the performance and affordability of LCDs have significantly narrowed the gap between these two technologies.

Here's a rundown of how modern CRTs and TFT (Thin Film Transistor) LCDs compare in several important functional areas, as of late 2002 (a nice chart summarizing the characteristics of the two technologies can be found at <http://www.tomshardware.com/display/02q1/020114/lcd-03.html>):

Image quality

- CRTs typically have better contrast, more accurate color rendition, a wider range of colors, more satisfactory viewing from off-axis (i.e. not looking straight on). They are better at displaying fast changing images, as in movies or animations. CRTs can display a quality image at a variety of pixel dimensions (LCD quality falls off considerably when not used at primary design resolution, called native resolution). CRTs are not subject to dead or stuck pixels, in which spots on the display may be permanently dark or bright.
- LCDs typically have brighter images, better focus, less distortion, absence of convergence problems and no flicker.

Ergonomics

- LCDs tend to excel across the board in ergonomic factors, being smaller, lighter, and producing less heat and other harmful emissions.

Economics

- CRTs are less expensive to buy, especially for large size displays (17" and above).
- LCDs are cheaper to operate (lower power consumption), and may have a lower total cost of operation.

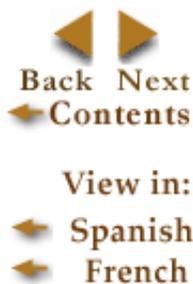
Flat panel technologies such as LCD have been under development for decades, and have improved substantially in that time. As noted above, the

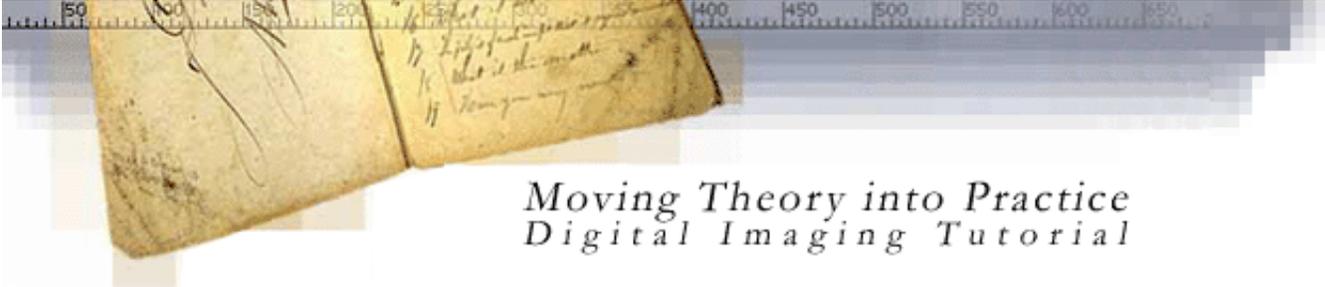
current generation of LCDs (called TFT or active matrix) can now outperform CRT monitors in many areas. For many routine uses not requiring pixel dimensions above 1024 x 768, the small price premium can be easily justified by other advantages.

However, the accurate display of digital images, especially continuous tone images with large pixel dimensions, remains one of the few areas where CRTs still offer superior performance. In deciding whether to deploy LCDs for the viewing of digital image collections, careful consideration should be given to whether image presentation will suffer from the loss of color fidelity and dynamic range, and whether that loss is of significance to users. Side-by-side comparisons may be the best way to judge. In addition, given that many end users are now purchasing LCDs for personal use, it is prudent to assess how your images will look to users employing them.

Whether a monitor is being used for image editing, quality control, or for end user delivery, the more complete the user controls provided, the greater the ability to optimize performance. Monitors used to come with only brightness and contrast controls. Modern monitors allow considerably more fine tuning. Check a monitor's specifications to determine if the settings critical to your intended use are user-controlled.

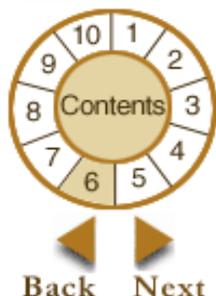
© 2000-2003 Cornell University Library/Research Department





Moving Theory into Practice

Digital Imaging Tutorial



6D. Technical Infrastructure: DELIVERY

Key Concepts

introduction

networks

concerns

speed

trends

monitors

evaluation

image quality

printers

technologies

evaluation

MONITORS: DETERMINANTS OF IMAGE QUALITY

- Screen resolution
- Screen size
- Dot pitch
- Refresh rate
- Bit depth
- Monitor performance
- Video card performance

Obviously, the best display device can't correct image problems resulting from the use of inadequate equipment or poor decision-making in the capture or processing steps. Limitations in the color handling of operating systems and image viewing software (particularly Web browsers) can also affect final image quality.

Assuming, however, that considerable effort has been invested in image capture, it makes sense to choose a display device that will show off your images to best effect. Not all displays are created equal. Amongst CRT monitors, for example, the shadow mask design excels for text, while the aperture grille design produces better images, though the thin horizontal lines near the top and bottom of the screen may be annoying.

LCDs can also vary substantially in quality. Look for displays that are uniformly bright, high-contrast and that are viewable across a wide angle without falloff of brightness or color distortion.

When purchasing monitors, 17" CRTs should now be considered the minimum for most viewing purposes, though 19" monitors have dropped so much in price that they should be preferred unless desk space is limited. 21" monitors supporting up to 1920 x 1440 pixels have also declined in price and might be considered, despite their size and weight, if image completeness and/or dimensional fidelity are critical concerns. On the production side of digital imaging, larger monitors reduce eyestrain for staff doing quality control work.

15" LCDs offer almost the same viewing area as 17" CRTs, and may serve well for images that can be displayed in full at 1024 x 768 pixel dimensions and do not have demanding color accuracy requirements.

Dot pitch refers to the distance between the phosphor dots that the CRT's electron beam excites to create an image. That distance determines the finest detail that the CRT can resolve. Better CRTs have dot pitch specifications in the .24-.25mm range.

Other aspects of image quality are determined by the video card that drives the monitor. Many monitor specifications reflect the assumption that a sufficiently capable video card is used.

Most modern CRTs can support multiple resolutions, though one or two will be optimal, depending on the monitor's size. The "sweet spot" for 17" monitors is in the 800 x 600 to 1024 x 768 range. For 19" monitors, it's 1024 x 768 to 1280 x 1024. Most monitors support higher resolutions, though the highest resolution usually sacrifices some image quality and often results in text that is too small to read comfortably. LCDs are much more limited in this area, producing a truly quality image at only one resolution.

Refresh rate refers to how frequently the entire image is updated. If the refresh rate is too low, the viewer detects a subtle flicker in the image. Flicker-free images require a refresh rate of at least 75 Hz, though rates as high as 85 Hz improve viewing on some monitors. Excessively high refresh rates can also compromise image quality. To check for flicker, use your peripheral vision to view an all white screen. Due to the way an image is created on an LCD, refresh rate isn't a factor in the viewability of still images.

Bit depth support determines the number of colors or grays a monitor can reproduce. Virtually all CRTs and video cards now support 24- or 32-bit display at the highest pixel dimensions. A few LCD monitors are still limited to 18-bit display (rather than the typical 24-bit) and thus cannot produce as wide a range of colors.

Computer considerations

Beyond what's already been discussed, other aspects of the computer related to display center around additional hardware enhancements. Specialty hardware can provide accelerated compression and decompression and/or file format conversion. A second video card can, on certain platforms, support a second monitor on the same computer. This can be useful in situations when even the largest monitors don't provide adequate screen real estate. For example, all the menus and palettes for an image-editing package can be displayed on one monitor, leaving the second one for just the image. Or metadata can be input on one screen, with the image occupying another.

Some LCD monitors will take either analog or digital input signals. Running with digital input avoids the need to convert digital to analog (and back again) and may result in a slightly better image. Be aware that in order to utilize an LCD with digital inputs, the computer driving it must have a video card with digital outputs (usually a DVI port) and the correct cable must be used.

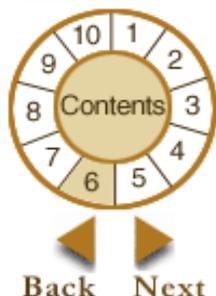
© 2000-2003 Cornell University Library/Research Department





Moving Theory into Practice

Digital Imaging Tutorial



6D. Technical Infrastructure: DELIVERY

Key Concepts

introduction
networks
concerns
speed
trends
monitors
evaluation
image quality
printers
technologies
evaluation



View in:

← Spanish
← French

PRINTERS

As long as computers are bulky, display devices low resolution and hard on the eyes, battery technology in its infancy, and the communications infrastructure bound by cables, the desire to create printouts from digital images will endure. However, the costs of actually making high-resolution images available online, in formats that can be printed via a number of platforms and a range of printers, should not be underestimated. Before making promises to deliver print-quality images in a networked environment, verify that the technical infrastructure is up to the task, and consider the additional storage costs associated with online access.

PRINTER TECHNOLOGIES

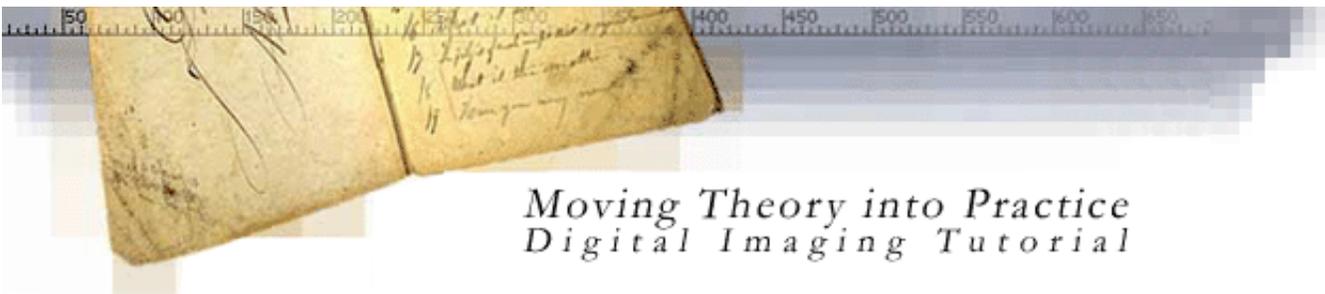
Today, black and white printing is dominated by two technologies: Inkjet printers, which squirt liquid ink through tiny nozzles onto the paper, and laser printers, which use a light source to create charges on a photoconductive drum, allowing it to attract dry ink particles (toner) that fuse onto paper. Inkjet printers have become quite inexpensive, but are slower than lasers and generally not designed for high volume printing. High end production laser printers can produce well over 100 pages per minute at 600 dpi.

Both technologies have been adopted for color. Color inkjet printers come in 3- and 4-color models. Color lasers are much more expensive, both for initial purchase and for the cost of consumables. Color inkjets and lasers are both substantially slower than their black and white counterparts, color inkjets average about 5 pages per minute for text and 1 page per minute for full page graphics. Color lasers are faster, averaging 12 pages per minute for text, and 2 pages per minute for full page graphics.

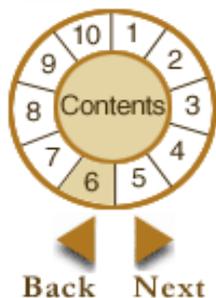
Several other technologies are available for color printing. These include dye sublimation, solid ink, and thermal wax. Dye sublimation is especially significant in that it can produce true continuous tone color printouts, though it is extremely slow and requires special coated paper.

For larger scale color printing, Electronics for Imaging makes the Fiery line of print servers, which enable digital color photocopiers and digital presses to be networked to serve as high volume, high quality color printers. The resulting combination is called a copier-printer. Resolution is generally 400 dpi maximum, but is supported for whatever paper sizes the copier normally uses.

© 2000-2003 Cornell University Library/Research Department



Moving Theory into Practice Digital Imaging Tutorial



PRINTERS: CRITERIA FOR EVALUATION

- Resolution and dot spacing
- Color reproduction
- Tonal representation
- Image enhancement capabilities
- Document size supported
- Single versus double-sided printing (simplex/duplex)
- Media supported (plain paper, coated paper, transparencies, envelopes)
- Speed and capacity
- Page description languages and raw image formats supported
- Networking capabilities
- Cost

6D. Technical Infrastructure: DELIVERY

Key Concepts

introduction
networks
concerns
speed
trends
monitors
evaluation
image quality
printers
technologies
evaluation

Computer Considerations

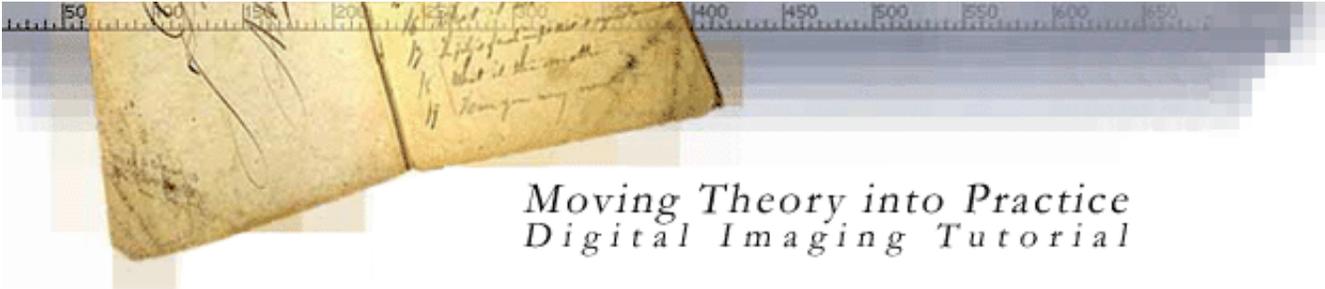
Not all printers are supported on all computing platforms, so check for compatibility. Also, make sure software drivers are available for the specific operating system version you are using. Check for availability of the correct networking or direct connection capability. Print accelerators can offload some of the burden of printing from the computer's CPU.

© 2000-2003 Cornell University Library/Research Department

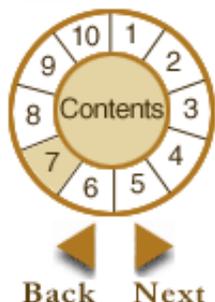


View in:

← Spanish
← French



Moving Theory into Practice Digital Imaging Tutorial



INTRODUCTION

Using the Web to make retrospective resources accessible to a broad public raises issues of image quality, utility, and delivery at the user's end. User studies have concluded that researchers expect fast retrieval, acceptable quality, and complete display of digital images. This leads cultural institutions to confront a whole host of technical issues that do not exist in the analog world.

7. Presentation

Key Concepts

introduction

formats/compression

web browsers

network

scaling

monitors

image quality

guidelines

additional reading

Technical Links Affecting Display-

- File format and compression used
- Web browser capabilities
- Network connections
- Scaling routines and programs
- End user computer and display capabilities

© 2000-2003 Cornell University Library/Research Department



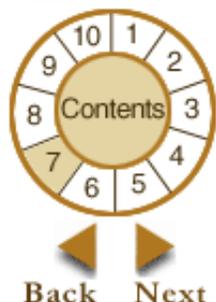
View in:

← Spanish

← French



Moving Theory into Practice Digital Imaging Tutorial



7. Presentation

Key Concepts

introduction
formats/compression
web browsers
network
scaling
monitors
image quality
guidelines

additional reading

FILE FORMATS AND COMPRESSION

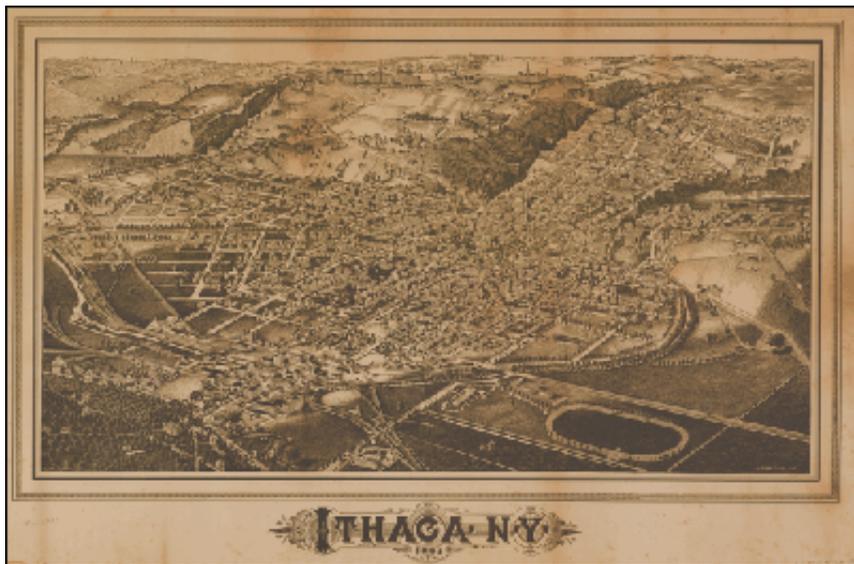
Factors in choosing a file format for display include the following:

- Bit depths supported
- Compression techniques supported
- Color management
- Proprietary vs. standard file format
- Technical support (Web browser, user computer and display capabilities)
- Metadata capability
- Fixed vs. multi-resolution capability
- Additional features, e.g., interlacing, transparency

Although there is a multitude of file formats available, the [Table on Common Image File formats](#) summarizes important attributes for the eight most common image formats in use today.

Despite interest in finding alternative formats for master files, TIFF remains the defacto standard. For access images, GIF and JPEG files are the most common. PDF, while not technically a raster format, is used extensively for printing and viewing multi-page documents containing image files. PDF also offers a zooming feature that supports variant views of an image. PNG has been approved by the [World Wide Web Consortium](#) (W3C) for Web use, and as browser support for the format becomes more complete, PNG may replace GIF for network access. (See an [RLG DigiNews FAQ](#) on the future of PNG.)

As larger and more complex images are being intended for Web access, there is increasing interest in file formats and compression techniques that support multi-resolution capabilities, such as [FlashPix](#), [LuraWave](#), [JTIP](#) and wavelet compression, such as MrSID from [LizardTech](#) or Enhanced Compressed Wavelet from [ER Mapper](#). [JPEG 2000](#) also utilizes wavelet compression and supports multi-resolution capabilities. [DjVu](#) is a recently-developed format optimized for scanned documents. It offers efficient compression of both bitonal images (using the JBIG2 variant, JB2), as well as of full color images, using wavelet compression. Unfortunately, all of these formats require users to download and install plug-ins in order to view them on the Web.



Resolution on Demand: Several new file formats and compression techniques allow users to zoom in by clicking on a section to view at a higher resolution. Click on the image above to view an example of a Zoom Feature.

The compression technique used and level of compression applied can affect both speed of delivery and resulting image quality. The [Table on Compression](#) summarizes important attributes for common compression techniques. AIIM offers a questionnaire ([AIIM TR33-1998](#)) to assist in choosing a compression method to match user requirements.

The following Table compares file sizes resulting from using various compression programs on a 300 dpi, 24-bit image of an 8.45 x 12.75-inch color map.

Table: File Size and Compression Comparison

Compression Type	File Size	Compression Ratio
Uncompressed TIFF	28.4 MB	--
TIFF-LZW	21.2 MB	1:1.34
GIF (8 bit)	4.0 MB	1:6
JPEG-low	10.4 MB	1:2.7
JPEG-high	1.2 MB	1:24
PNG	20.8 MB	1:1.37



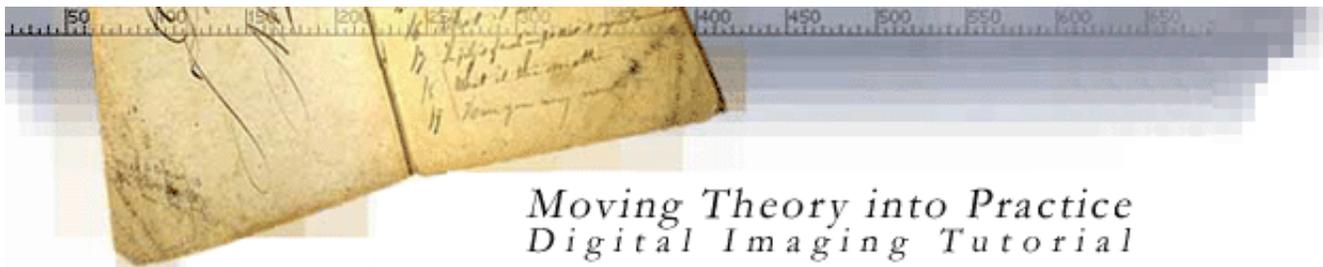
View in:

- ➡ Spanish
- ➡ French

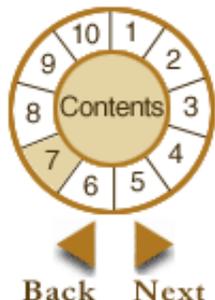


Effects of Lossy Compression on Text/Line Documents: Click on these images to view a close-up version. The left image is saved in GIF format, the one on the right as JPEG. The compression artifacts are most evident around the sharp-edged characters in the enlarged version of the right image. Courtesy of Bob Rosenberg, The Edison Papers Project.

© 2000-2003 Cornell University Library/Research Department



Moving Theory into Practice Digital Imaging Tutorial



7. Presentation

Key Concepts

- introduction
- formats/compression
- web browsers**
- network**
- scaling
- monitors
- image quality
- guidelines

additional reading

WEB BROWSER CAPABILITIES

The Web supports few raster file formats: JPEG, GIF, and incomplete support for PNG. Other formats require use of a specialized viewer, such as a plug-in, applet, or some other external application. This limitation tends to dampen use as it places more demand on the user's end. In some circumstances, the value of the format is sufficiently compelling to overcome user resistance, as is the case with PDF files. Adobe lessens user constraints by supplying a browser plug-in with its PDF reader. If the stand-alone Acrobat Reader is already available when the browser is installed, most will self-configure to launch it when a PDF file is encountered. Some institutions convert non-supported formats or compression schemes on the fly to ones that are Web-supported (e.g., wavelet to JPEG) in response to user request.

NETWORK CONNECTIONS

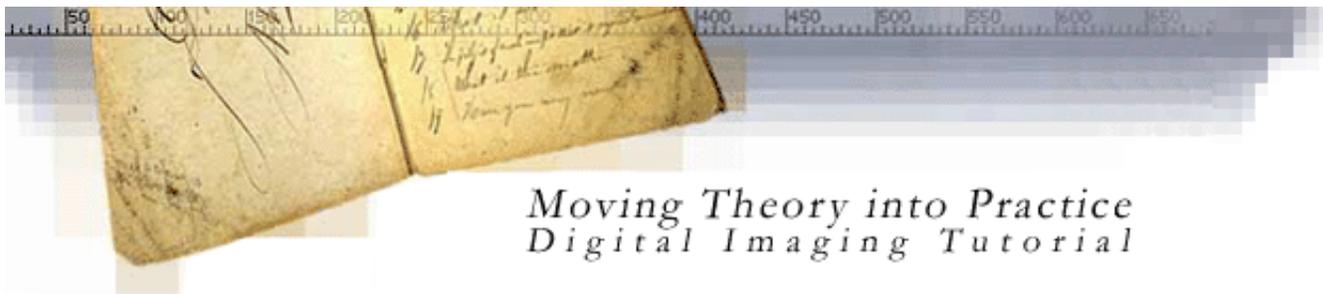
Users probably care most about speed of delivery, as noted [earlier](#). Several variables control access speed, including the file size, network connections and traffic, and the time to read the file from storage and to open it on the desktop.

© 2000-2003 Cornell University Library/Research Department

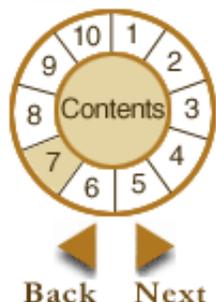


View in:

- ← Spanish
- ← French



Moving Theory into Practice Digital Imaging Tutorial



SCALING ROUTINES AND PROGRAMS

Institutions have constrained file size by reducing resolution, bit depth, and/or by applying compression. The goal is to speed delivery to the desktop without compromising too much image quality. Scaling refers to the process of creating access versions from a digital master without having to rescan the source document. The program and scripts used for scaling will affect the quality of the presentation. For instance, scaling can introduce moiré in illustrations, such as halftones, when resolution is reduced without attention paid to screen interference.

7. Presentation

Key Concepts

- introduction
- formats/compression
- web browsers
- network
- scaling**
- monitors
- image quality
- guidelines

additional reading

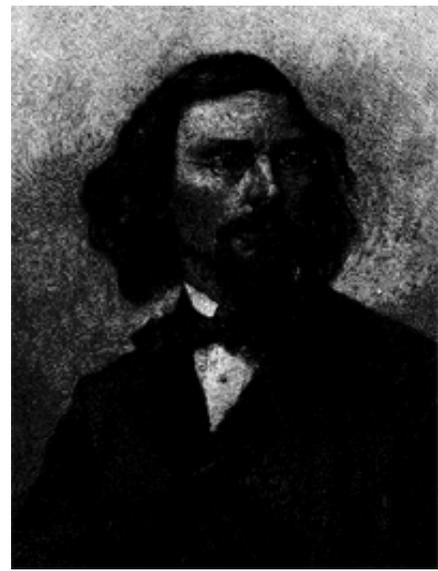


Effects of Scaling on Image Quality: The left image is scaled by using a blur filter, resizing, and reducing the bit depth. The right image is scaled without the use of blur filter, resulting in moiré.

Scaling programs are also used to reduce the bit-depth of an image and different processes result in substantially different quality.

◀ ▶
Back Next
← Contents

View in:
← Spanish
← French



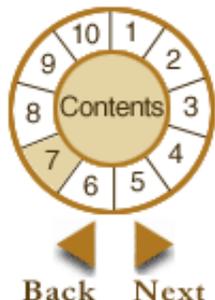
Effects of Scaling Programs: Note the image quality difference between these two derivatives created by different conversion software.

Several Web sites listed at the end of this section provide helpful information on scaling programs, optimizing graphics, and choosing file formats to enhance image quality. Consider also whether the program offers batch processing and user-defined scripting capabilities, and get a sense of the total processing times. Minutes spent on one image quickly add up to days, weeks, and months, depending on the size of your image collection.

© 2000-2003 Cornell University Library/Research Department



Moving Theory into Practice Digital Imaging Tutorial



MONITOR CAPABILITIES

User satisfaction with on-screen images will depend on the capabilities of display systems. In addition to speed of delivery, users are interested in image quality (legibility and color fidelity adequate to a task); full display of images on screen; and to a lesser degree accurate representations of the dimensions of original documents. Unfortunately, given current monitor technology, meeting all these criteria at the same time is often not possible.

7. Presentation

Key Concepts

- introduction
- formats/compression
- web browsers
- network
- scaling
- monitors**
- image quality
- guidelines

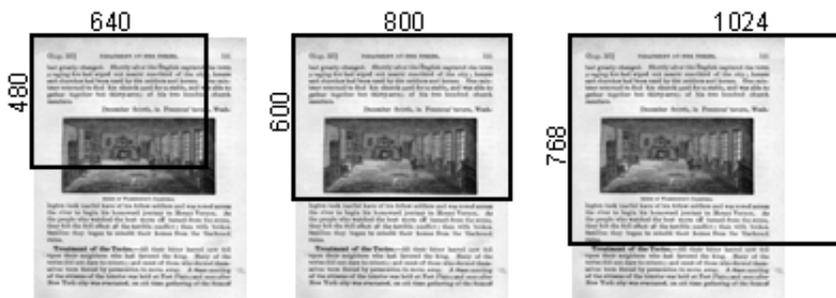
additional reading

Screen size and pixel dimensions

In contrast to scanners and printers, current monitors offer relatively low resolution. Typical monitors support desktop settings from a low of 640 x 480 to a high of 1,600 x 1,200, referring to the number of horizontal by vertical pixels painted on the screen when an image appears.

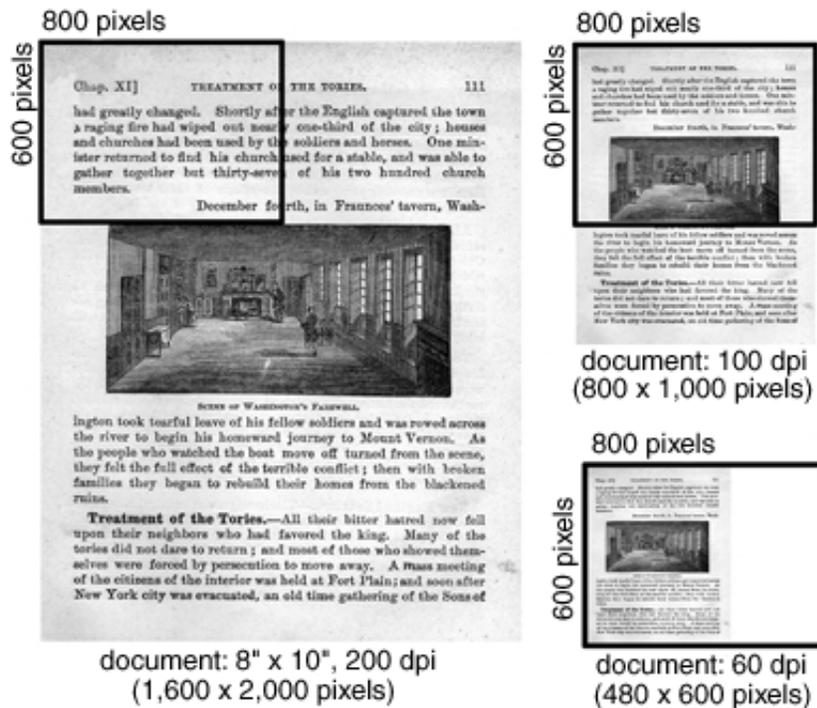
The amount of an image that can be displayed at once depends on the relationship of the image's pixel dimensions (or dpi) to the monitor's desktop setting. The percentage of an image displayed can be increased several ways: by *increasing* the screen resolution and/or by *decreasing* the image resolution.

Increasing screen resolution. Think of the desktop setting as a camera viewfinder. As the monitor setting dimensions increase, more of an image may be viewed. The figure below illustrates the viewing area for an image at various monitor settings.



Increasing Screen Resolution: Viewing area comparison for a 100 dpi (original document size 8"x10") image displayed at different monitor settings. The pixel dimension for the image is 800 x 1,000.

Decreasing image resolution. One can also increase the amount of an image displayed by reducing the resolution of the image through scaling. This figure illustrates the relationship of a monitor's desktop setting at 800 x 600 to an image scaled to various resolutions.



Balancing Legibility and Completeness: Displayed at 200 dpi on a 800 x 600 monitor, one can only see a small portion of the page (left). At 60 dpi, the whole page is fully displayed, but at the expense of legibility (bottom-right). Scaling the image to 100 dpi offers a compromise by maintaining legibility and limiting scrolling to one dimension (top-right).

You can calculate the percent of display if you know the following variables:
 1) document dimensions and image dpi, or pixel dimensions of image, and 2) desktop setting.

Calculating percentage displayed

Enter document dimensions in inches: (width) and (height) and enter resolution of image: dpi

OR

Enter image pixel dimensions horizontal vertical

Screen size	% displayed	% width displayed	% height displayed
640 x 480			
800 x 600			
1024 x 768			

Dimensional Fidelity

At times, it may be important to represent an image on-screen at the actual size of the original scanned document. This can only be achieved when the digital image resolution equals the monitor's resolution (dpi). The Blake Archive Project has developed a Java applet, called the [Image Sizer](#), for representing images at the actual size of the original.

Reality Check

If representation of dimensional fidelity on-screen is important, what is the likely impact on image quality?

Image quality will not be affected, only the size of the image

Image quality will often increase, as the document will be presented at its native size with details fully presented

Image quality will often decline as the screen resolution is typically lower than the digital image resolution



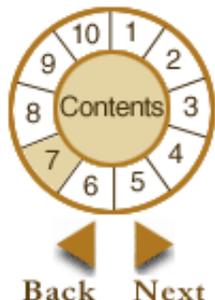
© 2000-2003 Cornell University Library/Research Department

View in:





Moving Theory into Practice Digital Imaging Tutorial



ON-SCREEN IMAGE QUALITY

We have noted the effects of various scaling programs and routines on image quality. Two other factors should be considered as well:

1. Is the resolution of the image sufficient to ensure legibility or support detailed study of an image?
2. Can color and tone be conveyed effectively?

7. Presentation

Key Concepts

- introduction
- formats/compression
- web browsers
- network
- scaling
- monitors
- image quality**
- guidelines

additional reading

Legibility of Text

As we have seen, legibility and completeness often conflict. For instance, when an 8 x 10-inch text-page scanned at 200 dpi is scaled to fit on a monitor set to 800 x 600, over 90% of the pixels have been discarded. The image fits, but the text may no longer be readable.

Cornell has developed a benchmarking formula for the display of text-based materials that correlates image quality, resolution, and the required level of detail:

Benchmarking On-Screen Legibility Formula

$$\text{dpi} = \text{QI} / (.03\text{h})$$

$$\text{QI} = \text{dpi} \times .03\text{h}$$

$$\text{h} = \text{QI} / (.03\text{dpi})$$

In the formula, **dpi** refers to the resolution of an image (not to be confused with the monitor's dpi); **h** refers to the height of the smallest character in the original (in mm); and **QI** refers to levels of legibility (*Note: if h is measured in inches, multiply by 25.4 before using formula*). This formula presumes that bitonal images are presented with 3 bits or more of gray and that filters and optimized scaling routines improve image presentation. Using this formula, establish your own levels of acceptable quality. Cornell benchmarks readability at a QI of 3.6, although 3.0 often suffices for cleanly produced text, particularly if it has been scanned in grayscale or color.

Reality Check

A 4 x 5-inch text page contains characters as small as 1mm in height and has been scanned at 600 dpi-1bit. At what resolution can you scale this image for on-screen presentation and still maintain character legibility (at a QI of 3.6)?

dpi

At that resolution, what percentage of your document could be displayed on a monitor set to 800 x 600?

%

Color and Tone

Presenting [color and tone](#) depends on monitor and system capabilities. Color appearance is most problematic since it is affected by different browsers, [monitor](#) settings, and transfer between color spaces.

Several Web sites provide useful information on Web palettes for access (see [additional reading](#)). Some recommend using file formats such as PNG, which supports both a Web-safe palette and sRGB, designed to ensure color consistency across platforms. Others are including grayscale/color targets with the images to enable the end user to adjust the color. Still others, including the [National Archives](#) and the [Denver Public Library](#), have developed monitor adjustment targets to help users calibrate their monitors.



Monitor Adjustment: This portion of the [NARA Monitor Adjustment Target](#) illustrates the full range of tones that a computer monitor can represent when set to 256 or more colors (8 bits or higher). The shades should be just distinguishable from one another. Courtesy of the National Archives and Records Administration.

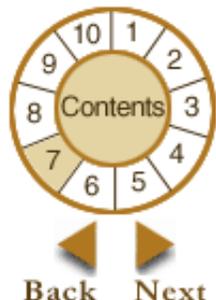
© 2000-2003 Cornell University Library/Research Department

Back Next
← Contents

View in:
← Spanish
← French



Moving Theory into Practice Digital Imaging Tutorial



GUIDELINES

To see what some institutions recommend for display, click on:

[Table: Representative Institutional Requirements for Access](#)

7. Presentation

Key Concepts

introduction
formats/compression
web browsers
network
scaling
monitors
image quality
guidelines

additional reading

ADDITIONAL READING

Anne R. Kenney, "Digital Benchmarking for Preservation and Access," in *Moving Theory into Practice: Digital Imaging for Libraries and Archives*, Mountain View, CA : Research Libraries Group, 2000; pp. 24-60
<http://www.rlg.org/preserv/mtip2000.html>

John Price-Wilkin, "Enhancing Access to Digital Image Collections: System Building and Image Processing," in *Moving Theory into Practice: Digital Imaging for Libraries and Archives*, Mountain View, CA: Research Libraries Group, 2000; pp. <http://www.rlg.org/preserv/mtip2000.html>

Color and Tone

Lynda Weinman, "The Browser Safe Color Palette,"
<http://www.lynda.com/hex.html> and <http://the-light.com/netcol.html>. Both of these have links to other useful sites on browser palettes and other Web graphics information.

Digital Images in Multimedia Presentation, "Image Manipulation and Preparation," <http://www.tasi.ac.uk/advice/using/dimpmanipulation.html>

The Bandwidth Conservation Society, <http://www.tbcr.org/>

Scaling Routines and Programs

Patrick J. Lynch and Sarah Horton, "Web Style Guide,"
<http://info.med.yale.edu/caim/manual/contents.html>

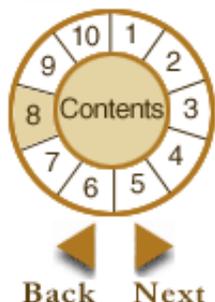
Wotsit's Graphic File Formats <http://www.wotsit.org/search.asp?s=graphics>

Anne R. Kenney and Louis H. Sharpe II, *Illustrated Book Study: Digital Conversion Requirements of Printed Illustrations*, 1999,
<http://lcweb.loc.gov/preserv/rt/illbk/ibs.htm>

TASI, "DIMP WWW - Image Incorporation Case Study,"
http://www.tasi.ac.uk/advice/using/web_case.html



Moving Theory into Practice Digital Imaging Tutorial



DEFINITION

The goal of digital preservation is to maintain the ability to display, retrieve, and use digital collections in the face of rapidly changing technological and organizational infrastructures and elements. Issues to be addressed in digital preservation include:

- Retaining the physical reliability of the image files, accompanying metadata, scripts, and programs (e.g., make sure that the storage medium is reliable with back-ups, maintain the necessary hardware and software infrastructure to store and provide access to the collection)
- Ensuring continued usability of the digital image collection (e.g., maintain an up-to-date user interface, enable users to retrieve and manipulate information to meet their information needs)
- Maintaining collection security (e.g., implement strategies to control unauthorized alteration to the collection, develop and maintain a rights management program for fee-based services)

8. Digital Preservation

Key Concepts

definition

challenges
technical strategies
organizational
strategies

additional reading

Although this section is one of the last of the tutorial, issues associated with longevity need to be discussed at the onset of any imaging initiative. Many of the issues that become impediments to long-term preservation are rooted in early decisions centering on selection and conversion. Digital preservation decisions and strategies should be developed as an integral part of a digital imaging initiative as many decisions will be closely coupled with an institution's long-term retention plans.

© 2000-2003 Cornell University Library/Research Department

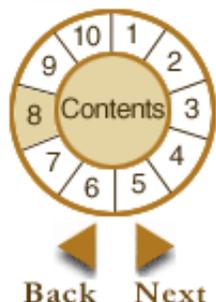


View in:

← Spanish
← French



Moving Theory into Practice Digital Imaging Tutorial



8. Digital Preservation

Key Concepts

definition

challenges

technical strategies

organizational
strategies

additional reading

WHY IS DIGITAL PRESERVATION SO CHALLENGING?

Challenges are multi-faceted and can be grouped into two categories:

Technical Vulnerabilities

- Storage media, due to physical deterioration, mishandling, improper storage, and obsolescence.
- File formats and compression schemes, due to obsolescence or over-reliance on proprietary and unsupported file and compression formats.
- Integrity of the files, including safeguarding the content, context, fixity, references, and provenance.
- Storage and processing devices, programs, operating systems, access interfaces, and protocols that change as technology evolves (often with limited backward compatibility).
- Distributed retrieval and processing tools, such as embedded Java scripts and applets.

Organizational and Administrative Challenges

- Insufficient institutional commitment to long-term preservation
- Lack of preservation policies and procedures
- Scarcity of human and financial resources
- Varying (and asynchronous) stakeholder interests in the creation, maintenance, and distribution of digital image collections
- Gaps in institutional memory due to staff turnover
- Inadequate record keeping and administrative metadata
- Evolving nature of copyright and fair-use regulations that apply to digital collections

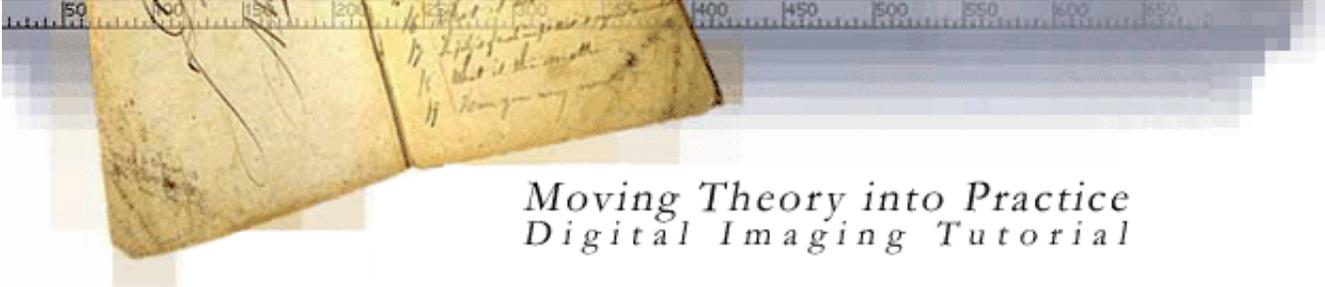
© 2000-2003 Cornell University Library/Research Department



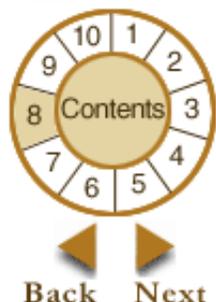
View in:

← Spanish

← French



Moving Theory into Practice Digital Imaging Tutorial



TECHNICAL STRATEGIES

Enduring Care should be seen as an ongoing strategy for monitoring the wellbeing of digital resources. Vigilant management of the collection includes housing images and accompanying files in secure, reliable media and locations; storing and handling media according to industry guidelines to optimize their life expectancy; and implementing periodic and systematic integrity checks and backups.

8. Digital Preservation

Key Concepts

definition

challenges

technical strategies

organizational
strategies

additional reading

Refreshing involves copying content from one storage medium to another. As such it targets only media obsolescence and is not a full-service preservation strategy. An example of refreshing is copying a group of files from CD-ROMs to DVDs. Refreshing should be seen as an integral part of an enduring care policy.

Migration is the process of transferring digital information from one hardware and software setting to another or from one computer generation to subsequent generations. For example, moving files from an HP-based system to a SUN-based system involves accommodating the difference in the two operating environments. Migration can also be format-based, to move image files from an obsolete file format or to increase their functionality.

Emulation involves the re-creation of the technical environment required to view and use a digital collection. This is achieved by maintaining information about the hardware and software requirements so that the system can be reengineered.

Technology Preservation is based on preserving the technical environment that runs the system, including software and hardware such as operating systems, original application software, media drives, and the like.

Digital Archeology includes methods and procedures to rescue content from damaged media or from obsolete or damaged hardware and software environments.



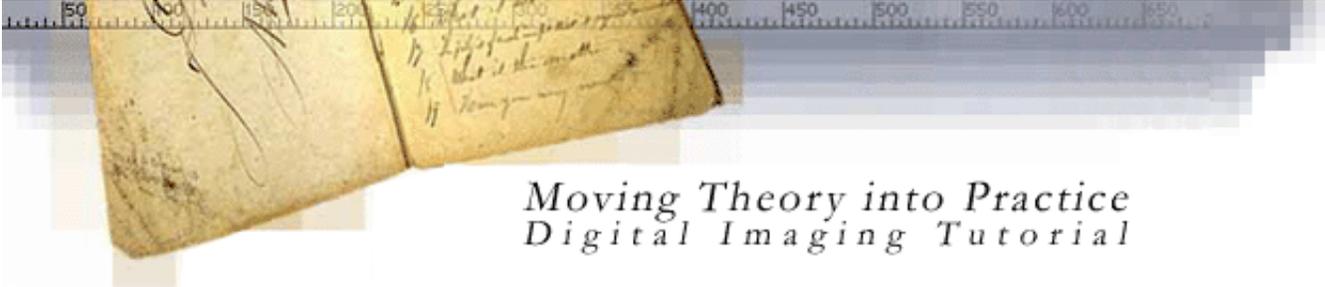
View in:
← Spanish
← French

Reality Check

Which strategy(ies) assumes the copying of data files from one storage medium to another in response to media obsolescence or deterioration?

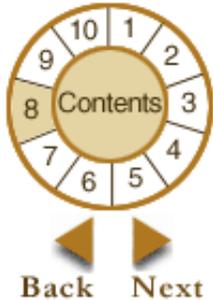
- Refreshing
- Migration
- Emulation
- Technology preservation
- Digital archeology

© 2000-2003 Cornell University Library/Research Department



Moving Theory into Practice

Digital Imaging Tutorial



8. Digital Preservation

Key Concepts

definition
 challenges
 technical strategies
organizational strategies

additional reading

ORGANIZATIONAL STRATEGIES

Technical solutions alone are insufficient to ensure the longevity of digital resources. A holistic approach is called for that recognizes the interdependencies between technical and organizational components. Among issues to be addressed in such a strategy are staffing and training needs, financial requirements, criteria for re-selection, and preservation metadata needs.

Although it is useful to examine each issue in detail, successful solutions require the integration of administrative and technical considerations. For example, an institution may have a well-developed strategy for day-to-day maintenance of image collections that codifies how to monitor, test, and refresh files. However, unless there is a concomitant financial and administrative plan that outlines how to staff and finance these activities over time, the maintenance plan may not succeed in the long-term. Likewise, having dedicated and qualified staff will not suffice unless there is a technical appreciation for the lifecycle management of digital assets. Effective management of digital collections will require institutions to develop and follow a business plan for evaluating long-term preservation and access requirements, identifying costs and benefits, and assessing risks.

Examples of initiatives that support such an approach include:

- [Arts and Humanities Data Service](#) (AHDS) in the UK is developing a decision-making tree to be used in cost-benefit analysis of digital preservation options (Administrative and Managerial Frameworks Preservation Management of Digital Materials).
- Cornell's [Risk Management of Digital Information](#) project examined the risks involved in file format migration (e.g., TIFF 4.0 to TIFF 6.0) and developed an assessment tool to evaluate the risks involved in migration. This tool also helps to assess institutional readiness for any digital preservation action.

The following initiatives are examples of promising, practical approaches to digital preservation:

[OAIS](#) (Open Archival Information System) reference model provides a framework for long-term digital preservation and access, including terminology and concepts for describing and comparing archival architectures. Both the NEDLIB and Cedars 1 projects have adopted the OAIS reference model as a basis for their explorations.

[Cedars 1](#) (CURL Exemplars in Digital Archives) project aims to produce strategic frameworks for digital collection management policies, and to promote methods appropriate for long-term preservation of different classes of digital resources, including the creation of appropriate metadata.

[Networked European Deposit Library](#) (NEDLIB) is a collaborative project of

European national libraries to build a framework for a networked deposit library. Among the key issues it explores are archival maintenance procedures and the link between metadata requirements and preservation strategies.

[PANDORA](#) (Preserving and Accessing Networked Documentary Resources of Australia) project has successfully established an archive of selected Australian online publications, developed several digital preservation policies and procedures, drafted a logical data model for preservation metadata, and outlined a proposal for a national approach to the long-term preservation of these publications.

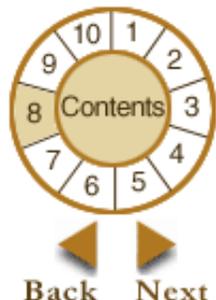
© 2000-2003 Cornell University Library/Research Department


Back Next
← Contents

View in:
← Spanish
← French



Moving Theory into Practice Digital Imaging Tutorial



8. Digital Preservation

Key Concepts

definition
challenges
technical strategies
organizational
strategies

additional reading

ADDITIONAL READING

Oya Y. Rieger, "Projects to Programs: Developing a Digital Preservation Policy," in *Moving Theory into Practice: Digital Imaging for Libraries and Archives*, Mountain View, CA : Research Libraries Group, 2000; pp. 135-152.
<http://www.rlg.org/preserv/mtip2000.html>

N. Beagrie and D. Greenstein, A Strategic Policy Framework for Creating and Preserving Digital Collections. Version 5.0. Arts and Humanities Data Service, July 1998/July 2001. <http://ahds.ac.uk/strategic.htm>

Task Force on Archiving of Digital Information, *Preserving Digital Information: Report of the Task Force on Archiving of Digital Information*, (Washington, DC: Commission on Preservation and Access, 1996.
<http://www.rlg.org/ArchTF/index.html>

National Library of Australia, Preserving Access to Digital Information.
<http://www.nla.gov.au/padi/>

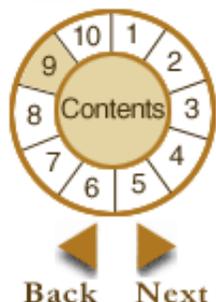
© 2000-2003 Cornell University Library/Research Department



View in:
← Spanish
← French



Moving Theory into Practice Digital Imaging Tutorial



9. Management

Key Concepts

introduction

- project life cycle
- in-house vs. outsource
- in-house facility
- project budgets
- communication
- project monitoring
- looking beyond

additional reading

INTRODUCTION

Institutions inaugurating digital imaging initiatives must address managerial issues. These can be characterized variously, but they all boil down to correlating resources and processes with project goals. Project *goals*, such as enhancing access or promoting efficiencies must be translated into project *deliverables*, such as digital image files, accompanying metadata, and Web-accessible databases. A manager will have a greater chance of completing the project successfully if she has a hand in defining project goals and deliverables. The figure below places goals and deliverables at the center of project management. Radiating out from them are institutional resources, including collections, personnel, finances, space, time, and technical capabilities. These elements will enhance or constrain digitization efforts. The outer circle represents the processes or steps that encompass digital imaging initiatives.



Management Wheel: The figure demonstrates the organic nature of digital imaging, with interdependencies connecting goals, resources, and processes.

Among responsibilities that fall to project managers are the following:

- Setting realistic timelines, objectives, and expectations

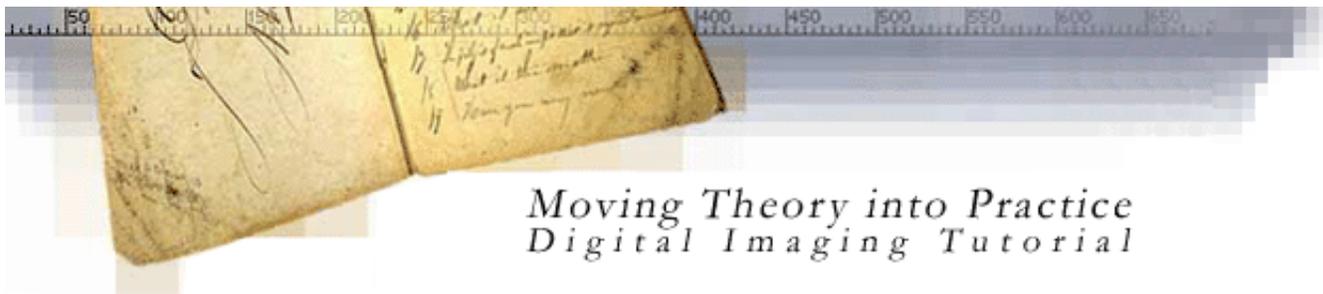


- Determining the best approach for accomplishing project goals
- Developing and defending budgets
- Facilitating communication among project participants, including outside vendors
- Monitoring production, quality, and costs
- Looking beyond project's end

© 2000-2003 Cornell University Library/Research Department

View in:





Moving Theory into Practice Digital Imaging Tutorial



Back Next

9. Management

Key Concepts

- introduction
- project life cycle**
- in-house vs. outsource
- in-house facility
- project budgets
- communication
- project monitoring
- looking beyond

- additional reading

SETTING REALISTIC TIMELINES, OBJECTIVES, AND EXPECTATIONS

It is the manager's responsibility to recognize and plan for a project's life cycle, which encompasses the following stages:

- *Pre-project activities*, including identifying goals and methodologies, securing resources and institutional commitment
- *Ramping up*, the stage from project initiation to first scanning batch
- *Production*, where the greatest productivity occurs in the middle of this phase
- *Project wind down*, a time to conclude the effort and for dealing with problems that have been set aside
- *Post-project activities*, principally associated with mainstreaming maintenance responsibilities for digital products

Recognizing the life cycle of a project enables a manager to develop a project timeline, where the beginning and end are clearly defined. In between, the manager must marshall resources to create project deliverables on time and within budget. Project steps and workflow must be characterized, and the several Web sources listed at the end of this section provide useful information that may be adapted to your particular circumstances.

Timeline development is facilitated if the institution has experience with similar efforts or can undertake a pilot phase where time and resources associated with project steps can be quantified. Creating a base level timeline using a software program capable of generating a Gantt chart such as Microsoft Project enables the manager to note process sequences and dependencies that will be affected by unanticipated delays in production. A common mistake is to overestimate production capabilities, especially in the early phases of a project. These tools facilitate project monitoring, enabling managers to respond more effectively to bottlenecks, competing requirements, and the like.

© 2000-2003 Cornell University Library/Research Department



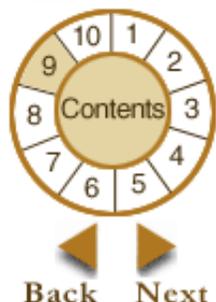
View in:

- ← Spanish
- ← French



Moving Theory into Practice

Digital Imaging Tutorial



DETERMINING THE BEST APPROACH: OUTSOURCING VS. IN-HOUSE PROGRAMS

There are pros and cons to outsourcing or creating in-house capabilities for digital imaging efforts. Even when the decision is made to outsource certain functions, the institution must support many aspects of the [digitization chain](#) as defined in the technical overview. For instance, if digitization is outsourced, an institution still needs to establish an in-house inspection program.

Outsourcing

9. Management

Key Concepts

introduction

project life cycle

in-house vs. outsource

in-house facility

project budgets

communication

project monitoring

looking beyond

additional reading

Advantages

- Cost containment and limited risk; institution pays for deliverables, usually a set price/image, which facilitates project planning and budgeting
- Costs typically lower than in-house figures, although prices vary widely
- Vendors can handle large volume and high throughput
- Expertise, training, technology obsolescence costs absorbed by vendor
- Broad range of options and services available, including imaging, metadata creation, enhancements, processing, encoding, derivative creation, printing, storing and backup, database development

Disadvantages

- Institution removed one step from imaging functions; services most often performed offsite or even off shore
- Vulnerability due to vendor instability
- Hard sell for existing products and services that are typically designed for business market
- Vendor inexperience with needs of cultural institutions
- Lack of standards and best practices with which to define requirements or negotiate for services
- Challenges in communication, from RFP development to contracting, to production and quality requirements
- Security, handling, transportation issues

Outsourcing is viable if an institution has a good understanding of the near- and long-term goals of an imaging initiative, and can fully specify imaging, metadata, and derivative requirements; locate reliable vendors; evaluate products and services; adopt policies and procedures for various functions; and define institutional and vendor responsibilities. Some service providers offer a [questionnaire](#) or checklist for institutions to clarify project requirements as well as determine products and costs.

The [Colorado Digitization Project](#) provides a list of US service providers. You can also search the [AIIM Products and Services Guide](#) by service required. Vendors of film scanning and COM recording are listed in an [RLG DigiNews FAQ](#).

Note: if you know of similar Web-accessible lists for other countries, [drop us a line](#).

A detailed Request for Proposals (RFP) must be developed, which clearly outlines content and requirements. A good starting point is the [RLG Guidelines for Creating a Request for Proposal for Digital Imaging Services](#). In addition, the [Library of Congress](#) has posted its RFPs for digital conversion.

The evaluation process must represent a consistent and well-documented methodology for three important reasons:

1. to assist the institution in choosing the right service provider
2. to justify the selection to institution purchasing agents especially if the lowest bidder is not chosen
3. to defend the choice to losing bidders and protect the institution against potential suits for unfair practice.

In-House Approach

Advantages

- Learn by doing
- Define requirements incrementally rather than up front
- Retain direct control over entire range of imaging functions
- Provide for security, proper handling, and accessibility to materials
- Ensure primacy of library/archives requirements
- Maintain consistent and high quality assurance requirements

Disadvantages

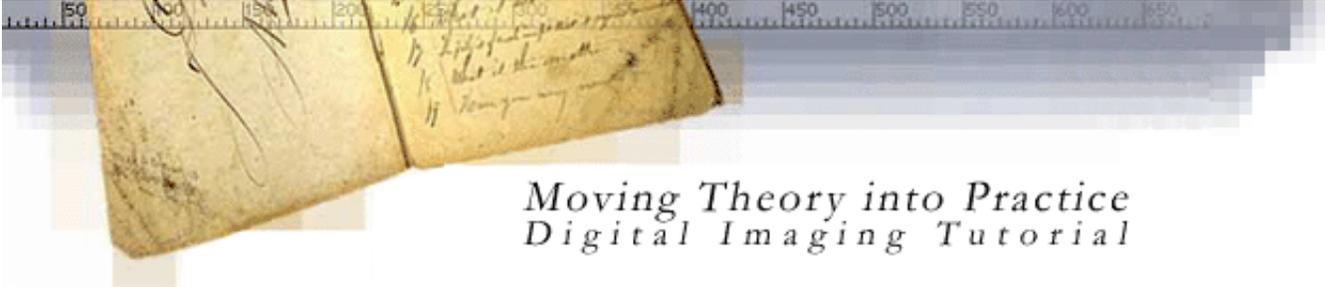
- Large investment and ramp-up time
- No set per-image cost
- Institution pays for expenses instead of products, including costs of downtime, training, and technology obsolescence
- Limited production capabilities and facilities
- Range of staffing expertise required



View in:

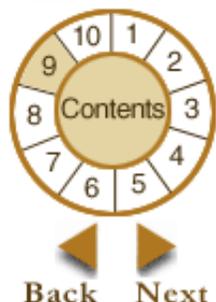


© 2000-2003 Cornell University Library/Research Department



Moving Theory into Practice

Digital Imaging Tutorial



9. Management

Key Concepts

introduction
 project life cycle
 in-house vs. outsource
in-house facility
 project budgets
 communication
 project monitoring
 looking beyond

additional reading

ESTABLISHING AN IN-HOUSE FACILITY

Establishing an in-house facility requires an institution to support the full digitization chain with appropriate staff, space and facilities, equipment and supplies, and to absorb time and expense associated with ramping up. Many of these resources will have to be provided, albeit to a lesser degree, when outsourcing some or all of production.

Staff will be needed for the following tasks: identification, selection, preparation, digitization, metadata creation, quality control, cataloging, data loading, systems support, and management. Depending on the institutional configuration and the extent of the imaging program, staff will also need to be hired to develop and maintain the image database and Web delivery system. Staffing the project requires decisions on the types, levels, and numbers of staff, the ratio of managers to workers to students, a program to train staff, and the identification of an administrative home. Sample job descriptions for imaging staff can be located by searching the archives of various mailing lists, such as [Conservation DistList](#), [DIGLIB](#), and [IMAGELIB](#) or by searching employment postings for various professional organizations, such as [Library and Information Technology Association](#), [ALA](#), or the [Society of American Archivists](#).

Dedicated Facilities must be identified and furnished to support the imaging effort. Consider hiring a consultant or choose a value-added retailer who can advise on facility requirements as well as hardware/software components and system integration. Allow 75 to 150 square feet per person, depending on the work to be performed. There must also be adequate and secure workspace to prepare and store materials for scanning (e.g., tables, shelves). The Library of Congress calculates table space at 6 times the size of the largest object to be digitized to promote safe handling and ordering of materials. Consider too equipment "footprints," especially if a staff member is responsible for operating more than one machine (e.g., multiple scanners). The facility must also provide the requisite communications—phone/data lines, LAN connections, and UPS (uninterrupted power supply) protection. It must support appropriate environmental controls, including proper HVAC, air filtration, and controlled lights (overhead and ambient). Scanning equipment and lights can raise temperatures, especially in confined areas. Consider workflow in designing the room configuration. Equipment includes the requisite hardware, software, and supplies to support the [digitization chain](#):

- Hardware
 - Scanning devices
 - High resolution monitors
 - Workstations
 - Peripherals
 - Servers and storage devices
 - Printers
- Software to support the following
 - Operating system, networking/server/graphics support, programming packages

- Scanning, image editing, viewing, color management, quality control
- Derivative creation
- File management, workflow management
- Indexing, OCRing, structuring
- Database management system

Other equipment and supplies

- Copy stands/cradles/lights/lenses
- Quality control equipment and supplies
- Routine office supplies
- Storage media, paper, ink cartridges
- Documentation, technical manuals, reference publications

Written procedures for handling, scanning, metadata creation, quality control, and other functions should be developed and consistently applied. The Library of Congress [National Digital Library Program](#), The [Technical Advisory Service for Images](#), and the [Arts & Humanities Data Service](#) in the UK provide various papers, reports and procedural outlines that can serve as models. See also [A Feasibility Study for the JISC Imaging Digitisation Initiative](#).

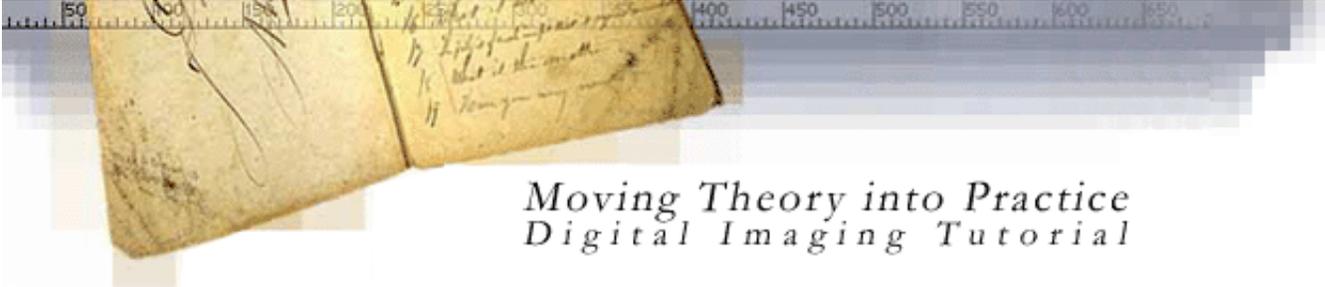
Reality Check

Correlate the Activity with the Project Phase (select more than one, where appropriate):

Project Phase	Activity
(1) pre-project (2) ramp up (3) production (4) conclusion (5) post-project	hiring staff purchasing equipment grant writing developing procedures locating a vendor contracting for services cataloging digital products resolving remaining problems preparing final report transferring custody of digital products to other units quality control

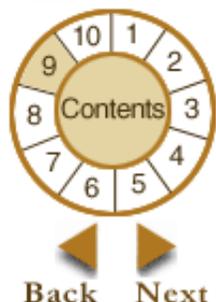


View in:
 ← Spanish
 ← French



Moving Theory into Practice

Digital Imaging Tutorial



9. Management

Key Concepts

introduction
 project life cycle
 in-house vs. outsource
 in-house facility
project budgets
 communication
 project monitoring
 looking beyond

additional reading

DEVELOPING AND DEFENDING PROJECT BUDGETS

A major concern of management is to project costs and develop budgets representing the full range of costs:

- *Direct expenses*, including staff salaries/wages and benefits; management overhead; equipment/software; supplies; services and contracts; maintenance, licenses, copyright clearances and use fees, and communication fees; and replacement costs.
- *"Ramp-up" costs*, which can be considerable, especially for first time projects or those involving untested methods. These costs include developing RFPs, establishing workflow processes and documentation, system configuration, training and other expenses incurred by the institution prior to project launch. Typically not supported by outside funders, but they should be documented. Factor in a "generous curve" from project inception to production.
- *Contingency* covers unanticipated expenses; not traditionally an "allowable" expense with US funding agencies, but increasingly recognized by UK and European funders. Contingency will vary from project to project, depending on complexity, staff experience, and size of effort.
- *Indirects/overhead*, often a negotiated rate, which includes space, utilities, services, and general and administrative support, calculated on the total direct costs. Cornell's federally negotiated rate for 1999-2003 is 57%. In some countries a value-added tax (VAT) is calculated on the direct costs.
- *Cost-share*. Institutions are often required to or voluntarily cover some of the costs associated with imaging projects, such as all or part of indirects. Cost-share is a real expense and should be calculated.
- *"Hidden" costs*. Institutions typically support imaging projects out of other projects or programs. Under-reporting such contributions conveys a false sense of project costs.

There is no consensus on what it costs to create digital image files, much less maintain them and make them accessible. Available figures vary tremendously with the types of material being scanned, the image-conversion and metadata requirements, hardware/software used, and the range of functions covered in the calculations. Some institutions provide cost figures, percentage breakdowns, and future estimates. Note, however, that their cost assessments will differ. Specific costs must be calculated based on local conditions. RLG maintains a [Worksheet for Estimating Digital Reformatting Costs](#) that details various components to include in estimating budgets for image creation. See Additional Reading for reports and articles on costing.

The following sites provide information on digital funding sources, although the information may be slightly dated:

United States

- [Colorado Digitization Project](#)

- [Amigos Library Services](#)

Europe & the United Kingdom

- [Content creation in Europe](#)
- [TASI \(UK\)](#)

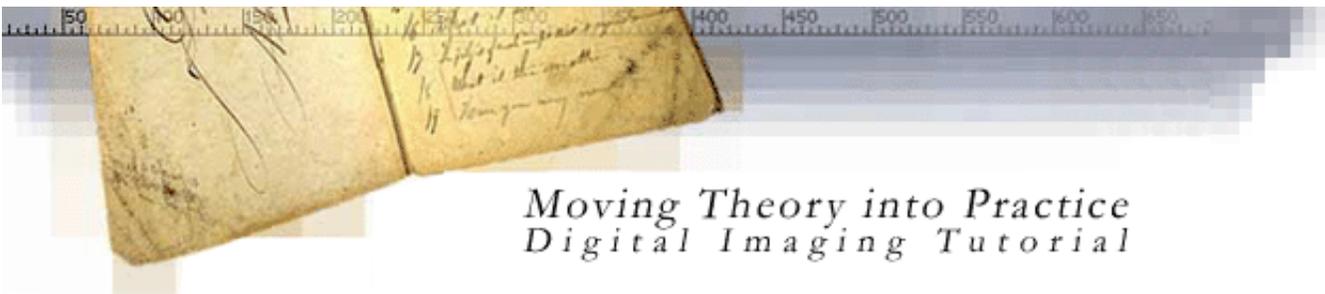
Australia

- [National Library of Australia, Community Heritage Grants](#)

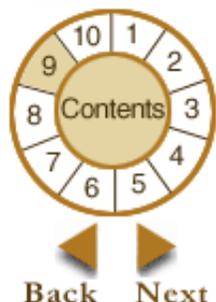
Note: we are interested in referencing lists of funding sources for other countries; if you can help, [drop us a line](#).

© 2000-2003 Cornell University Library/Research Department





Moving Theory into Practice Digital Imaging Tutorial



9. Management

Key Concepts

introduction
 project life cycle
 in-house vs. outsource
 in-house facility
 project budgets
communication
project monitoring
 looking beyond

additional reading

FACILITATING COMMUNICATION

Digital imaging projects will involve staff beyond those specifically assigned to the project. Regularly scheduled all-staff meetings provide a useful means for maintaining open communication. Decisions made, issues raised, and resolution of concerns should all be documented in writing, through meeting minutes or summaries, with process/product decisions ultimately finding their way into procedural manuals or guidelines. Narrow concerns may best be addressed in meetings that include only affected staff members. However, conflict resolution or changes in process should be reported to the broader group, as decisions made may have ramifications for the work of others. Regular communication is critical when dealing with outside service providers, especially if quality or production is adversely affected. It may be wise to formalize communication points by building in conference calls periodically or at critical junctures in the production schedule.

PROJECT MONITORING

Project tracking establishes a system to gather and analyze information about the source materials and digital files as well as performance, quality, and costs. A consistent methodology is critical when outsourcing any part of the project, as it provides the most direct way of ensuring contract compliance. Some service bureaus are encouraging institutions to develop joint production-tracking systems. For those functions performed in-house, project monitoring is the principal means for improving efficiency, effectiveness, and product reliability. Information gathered in one project can be used to project costs and workflow procedures in subsequent ones. Project monitoring involves data gathering and assessment on production processes, source materials and digital products, and project administration.

© 2000-2003 Cornell University Library/Research Department

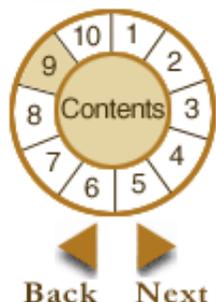


View in:

← Spanish
 ← French



Moving Theory into Practice Digital Imaging Tutorial



9. Management

Key Concepts

introduction
 project life cycle
 in-house vs. outsource
 in-house facility
 project budgets
 communication
 project monitoring
looking beyond

additional reading

LOOKING BEYOND PROJECT'S END

Digital imaging projects do not just end. Provisions must be made for monitoring the health of digital files and for ensuring their continuing accessibility. Projects may be undertaken with temporary staff and outside funds, but as the project winds down, digital products will become the responsibility of the institution. Project management extends to facilitating a shift from projects to mainstreamed production. This is easier said than done, especially when a project has been viewed as outside the core institutional mission. Little hard evidence suggests that digitization results in institutional economies, and thus it will compete with core programs for institutional support. Some simple truths regarding digital imaging projects:

- pilot projects are not production projects
- it's easy to initiate a digital project; it's hard to implement an on-going program
- a series of digital projects does not constitute a digital program
- maintaining digital collections is harder than you might initially think (see [Digital Preservation](#))

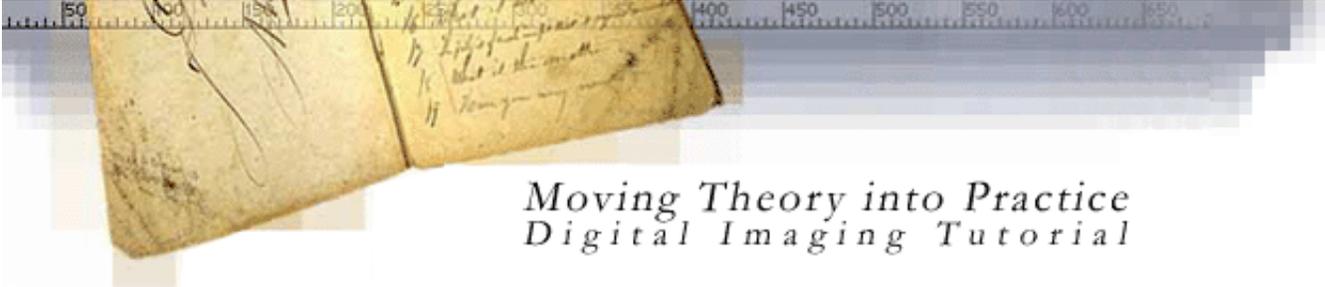
Libraries and archives should view digital conversion as a means to other goals, not an end in itself. If institutions are convinced of the value of digitization, their efforts may have a greater chance of becoming sustainable when projects turn into programs. A transition strategy for mainstreaming digital imaging initiatives is presented in the final chapter of [Moving Theory into Practice: Digital Imaging for Libraries and Archives](#).

© 2000-2003 Cornell University Library/Research Department

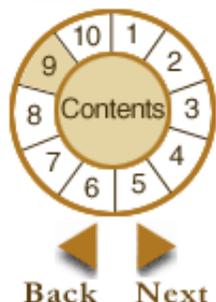


View in:

← Spanish
 ← French



Moving Theory into Practice Digital Imaging Tutorial



9. Management

Key Concepts

- introduction
- project life cycle
- in-house vs. outsource
- in-house facility
- project budgets
- communication
- project monitoring
- looking beyond

additional reading

ADDITIONAL READING

Anne R. Kenney, "From Projects to Programs: Mainstreaming Digital Imaging Initiatives," in *Moving Theory into Practice*, Mountain View, CA : Research Libraries Group, 2000; pp. 153-175.

<http://www.rlg.org/preserv/mtip2000.html>

Budgeting

Maria Bonn, "Benchmarking Conversion Costs: A Report from the Making of America IV Project," *RLG DigiNews*, Oct. 2001.

<http://www.rlg.org/preserv/diginews/diginews5-5.html#feature2>

Steve Puglia, "The Costs of Digital Imaging Projects," *RLG DigiNews*, Oct. 1999.

<http://www.rlg.org/preserv/diginews/diginews3-5.html#feature>

Simon Tanner and Joanne Lomax Smith, "Digitisation: How Much Does It Really Cost?" (paper for the Digital Resources for the Humanities 1999 Conference, Sept. 12-15, 1999).

<http://heds.herts.ac.uk/resources/papersl.html>

Facilities

Library of Congress, National Digital Library Program and Conservation Division "Conservation Implications of Digitization Projects," (Section 2. Consultation of Space and Environment)

<http://memory.loc.gov/ammem/ftpfiles.html>

Lisa L. Macklin and Sarah L. Lockmiller, *Digital Imaging of Photographs, A Practical Approach to Workflow Design and Project Management*, LITA Guides #4 ALA, Chicago 1999. Ordering information at

<http://www.lita.org/litapubs/lg4.html>

Paul Conway, *Conversion of Microfilm to Digital Imagery: A Demonstration Project*, Yale University Library, 1996.

Project Monitoring

Paul Conway, "Production Tracking," *Moving Theory into Practice: Digital Imaging for Libraries and Archives*, Mountain View, CA : Research Libraries Group, 2000; pp. 160-161. <http://www.rlg.org/preserv/mtip2000.html>

Workflow and Project Steps

Arts & Humanities Data Service, "Digitisation. A Project Planning Checklist," <http://ahds.ac.uk/checklist.htm>

Howard Besser, "Procedures and Practices for Scanning,"

<http://sunsite.Berkeley.EDU/Imaging/Databases/Scanning/>

Linda Serenson Colet, *Planning an Imaging Project, Guide 1 to Quality in Visual Resource Imaging*, <http://www.rlg.org/visguides/visguide1.html>

Stuart Lee, "Scoping the Future of Oxford's Digital Collections, Appendix B," <http://www.bodley.ox.ac.uk/scoping>

Library of Congress, <http://lcweb2.loc.gov/ammem/award/docs/stepsdig.html>
and <http://memory.loc.gov/ammem/prjplan.html>

Peter Noerr, The Digital Library Tool Kit,
<http://www.sun.com/products-n-solutions/edu/whitepapers/digitaltoolkit.html>

TASI, "An Introduction to Making Digital Image Archives,"
<http://www.tasi.ac.uk/advice/overview.html>

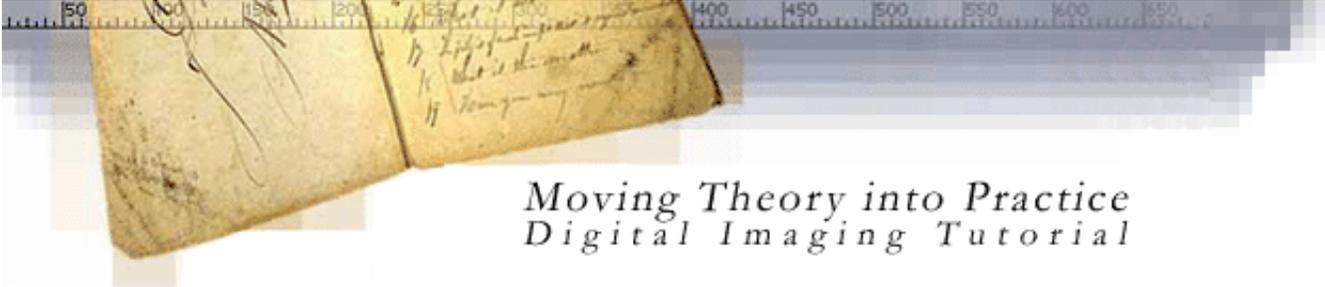
Visual Arts Data Service, "Creating Digital Resources for the Visual Arts,"
(Section 5: Project and Collections Management)
http://vads.ahds.ac.uk/guides/creating_guide/contents.html



View in:



© 2000-2003 Cornell University Library/Research Department



Moving Theory into Practice Digital Imaging Tutorial



Back Next

10. Continuing Education

Key Concepts

introductory information

web-based journals
mailing lists

DIGITAL IMAGING INTRODUCTORY INFORMATION

Besser, Howard and Jennifer Trant. *Introduction to Imaging*. Santa Monica, CA: The Getty Art History Information Program, 1995. Introduces digital imaging technology and vocabulary as they relate to the development of image databases, and outlines the areas in which institutional strategies regarding the use of imaging technologies must be developed.

Kenney, Anne R. and Stephen Chapman. *Digital Imaging for Libraries and Archives*. Ithaca, NY: Cornell University Library, 1996. (available only in print - [order information](#))

Provides an introduction to the central issues associated with the use of digital imaging technology in libraries and archives, including a theoretical and technological overview. Advocates a common vocabulary and set of perspectives from conversion to presentation.

Arts and Humanities Data Service, [AHDS Publications](#)

Offers several series that address creation, management, and distribution of digital image collections. The Guides to Good Practice series is particularly useful, providing practical instruction in applying standards and good practice to the creation and use of digital resources.

Colorado Digitization Project, [Digital Toolbox](#)

Provides links to general resources, bibliographies, initiatives, and clearinghouses on selection, scanning, quality control, metadata creation, and other project management issues. Also offers a glossary of digital imaging terms.

eLib Supporting Studies, [Preservation Studies](#)

Managed by the British Library Research and Innovation Centre, offers several reports on creating and preserving digital image collections. One of the goals is to compare various digital preservation strategies for different data types and formats.

Northeast Document Conservation Center. *Handbook for Digital Projects: A Management Tool for Preservation and Access*. Andover, MA, 1996-2000.

Given at NEDCC's school for scanning conferences.

[PADI](#): Preserving Access to Digital Information

The National Library of Australia's PADI site, offers a subject gateway to digital preservation resources. Includes current information on digital preservation-related events, organizations, policies, strategies, and guidelines. Also includes glossaries of terms that are relevant to digital information.

[PRESERV](#)-The RLG Preservation Program

Offers supporting materials, such as RLG project reports and the bimonthly *RLG DigiNews* to support institutions in their efforts to preserve and improve

access to endangered research materials. The "RLG Tools for Imaging" section includes a worksheet for estimating digital reformatting costs, and guidelines for creating RFPs for digital imaging services.

[TASI](#) (Technical Advisory Service for Images)

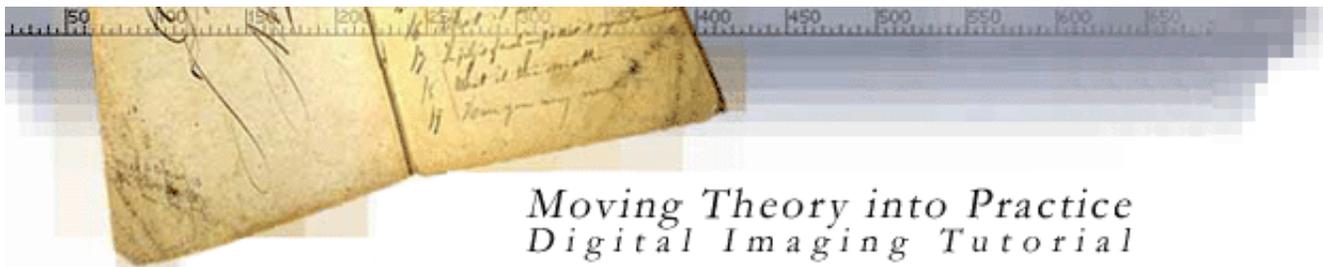
Funded by the Joint Information Systems Committee (UK), provides information on creating, storing, and delivering digital image collections. Also lists events and information resources of interest to those involved in digital imaging initiatives.



View in:



© 2000-2003 Cornell University Library/Research Department



Moving Theory into Practice Digital Imaging Tutorial



Back Next

10. Continuing Education

Key Concepts

introductory information
web-based journals
mailing lists

WEB-BASED JOURNALS, NEWSLETTERS, AND PUBLICATIONS

[Ariadne](#)

Published quarterly by the UK Office for Library and Information Networking (UKOLN), reports on progress and developments within the Electronic Libraries Programme, often covering issues related to digital imaging.

[CLIR](#) (Council on Library and Information Resources) Publications, Frequent reports and research briefs on national and international digital imaging and preservation initiatives.

[Current Cites](#)

Annotated monthly bibliography of selected articles, books, and electronic documents on information technology. While broad in scope, helps keep up with the changes and trends in the digital library realm. To subscribe, send the message "subscribe Cites your name" to listserv@library.berkeley.edu.

[D-Lib Magazine](#)

Monthly, offers articles, commentaries, and briefings that support digital library research. Like Current Cites, covers cutting-edge digital library research and often includes articles related to digital imaging.

[Journal of Electronic Publishing](#)

Focuses quarterly on current issues and trends in electronic publishing, covering issues from creation to delivery of electronic information. Many of the issues are also of interest to those involved in digital imaging initiatives. Published by the University of Michigan Press.

[RLG DigiNews](#)

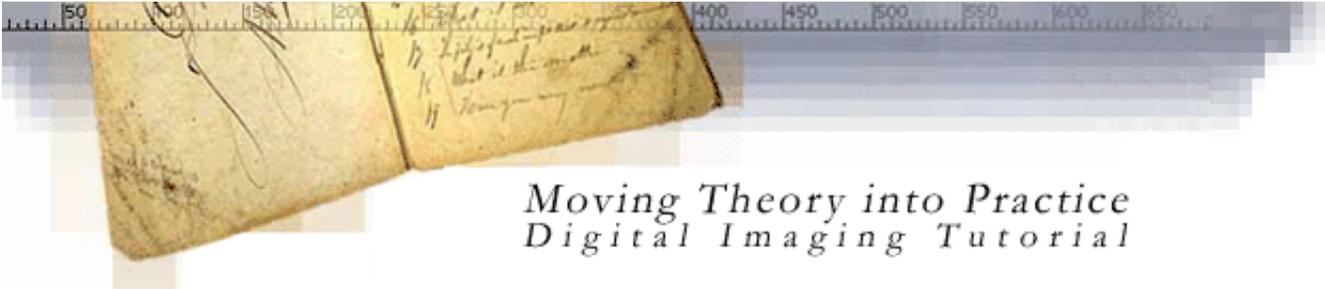
Produced for RLG by the Cornell University Library Research Department, *RLG DigiNews* is a bimonthly Web-based newsletter focused on issues of vital interest to managers of digital initiatives. Now in its sixth year of publication, *RLG DigiNews* provides filtered guidance and pointers to relevant projects, improving awareness of evolving practices in image conversion and digital archiving, while featuring announcements for related publications (in any form) that will help staff attain a deeper understanding of digital issues.



View in:

Spanish
French

© 2000-2003 Cornell University Library/Research Department



Moving Theory into Practice Digital Imaging Tutorial



Back

10. Continuing Education

Key Concepts

introductory information
web-based journals
mailing lists

ELECTRONIC MAILING LISTS

These electronic discussion groups frequently announce new digital imaging projects and report on ongoing initiatives. They also include information on digital imaging-related conferences, meetings, and training programs.

Digital Libraries Research Forum (DigLib)

To subscribe, send the message "subscribe DIGLIB YourFirstname YourLastname" to ifla.listserv@infoserv.inist.fr or visit <http://infoserv.inist.fr/wwwsympa.fcgi/subrequest/diglib>

IMAGELIB

To subscribe, send the message "SUB imagelib Your Full Name" to listserv@listserv.arizona.edu, or visit <http://listserv.arizona.edu/cgi-bin/wa?SUBED1=imagelib&A=1>.

PADI Forum

Specifically dedicated to the exchange of news and ideas about digital preservation issues. To subscribe, send the message "SUBSCRIBE padiforum-I Your Full Name" to listproc@nla.gov.au.

© 2000-2003 Cornell University Library/Research Department



Back

← Contents

View in:

← Spanish

← French

Using This Tutorial

Recommended Hardware and Software Configuration and Settings

- Monitor setting: 800 x 600 display
- 16-bit (thousands) color or higher bit depth
- Web browser: Netscape Communicator v.4.6 or up; Internet Explorer v.5.0 or up
- JavaScript must be enabled

Note: We have done some testing of this tutorial with older browser versions, such as Netscape Navigator v.4.0 and Internet Explorer v.4.0 under both Windows and MacOS. As long as JavaScript is enabled, the tutorial is usable with these versions, although we have noted some problems with proper display of fonts and character attributes such as italics and bold. Also, some text and some graphics appear incorrectly positioned. We would appreciate [notification](#) of any other problems in using older versions of either Netscape or Internet Explorer. Please describe the problem as specifically as possible and tell us what browser/version and what operating system/version you are using.

Navigation

- Each page of the tutorial provides two navigational tools:
 - At the top right corner of the page is a navigation wheel with navigation arrows. The highlighted number in the outer ring lets you know to which section the currently displayed page belongs. Click on any section number to move to the first page of that section. Click on "Contents" in the center of the wheel to move to the table of contents listing, from which you can link to the beginning of any section in the tutorial. The table of contents also includes links to this help page and to a form that allows you to send questions or comments to the tutorial's designers. The arrows beneath the wheel will take you to either the immediately preceding page ("Back" arrow) or the immediately following page ("Next" arrow).
 - Along the left side of the page is a navigation bar. At top is the section number and title, followed by a list of subsections. The subsection names are all hot links and can be used to navigate within the section. At the very bottom of each page is a set of arrows labeled "Back," "Next," and "Contents." These have the same function as the identically labeled features of the navigation wheel, described above.
- Links are used throughout the tutorial, both to show internal references, and to point to related information on other web sites. In order to avoid losing track of your place in the tutorial when you follow links, you need to understand a few things about the behavior of the links:
 - In most cases, links that point to material within the tutorial will appear in the current browser window. Use your browser's "Back" button (not the back arrow in the tutorial's navigation bar) if you wish to return to where you were.
 - External links will open in a separate browser window (a new session). To return to where you were, either close the new browser window or bring the tutorial window forward by clicking on it.

Printing

The tutorial is available in the [PDF format](#). If you do not have Adobe Acrobat Reader®, which is needed to view

PDF files, please click here 

Need Your Feedback

We are committed to updating and improving the presentation and content of this tutorial. Please send us [your comments](#).



© 2000-2003 Cornell University Library/ Research Department

Using This Tutorial

Recommended Hardware and Software Configuration and Settings

- Monitor setting: 800 x 600 display
- 16-bit (thousands) color or higher bit depth
- Web browser: Netscape Communicator v.4.6 or up; Internet Explorer v.5.0 or up
- JavaScript must be enabled

Note: We have done some testing of this tutorial with older browser versions, such as Netscape Navigator v.4.0 and Internet Explorer v.4.0 under both Windows and MacOS. As long as JavaScript is enabled, the tutorial is usable with these versions, although we have noted some problems with proper display of fonts and character attributes such as italics and bold. Also, some text and some graphics appear incorrectly positioned. We would appreciate [notification](#) of any other problems in using older versions of either Netscape or Internet Explorer. Please describe the problem as specifically as possible and tell us what browser/version and what operating system/version you are using.

Navigation

- Each page of the tutorial provides two navigational tools:
 - At the top right corner of the page is a navigation wheel with navigation arrows. The highlighted number in the outer ring lets you know to which section the currently displayed page belongs. Click on any section number to move to the first page of that section. Click on "Contents" in the center of the wheel to move to the table of contents listing, from which you can link to the beginning of any section in the tutorial. The table of contents also includes links to this help page and to a form that allows you to send questions or comments to the tutorial's designers. The arrows beneath the wheel will take you to either the immediately preceding page ("Back" arrow) or the immediately following page ("Next" arrow).
 - Along the left side of the page is a navigation bar. At top is the section number and title, followed by a list of subsections. The subsection names are all hot links and can be used to navigate within the section. At the very bottom of each page is a set of arrows labeled "Back," "Next," and "Contents." These have the same function as the identically labeled features of the navigation wheel, described above.
- Links are used throughout the tutorial, both to show internal references, and to point to related information on other web sites. In order to avoid losing track of your place in the tutorial when you follow links, you need to understand a few things about the behavior of the links:
 - In most cases, links that point to material within the tutorial will appear in the current browser window. Use your browser's "Back" button (not the back arrow in the tutorial's navigation bar) if you wish to return to where you were.
 - External links will open in a separate browser window (a new session). To return to where you were, either close the new browser window or bring the tutorial window forward by clicking on it.

Printing

The tutorial is available in the [PDF format](#). If you do not have Adobe Acrobat Reader®, which is needed to view

PDF files, please click here



Need Your Feedback

We are committed to updating and improving the presentation and content of this tutorial. Please send us [your](#)

Table 5.1 Metadata Types

TYPE	GOAL	SAMPLE ELEMENTS	SAMPLE IMPLEMENTATIONS
<i>Descriptive Metadata</i>	<p>describing and identifying information resources</p> <ul style="list-style-type: none"> ● at the local (system) level to enable searching and retrieving (e.g., searching an image collection to find paintings of animals) ● at the Web-level, enables users to discover resources (e.g., search the Web to find digitized collections of poetry). 	<ul style="list-style-type: none"> ● unique identifiers (PURL, Handle) ● physical attributes (media, dimensions condition) ● bibliographic attributes (title, author/creator, language, keywords) 	<p>Handle PURL (Persistent Uniform Resource Locator) Dublin Core MARC HTML Meta Tags</p> <p><i>controlled vocabularies such as:</i> Art and Architecture Thesaurus Categories for the Description of Works of Art</p>
<i>Structural Metadata</i>	<p>facilitates navigation and presentation of electronic resources</p> <ul style="list-style-type: none"> ● provides information about the internal structure of resources including page, section, chapter numbering, indexes, and table of contents ● describes relationship among materials (e.g., photograph B was included in manuscript A) ● binds the related files and scripts (e.g., File A is the JPEG format of the archival image File B) 	<p>structuring tags such as title page, table of contents, chapters, parts, errata, index, sub-object relationship (e.g., photograph from a diary)</p>	<p>SGML XML Encoded Archival Description (EAD) MOA2, Structural Metadata Elements Electronic Binding (Ebind)</p>

<p>Administrative Metadata</p>	<p>facilitates both short-term and long-term management and processing of digital collections</p> <ul style="list-style-type: none"> ● includes technical data on creation and quality control ● includes rights management, access control and use requirements ● preservation action information 	<p>Technical data such as scanner type and model, resolution, bit depth, color space, file format, compression, light source, owner, copyright date, copying and distribution limitations, license information, preservation activities (refreshing cycles, migration, etc.)</p>	<p>MOA2, Administrative Metadata Elements National Library of Australia, Preservation Metadata for Digital Collections CEDARS</p>
---------------------------------------	---	--	---



Back

© 2000-2002 Cornell University Library/Research Department

Digitization Chain: Click on any text or image component of this illustration to go to the related section.

