BCR's CDP Digital Imaging
Best Practices Working Group

BCR's CDP Digital Imaging Best Practices Version 2.0

June 2008

discover.
share.
experience.

BCR can help your organization achieve greater success
by expanding your knowledge, reach and power.

www.BCR.org

Expand your knowledge.
Increase your power.

CONTENTS

Updating BCR's CDP Digital Imaging Best Practices, Version 2.0

BCR is pleased to make available the 2008 version of *BCR's CDP Digital Imaging Best Practices*. It is intended to serve as a guide for practitioners in cultural heritage institutions that seek to create images from physical objects. Originally published by the Colorado Digitization Program in 2003 as the *Western States Digital Imaging Best Practices,* this updated version reflects the many changes that have occurred in the field during the last five years.

The Colorado Digitization Program has seen many changes during that time as well, and soon after the publication of the original document, it became the Collaborative Digitization Program (CDP). In April 2007, the CDP merged into BCR. Through its newly formed Digital & Preservation Services unit, of which the CDP is a part, BCR continues to provide services to libraries and cultural heritage institutions throughout a multistate area focused in the West.

Many individuals made significant contributions to the development of BCR CDP Digital Imaging Best Practices. The Digital Imaging Best Practices Working Group took leadership for the revision, spending time researching new and updated standards, examining new technologies and evaluating new workflow options. Sarah Goodwin Thiel, Digital Services Librarian, University of Kansas, served as Chair. Other Working Group members included Steve Boss, head, Library Systems, Coe Library, University of Wyoming; Mary Ellen Ducey, associate professor, University of Nebraska, Lincoln Libraries; Leigh A. Grinstead, digital initiatives consultant, BCR; Robert Hickerson, photographer, Spencer Art Museum, University of Kansas; Steve McCann, digital initiatives librarian, University of Montana; Marlys Rudeen, deputy state librarian, Washington State Library; and Mark Shelstad, head, Archives and Special Collections, University of Texas, San Antonio. Their efforts are greatly appreciated. Our special thanks go out to Stanley Smith, manager, Imaging Services, J. Paul Getty Museum, for his insightful comments and suggestions on the document.

One of the biggest joys of working in a collaborative environment is seeing the great achievements — like this document — that collective efforts bring about. The biggest reward for this particular endeavor will be the rich digital content that will result from the many librarians, archivists and museum professionals who will use *BCR's CDP Digital Imaging Best Practices* as a guide to their work.
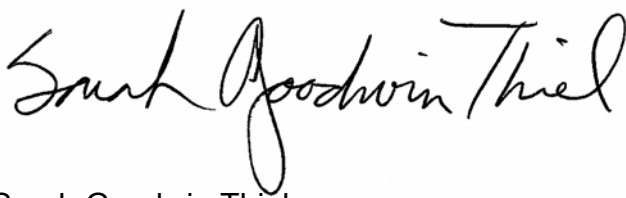
Brenda Bailey-Hainer
President and CEO

Introduction

Institutions of all types and sizes are digitizing collections. Their resources cover a broad spectrum ranging from small scanning operations to large digital studios with high-end image capture equipment, dedicated storage servers and long-term preservation plans. Imaging practitioners continue to face the overriding issues of digitization such as material selection, hardware and software acquisition and workflow development, but their choices have greatly expanded in both cost and complexity. Output expectations have also increased, making image quality and sustainability primary concerns for every institution.

BCR's CDP Digital Imaging Best Practices Working Group draws on the expertise of its members and digitization colleagues to address the many changes informing image digitization since the publication of Version 1.0 in 2003. While continuing to address the needs of a particular audience, beginning practitioners and those from institutions with limited resources and/or expertise, Version 2.0 is developed to assist practitioners from the cultural heritage sector to navigate the quickly evolving world of image digitization and to efficiently increase and improve access to digital scholarship.

In order to increase the scope of *BCR's CDP Digital Imaging Best Practices*, and to take best advantage of current conclusive, well-researched literature by experts in the field, Version 2.0 draws on the National Archives and Records Administration (NARA) *Technical Guidelines for Digitizing Archival Materials for Electronic Access — Raster Images*; the *Universal Photographic Digital Imaging Guidelines* (UPDIG); and the *Reference Model for an Open Archive Information System* (OAIS). Readers will find sections covering equipment selection, digital photography, image quality and storage and preservation issues that are either newly included or enhanced in this revision. These guidelines were prepared by BCR's CDP Digital Imaging Best Practices Working Group during the fall and winter of 2007/2008.

Sarah Goodwin Thiel
Chair, BCR's CDP Digital Imaging Best Practices, Version 2.0
2008

Purpose

The purpose of this document is to offer guidance and to provide digital imaging recommendations for a variety of institutions and collections that are planning for or are involved in digitization projects. These best practices are not intended to be seen as the de facto standard for digital imaging, but rather as a guide for image capture, presentation, storage and preservation. Inherent and unique characteristics of different source materials necessitate different approaches to digitizing, and conversion requirements for digital projects should be considered on a case-by-case basis.

This document addresses text, photographs, maps and graphic materials and is written for institutions that have the equipment and expertise to digitize in-house. If you are planning to digitize oversized materials, bound materials or materials in nonstandard formats and sizes, you may want to consider outsourcing these materials to imaging vendors. These guidelines may be used to define specifications for potential vendors.

These guidelines have been developed in order to:
- Increase the interoperability and accessibility of digital collections across the cultural heritage community through the use of widely accepted standards and formats
- Ensure a consistent, high level of image quality across collections
- Decrease the likelihood of redigitizing in the future by promoting best practices for conversion of materials into digital format
- Promote the long-term preservation of these digital resources

Because technology and industry standards are constantly improving and changing, we view this as a continually evolving document. We welcome your comments and suggestions.

Scope

What is addressed in this document:
- Digitizing and file format recommendations for:
    ◊ Text, photographs, maps, graphic materials, artwork and 3-D objects, film and born-digital files
    ◊ Suggested hardware configurations
    ◊ Software considerations
    ◊ Quality control
    ◊ File naming conventions
    ◊ Scanner and monitor calibration
    ◊ Targets and color bars
    ◊ Storing images

What is not addressed in this document:
- Scanning and file format recommendations for:
    ◊ Audio, video/moving images

◊   Project management
◊   Selection of collections for digitization

Revisions

The CDP Digital Imaging Best Practices Working Group will review and update the *Digital Imaging Best Practices*. BCR will serve as the host institution for the working group's efforts, and all documents and related materials are available through the BCR website (www.bcr.org/cdp/index.html).

Laying the Groundwork

Digital imaging procedures have evolved steadily over the past decade. Technological advancements have resulted in higher quality hardware and software, and now many imaging options exist. The following principles come from *Procedures and Practices for Scanning*, Howard Besser's highly regarded 1997 paper on digital imaging. These principles remain sound and provide a solid foundation for those embarking on digitization projects. The list has been revised here and uses digitization as the broader concept.

General Principles
- Digitize at the highest resolution appropriate to the nature of the source material
- Digitize at an appropriate level of quality to avoid redigitizing and rehandling of the originals in the future
- Digitize an original or first generation (i.e., negative rather than print) of the source material to achieve the best quality image possible. In the case of art prints, the developed print is considered the original piece. Conservation concerns may prevent digitizing original negatives
- Create and store a master image file that can be used to produce surrogate image files and serve a variety of current and future user needs
- Use system components that are nonproprietary
- Use image file formats and compression techniques that conform to standards within the cultural heritage community
- Create backup copies of all files on servers and have an off-site backup strategy
- Create meaningful metadata for image files or collections
- Store digital files in an appropriate environment
- Monitor data as necessary
- Document a migration strategy for transferring data across generations of technology
- Plan for future technological developments[1]

---

[1] Howard Besser. "Procedures and Practices for Scanning." Canadian Heritage Information Network, (1997). http://besser.tsoa.nyu.edu/howard/imaging/

Questions to Ask Before Starting a Digitization Project

Carefully thinking through the many components of a digitization project will go a long way towards ensuring a successful outcome. Consider how digitization fits into your institution's overall strategic plan, technology plan and project workflows.

The CDP Project Management "Questions to Ask" provides a full discussion of the following question. See Appendix A — Bibliographical and Online Resources.

- What is your purpose?
- Does the project support the institution's mission?
- Who is your audience?
- Who owns it?
- What are the physical characteristics of the collection?
- What is your timeframe?
- How is the project being funded?
- Who will be responsible at different stages of the project?

Documentation

Documentation of the choices your project has made can be a key factor in the long-term success of digitization efforts. Good documentation can offset the impact of staff turnover and allow future staff an ability to deal with digital collections created by their predecessors. Among the items to consider documenting:

- Local guidelines and benchmarks for image quality and resolution
- Resources that contributed to local practice guidelines
- Types of metadata captured
- File naming schemes
- Sustainability plans and procedures (storage, archiving, refreshing media, etc.)

Staffing

In practice, many digital imaging projects will not have unique staff working on the project, but will utilize existing staff from other areas in the organization, student assistants or volunteers. It may benefit the project coordinator(s) to look at "transferable skills" that project staff members already possess that would be useful in any digitization project. Sufficient time for training, and opportunities to receive further education, should also be provided. Metadata creation and operating high-end scanners and/or digital cameras are labor-intensive activities.

Digitization projects require a combination of skills from a variety of staff with different areas of expertise. The following areas and skills may be important to any digitization project:

- Project management skills
- Knowledge of cataloging, registration methods or metadata schema
- Familiarity with conservation methods
- Understanding of photographic techniques and methods
- Subject matter specialists (curators, archivists, scholars, librarians, faculty, etc.)
- Database development and administration skills

- Computer programming skills
- Web design and development skills
- Artistic/graphic design skills

By nature, digitization projects require a team approach, bringing together diverse sets of skills from different areas of the organization, perhaps more than any other project. Administration, technical services staff, cataloging specialists, the information technology department, subject specialists, curators, librarians, preservation/conservation staff, faculty and others may all be involved.[2] Depending on the size of the institution and/or project, a single individual may fill several or all of these roles.

Sample project staff and their roles:

- Project Manager
- Selector, Conservator, Preparations Technician
- Cataloger, Metadata Analyst
- Scanning Technician or Photographer, Quality Control Technician
- Programmer or other Database Developer to integrate metadata and images
- Systems Administrator, Network Administrator
- User Interface Developer or Designer

Training

Many organizations around the country offer workshops and training on digital imaging, and many conferences are held each year addressing imaging issues.

- BCR offers workshops and seminars in digital imaging and metadata creation
- "School for Scanning," Northeast Document Conservation Center
- American Association for State and Local History (AASLH) offers Digitizing Historic Collections around the country
- AMIGOS Bibliographic Council provides digital imaging workshops to institutions in the Southwest.
- *Moving Theory Into Practice: Digital Imaging Tutorial*, Cornell University Library/Research Department, 2000-2003.
- A good resource for Photoshop training can be found at the National Association of Photoshop Professionals (NAPP). See http://www.photoshopuser.com
- Several professional organizations often host workshops and conferences on issues related to digitization:
    ◊ American Library Association (ALA)
    ◊ Library and Information Technology Association (LITA)
    ◊ Association for Library Collections and Technical Services (ALCTS)

---

[2] Stephen Chapman, "Considerations for Project Management," in *Handbook for Digital Projects: A Management Tool for Preservation and Access*, ed. Maxine K. Sitts, (Andover, Massachusetts: Northeast Document Conservation Center, 2000). http://www.nedcc.org/oldnedccsite/digital/dman2.pdf

◊   Society of American Archivists (SAA)
◊   American Association of Museums (AAM)
◊   Museum Computer Network (MCN)
◊   Archives & Museum Informatics (A&MI)
◊   Institute for Museum and Library Services (IMLS)

In-house vs. Outsourcing

Every organization should carefully consider the pros and cons of outsourcing digitization projects or conducting them in-house. For projects being considered for outsourced digitization, a visit to vendor facilities is recommended to ensure that the workspaces provided meet the criteria above.

Following are some points to consider for both strategies:

In-house pros:
- Development of digital imaging project experience by "doing it" (project management, familiarity with technology, etc.)
- More control over the entire imaging process as well as handling, security and storage of originals
- Requirements for image quality, access and scanning can be adjusted as you go instead of defined up front
- Direct participation in development of image collections that best suit your organization and users

In-house cons:
- Requires large initial and ongoing financial investment in equipment and staff
- Longer time needed to implement imaging process and technical infrastructure
- Limited production level
- Staffing expertise not always available
- Institution must accept costs for network downtime, equipment failure, training of staff, etc.
- Need to enforce standards and best practices

Outsourcing pros:
- Pay for cost of scanning the image only, not equipment or staffing
- High production levels
- On-site expertise
- Less risk
- Vendor absorbs costs of technology obsolescence, failure, downtime, etc.

Outsourcing cons:
- Organization has less control over imaging process, quality control
- Complex contractual process: image specifications must be clearly defined up front, solutions to problems must be negotiated, communication must be open and problems must be accommodated
- Vendor may know more than client or may presume a level of understanding on part of library/museum/archives that they may not have

- Lack of standards with which to negotiate services and to measure quality against
- Originals must be transported, shipped and then also handled by vendor staff
- Possible inexperience of vendor with library/archival/museum/historical society[3]

Costs

It is difficult to predict just how much a digital imaging project is actually going to cost, and little hard data on the cost, cost effectiveness and costs over time of digital projects is readily available. Generally, capture and conversion of data often comprises only one-third of the total costs, while cataloging, description and indexing comprise two-thirds of the total costs. Upfront and ongoing costs can be significant, and economic advantage may be better realized through collaborative initiatives or cooperative/regional digitization initiatives, where costs, resources, goals and expertise can be shared.

Initial investment in equipment, staff training, capture and conversion, handling, storing and housing originals, producing derivative files, cataloging, building the image database system and developing web interfaces are all possible areas of cost for any digitization project.

However, the costs of a project do not end after conversion. Some ongoing costs that an institution must commit to, include the costs of maintaining data and systems over time, including media migration costs and infrastructure costs.

Rights Management

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

Before beginning a digitization project, establish which objects are in the public domain and which objects will require permission from the copyright holder. Items in the public domain may in fact drive the selection of digital images for a digitization project because the issues of copyright are already resolved.

If a collection has been chosen for digitization and copyright is not yet cleared, locating the copyright holder and obtaining permission can be a lengthy and costly process. It is important to allot staff time and to document the procedures followed, along with any results in order to demonstrate due diligence.

---

[3] Kenney, Anne and Steven Chapman. "Digital Imaging for Libraries and Archives." Ithaca: New York, Department of Preservation and Conservation, Cornell University Library, June 1996.

Due to the rapidly changing U.S. and international rights management laws, monitoring of legislation is an important activity.

See Appendix A — Bibliographical and Online Resources, for additional resources, *Legal Issues to Consider When Digitizing Collections.*

Digital Preservation

Digital preservation is one of the components in the life cycle of digitization, along with creation, storage, search and discovery. Commonly referred to as digital curation in the United Kingdom, the goal of digital preservation is to assure long-term accessibility of born digital and reformatted objects.

There are numerous definitions of digital preservation. For the purpose of this document, we will be using the RLG-OCLC Report, Trusted Digital Repositories: Attributes and Responsibilities (TDR) definition. Digital preservation is the managed activities necessary for ensuring both the long-term maintenance of digital information and the continued accessibility of its contents. Other definitions that might be of interest include the American Library Association's ALCTS Preservation and Reformatting Section's Digital Preservation Definition (www.ala.org/ala/alcts/newslinks/digipres/index.cfm).

The TDR defines the attributes of a digital archive/repository within the context of cultural heritage content, building the Reference Model for an Open Archive Information System (OAIS). OAIS provides a conceptual framework for the understanding and increased awareness of archival concepts needed for long-term digital information preservation and access. Tools and software applications are being developed to support the range of activities necessary in a comprehensive digital preservation plan.

To support digital preservation, it is critical to capture administrative, structural and technical data associated with each object. It means following industry-accepted best practices and open standards, and it relies on sustainable storage strategies.[4]

---

[4] Conway, Paul, "The Relevance of Preservation in a Digital World." Northeast Document Conservation Center. (2007) http://www.nedcc.org/resources/leaflets/6Reformatting/04RelevanceOfPreservation.php

Creation Basics

PPI/DPI/SPI

Pixel is short for picture elements, which make up an image, similar to grains in a photograph or dots in a half-tone. Each pixel can represent a number of different shades or colors, depending on how much storage space is allocated for it.

Pixels per inch (PPI) refers to the number of pixels captured in a given inch and is used when discussing scanning resolution and on-screen display. *When referring to digital capture, PPI is the preferred term, as it more accurately describes the digital image.* This document will refer to PPI when discussing capture resolution. For more information, see the Appendix B — Digital Imaging Glossary.

Digitization guidelines and hardware manufacturers' documentation frequently use the measurement of dots per inch (DPI) when discussing optical resolutions for images and hardware. DPI more accurately refers to output devices, how many dots of ink per inch a printer puts on the paper or onscreen monitor display.

SPI stands for number of samples or units measured per inch. SPI is considered interchangeable with the preferred term PPI.

Modes of Capture

Most imaging equipment offer four modes for capturing a digital image:

- Bitonal — One bit per pixel representing black and white. Bitonal scanning is best suited to high-contrast documents such as printed text
- Grayscale — Multiple bits per pixel representing shades of gray. Grayscale is suited to continuous tone items, such as black and white photographs
- RGB — Multiple bits per pixel representing color. Color capture is suited to items with continuous tone color information
- CMYK — Short for Cyan-Magenta-Yellow-Black and pronounced as separate letters. CMYK is a color model in which all colors are described as a mixture of these four process colors. CMYK is the standard color model used in offset printing for full-color documents. Because such printing uses inks of these four basic colors, it is often called four-color printing. CMYK is a derivative color space that is device-specific and should never be used for archival purposes

These three modes of capture — bitonal, grayscale and RGB — require some subjective decisions. For example, a black and white typed document may have annotations in red ink. Although bitonal scanning is often used for typed documents, scanning in color may be preferable in this case, depending on how the image will be used. Manuscripts, older printed matter and sheet music may be better served by capturing as continuous tone in grayscale or color to bring out the shade and condition of the paper and the marks inscribed on it. Projects interested in capturing the current condition of source materials should consider capturing in color.
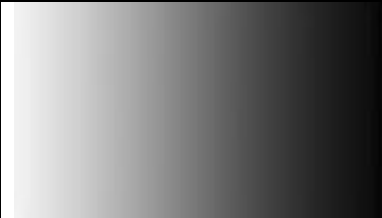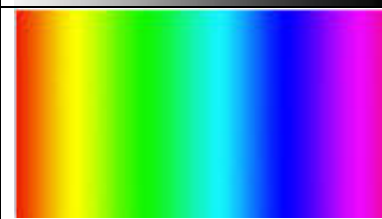
Bit Depth

Bit depth measures the number of colors (or levels of gray in grayscale images) available to represent the color/gray value in the original work. A bit is the basic digital building block with a value of either 1 or 0. Every pixel sampled is assigned a value that corresponds to the color/shade it represents.

"An image with a bit depth of 1 has pixels with two possible values: black and white. An image with a bit depth of 8 has 2^8, or 256, possible values. Grayscale mode images with a bit depth of 8 have 256 possible gray values.

RGB images are made of three color channels. An 8 bit per pixel RGB image has 256 possible values for each channel, which means it has over 16 million possible color values. RGB images with 8 bits per channel (bpc) are sometimes called 24 bit images (8 bits x 3 channels = 24 bits of data for each pixel)."[5] RGB images with 16 bits bpc are called 48 bit images (16 bits x 3 channels = 48 bits of data for each pixel).

In addition to 8 bpc images, Photoshop can also work with images that contain 16 bpc or 32 bpc. Images with 32 bpc are also known as high dynamic range (HDR) images.

1 Bit Bitonal	2 colors; usually black or white		
8 Bit Grayscale	256 shades of gray		
48 Bit RGB	2800 times one trillion colors and shades of gray		

[5] Adobe Help Resource Center
http://livedocs.adobe.com/en_US/Photoshop/10.0/help.html?content=WSfd1234e1c4b69f30ea5 3e41001031ab64-73da.html

Color Space

Setting the color space is critical to digital capture and can be set within your photo editing software prior to digital capture, unless you are using a digital camera and capturing a RAW file. A device color space simply describes the range of colors, or gamut, that a camera can see, a printer can print or a monitor can display. Editing color spaces, on the other hand, such as ProPhoto RGB, Adobe RGB or sRGB, are device-independent and determine a color range you can work in. Their design allows you to edit images in a controlled, consistent manner. Choosing a wide gamut space, such as ProPhoto, RGB or Adobe RGB, allows for greater color information to be captured about the original item and will allow you to convert to a narrow space later. Narrow gamut space such as sRGB captures less information and is intended for images used on the web.

Resolution

Resolution determines the quality of an image. It is described either by pixel dimensions (height and width) for on-screen use or by physical size and PPI. Increased PPI take more frequent samples of the original and contain a more accurate representation of the original. Since higher resolutions are capturing more information, file sizes also increase. There is no one "perfect" resolution to scan all collection materials. Resolution should be adjusted based on the size, quality, condition and uses of the digital object. The combination of PPI and size of the original object determine the resolution needed to accurately capture as much information about the original object as is available. See Guidelines for Creating Digital Images for specific spatial resolution targets.

Captured at 50 PPI

Captured at 150 PPI

Captured at 600 PPI

There is a point at which adding more pixels per inch no longer adds content, because the original source object has a finite amount of information available based on the way that it was produced.

Tonal Dynamic Range

One of the most significant factors affecting image quality is the tonal dynamic range. The detail and range of tones that the capture device can record is often expressed in f-stops. A digital camera can record 11 stops of data — meaning that there is detail in the

shadows and highlights over an 11-stop range. Tonal range can be expanded by doing a high dynamic range (HDR) capture. Using Photoshop, one can take widely bracketed exposures and combine them into a single image that records detail in much greater range than any single exposure could.

The color space an image occupies between pure black, represented graphically as the number 0, and pure white, is also represented graphically as the number 255. Tonal dynamic range can be displayed in Adobe Photoshop as a histogram. The histogram graphically displays the number of pixels in the image at selected values from black to white. Reviewing histograms at the time of capture can ensure that all of the image's information is being recorded.

Compression

- Lossy — In lossy compression, a certain amount of information is discarded during the compression process. Although the discarded information may be invisible to the human eye, a loss of quality occurs. Lossy compression formats also introduce generational loss — each time a lossy image is manipulated or edited, the quality of the image decreases. Generational loss is one of the reasons master and service master images are not stored using compression.

  Converting images from a bit depth of 16 to a bit depth of 8 is also considered a lossy compression method as color information is discarded.

| JPEG High Compression Low Quality 5 kilobytes | JPEG Medium Compression Medium Quality 10 kilobytes | JPEG Low Compression High Quality 82 kilobytes |

Detail of Serene Donner in a Picture Hat, 1910
courtesy Ira M. Beck Memorial Archives, University of Denver
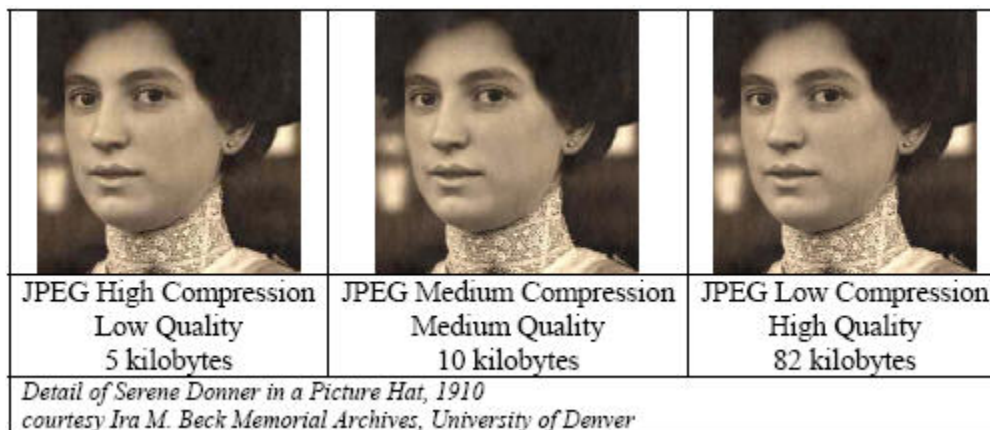
  The Joint Photographic Experts Group (JPEG) format is most frequently used for access images requiring lossy compression. The JPEG compression algorithm was designed for continuous tone images.

- Lossless — Lossless compression results in a file similar to the original image, with no loss of information. The Tagged Image File Format (TIFF) and Portable Network Graphic (PNG) formats support lossless compression.

File Formats

There are proprietary and nonproprietary formats for image files. The recommendation for master file image capture is to use a non-proprietary format. See Guidelines for Creating Digital Images.

- RAW Files — A RAW file is an image file that contains unprocessed data. Digital Single Lens Reflex (DSLR) cameras and some high-end scanners allow users to capture images in a RAW or native file format that is unique to each manufacturer. The proprietary nature of these files is of concern for the long-term preservation and access of these digital files. RAW files have advantages for photographers, including smaller file size and fine processing controls. After processing or editing and before use, RAW files must be converted to an open standard format such as JPEG or TIFF. Processing RAW files creates an additional step in the imaging workflow and may require sophisticated photographic skills or expertise.

- DNG — In response to concerns over the future support for proprietary RAW files, Adobe has produced the DNG (Digital NeGative) format. DNG is supported by a growing number of camera and software manufacturers and, as of this writing, has been submitted to the ISO as a vendor-independent standard. A free Adobe DNG Converter is available to translate many different proprietary RAW formats to DNG.

  As explained in the UPDIG Guidelines, DNG files "store the RAW image data, metadata, and a color-corrected JPEG preview of the image. The DNG file format provides a common platform for information about the file and adjustments to the image. Because of this, cataloging applications like iView MediaPro and Extensis Portfolio can see metadata that has been entered in Photoshop, and these programs can see the image adjustments made in Photoshop. DNG files can be re-opened in Photoshop as though they were RAW files and offer the full range of conversion options in Photoshop that the original RAW file offered."

  "DNG is likely to be readable long after the original RAW format becomes obsolete, simply because there will be so many more of them than any particular RAW file format. Additionally, DNG offers significant file-size savings through a lossless compression that can reduce the file size by up to one third. DNG also offers the possibility of embedding the RAW file itself, so that it can be re-converted later if desired. Of course, this particular option makes the DNG almost double the size of the original RAW file, so it isn't always to be recommended."[6]

- TIFF — Tagged Image File Format (TIFF) is the format of choice for archival and master images. It is a flexible, highly portable, widely accepted, open standard image format and considered the professional image standard. TIFF files may or may not use lossless compression such as the LZW algorithm. Due to the large file size of TIFF images, they are not suitable for web delivery.

- JPEG 2000 — JPEG 2000 is a wavelet-based standard for the compression of still digital images. It was developed by the ISO JPEG committee to improve on the performance of JPEG while adding significant new features and capabilities

---

[6]   UPDIG Guidelines, Version 3.0. "Archiving, Archiving DNG files,"
http://www.updig.org/guidelines/archiving.php

to enable new imaging applications. Beyond image access and distribution, JPEG 2000 is being used increasingly as a repository and archival image format. Many repositories are storing "visually lossless" JPEG 2000 files: the compression is lossy and irreversible, but the artifacts are not noticeable and do not interfere with the performance of applications. Compared to uncompressed TIFF, visually lossless JPEG 2000 compression can reduce the amount of storage by an order of magnitude or more. JPEG 2000 is still a lossy, compression technique, but may have potential for becoming the file format of choice for archival master images in the near future.

In October 2007, the Library of Congress, announced collaboration with Xerox Corporation to study the use of the JPEG 2000 format in large repositories of digital cultural heritage materials. This study will build on the work that led previously to the JPEG 2000 profiles for newspapers and extend it to cover prints, photographs and maps. It will pay attention to the preservation, access and performance issues associated with large image repositories and how the JPEG 2000 standard can address those issues. The work, due to be completed in 2008, is expected to lead to specifications and best practices for the use of JPEG 2000[7].

- JPEG — Joint Photographic Experts Group. A compression algorithm for condensing the size of image files. JPEG image files allow online access to full screen image files because they require less storage and are therefore quicker to download into a web page.

- PDF and PDF/A — In April 2008, the United Kingdom's Digital Preservation Coalition (DPC), named Portable Document Format (PDF) as one of the best file formats to preserve electronic documents and ensure their survival for the future. This decision allows information officers to follow a standardized approach for preserving electronic documents. The DPC report suggests adopting PDF/Archival (PDF/A) for archiving electronic documents as the standard that will help preservation and retrieval in the future. It concludes that it can only be done when combined with a comprehensive records management program and formally established records procedures[8].

- Proprietary Formats — Proprietary formats are controlled or owned by a particular entity that licenses the format for use by others. These formats often require special plug-ins or software for viewing. Proprietary formats are not recommended for master images because licensing requirements may prevent

[7] Buckley, Ph.D., Robert. "Technology Watch Report JPEG 2000 - a Practical Digital Preservation Standard?" DPC Technology Watch Series Report 08-01 (February 2008), http://www.dpconline.org/docs/reports/dpctw08-01.pdf

[8] Fanning, Betsy A., "Technology Watch Report Preserving the Data Explosion: Using PDF," DPC Technology Watch Series Report 08-02 (April 2008), http://www.dpconline.org/docs/reports/dpctw08-02.pdf

the long-term access and preservation of images. One example of a proprietary file format is Encapsulated PostScript (eps).

Hardware Considerations

When creating digital content, one must consider the complete workflow — from direct digital capture to building a sustainable archive and finally, to public presentation. In order for a digitization project to be successful, all aspects of capture, preservation and presentation must be accounted for. The number of objects, their format and condition and the timeline to completion, will determine the scale and rate of the project.

Computers

- Critical to the success of any digitization initiative is the purchase of a computer with a balance of reliable components, speed and storage that will increase productivity and overall effectiveness. Projects that require the purchase of computer hardware should consider the following general principles:
  ◊ Purchase a computer dedicated solely to digitization initiatives.
  ◊ Purchase as much Random Access Memory (RAM) as your budget allows More memory allows the computer to more quickly process large amounts of image data
  ◊ Purchase computers with processors optimized for image manipulation
  ◊ Purchase computers that support high-speed data input through serial connections, USB 2.0 or IEEE 1394 Firewire
  ◊ Purchase as much hard drive space as possible. Portal hard drives that attach via the USB port can be used to supplement the hard drive workspace on the digitization PC

Monitors/Display

The monitor or display is the most-used output device on a computer. Monitors give the end-user instant feedback showing text or graphics. For that reason, it is very important to keep your monitor regularly calibrated. Most computers now use liquid crystal display (LCD) as opposed to the older cathode ray tube (CRT) technology. Both types of monitors use the same type of connection to a computer. Monitors are increasing in viewing size, color depth, bit depth and resolution pixels. Other areas of improvement include adaptive contrast enhancement, texture enhancements and color correction.

LCDs are smaller and lighter than CRT monitors. They are more energy efficient and are now the default option with most computers. LCD's are attractive because they are thin and take up less space on a desk than the bulkier CRT units. Their fixed resolution has been known to lessen adaptability and color rendering; however, LCD image quality has greatly increased as the price for these monitors has decreased.

CRTs create a high quality of color rendering and a high range of color, and they have more options for handling graphics. CRT popularity has fallen partly because they do issue higher radiation emission, produce more heat and use more power and energy. They are also large, heavy and can produce "screen flicker."

It is possible to purchase monitors designed for critical color work. To learn more about choosing a monitor, you may want to review the current literature. See Appendix A — Bibliographical and Online Resources for online articles.

Image Capture Devices

One element essential to a successful digitization project is the selection of image capture devices. Recent developments have increased the challenges in selecting these devices by increasing variety and availability while reducing the costs of equipment. What device is right for your project depends on numerous factors including overall project goals, format, size, condition of materials to be captured and available budget. Several technical factors will also influence your purchase including cost, required expertise, size of scan area, speed, connectivity and ability to handle different formats and materials in your collection. In order to decide which device to use and when, and to select the correct model with the appropriate features, many considerations exist:

What do I need?

Consider the collections to be digitally captured. The physical nature of the objects will impose restrictions and direct your decisions regarding which type of capture device to use. For example, if your collection is comprised of letters, newspaper clippings and small photographs, perhaps a flatbed scanner would be the best option. They provide quality scans of two-dimensional objects that fit on the surface of the scanner; they require no external studio lighting; most fit on a table top; and they can easily be used by someone with only a basic understanding of the digitization process. If you are scanning complete books, a dedicated book scanner may be the ideal choice. If your entire collection has been photographed on film, then a dedicated film scanner is recommended. If you have a collection that includes oversized materials, original art or a variety of items that must be captured overhead and in color, you will need a digital camera that can be used to capture any object, two- or three-dimensional, small or large. Keep in mind that as image capture equipment becomes more complex and sophisticated, a higher skill level may be required. Institutions may need to consider further training for in-house imaging practitioners.

When considering a capture device for your digitization project, budget will be one of many factors to consider. Fortunately, there are capture devices in every price range. Of course, you get what you pay for. A $100 scanner will not have the same quality as a $1,000 scanner, so consider your project needs and then budget accordingly. Investing in digital imaging technology will repay itself many times over in the future.

When it comes to selecting specific equipment, talk to your colleagues. Look at established imaging facilities at other cultural institutions; they will often list equipment used, both hardware and software. Review specific equipment online, in trade publications such as *MacWorld, PC World* or *PC Magazine.* Join imaging listservs such as IMAGELIB@LISTSERV.ARIZONA.EDU and TASI@JISCMAIL.AC.UK for ongoing discussions of digital imaging issues and concerns.

Optical Density

The optical density measures the "brightest bright" and the "darkest dark" that a piece of equipment can capture. A scanner or camera's optical density will impact the overall quality of the image's tonal dynamic range (see below) and ability to capture highlights and shadows in an image.[9]

Scanner Density Ranges	
Scanner	**Typical Dynamic Range**
Handheld scanners	<2.3
Office grey-scale scanners	<2.5
24 Bit Colour Flatbed scanners (older models & low-end)	2.2 - 2.6
30 Bit Colour Flatbed scanners (midrange)	2.8 - 3.2
34-36 Bit Colour Flatbed scanners (high-end)	3.3 - 3.6
Desktop Drum Scanners	3.3 - 3.7
High-end Drum Scanners	3.4 - 3.8

Note: manufacturers often exaggerate the dynamic range values of their products. Values over 4 describe the maximum range that can potentially be contained within the scanned image rather than the dynamic range that can be captured by the scanner's CCD

Media Density Ranges	
Source Media	**Typical Dynamic Range**
Images on newsprint	0.9
Images on coated stock	1.5 - 1.9
Photographic Prints - normal (C-type)	1.6 - 2.0
Photographic Prints - High Contrast (R-type, cibachrome)	2.0 - 2.3
Negative Film	2.4 - 2.8
Colour Transparency	2.8 - 3.2
High Quality Transparency and Dupe film	3.4 - 3.8

Scanners that are able to capture higher bit depths generally are able to capture higher optical densities. Newer equipment specifications include the measurement for the "darkest dark" or maximum density (dMax). Equipment with a higher dMax is able to capture deeper shadows. A higher dMax is particularly critical for quality capture of transparencies, negatives and slides that require an external light source to be captured.

Speed and Connectivity

An important factor to consider for the most efficient projects is the speed at which equipment can capture images and transmit them to the host computer. Most scanners include specifications on scanning speed. Higher scan speeds and transfer rates will reduce the exposure of sensitive materials to light, particularly when using digital scan-backs and cameras on a copy stand.

To ensure efficient transfer of image data, select equipment (both scanner and computer) that uses high speed data transfer standards, such as Universal Serial Bus (USB) 2.0, 2.1 or 3.0; Small Computer Serial Interface (SCSI) cards and cables; or IEEE 1394 Firewire. Avoid equipment that uses slower methods such as parallel or serial ports or USB 1.0. Other successful transfer protocols include FTP or SFTP/SSH. Software applications that support these transfer protocols include WS_FTP Pro or Filezilla.

---

[9].Kenny & Reiger .Density charts from Technical Advisory Service for Images. "Moving Theory into Practice." http://www.tasi.ac.uk/advice/creating/scanners.html

Scanners

Flatbed Scanners — Flatbed scanners are one of the most popular types of scanners used in libraries, archives and museums. Depending on the condition and format of the selected material, flatbed scanners can be used to scan 2-dimensional materials such as papers, some photographs, printed material, etc. Flatbed scanners are not recommended for original art, including art prints. An important consideration when selecting a flatbed scanner is the size of the scan area. Large scan areas of 11" by 17" are needed to accommodate a variety of materials and formats and are widely available in recommended professional grade models.

Accessory items available with some models include transparency adaptors and automatic document feeders (ADF). Transparency adaptors are needed when scanning 4"x 5" or 8"x 10" negatives and transparencies. There can be focus issues when using flatbed scanners for scanning film, unless using a very high-end scanner. ADFs are used when scanning multiple pages of contemporary material, but cannot be used for historic materials because of danger of damage.

Slide/Film Scanners — Dedicated slide/film scanners are specifically designed to digitize slides and film. Although a flatbed scanner with a transparency lid can be used for this purpose, a dedicated film scanner has much higher quality scanning capabilities: resolution, color density, film handling and focusing. Slide/film scanners have higher dynamic tonal ranges and optical resolutions. Optional slide feeders can be purchased to allow a batch scan of up to 50 slides. Slide feeder attachments have become more sophisticated, but still need monitoring to avoid jamming and potential damage. Slide scanners are highly recommended for projects with large numbers of slides.

Large Format Scanners — Large format scanners can be useful when scanning maps, blueprints, architectural drawings, site plans, posters, etc. They operate like a flatbed scanner, but look more like a small pool table. Professional grade models come with high optical resolutions; however, they remain potentially cost prohibitive for many projects. Institutions needing to digitize this type of material may want to consider outsourcing to a digital imaging vendor or using a digital camera for capture (see Digital Cameras below).

High-End Book Scanners — Book copiers or scanners are used in two ways. They allow for overhead copying of bound books and oversized and/or fragile materials that cannot be placed on a flatbed scanner. They can also be used for reformatting projects as they are quick and are built to support the structure of a book. Book scanners include software which compensates for any distortion caused by the curve of the page when digitizing complete books. These scanners are more complex and, consequently, much higher priced than flatbed scanners. As with the large format scanners, institutions needing to digitize overhead may want to consider using a digital camera for capture (see Digital Cameras below).

Wide-format Scanners — Wide-format scanners were developed to digitize large format materials such as engineering drawings and architectural blueprints and are frequently found in municipal engineering departments or local blueprint shops. Materials are drawn over the scanning sensor through a pair of drums. Due to the

danger of mechanical damage (ripping, tearing), these types of scanners are not recommended for cultural heritage materials.

Most scanners use a grid-like array of light sensors that translate light into the 1s and 0s of your digital image. The number of sensors in the array determines the optical resolution of a particular device. The optical resolution is normally expressed in scanner specifications as pixels per inch (PPI). The optical resolution of any equipment you purchase should exceed the maximum resolution needed to accurately capture the types of material in your collections (see Guidelines for Creating Digital Images). For example, a flatbed scanner with an optical resolution of 1200 PPI has sufficient optical resolution to scan an 8"x10" print at 600 PPI, but insufficient optical resolution to scan a 2x2 slide at 2000 PPI.

Many models of scanners are advertised with very high resolutions that represent the interpolated resolution. To increase the resolution the equipment uses a mathematical algorithm to "guess" what color and light values exist in the spaces that the light sensors can't see. Make sure to select equipment based on its optical resolution and not the interpolated resolution since scanners with adequate optical resolution will produce more accurate scans.

Scanner specifications often include the size of the array (1600 x 3200). The first value measures the optical resolution of the array and the second value represents the capacity of the array to capture information as it moves across the scan area (how many pixels does the array move before taking another sample). If the second number is smaller than the first number, the samples are interpolated. For most professional quality scanners, the second value will be higher than the optical resolution.

The Digital Camera System

Point and Shoot — Consumer-oriented "point and shoot" cameras are not suitable for digitization projects. The file size is generally insufficient; the lens quality is limited, and there is no studio flash synchronization. Consumer cameras rarely allow for shooting in Adobe RGB color space or in RAW mode.

35 mm Digital Single Lens Reflex — For many digitization projects, a 35mm digital single lens reflex (DSLR) type camera may be the best capture device. The digital DSLR is ideal for making high-quality images suitable for archiving and publication and is a simple, cost effective solution. The advantages of this type of camera include: high image quality; appropriate image resolution; ease of use; flexibility in shooting situations; option for tethered shooting; flash synchronization for quick capture; and modest price range. Tethered shooting is when the camera is connected directly to the computer via Firewire or USB cable. The image appears on the computer screen seconds after it is captured allowing for technical and aesthetic decisions to be made immediately. This insures the final capture will be archival quality.

Digital Scan Backs or Scanning Attachments — Digital scan backs consist of a scanning array that attaches, in place of a film holder, to a 2"x3" or 4"x5" view camera body. Scan backs are used to create high resolution images of oversized, fragile or unwieldy items. Materials include: three-dimensional objects; photos or artwork bound in

albums; large photographs; original artwork; maps; and others. Institutions considering the purchase of a digital scan back should consult with manufacturers to select the correct lenses and lighting needed for the types of material being scanned.

Digital scan backs produce very large (100 to 500 MB), high-quality image files; consequently, high-resolution details can be generated from the master file. Because a scan back is used on a view camera, the photographer has more precise control over perspective, focus and sharpness of the image. The cost of a digital scan back remains high, and the image capture process is quite slow (5 to 11 minutes). Lighting concerns will need to be specifically addressed if purchasing a scan back.

Camera Stands and Lighting

Copy stand with lights — A copy stand is a stand for mounting a camera, complete with base board and arms to hold lights. The stand enables you to take photographs of flat art, books and three-dimensional objects. The lamps cast even lighting on the work being photographed. A copy stand with lights offers a controlled environment.

Mono stand — A mono stand is a column with a movable cross arm, for mounting a camera.

Book cradle — Book cradles are special supports for scanning bound documents. A book cradle allows a book to be fully supported and properly aligned for safe and accurate digital capture.

Tripods — A tripod is a three-legged stand for a camera, used to stabilize and elevate the camera. A tripod is especially useful when using slow shutter speeds and/or telephoto lenses.

Lighting — The key to taking quality images with a digital camera is adequate, proper and even lighting. Another important consideration is to avoid light sources that raise the temperature of the physical item being digitized. It is recommended that a professional photographer or camera manufacturer be consulted when selecting lights.

For many photographers, the transition from film to digital photography was made easier by the DSLR which could still use the same lenses and lighting equipment as film cameras. Interchangeable lenses are used to adapt to different shooting conditions, making it possible to photograph large buildings or the smallest macro details of any object with the same camera. Because the camera synchronizes with studio lights, greater control of lighting is possible, resulting in a better image. Image capture is 1/60th of a second, as apposed to 30 seconds to 2 minutes for flatbed scanners or 5 to 11 minutes for capture using a large format scan back. Because of the capture rate of DSLR cameras, it is possible to photograph objects at a very efficient rate. For example, it would not be uncommon to photograph 200 prints per day using a well-designed workflow.
- Minimum Specifications for Digital Camera
  ◊ Minimum 12 MB RAW file size
  ◊ Interchangeable lenses
  ◊ Flash synchronization

      ◊   Ability to shoot tethered to computer
      ◊   Uses Adobe RGB 1998 color space
      ◊   Shoots in RAW format

- Additional equipment for system:
  - ◊ Color/grayscale target
  - ◊ Studio flash system and camera stand
  - ◊ Dedicated portable computer optimized for digital imaging
  - ◊ Tabletop copy stand with proper lighting
  - ◊ Book cradle
  - ◊ Studio space

In order to capture the appropriate image file size necessary for creating a sustainable archive and appropriate for all output possibilities (print, web and multimedia), the camera should have a minimum of 10 to 14 megapixels. This will generally result in a master TIFF file of around 50 to 70 megabytes. If the project requires higher resolutions, a medium or large format digital capture device must be used to produce files from 200 to 500 megabytes. While image quality and resolution are superb with these types of cameras, hardware and data storage costs go up exponentially while production rates go down significantly.

Software Considerations

Scanning Software

The link between your scanner and computer hardware is the software that both controls the scanner and passes scan information directly to your storage media, image editing software or both. Most scanners come equipped with software to manage this transfer for you. However, ensuring control of your scanner software is crucial to capturing a clean, unaltered master image. This generally means you will not make use of the scanning software's automatic settings. If available, opt for the "advanced" or "custom" settings mode, and disable features such as automatic unsharp masking, preset output levels for black and white, moire descreening and anything else which alters the image in any way. Advanced settings mode will allow you to control both the image captured and file formats generated by your scanner. If your scanner software does not allow advanced or custom settings, consider using your image editing software (i.e. Adobe Photoshop or similar) to import scanned images directly from the scanning hardware.

Image Editing Software

Once you have captured an image from your scanner and have saved an unaltered master file to your storage media, you will need to manipulate surrogate copies of that image using image editing software. Project managers will need professional image editing software. Adobe Photoshop is the industry standard for the creation of surrogates which can be delivered via the web, print publication or for in-house uses such as exhibits.

When selecting image editing software consider the following features:

- Ability to work directly with scanner software via TWAIN or other plug-ins
- Support for common nonproprietary file formats (see File Formats below)
- Tools for image optimization and repair (color adjustment, tonal adjustment, sharpen, resize, crop, rotate, clone/retouch, etc.)
- Ability to convert color spaces (RGB to CMYK for print output)
- Ability to undo edits, either one-at-a-time or multiple
- Ability to watermark images
- Usable interface and documentation, reliable technical support and/or an active community of users
- Ability to extend functionality through custom plug-ins
- Ability to create a batch sets of actions with macros for frequently applied functions
- Ability to apply a batch set of actions to increase productivity

Digital Asset Management (DAM)

DAM software assists project managers in managing simple to complex digital assets. Digital assets include image, text and/or multimedia files. DAM software saves time by providing a structured environment for projects. It is used to centralize the discovery and access of project files, track multiple projects, simplify workflow management and assist with rights management. Software is available for both stand-alone and network management of digital files or as add-on modules to existing library and collections management software.

Project managers considering digital asset management software should ask these questions:

- How does it fit with the overall project goals and workflows?
- Is it scalable to meet future needs?
- Does my current technology infrastructure support the software?
- Does the system work with all the current file formats and metadata schemes?
- What are the startup, maintenance and staffing costs?
- Is it possible to easily extract and migrate metadata and files?
- Does it support multiple users?
- Does it offer multiple user security levels?
- What search and retrieval functions does the system offer?
- Does the system offer customization of the public interface for image size and display of metadata fields?
- Does the system support emailing, saving or exporting images and accompanying metadata for web galleries, slide shows or reference?
- Does the system provide an e-commerce system?
- Does it support color management?
- Does the security schema support institutional guidelines?
- Does it have good transformation functionality to deliver surrogate images from master files?

Workspace

Providing a comfortable, safe and secure workspace for a digitization project can increase productivity and quality of images by reducing operator fatigue and potential damage to collections. Proper climate control and security are important if collections will stay in the lab for extended periods of time. A workspace for digitization should offer a controlled lighting source to maintain consistency and quality of images. Changes in room lighting can affect how images are represented on computer monitors and may introduce challenges to accurate calibration. Walls should be neutral-gray; room light levels should be low, but consistent; workers should not wear brightly colored clothing while working on images.

Optimal space for a small digital photography studio is three- to four-hundred square feet. If necessary, considerations for art handling must be taken into account when designing a studio workflow. If at all possible, the camera system, lighting and equipment should be mobile. There are times when it is better to move the studio to the collection rather than the other way around.

Guidelines for Creating Digital Images

We recognize that all collections differ in the ways they are used and accessed and that institutions have differing purposes and clientele, which will likely have an impact on how and for what purposes and reasons collections are digitized. Decisions on image quality and resolution should be based on the needs of users, how the images will be used and the nature of the materials you are scanning (dimensions, color, tonal range, format, material type, etc.). The quality and condition of the original (such as the quality of the shooting or processing technique in the case of photographs) impacts on the resolution at which you capture and the resulting quality of the digital image. These are not hard and fast recommendations for every collection and every institution. As a rule, the key to quality imaging is not to capture at the highest resolution possible, but to scan at a level that matches the informational content of the original.

The following guidelines contain the recommended settings for the creation of master, access and thumbnail images. Be aware that the combination of an increased bit depth and resolution will result in large to extremely large image files. These files may be difficult and/or costly to manage. Because of storage concerns and time considerations, it may be necessary to reduce the recommended resolution and/or bit depth when creating master images. Carefully consider institutional needs and capabilities, however, before departing from recommended best practices.

Institutions choosing to capture images at the recommended bit depth will need the CS (Creative Suite) versions if using Photoshop. Older Photoshop versions do not offer the full suite of editing tools for images captured at 16 bit or greater.

Master Files

Good digital imaging projects scan a high-quality master or archival image and then derive multiple versions in smaller sizes or alternative formats for a variety of uses. There are compelling preservation, access and economic reasons for creating an archival-quality digital master image: it provides an information-rich, research-quality surrogate. A high-quality master image will make the investment in the image capture process worthwhile. Since user expectations and technologies change over time, a digital master must be available and rich enough to accommodate future needs and applications. The master image should be the highest quality you can afford; it should not be edited or processed for any specific output; and it should be uncompressed. Intensive quality control should be applied in creating master image files. Any errors made while creating the master file will necessitate going back to the scanner or camera to capture another image.

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

- Nonproprietary / open source
- Large color space
- Uncompressed
- Technical metadata captured, as part of file structure

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

Service Master Files

The service master is an optimized working copy of the master file that can be used as a source for all subsequent derivatives. They are also used to create print publications. Like master files, service masters should also be stored in a nonproprietary, uncompressed file format that supports the fidelity and long-term preservation of the image. The service master is generally optimized in the following ways:

- Rotated, if necessary
- Cropped, if necessary
- Tone levels optimized
- Colors management applied, if necessary
- Sharpened using "unsharp mask," depending on the goals of the project
- Target cropped, when applicable, if included with the master file

Derivative Files

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

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

File Naming Conventions

Systematic file naming is important for system compatibility, interoperability and to demonstrate ownership of the digital asset. It is critical that your file names are unique, and it is recommended that they follow an established convention to assure consistency and ease of use. File naming recommendations include:

- Use lowercase letters of the Latin alphabet and the numerals 0 through 9.
- Avoid punctuation marks other than underscores and hyphens
- Begin each file name with a two- to three-character acronym representing the institutional name followed by a second two- to three-character acronym representing the department or unit name (when applicable).
- Follow the institutional and departmental acronyms with an object ID. The object ID consists of any unique numbering scheme already in use to represent the object or, if no such number exists, a short description representing the item
- File names should be limited to 31 characters, including the three character file extension
- If burning to CD-ROM, file names should be limited to 11 characters, including the three character file extension, in case a recipient's computer does not support long filenames
- Use a single period as a separator between the file name and the three letter extension
- Include a part designator after the object ID, when applicable

When selecting a file naming convention, think long-term. Select a system that will outlast staff involved in the current project. Consider the number of files your institution will ultimately be managing. Remember human error — if technicians will manually be assigning file names, how simple or easy will it be to make a mistake? Remember, file names do not take the place of metadata. Keep them simple and straightforward.

Watermarks

What is a watermark? A watermark is information stored in or on a digital image, which allows image creators to store copyright or branding information to images, audio and video files and documents. Watermarks are applied to images in hopes of reducing misuse or unauthorized distribution of images.

There are two types of watermarks in use for digital image files, visible and electronic:
- Visible watermarks are applied on top of the image, very much like a seal is applied to an official document. Often these watermarks consist of the name of the institution who owns the file, that institution's official seal or some other identifying logo. In all cases, visible watermarks cover a portion of the image file. Visible watermarks do not stop users from downloading files, and they can be removed, depending on the complexity, size and color value of the design. The biggest drawback to visible watermarks is the obstruction to parts of the image, making the use of that image file less appealing to some researchers.

- Electronic watermarks are imbedded in the image file, and they are invisible. They usually use a numeric code licensed by an electronic watermarking firm. The numeric code is specific to the institution that owns the files. Electronic watermarks are usually applied as part of the filter mechanism in programs like Adobe Photoshop. In some cases, upon very close inspection, the file will appear grainy after an electronic watermark is applied. Electronic watermarks do not stop users from downloading files, and they can be easily overcome through manipulation.

The cost of watermarking varies. Visible watermarking is virtually free, but invisible watermarking can be costly. If you are debating whether or not to watermark, consider controlling the use of your images by limiting the quality and size of publicly accessible files. Be sure metadata concerning ownership and copyright information travels with your images, either in embedded file information or associated metadata records.

An advantage to using watermarks, visible or electronic, is that they can assist in controlling the use and distribution of images. A disadvantage is that, when embedded they will degrade the image and when layered on top, will most certainly obscure image content. Watermarks may suggest the institution's intent to protect its collections, but they do not prevent theft or misuse.

Guidelines for Source Type

For the following source types, recommendations are made for master, access and thumbnails. Be aware that access files should be made from an optimized master file — see Service Master Files described above.

- **Text** — When scanning text documents, spatial resolutions should be based on the size of text found in the document and resolutions should be adjusted accordingly. Documents with smaller printed text may require higher resolutions and bit depths than documents that use large typefaces. Projects that will have Optical Character Recognition (OCR) applied, may wish to test pages at several resolutions to find the most satisfactory results. Images that produce the best results for OCR may not be pleasing to the eye and may require separate scans for OCR and human display.

  Projects with large amounts of textual materials, particularly hard-to-read materials such as manuscripts, should provide transcriptions of the materials in addition to the digital image. Access to textual material can be further enhanced through SGML/XML markup schemes such as the Text Encoding Initiative (TEI). As rekeying text can be cost prohibitive, projects considering transcriptions should investigate including Optical Character Recognition (OCR) software in their toolkit.

Text			
	Master	**Access**	**Thumbnail**
File Format	TIFF	JPEG	JPEG
Bit Depth	1 bit bitonal 8 to 16 bit grayscale 48 bit color	8 bit grayscale 24 bit color	8 bit grayscale 24 bit color
Spatial Resolution	Adjust scan resolution to produce a minimum pixel measurement across the long dimension of 6,000 lines for 1 bit files and 4,000 lines for 8 to 16 bit files	150 – 200 PPI	144 PPI
Spatial Dimensions	4000 to 6000 pixels across the long dimension	600 pixels across the long dimension	150 to 200 pixels across the long dimension

- **Photos** — Photographs can present many digitization challenges. We recommend digitizing from the negative (or the earliest generation of the photograph) to yield a higher-quality image. However, in the case of photographs developed according to artist specifications, the photograph itself should be digitized rather than the negative.

When considering whether to capture sepia-tone photographs in color or black and white, we

Cowboys on a ridge in northern Wyoming, undated. Charles J. Belden Papers, American Heritage Center, University of Wyoming.

recommend digitizing them as color images to create a more accurate image. Digitize the backs of photographs as separate image files if there is significant information on the back of the photo (which may be of interest to users) that may not be included elsewhere. If an image of the verso of the photograph is available, the digital image will serve as a more successful surrogate for the original.

Photographs			
	Master	**Access**	**Thumbnail**
File Format	TIFF	JPEG	JPEG
Bit Depth	16 bit grayscale 48 bit color	8 bit grayscale 24 bit color	8 bit grayscale 24 bit color
Spatial Resolution	400 to 800 PPI	150 to 200 PPI	144 PPI
Spatial Dimensions	4000 to 8000 pixels across the long dimension, depending on size of original, excluding mounts and borders	600 pixels across the long dimension	150 to 200 pixels across the long dimension

- **Graphics** — Graphics include the various techniques used to reproduce words and images from originals such as engraving, lithography, line art, graphs, diagrams, illustrations, technical drawings and other visual representations. Nearly all graphics will be two dimensional and should be scanned using the following guidelines.

Graphic Materials			
	Master	**Access**	**Thumbnail**
File Format	TIFF	JPEG	JPEG
Bit Depth	16 bit grayscale 48 bit color	8 bit grayscale 24 bit color	8 bit grayscale 24 bit color
Spatial Resolution	600 to 800 PPI	150 to 200 PPI	144 PPI
Spatial Dimensions	6000 to 8000 pixels across the long dimension, excluding mounts and borders	600 pixels across the long dimension	150 to 200 pixels across the long dimension

- **Artwork/3-Dimensional Objects** — For projects where the physical dimensions of the non-3D artwork matches the equipment available, the following standards can be used. If scanning photographic copies of objects and artifacts, see recommended requirements in the appropriate photo and film charts above.

Artwork/3-Dimensional Objects			
	Master	**Access**	**Thumbnail**
File Format	TIFF	JPEG	JPEG
Bit Depth	48 bit color	24 bit color	24 bit color
Spatial Resolution	Device Maximum	300 PPI	144 PPI
Spatial Dimensions	100% of original	600 pixels across the long dimension	150 – 200 pixels across the long dimension

- **Maps** — Scanning maps may involve items that vary widely in size, condition and amount of detail. Small maps may fit easily onto a flatbed scanner, while large plat maps may need to be scanned in sections using a large format scanner or captured by a camera. The size of the image can become a problem for storage, but also for viewing, serving over the web or processing.

  Smaller maps (less than 36 inches on the longest dimension) should be digitized at 600 PPI, 48-bit color or 16-bit grayscale if possible. For larger maps, 300-400 PPI may be more practical. If it becomes necessary to digitize a map in sections and stitch the image together in Photoshop, keep both the original images of the sections as well as the combined image.

Maps			
	Master	**Web**	**Thumbnail**
File Format	TIFF	JPEG	JPEG
Bit Depth	16 bit grayscale 48 bit color	8 bit grayscale 24 bit color	8 bit grayscale 24 bit color
Spatial Resolution	600 PPI 300 to 400 PPI for larger maps	150 to 200 PPI	144 PPI
Spatial Dimensions	6000 to 8000 pixels across the long dimension	1078 pixels across the long dimension	150 to 200 pixels across the long dimension

- **Film** — For duplicates (negatives, slides, transparencies), match the original size. However, if original size is not known, the following recommendations are supplied: For a copy negative or transparency, scan at a resolution to achieve 4000 pixels across the long dimension. For duplicates, follow the scanning

recommendations for the size that matches the actual physical dimensions of the duplicate.

Master scans of camera originals may be captured and saved in RGB, particularly those negatives that contain color information as a result of staining, degradation or intentional color casts. Derivative files could later be reduced to grayscale in the scanning software or during post-processing editing.

Film			
	Master	**Access**	**Thumbnail**
File Format	TIFF	JPEG	JPEG
Bit Depth	16 bit grayscale 48 bit color	8 bit grayscale 24 bit color	8 bit grayscale 24 bit color
Spatial Resolution	Resolution to be calculated from actual image format and/or dimensions - approx. 2800 PPI for 35mm originals, ranging to approx. 600 PPI for 8x10 originals	150 to 200 PPI	144 PPI
Spatial Dimensions	4000 to 8000 pixels across long dimension of image area, depending on size of original and excluding mounts and borders	600 pixels across the long dimension	150 to 200 pixels across the long dimension

Quality Control

Numerous factors play an important role in the final outcome of a digitization project. Original condition of materials, quality and maintenance of equipment, staff training and external lighting are some factors that can influence the quality of images.

A quality control program should be conducted throughout all phases of the digital conversion process. Inspection of final digital image files should be incorporated into your project workflow. Typically, master image files are inspected online for a variety of defects. Depending on your project, you may want to inspect 100 percent of the master images or 10 percent of the files randomly. We do recommend that quality control procedures be implemented and documented and that you have clearly defined the specific defects that you find unacceptable in an image. Images should be inspected while viewing at a 1:1 pixel ratio or at 100 percent magnification or higher. Quality is

evaluated both subjectively, by project staff (scanner operator, image editors, etc.) through visual inspection, and objectively, in the imaging software (by using targets, histograms, etc.).

Tonal Dynamic Range

As noted earlier, one of the most significant factors affecting image quality is the tonal dynamic range — the color space an image occupies between pure black (0) and pure white (255). Reviewing histograms at the time of capture can ensure that all of the image's information is being recorded. When the white and black points are not set on true white and black during the set up of the capture, clipping and spiking can occur.

White & Black Points

Optimum placement of whites and blacks is best observed through the histogram, although the image itself will be examined without the histogram on the screen. It is important to look at the number value assigned to the brightest highlight and the darkest shadow. Highlights should not read a number value higher than 247, and shadows should not be less than 7 or 8. If these numbers are exceeded, the scan must be redone. This is particularly vital if the original image has a short dynamic range. The white and black points must not be set on 0 and 255, as this will stretch the dynamic range of the image, creating gaps in the histograms, and thus unusable scans. These adjustments need to be made at the time of scanning, since adjusting images after the fact in image editing software introduces interpolation of the dynamic range (guessing what the points are in the image) and frequently results in gaps in the histogram rather than continuous curves.

Clipping and Spiking

Clipping and spiking result when the white and black points are not set on true white and black during the set up of the scan. If white and black are improperly set, everything above or below those points is clipped, or registers as the same tone. Spiking on the ends of the histogram usually indicates clipping. This problem also shows up in the image itself as blockage and pixelization in the shadows and blowouts in the highlights. Acceptable spikes can occur if the edge of the original negative has lost emulsion, for example, or the sky holds no detail and is one tone in the original. Such instances, however, are rare.
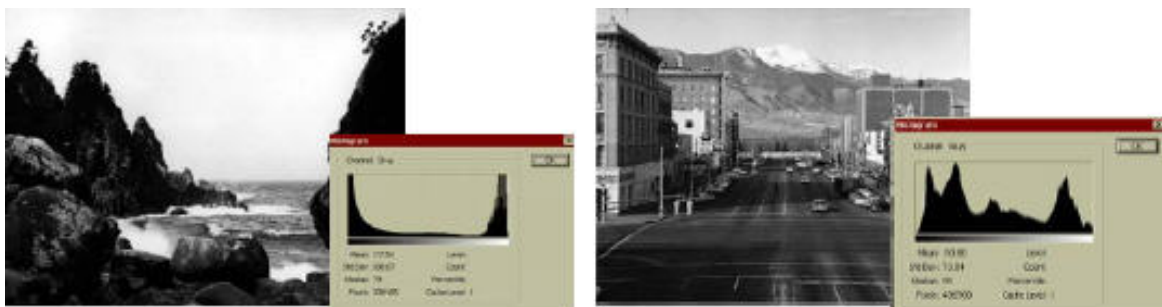
Detail in the image on the left has been lost due to improperly adjusted white and black points. The histogram shows the "clipping and spiking" associated with incorrect points (*Rocky coastline on Forrester Island, 1920; courtesy Denver Museum of Nature and Science.*). The image on the right shows properly adjusted white and black points with no clipping or spiking in the histogram (*Downtown Colorado Springs, 1964; courtesy Pikes Peak Library System*).

Color Management

Color management can be one of the most challenging aspects of the digitization workflow. Each piece of hardware in the path from source to digital file can introduce biases of color and tone. The goal of color management procedures is to accurately and predictably compensate for each of these biases across the entire workflow from scanner to print output or monitor display.

Applying the practices outlined below will allow you to fine tune the color captured in the image. This assumes, however, that prior to capture, accurate lighting is in place, capture equipment and display devices have been regularly calibrated and a proper light source is available to accurately view analog materials. Take time to evaluate and test all your equipment and accessories before image capture so that you can minimize, if not eliminate, the need for extensive post-capture color management.

Projects not undertaking color management should be aware that equipment will introduce color biases into any digitized materials. Attempting to make digitized materials "look good" on uncalibrated equipment may introduce these biases into the master images. Projects without a color management system should use available tools to perform basic monitor calibrations:

- Set to 24 millions of colors
- Set monitor Gamma at 2.2 (including Macintosh computers, which by default are set at 1.8 gamma)
- When using a CRT monitor, calibrate your color temperature to 6500° K
- If using an LCD monitor, calibrate your color temperature to 5500° K

Targets and Color Bars

Targets and color bars are used to measure system resolution, tonal range and color fidelity. Including targets in a digitization workflow allows color management systems to create profiles for each device or for later adjustment in projects not implementing color management during image capture. Targets are a way of predicting image quality and help ensure that the imaging system you are using is producing the best quality image it can and is operating at a consistent level of quality over time.

Color targets are a scientifically created set of color patches with established numerical values and neutral gray. They help to ensure that the imaging system you are using is effective and accurate and performing in a consistent level of quality over time. Targets for prints and transparencies exist, and targets appropriate for the materials being scanned should be used (paper, film, transparency, etc.).

Targets usually contain patches of color, black and white or shades of gray for verifying tone reproduction. Resolution targets allow projects to measure the level of detail a particular piece of equipment can capture. Resolution targets can be helpful in evaluating equipment before purchase or assessing the quality of output from a vendor.
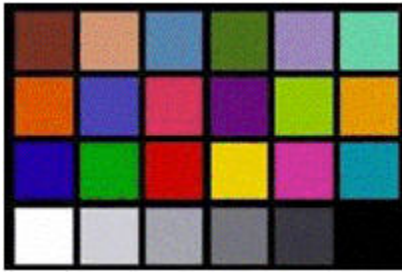
As equipment biases can shift over time, best practice is to include targets on a regular basis throughout the course of the project. Some digitization projects also scan a color bar along with the original, to be included in the final digital image, to aid users in verifying accuracy in color reproduction. See Appendix E — Data Center Basics for a tutorial on how to use targets and color bars.

- GretagMacbeth 24-patch ColorChecker (shown above)
- Kodak Q-60, IT8.7
- Kodak Q-13 and Q-14

Visual Inspection

Things to look for during archival master visual inspection may include: (Note: If these attributes, with the exception of file name, are met, then you do not need to recapture the item.)

- Image is the correct size
- Image is the correct resolution
- File name is correct
- File format is correct
- Image is in correct bit depth and color mode (i.e., color image has been scaled as grayscale)
- No loss of detail in highlight or shadows
- No excessive noise especially in dark areas or shadows
- Even tonal values, no flare
- Correct focus
- Not pixilated
- Excessive dust spots or other objects
- No digital artifacts (such as very regular, straight lines across picture)
- Image not cropped
- Image not rotated or reversed
- Correct color balance
- Histogram:
    ◊ No spikes or clipping
    ◊ No tonal values lower than 9 or higher than 247

If levels and curves are adjusted in the service master, then you should check the histogram again.

The service master will be checked for:
- Image is the correct size
- Image is the correct resolution
- File name is correct
- File format is correct

- Image is in correct bit depth and color mode (i.e., color image has been scaled as grayscale)
- No loss of detail in highlight or shadows
- No excessive noise especially in dark areas or shadows
- Even tonal values, no flare
- Correct focus
- Not pixilated
- No digital artifacts (such as very regular, straight lines across picture)
- No moire patterns (wavy lines or swirls, usually found in areas where there are repeated patterns, such as half-tone dots)
- Image cropped correctly
- Image rotated correctly and not reversed
- Image centered and not skewed
- Correct color balance
- Histogram:
  ◊ No spikes or clipping
  ◊ No tonal values lower than 9 or higher than 247

Describing Digital Images

Metadata Theory and Practice

The creation of good, high-quality metadata is a key component for the responsible management and long-term preservation of digital files. The term metadata is inclusive of the five different types of metadata information — descriptive, technical, administrative, structural and preservation.

Metadata issues must be given a high priority during the design of projects or at the outset of a program to ensure information about the collections can be found by users, shared with other institutions and managed over the life of the object. Metadata creation for digital objects can reuse preexisting descriptive information from its physical counterpart, such as accession records, finding aids or catalog records. For some digital objects, metadata creation will need to be created from scratch. There are numerous metadata schema; therefore, the institution will need to select the schema most appropriate to the collection, audience and institution. The most common metadata schemes include the following sets of information, with some overlap:

Descriptive Metadata	Metadata that describes the intellectual content of a resource and used for the indexing, discovery and identification of a digital resource.
Administrative Metadata	Metadata that includes management information about the digital resource, such as ownership and rights management.
Structural Metadata	Metadata that is used to display and navigate digital resources and describes relationships between multiple digital files, such as page order in a digitized book or diary.
Technical Metadata	Metadata that describes the features of the digital file, such as resolution, pixel dimensions and hardware. The information is critical for migration and long-term sustainability of the digital resource.

The Dublin Core metadata standard has been widely implemented, containing 15 elements. Different communities have developed Dublin Core profiles for their unique needs, adding elements to reflect those unique needs. Another standard which may be applicable to your program is VRA Core 4.0, a data standard for the cultural heritage community that was developed by the Visual Resources Association. See http://www.vraweb.org/projects/vracore4/index.html.

The CDP's *Dublin Core Metadata Best Practices* provides guidance in using Dublin Core, in a collaborative environment. See http://www.bcr.org/cdp/best/dublin-core-bp.pdf.

Preservation Metadata and METS

Preservation metadata is information that supports and documents activities related to digital preservation.[10] It is information that is used in supporting the processes of ensuring the core preservation processes of availability, identity, understandability, authenticity, viability and renderability. While some of these activities require descriptive and structural metadata, the majority of it is administrative metadata.

The preservation of digital objects requires extensive documentation that accumulates over the life of the digital object. Management of this metadata is handled through the PREservation Metadata Implementation Strategies (PREMIS), a set of core preservation metadata likely to be needed to be used to manage digital resources in a digital repository. Due to the limited implementation of PREMIS, Priscilla Caplan, Chair of the PREMIS working group, notes, "…nobody knows whether it works. That is, there has not been enough experience applying preservation strategies to know whether today's preservation metadata schemes actually support the process of long term preservation." [11]

OAIS requires packaging information to describe the various types of OAIS information — ingest dissemination, etc. Within the cultural heritage community, Metadata Encoding and Transmission Standard (METS), is the mostly commonly used standard. An XML-based schema, METS is designed to be an overall framework in which all metadata associated with a digital object can be stored. The following describes the components of METS.

Descriptive Metadata — The descriptive metadata section may point to descriptive metadata external to the METS document (e.g., a MARC record in an OPAC or an EAD finding aid maintained on a WWW server) or contain internally embedded descriptive metadata or both. Multiple instances of both external and internal descriptive metadata may be included in the descriptive metadata section.

Administrative Metadata — The administrative metadata section provides information regarding how the files were created and stored, intellectual property rights, metadata regarding the original source object from which the digital object derives, and

---

[10] Caplan, Priscilla, "The Preservation of Digital Materials," *Library Technology Reports* 44:2 (2008):15.

[11] Caplan, 16.

information regarding the provenance of the files comprising the digital object (i.e., master/derivative file relationships, and migration/transformation information). As with descriptive metadata, administrative metadata may be either external to the METS document or encoded internally.

File Groups — The file group section lists all files comprising all electronic versions of the digital object. File group elements may be nested, to provide for subdividing the files by object version.

Structural Map — The structural map is the heart of a METS document. It outlines a hierarchical structure for the digital library object and links the elements of that structure to content files and metadata that pertain to each element.

Behavior — A behavior section can be used to associate executable behaviors with content in the METS object. A behavior section has an interface definition element that represents an abstract definition of the set of behaviors represented by a particular behavior section. A behavior section also has a behavior mechanism which is a module of executable code that implements and runs the behaviors defined abstractly by the interface definition.

The Library of Congress provides maintenance services both PREMIS and METS working closely with national and international organizations.

Storage

Best Practices for Image Scalability and Long-term Viability

Any discussion of sustainable digital projects must include an overview of the infrastructure basics of a data center. While digital projects, i.e. scanning and digital photography, can be done by smaller institutions, it is in their favor when they are able to partner with a larger organization that is able to take their "work product" (digital objects and descriptive metadata) and place the content in a data center setup to manage data backups and has a plan to preserve the digital objects and metadata. See Appendix E — Data Center Basics for additional and more detailed. See .Appendix F — Network Design for Data Protectionfor information about Network Design for Data Protection.

Solid network design is critical to allow access to and to protect the digital objects and the metadata that go into the making of quality digital projects. An exhaustive treatise on network design is not in the scope of this document, but a general overview is important to give the reader the basics to protect their data. Most complex, well-designed networks are created by network engineers in organizations.

The design of a basic TCP/IP network model consists of the following five layers. They are: the application layer; the transport layer; the network layer; the data link layer; and the physical layer. The layers near the top are closer to the user application. Those layers located near the bottom of the network are logically closer to the physical storage and transmission of the data. For additional information about Network Architecture, see Appendix F — Network Design for Data Protection.

RAID Array

Storing scanned images on "live" servers is now a viable alternative to archiving high resolution TIFF images on CD-ROMs. Proper configuration of a server or workstation with RAID arrays can help sustain data without loss. However, this must be paired with an off-site archive to guard against catastrophic events.

RAID is an acronym for Redundant Array of Inexpensive (or Independent) Disks. A RAID array consists of a number of drives which collectively act as a single storage system and can tolerate the failure of a drive without losing data These drives operate independently of each other. The number of disk drives on the server or even a workstation PC set up as a storage device, will help determine how the RAID array on the equipment will be formatted.

A research group at UC-Berkeley coined the term RAID, initially defining six RAID levels. Each level is a different way to spread data across multiple drives — a compromise between cost and speed. Understanding these levels is important, because each level is optimized for a different use. Even though a RAID array offers another layer of data protection, one still needs to do regular backups and have an overall plan to transition and preserve the digital objects and metadata over time. Setting up the proper RAID array on a large-scale storage device or a workstation PC setup as a storage device, is a basic way to insure image and metadata sustainability.

Note: There are about 10 types of RAID configurations. New RAID configurations are constantly being presented to the standards association, the RAID Advisory Board. RAID levels 0 through 5 are the most commonly used configurations. The different RAID levels have different strong points, with some configurations better suited for speed, others for capacity and still others for fault tolerance. For examples and descriptions of RAID 0 – 5, see Appendix G — RAID Array.

Online Storage/Storage

Recently, a number of new developments in hardware and software applications have impacted the way images are stored, made accessible and preserved for the long term. The amount of affordable "live storage" one can purchase has increased while the price has come down.

Network security layer options have proliferated as have the options to secure both the metadata and digital objects. Improvements to data center workflows in the form of automated software applications able to manage routine tasks, such as backups or the movement of data/digital objects across platforms on layers of servers running different operating systems, have made it easier and less costly for data centers to conduct business. Routine machine setup or maintenance and data migrations that used to take days or hours to accomplish now can be run in very short time periods on software/hardware combinations that enable cross-operating systems management and server virtualization. While running large-scale data centers still requires a highly, well-trained technical staff, smaller organizations can take advantage of some of these image and metadata sustainability software/hardware tools and practices to ensure that

their hard work remains viable. For more detailed information about online storage systems, see Appendix H — Online Storage/Storage.

Appendix A — Bibliographical and Online Resources

Beal, Vangie 'Aurora.' "All About Monitors CRT vs LCD." (February 18, 2005), http://www.webopedia.com/DidYouKnow/Hardware_Software/2005/all_about_monitors.asp (accessed May 21, 2008)

Boyle, Francis. "What's All This About Digital Preservation: Challenges, Methods, Realities." http://www.rms-gb.org.uk/download/741 (accessed June 26, 2008)

BCR's Collaborative Digitization Program. "Legal Issues to Consider When Digitizing Collections" BCR. http://www.bcr.org/cdp/digitaltb/laws/index.html (accessed May 21, 2008)

BCR's Collaborative Digitization Program. "Questions to Ask Before Your Begin a Digitization Project." BCR. http://www.bcr.org/cdp/digitaltb/getting_started/faq.html (accessed May 21, 2008)

BCR's Collaborative Digitization Program. "Standards" http://www.bcr.org/cdp/digitaltb/metadata/standards.html (accessed May 21, 2008)

Buckley, Ph.D., Robert. "Technology Watch Report JPEG 2000 - a Practical Digital Preservation Standard?" DPC Technology Watch Series Report 08-01 (February 2008), http://www.dpconline.org/docs/reports/dpctw08-01.pdf (accessed June 23, 2008)

Caplan, Priscilla, "Preservation Pyramid – Authentication." http://www.rms-gb.org.uk/download/741 (accessed June 26, 2008)

Carter, John. "ABC for Book Collectors." http://digitalcommons.unl.edu/ir_information/14/ (accessed June 23, 2008)

CDP Metadata Working Group, "Dublin Core Metadata Best Practices Version 2.1.1. (September 2006), http://www.bcr.org/cdp/best/dublin-core-bp.pdf (accessed June 23, 2008)

CDP Metadata Working Group, "Final report of the Collaborative Digitization Program's Metadata Standards Working Group Task Force on the Metadata Encoding and Transmission standard (METS)." (May 2006). http://www.bcr.org/cdp/projects/soundmodel/docs/AppendixD.pdf (accessed June 23, 2008)

The Cedars Project. "CURL exemplars in digital archives" Cedars Document Number: PSW01" (May 13, 1999), http://www.leeds.ac.uk/cedars/documents/PSW01.htm (accessed June 23, 2008)

Conway, Paul. "Reformatting: The Relevance of Preservation in a Digital World."
http://www.nedcc.org/resources/leaflets/6Reformatting/04RelevanceOfPreservation.php
(accessed June 23, 2008)

Cornell University Library/ Research Department. "Moving Theory into Practice: Digital Imaging Tutorial" (2000-2003),
http://www.library.cornell.edu/preservation/tutorial/contents.html (accessed June 23, 2008)

CRL, The Center for Research Libraries. "Trustworthy Repositories Audit & Certification: Criteria and Checklist (TRAC) Version 1.0." (February 2007),
http://www.crl.edu/PDF/trac.pdf (accessed June 23, 2008)

Digital Commons @UNL. http://digitalcommons.unl.edu/ir_information/14/ (accessed June 30, 2008)

"Digital Negative (DNG) Specification, version 1.1.0.0." (February 2005).
http://www.adobe.com/products/dng/pdfs/dng_spec.pdf (accessed June 23, 2008)

Fanning, Betsy A., "Technology Watch Report Preserving the Data Explosion: Using PDF" DPC Technology Watch Series Report 08-02 (April 2008).
http://www.dpconline.org/docs/reports/dpctw08-02.pdf (accessed June 23, 2008)

Gookin, Dan. "Deciding Between a CRT and an LCD Monitor." YAHOO Tech (2006),
http://tech.yahoo.com/xg/deciding-between-a-crt-and-an-lcd-monitor/153241 (accessed May 21, 2008)

Kyrnin, Mark. "CRT vs LCD Monitors: Which Monitor is the Best to Buy." About.com: PC Hardware Reviews (2008),
http://compreviews.about.com/od/multimedia/a/CRTvsLCD.htm (accessed May 21, 2008)

Lavoie, Brian, and Richard Gartner. "Preservation Metadata." DPC Technology Watch Series Report 05-01. (September 2005),
http://www.dpconline.org/docs/reports/dpctw05-01.pdf (accessed June 23, 2008)

Ludington, Jake. "Optimize your PC for Audio and Video"
http://www.jakeludington.com/ask_jake/20050225_optimize_your_pc_for_audio_and_video.html (accessed May 21, 2008)

Management Council of the Consultative Committee for Space Data Systems (CCSDS). "Reference Model for an Open Archival Information System (OAIS) CCSDS 650.0-B-1 Blue Book, Issue 1." (January 2002).
http://public.ccsds.org/publications/archive/650x0b1.pdf (accessed June 23, 2008)

NTAC (Nebraska Technical Assistance Center),
http://ianrpubs.unl.edu/consumered/ec808.htm. (accessed June 30, 2008)

Puglia, Steven, Jeffrey Reed, and Erin Rhodes. "U.S. National Archives and Records Administration (NARA) Technical Guidelines for Digitizing Archival Materials for Electronic Access: Creation of Production Master Files – Raster Images." (June 2004), http://www.archives.gov/preservation/technical/guidelines.pdf (accessed June 23, 2008)

Rog, Judith. "PRF Guidelines Recommendations for the creation of PDF files for long-term preservation and access. " National Library of the Netherlands/Koninklijke Bibliotheek (KB) Version 1.7 (May 31, 2007). http://www.kb.nl/hrd/dd/dd_links_en_publicaties/PDF_Guidelines.pdf (accessed June 23, 2008)

"Sustainability of Digital Formats: Planning For Library of Congress Collections." Adobe Digital Negative (DNG), version 1.1. http://www.digitalpreservation.gov/formats/fdd/fdd000188.shtml (accessed June 23, 2008)

The Technical Advisory Service for Images (TASI) Online resource providing advice and guidance to assist in and encourage the creation of high quality digital image collections. http://www.tasi.ac.uk/advice/creating/creating.html (accessed June 23, 2008)

"The Universal Photographic Digital Imaging Guidelines Version 3.0." edited by Michael Stewart and Greg Smith. (September 2007), http://www.updig.org/guidelines/index.php (accessed June 23, 2008)

Appendix B — Digital Imaging Glossary

Archival or Master Image

A digital image that is uncompressed, unmanipulated and high-quality, retaining the best possible color information. It is meant to have lasting utility. Archival images are usually saved in a nonproprietary format and kept off-line in a secure environment. Archival or Master images are of a higher resolution and quality than the digital image delivered to the user on-screen.

An archival master accurately represents the original analog item and has an associated metadata record which provides descriptive, administrative, and technical information concerning the original and digital items.

BMP

Microsoft Windows bitmap format. This is the only graphics format where compression actually enlarges the file

Best Practices

Guiding practices based around continuous learning and continual improvement.

Bit Depth

Bit depth quantifies how many unique colors are available in an image's color palette in terms of the number of 0's and 1's, or "bits."

CCD Array

Charge-Coupled Device array. Light sensitive diodes used in scanners and digital cameras that sweep across an image during capture and, when exposed to light, generate a series of digital signals that are converted into pixel values.

Clipping and Spiking

Clipping and spiking result when the white and black points are not set on true white and black during the set up of the capture. If white and black are improperly set, everything above or below those points is "clipped" or registers as the same tone.

Spiking on the ends of the histogram usually indicates clipping. This problem also shows up in the image itself as blockage and pixelization in the shadows and blowouts in the highlights. Acceptable spikes can occur if the edge of the original negative has lost emulsion, for example, or the sky holds no detail and is one tone in the original. Such instances, however, are rare.

Color Gamut

The complete range of hues and strengths of colors that can be achieved with a given set of colorants such as cyan, magenta, yellow, and black inks on a specific substrate.

CMYK

Short for Cyan-Magenta-Yellow-Black, and pronounced as separate letters. CMYK is a color model in which all colors are described as a mixture of these four process colors. CMYK is the standard color model used in offset printing for full-color documents. Because such printing uses inks of these four basic colors, it is often called four-color printing.

In contrast, display devices generally use a different color model called RGB, which stands for Red-Green-Blue. One of the most difficult aspects of desktop publishing in color is color matching -- properly converting the RGB colors into CMYK colors so that what gets printed looks the same as what appears on the monitor.

Compression/Decompression

The reduction of image file size for processing, storage, and transmission. The quality of the image may be affected by the compression techniques used and the level of compression applied. Decompression is the process of retrieving compressed data and reassembling it so that it resembles its original form before compression.

There are two types of compression:

- Lossless compression is a process that reduces the storage space needed for an image file without loss of data. If an image has undergone lossless compression, it will be identical to the image before it was compressed.
- Lossy compression is another process that reduces the storage space needed for an image file, but it discards information (information that is "redundant" and not perceptible to the human eye). If an image that has undergone lossy compression is decompressed, it will differ from the image before it was compressed, even though the difference may be difficult for the human eye to detect.

There are both standard and nonstandard compression techniques available. In general, it is better to employ a compression technique that is supported by standards, is nonproprietary, and maintained over time. In selecting a compression technique, it is necessary to consider the attributes of the original. Some compression techniques are designed to compress text, others are designed to compress pictures.

Cultural Heritage Institutions

The term cultural heritage is used in a broad sense to include scientific, industrial, social, and cultural heritage. Libraries, museums, archives, societies and other repositories that house and care for material fall within these broad categories.

Derivative, Surrogate or Access Images

Digital images created from another digital image through some kind of automated process, usually involving a loss of information. Techniques used to create derived images include sampling to a lower resolution, using lossy compression techniques, or altering an image using image processing techniques.

Digital Artifacts

Visual digital effects introduced into an image during scanning that do not correspond to the original image being scanned. Artifacts might include pixilation, dotted or straight lines, regularly repeated patterns, moire, etc.

Digital Image

An electronic photograph, made up of a set of picture elements ("pixels"). Each pixel is assigned a tonal value (black, white, a shade of gray, or color) and is represented digitally in binary code (zeros and ones). The term "image" does not imply solely visual materials as source material; rather, a digital image is simply a representation of whatever is being captured, whether it be manuscripts, text, photographs, maps, drawings, blueprints, halftones, musical scores, 3-D objects, etc.

Digital Preservation

Digital Preservation "is a set of managed activities necessary for ensuring both the long-term maintenance of a byte stream and continued accessibility of its contents." Digital preservation involves the use of technology that guarantees the continued viability of the digital objects and metadata as well as the administrative policies and procedures to make sure steps are followed.

The following selected digital preservation definitions are based on OAIS terminology:
- Data: Information of any kind represented in computer storage.
- Digitization: The process of converting a nondigital object into a digital object with associated metadata.
- Digital Image: Any form of "electronic photograph" composed of a grid of pixels or other binary data directly representing the appearance or sound of a non-digital resource.
- Digital Imaging: The process whereby non-electronic materials (books, manuscripts, newspapers, pictures, photographs, sounds, performances...) are scanned, photographed, or otherwise automatically processed to generate digital files.
- Digital Object: An object composed of a set of bit sequences. A digital object is an object that can be represented by a computer; this may include databases, spreadsheets, word processor documents, web pages, video, audio, images, maps, 2 and 3-d models etc.
- Digital Preservation/Digital Archiving: Storage, maintenance and access to digital objects/materials over the long term.
- Digital Archive/Repository: A collection of digital objects stored for preservation purposes.
- Long Term: A period of time long enough for there to be concern about the impacts of changing technologies, including support for new media and data formats, and of a changing user community, on the information being held in a repository. This period extends into the indefinite future.[12]

---

[12] The Cedars Project. "CURL exemplars in digital archives" Cedars Document Number: PSW01" (May 13, 1999), http://www.leeds.ac.uk/cedars/documents/PSW01.htm

The following digital preservation concepts published by NEDCC provide a starting point for institutions developing preservation plans:

- Longevity — Digital objects on media will probably outlast the system's ability to access, retrieve and interpret the media. Digital librarians and Archivists must be prepared to migrate the valuable image data, indexes, and software to future generations of the technology. A migration and scheduling plan as well as a budget should be developed to accommodate these data. Having access to well designed and robust live storage media is key to influencing the life expectancy of the information.
- Choice — Curators, Librarians and Archivists ultimately decide, based on their experiences and customer demands and needs, which materials to convert from analog into digital objects. Since the cost of doing this is high and time consuming, deciding on which items in their collections make sense to digitize is an important collection management decision. A digital preservation plan should include a digital collection development plan.
- Quality — Practitioners should treat image and data quality as the heart and soul of digital preservation. This means that those involved in the practice should work to maximize the amount of data captured in the digital imaging process, document image enhancement techniques, and specify file compression routines that do not result in the loss of data as the digital objects and descriptive metadata travel across the network.
- Integrity — In the digital world the primary concern of integrity has to do with documenting the real or potential loss of information when a file is originally created and then compressed mathematically or sent across a network. In the digital world preserving intellectual integrity involves an authentication processes. A digital audit trail procedure should be in place to make sure files are not altered either intentionally or by accident. Practitioners can control integrity of digital image files by authenticating access procedures and documenting successive modifications to a given digital record. Librarians and Archivists should also play a role in influencing the development of metadata interchange standards -- including the tools and techniques that will allow structured, documented, and standardized information about data files and databases to be shared across platforms and systems.
- Access — In the digital world, access is transformed so that it is no longer a byproduct of the preservation process but its central theme. Control over the access requirements of digital preservation, especially, the ability to migrate digital image files to future generations of new technology, can be exercised by purchasing nonproprietary hardware and software. Practitioners should encourage vendors to adopt open system architectures and/or equipment that is "backwardly compatible" with existing systems. [13]

---

[13] Conway, Paul. "The Relevance of Preservation in a Digital World," Northeast Document Conservation Center. (2007). http://www.nedcc.org/resources/leaflets/6Reformatting/04RelevanceOfPreservation.php

Diodes

Light-sensitive electronic components used by the scanner during image capture. Diodes sense the presence or absence of light and create a digital signal that the computer then converts into pixel values.

Dots Per Inch or DPI

A measure of resolution used for printed text or images and monitor display.

Dynamic Range (Bit Depth)

The number of colors or shades of grey that can be represented by a pixel. The smallest unit of data stored in a computer is called a bit. Dynamic range is a measurement of the number of bits used to represent each pixel in a digital image. 1 bit or bitonal means that a pixel can either be black or white. Bitonal imaging is good for black and white images, such as line drawings and text. However, scanning in grayscale rather than bitonal may produce a better looking image. 8 bit color or 8 bit grayscale means that each pixel can be one of 256 shades of color or one of 256 shades of gray. 24 bit color means that each pixel can be one of 16.8 million colors. When moving to 48 bit color the available colors in the display will number 2800 times one trillion. This is an enormous number of colors. Plus a higher bit resolution can display more shades of gray. With 30 bit color depth, four times more gray can be represented in the display. Eight times more gray, or even higher, can be represented by a 36 or 48 color bit-depth display.

File Size

The file size of a digital image is proportional to its resolution. The higher the resolution, the bigger the file size — file size is different from image size.

Flatbed Scanner

An image capture device much like a photocopier. The object to be scanned is placed face-down on a glass plate. The CCD array passes beneath the glass and captures information from the original object and transmits it electronically to capture software.

Grayscale

A range of shades of gray in an image or the values represented between black and white.

Image Manipulation or Alteration

Making changes (such as tonal adjustments, cropping, moire reduction, etc.) to a digital image using image editing and processing software such as Adobe Photoshop

Image size

Describes the actual physical dimensions of an image, not the size it appears on a given display device.

Interpolation

In the image manipulation context, this is the increase of image resolution by the addition of new pixels throughout the image, the colors of which are based on neighboring pixels.

JPEG

Joint Photographic Experts Group. A compression algorithm for condensing the size of image files. JPEGs are helpful in allowing access to full screen image files online because they require less storage and are therefore quicker to download into a web page.

JPEG-2000

At its core, JPEG 2000 is an international standard for the compression of still digital images. JPEG 2000 improves on the compression performance of previous methods while offering significant new features and capabilities. The format is multi-resolution, which allows an application to access and decode only the amount of image needed. This allows the user to quickly view an extremely large JPEG 2000 image almost instantly by retrieving and decompressing a low resolution, display sized potion of the file. In addition, the format allows for new access opportunities such as zoom, pan and rotate. JPEG 2000 is still a lossy, compression technique but may have potential for becoming the file format of choice for archival master images in the near future.

Metadata

Data about data, or information provided about a digital object in order to provide access to that image. Usually includes information about the original object, intellectual content of the image, digital representation data, the creation of the digital files, and security or rights management information.

Migration

Migration refers to preserving the integrity of digital images by transferring them across hardware and software configurations and across subsequent generations of computer technologies. Migration includes refreshment (copying digital files from one media to another) as a means of preservation and access. However, migration differs from refreshment in that due to changes in hardware and software, it is not always possible to make an exact copy of a database or even an image file and still maintain compatibility with the new generation of technology. (See also, Refreshment)

Noise

Data or unidentifiable marks picked up in the course of scanning or data transfer that do not correspond to the original.

Optical Resolution

The number of pixels (in both height and width) making up an image, the more pixels in an image, the higher the resolution, and the higher the resolution of an image, the greater its clarity and definition (and the larger the file size).

Resolution can also refer to the output device, such as a computer monitor or printer, used to display the image. Image file resolution is often expressed as a ratio (such as 640x480 pixels), as is monitor resolution; however, resolution is also expressed in terms of PPI. The assumed universal monitor resolution for web users is 144 DPI. Image file resolution and output (print or display) resolution combine to influence the clarity of a digital image when it is viewed.

Pixel

Pixel is short for picture elements, which make up an image, similar to grains in a photograph or dots in a half-tone. Each pixel can represent a number of different shades or colors, depending on how much storage space is allocated for it.

Pixels per inch refers to the number of pixels captured in a given inch. When referring to digital capture pixels per inch (PPI) is the preferred term, as it more accurately describes the digital image.

Quality Control

The process of making sure that the digital images are made to consistently high standards.

RAID

Redundant Array of Inexpensive (or independent) Disks — A project at the computer science department of the University of California at Berkeley, is reaching its culmination with the implementation of a prototype disk array file server with a capacity of 40 GBytes and a sustained bandwidth of 80 MBytes/second.

Refreshment

Refers to the transfer of digital files to new media on a regular basis. This is the most important part of an institution's long-term commitment to digitization. The data we generate today must be retrievable five, fifty, and a hundred years from now. In order to ensure long-term access to the data, it must be transferred to the most recent and stable type of media storage. In a hundred years, the computers on our desks today will not function. We must make sure that the data can be retrieved by future systems. (See also, Migration)

RGB

Red, Green, Blue. (Similar to style of definition for CMYK.) Colored images are often stored as a sequence of RGB triplets or as separate red, green and blue overlays though this is not the only possible representation. (See CMYK)

Standards

Standards are defined by the International Organization of Standardization (ISO) as "documented agreements containing technical specifications or other precise criteria to be used consistently as rules, guidelines or definitions of characteristics, to ensure that materials, products, processes and services are fit for their purpose."

TEI

Text Encoding Initiative is an international project to develop guidelines for the preparation and exchange of electronic texts for scholarly research. The TEI has created a set of Standard Generalized Markup Language Document Type Definitions (SGML DTD) for the encoding of humanities and social science-related texts.

Documents are regarded as having types, just as other objects processed by computers do. The type of a document is defined by its constituent parts and structure. A DTD defines the structure of an SGML (Standard Generalized Markup Language) document. SGML is an international standard for the definition of device-independent, system-independent methods of representing texts in electronic form. SGML emphasizes descriptive rather than procedural markup. While HTML is a markup language which deals primarily with the appearance of a document, SGML is a more complex system for describing structural divisions in a text (title page, chapter, scene, stanza), typographical elements (changing typefaces), and other textual features (grammatical structure, etc.). The "tags" in SGML preserve the structure of a text, enable the user to constrain searches to particular structural features of the text and aid in the navigation and use of the text.

TIFF

Tagged Image/Interchange File Format a file storage format implemented on a wide variety of computer systems, usually used for archival or master digital capture.

Watermark

A watermark is information stored in or on a digital image, which allows image creators to store copyright or branding information to images, audio and video files and documents. Watermarks are applied to images in hopes of reducing misuse or unauthorized distribution of images.

Appendix C — Color Management

The color reference chart used for the following color management procedure is the Greytag MacBeth ColorChecker chart. The MacBeth ColorChecker is a scientifically created set of color patches with established numerical values and a neutral gray 6-patch gray scale. Neutral gray is composed of equal parts of red, green and blue. When these colors are mixed in equal portions, neutral grey is the result. By using neutral gray as the standard, anyone can prepare images, regardless of individual color vision, monitor calibration or the conditions under which the image was made. Because this process is device independent, the images will be consistent in exposure and color balance and will produce high quality reproductions.

The process of color management begins at the point of digital capture. It is essential that practitioners begin with calibrated equipment, proper lighting and correct color settings before you begin your capture. You can use the MacBeth ColorChecker with flatbed scanners as well as digital cameras, although the procedure for each type of capture device varies slightly. The one adjustment necessary to use the ColorChecker on a flatbed scanner is that the paper frame prevents the color squares from sitting directly on the glass of the scanner, resulting in false readings. You should remove the paper frame or cut out the squares so that they sit in direct contact with the glass bed on the scanner. Process is the same as digital camera after that. With either device, the result will be a color-balanced and properly exposed image.

The first step in managing color properly with a digital camera is to set the white balance on the camera. The procedure is slightly different on all cameras, but the concept is to set the white balance to match to lighting source, whether it is daylight, tungsten, flash or fluorescent. Professional digital cameras allow the user to fine-tune the white balance through custom settings on the camera. Use the ColorChecker as a reference for fine-tuning.

The next step is to set the exposure. Photograph the MacBeth ColorChecker and open the image in PhotoShop. Sample the white square. The red-channel value should be as close to 242 as possible. Do not worry about the other channels yet. Adjust the exposure and reshoot until you achieve the correct value. This will guarantee that you set the exposure properly.

When you have achieved the proper exposure, photograph the ColorChecker, and then compare the RGB values on the image of the ColorChecker to the published RGB values. You should then adjust curves and levels in Photoshop on the individual red, green or blue channels until the values of the six gray patches on the image of the ColorChecker match the published RGB values (242, 200, 159, 121, 84, and 53, for each channel). To automate the color adjustments on images made under the same lighting and exposure conditions, you should record these adjustments to make a PhotoShop action which will allow you to use batch processing. As long as the lighting and exposure remain the same, each image can be adjusted with one action. You can choose to add a target to each item or have just one target image per session folder.

Avoid adjusting color beyond this procedure. Any discrepancy that you perceive by viewing the monitor or printed output could be your personal perception of color or a monitor or printer that needs calibration. Do not get lured into the trap of adjusting hue and saturation or color balance based on perception. That type of adjustment may be necessary for images taken without a reference standard, but not for critical archival work.

This process for exposure and color management will produce images that look correct not only by themselves, but also in relation to each other.

Appendix D — Master and Working Master Scanning Procedures

Preparation
- Create a network directory to hold master and submaster files. Consider creating duplicate directory on desktop to serve as a working space for the project (to avoid the slow pace of working on the network)
- Calibrate the scanner and monitor (this should be done on a regular basis)
- Using soft bristle brush, brush off scanner glass
- Open scanning software and set the following specifications:
  - ◊ 600 PPI (800 PPI if scanning very small items) (PPI may vary)
  - ◊ 100% image size
  - ◊ RGB – 48 bit (bit depth may vary)
  - ◊ Adobe RGB color space

Scanning
- Place item on scanner with top of photo facing forward
- Place reference target below or to the left of the item, depending on orientation
- Scan item

Quality Check
- Correct orientation
- Image straight
- Try to place each item squarely on the scanner glass with the correct orientation The image can be rotated later in Photoshop but any rotating you do will slightly degrade the image so it's best to avoid if possible. Rescan if the image has not been correctly placed

Review image at 100%
- Any dust or debris on scanner bed
- Overall image quality — color, exposure, contrast and focus
- Noise (odd patterns, dots, reflections, etc.) in scanned file
- Moire patterns resulting from the scanning of printed material
- If any of the above quality issues are noted in the master scan, correct scanning software settings and rescan rather than manipulating the master post-capture

Image Manipulation
- Crop image, with target included, leaving ¼" border

Saving
- Conduct final visual review of the master before saving
- Name each master file according to predetermined naming conventions
- Save to Masters folder in established directory

Working Masters
- Open each file in Masters folder

- Review image at 100%
- Apply color manipulation, if deemed necessary
- Apply unsharp mask sharpening filter
- Crop out target leaving 1/8" border around item
- Name each working master according to pre-determined naming conventions
- Save to Working Master folder in established directory

Appendix E — Data Center Basics

Server Room Space and General Recommendations

The server room in the data center must be built to accommodate all of the equipment with plenty of room for growth. It must allow enough space for cabling and access to the sides and backs of server racks and other equipment. It must have enough height to accommodate a raised floor for distribution of ventilation ducts, conditioned cold air and/or wiring that runs underneath, avoiding clutter. Packing servers racks too close together will not only make it difficult to work in the server room, it will restrict air flow as well.

Review and set clear standards early to ensure that all of the equipment racks holding the servers are configured the same way in the data center. It should estimate the weight, power requirements and heat generated by each rack in the data center in order to manage expansion and/or reorganization.

All racks and equipment within them should be labeled with sever names, IP address and the name or contact information of the administrator responsible for the servers within that rack.

Electrical Power

- Determining how much power the server room needs can be complex, and it is best to consult with an electrician to determine the maximum load ratings for every piece of hardware that will be located in the data center.

- On average, 45 to 60 watts per square foot is typical. However, a densely packed server room may require 80 to 100 watts per square foot or more. Power requirements should be factored in to accommodate:
  ◊ Overhead lighting
  ◊ UPS (Uninterruptible Power Supplies)
  ◊ Rack mounted fans
  ◊ Portable backup Air-conditioning units

Electrical Outlets

- The estimated number of outlets required and their approximate placement in the server room should be considered. For any floor-mounted outlets, consider using power whips (outlets attached to flexible cabling) instead of fixed outlets to allow some flexibility when moving server racks. Server electrical loads should be distributed across the circuits. This flexibility should be built into the expansion planning guidelines.
- Cooling and ventilation should be given their own circuitry to prevent power fluctuations. All power cables should be shielded in a flexible steel conduit encased in a cooper shield to prevent electrical noise (EMI) from reaching the network cables and equipment.

- The power redundancy requirements of the server room will depend on uptime requirements. Data centers with domain controllers, DNS servers, web servers and other mission critical systems will require backup generators as well as UPS units. Decide early which servers (and circuits) will be tied to generator power and which will simply be tied to Uninterruptible Power Supplies (UPS).
- Finally, check the quality of the power coming into the building, and make sure that it's within the American National Standards Institute (ANSI) specifications. If not, consider installing power conditioning device that helps clean the signal thus eliminating line interference.

Air Conditioning/Environmental Control
- Computer equipment ages faster when it gets hot. Servers generate a fair amount of heat and are sensitive to high temperatures and fluctuations in humidity. A stable environment can improve uptime and extend the life of the hardware. An ideal environmental control system should be modular, expandable and flexible. Air temperature is an important consideration. Air conditioning capacity will depend on the size of the server room, lighting, and numbers of people working in the room, the quantity of electrical equipment and the amount of heat generated by that equipment. In short, one will need to determine the total load power in watts generated by all of these devices in order to determine their thermal output.
- Environmental engineering specialists can determine cooling requirements based on these factors, as exact calculation methods are beyond the scope of this report. If possible, build redundancy and overcapacity into the environmental systems to allow for failures, maintenance downtime and temporary capacity spikes.
- *Note:* Server rooms require large amounts of power and air conditioning. When adding more servers to the server room one must consider and plan for the impact of these new nodes on the air conditioning and power grid. One way smaller (and larger) organizations can reduce their server room footprint is to use virtualization software. (Please see the section on "Machine Virtualization" below.)

Air Flow
- Temperature, although very important, is not the only environmental issue to take into account. Good air flow in and around server racks is critical. It is necessary to work with environmental engineers to work out the best solution in terms of air conditioning, duct placement and ceiling vs. floor-mounted ventilation.
- Keep on the lookout for hotspots and stagnant air flow. Temperature and humidity sensors can be placed throughout a data center to monitor conditions.

Here are some other factors to consider and implement:
- Moving air cools more than still air. Maintaining good airflow is important.
- Keep an eye on the fans in the server room. Fan failure can cause localized hot spots and rises in temperature.
- Repositioning machines can have a positive effect. Designing and placing baffles or deflectors can avoid channeling hot air from one machine into another.

Humidity

- Another important factor to consider is the level of humidity. Low humidity levels increase the risk of static electricity, and high amounts of relative humidity can contribute to corrosion as well as lowering the heat removal capacity of the equipment. The relative humidity in a data center should generally be kept between 40 to 55 percent.
- ASHRAE publishes specific guidelines for data centers and server rooms (http://www.ashrae.org).

Cable Management

- Proper cable management is a science, and if implemented and managed properly it can reduce troubleshooting time dramatically. Decide if the cabling will run under the floor or overhead (or both). Determine specific zones for cable runs and keep them away from electrical lines. Keep fiber and copper cables on separate runs as the weight of copper cables can potentially crush the fiber cables. Also be sure to allow enough room in your cabling design to accommodate the proper bend radius for your cable types.
- Color coding cables and labeling greatly simplifies management, and properly labeling individual cables will save countless hours pulling up floor tiles or tracing overhead wires trying to find a bad cable.

Security & Fire Suppression

The final piece of putting together a data center is designing security into the server room. At a minimum, one must control and log access to the room via security cards or other auditable methods.

- Fire suppression systems will vary depending on local ordinances, and their design and placement may affect the final server room design as well.

Appendix F — Network Design for Data Protection

Solid network design is critical to allow access to and to protect the digital objects and the metadata that go into the making of quality digital projects. An exhaustive treatise on network design is not in the scope of this document but a general overview is important to give the reader the basics to protect their data. Most complex, well designed networks are the created by network engineers in large organizations. The design of a basic TCP/IP network model consists of the following five layers. They are: the application layer, the transport layer, the network layer, the data link layer and the physical layer. The layers near the top are closer to the user application. Those layers located near the bottom are logically closer to the physical storage and transmission of the data.

DMZ

The gateway to the data available to the general public is via the DMZ (demilitarized zone). In a computer network, the DMZ is a computer host or small network implemented and inserted into the system as a "neutral zone" between the private (back office) network and the outside public network. The DMZ prevents outside users from getting direct access to a server that has the critical digital objects and the descriptive metadata. Users of the public network outside the institution can access only content that the institution wants the outside world to see via the DMZ host.

The DMZ typically gives users access to the institution's web pages via a series of "public IP addresses" that are accessible by web servers located within the DMZ. (See Figure 1 – Example of an Institutional Network) The DMZ does not allow access to the institution's internal data. The different layers in the instruction's network include: machines in the DMZ with separate IP ranges, the internal institutional network or "Intranet" that is located on machines located in a separate IP range behind the DMZ and firewall. Lastly, the portion of the network that houses critical machines holding critical data including the archived master digital objects (TIFF images) and descriptive metadata (stored in the .txt format) is located even deeper in an institution's network on another separate, and very private IP range.

In the event that an outside user penetrated the DMZ host's security and the Web pages or digital objects on the host site were corrupted or compromised no other critical information beyond the DMZ would be exposed. The servers hosting content in the DMZ could be rebuilt from backed up or mirrored sites, or content stored on servers deep within the institutional network.

Firewall

A firewall is a system designed and setup to prevent unauthorized users from accessing an institution's internal network and a private internal network. Firewalls can be setup on a hardware or software environment or some combination of both. As an example, a firewall protects the critical backbone of a digital collection archive by preventing users from the open Internet from accessing the area in the internal network where the mater digital object masters (TIFFs) and archival descriptive metadata is stored. All traffic

either entering or leaving the internal network must pass through the firewall. The firewall examines the messages and then blocks those that do not conform to the institution's security criteria.

Internal Network

The "Internal Network" is an internal computer network that sits behind the DMZ and firewall that uses Internet protocols, network connectivity to securely share part of an organization's information or operations with its employees. At times the term refers only to the most visible service, the internal website. The same concepts, protocols and technologies of the Internet, such as clients and servers running on the Internet protocol suite, are used to build an internal intranet. HTTP and other Internet protocols are commonly used as well, such as FTP or SSH. There is often an attempt made to use the newer Internet technologies to provide new and improved interfaces with institutional "legacy" data and information systems.

Private Internal Network

The "Private Internal Network" (PIN) is a network within a network behind the firewall that is secure beyond the securities put in place for the "Internal Network" portion of the network. A PIN uses Internet protocols and public wires to connect nodes. For example, there are a number of systems that enable you to create networks using the Internet as the medium for transporting data. Networks use encryption and de-encryption and other security mechanisms to make sure that only authorized users can access this portion of the network and that the data cannot be compromised.
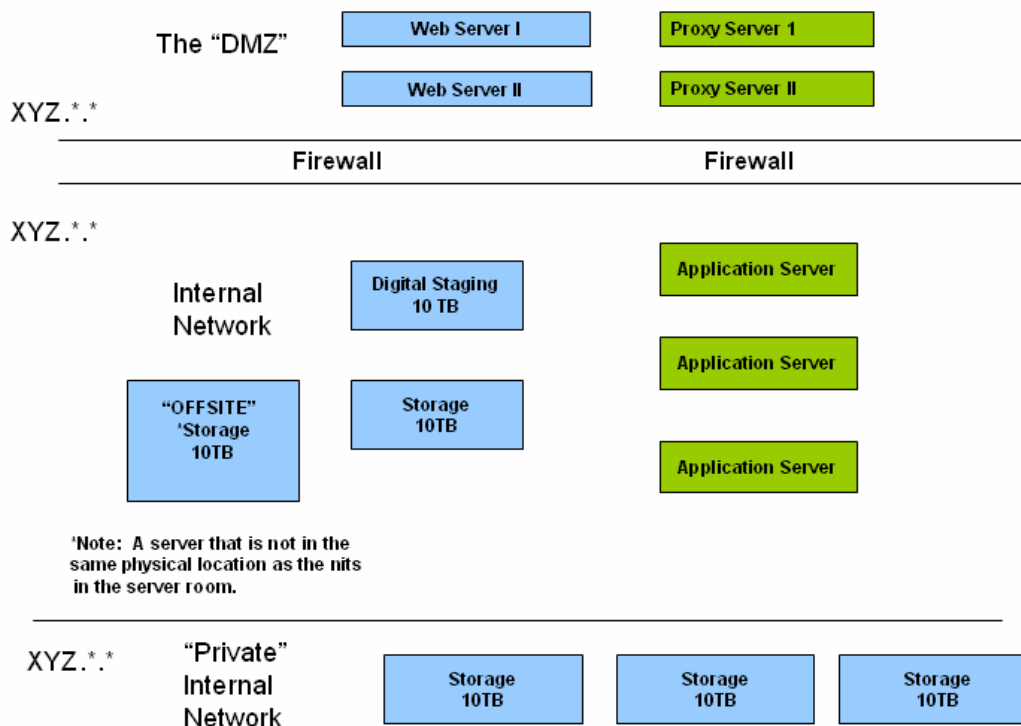

Figure 1 – Example of an Institutional Network

Machine Virtualization

The information technology world has seen rapid new developments during the past few years. Virtualization is a movement that is in the process of reorganizing the relationships between physical hardware, differing operating systems, application software and in general, how the work in large IT data centers gets done. Virtualization is a term that describes the separation of a computer resource or request for a service from the underlying physical delivery of the resource or service. An example of virtualization in action is virtual memory. With virtual memory, computer software applications gain access to more memory than has been physically installed. This happens because a background application is running that swaps data in memory to disk storage. Virtualization can be applies to other IT infrastructure layers – including networks, storage, laptop, desktop or server hardware, operating systems and applications.

There are a number of sound reasons why it is to one's benefit to implement a virtualization program at an institution. These reasons include:

- Server Consolidation/Containment – Contain server sprawl by running software applications in virtual machines on fewer, highly scaleable, reliable enterprise-class servers. Some virtualization software applications enable administrators to virtualize up to 10 or more servers per physical server. This drastically increases server utilization while containing server sprawl.

- Software vendors like to have their customers run their application on a specific server, on a specific operating system. Often that is because the vendor likes to maintain complete control over the application so the application is not competing for server processing time and power. Perhaps the application is not well built so the vendor wants all of the server's processing power so the application runs smoothly. Throwing machinery at a poorly designed application is at times the preferred "vendor support" model.

  ◊ Reduce the cost of server rack space
  ◊ Reduce the electrical power demand — fewer machines are being run in the data center
  ◊ Reduce the need for cooling – fewer machines require less cooling, a major cost for IT operations
  ◊ Decrease labor costs due to the simplification automation of tasks across disparate operating systems

- Provide Institutional Continuity/Protection at a Lower Cost — Deliver high-availability for critical applications. Customers can implement a unified disaster recovery system that allows many production virtual machines without investing costly one-to-one mapping of production and disaster recovery hardware.

- Streamlined Software Testing and Development — Consolidate development, testing and staging environments involving multiple operating systems.
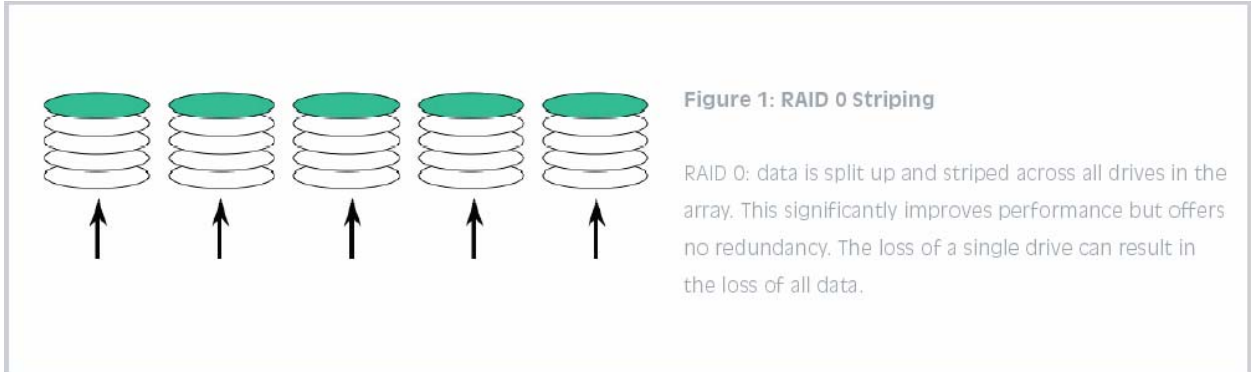
- Manage/Secure Enterprise Desktops – Secure and manage desktops of a geographically dispersed workforce. Provide a corporate standard desktop image via a virtual machine.

Virtualization allows an organization to do more computing with a much smaller IT computing footprint. It brings a great deal of flexibility into systems that in the past were both costly stand-alone and inflexible. Organizations needing to adapt quickly to new and changing business or academic scenarios are well served by virtualization. Applications can be changed in the background, and critical work can be done with very minimal down-time. For organizations doing digitization, business challenges are not as dynamic as in a product production environment, but digital organizations certainly can gain flexibility and virtualization can allow them to do more with less.
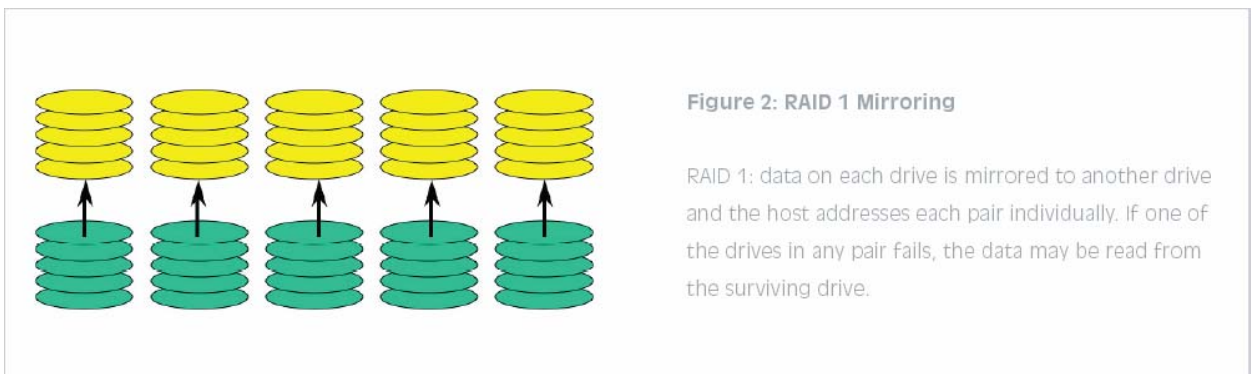
Appendix G — RAID Array

RAID 0

RAID 0 implements a striped-disk array to offer improved performance by spreading the data across multiple drives. However, RAID 0 is not "true" RAID because it provides no fault tolerance — the failure of a single drive will result in the loss of all data in the array.

Figure 1: RAID 0 Striping

RAID 0: data is split up and striped across all drives in the array. This significantly improves performance but offers no redundancy. The loss of a single drive can result in the loss of all data.

RAID 1

RAID 1 This solution is less complex and is easily implemented. The RAID 1 configuration requires as little as two disks. One disk simply mirrors the other disk. If one disk were to fail, the system could be restarted with the remaining drive. This configuration offers twice the read rate, but write time is the same as with single disks. RAID 1 has the highest overhead of all RAID types and usually does not support the hot-swapping of failed drives.

Figure 2: RAID 1 Mirroring

RAID 1: data on each drive is mirrored to another drive and the host addresses each pair individually. If one of the drives in any pair fails, the data may be read from the surviving drive.
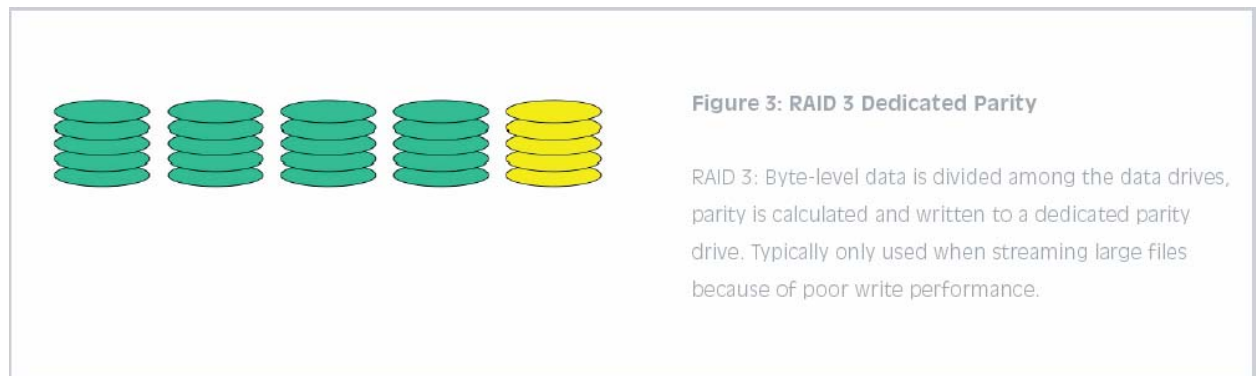
RAID 2

RAID 2 uses hamming code error checking and correction (ECC) for on-the-fly error correction. It offers a high level of data integrity by striping the data on the bit level. The

high ratio of ECC disks to data disks makes RAID 2 such an expensive alternative that there are no commercial offerings.
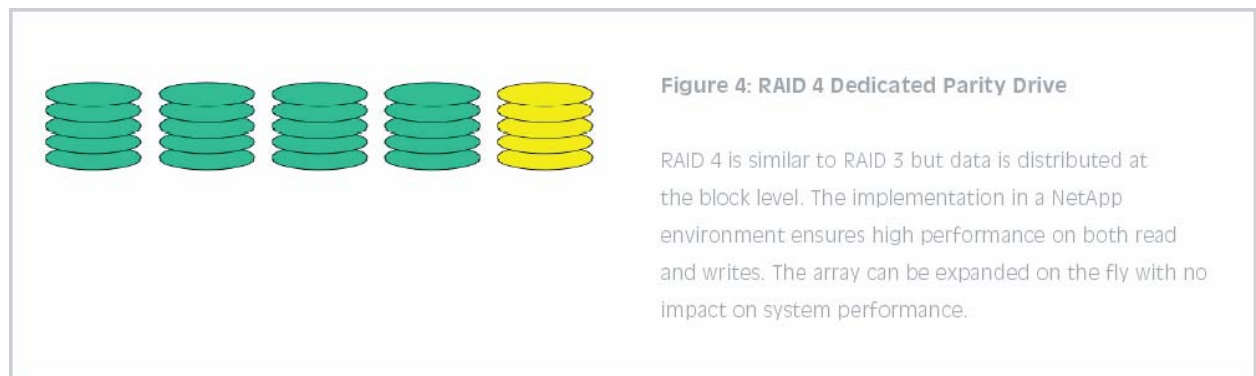
RAID 3

RAID 3 provides redundancy by striping the data on the byte level and storing the parity information on dedicated ECC disks. The low ratio of ECC disks to data makes RAID 3 a far more cost-efficient solution. RAID 3 offers a high read rate, especially with large sequential files, with little impact on throughput when an individual disk fails. Unfortunately, the write rate of RAID 3 is, at best, equal to a single drive due to the potential of an ECC bottleneck when storing the parity information.

Figure 3: RAID 3 Dedicated Parity

RAID 3: Byte-level data is divided among the data drives, parity is calculated and written to a dedicated parity drive. Typically only used when streaming large files because of poor write performance.
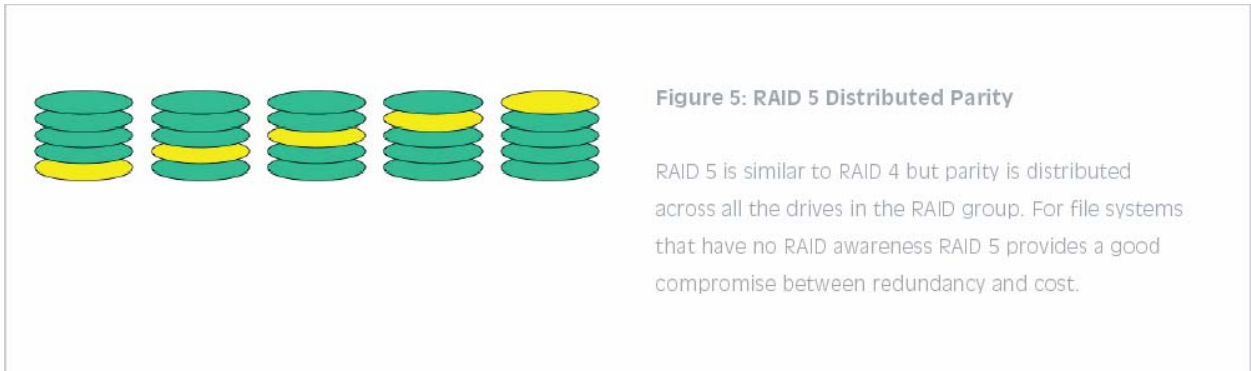
RAID 4

RAID 4 is similar to RAID 3 except that data is striped in blocks rather than bytes. When the data being updated is small and does not access all the data disks, four disk requests must be processed: two requests to read the old data and parity information to compute the new parity information, and two requests to write the new data and parity information.

Figure 4: RAID 4 Dedicated Parity Drive

RAID 4 is similar to RAID 3 but data is distributed at the block level. The implementation in a NetApp environment ensures high performance on both read and writes. The array can be expanded on the fly with no impact on system performance.

RAID 5

The RAID array that offers the best option for data sustainability is the RAID 5 configuration. RAID 5 requires 3 disk drives. RAID 5 distributes the parity information

evenly across all disks in the parity group. It lends itself well to transaction-oriented processing where many small chunks of information need to be transferred simultaneously. When a disk fails, there is an impact on performance with RAID 5 due to the on-the-fly computing of missing data. Moreover, when the faulty disk is replaced, the rebuild time can be lengthy, but the system remains operational.

Figure 5: RAID 5 Distributed Parity

RAID 5 is similar to RAID 4 but parity is distributed across all the drives in the RAID group. For file systems that have no RAID awareness RAID 5 provides a good compromise between redundancy and cost.

Appendix H — Online Storage/Storage

Storage Equipment — Large and Small Scale

In this document the authors are not endorsing specific products but are making comment on quality products in use. Currently, the University of Wyoming uses NetApp's StoreVault storage appliances, however, other competitors of NetApps include: EMC, Hewlett Packard, Storage Technology and Hitachi Data Systems. Smaller and very affordable storage units such as the Buffalo TeraStation PRO might be an option for a smaller organization wanting better storage and protection of their digital objects.

DAS

Networked storage technologies have become the standard for many who are doing large-scale digitization projects. The most common form of server storage today is direct attached storage or DAS. Storage disks on a DAS environment are either internal to the server or are in an array connected directly to the server. No sharing storage resources occur. Each application server (print, web, file) has its own disk storage. This setup can work for a smaller environment, but as application servers proliferate the backup of all of these separate drives can get to the point where it is a real challenge.

SAN

SANs or Storage Area Networks allow more than one application server to share storage. Application servers share the storage resources of the SAN. The interrelating components of the SAN array are all interlinked via a series of fiber channels. Data flows quickly over these channels. Backups can be performed simply and centrally without application interruption. SANs are scaleable and flexible. Storage can be added and different kinds of storage can be added without causing application disruptions. SANs have a number of advantages. The disadvantages are that setting up a SAN network can be very daunting. The technology is expensive and getting the fiber technology to work is challenging.

NAS

NAS units are simplified versions of file servers that allow optimization for file sharing in the organization. Clients can see files and folders on the server if they are authorized to do so. NAS units are easy to configure setup and install. Setup basically goes like this. Mount the device on a rack; connect the power and the Ethernet network hookup and configure the device using an easy-to-use web browser interface. Typical setup time can take less than an hour. NAS units are multi-protocol meaning that they can share files among clients using Windows or UNIX operating systems.

SAN/NAS Combinations -- "All-In-One" Storage

An "all-in-one" storage device combines the best of the SAN and NAS setups. A number of schools and universities are using StoreVault units that feature the all-in-one combination of SAN and NAS storage to help solve their large-scale storage needs. StoreVault units while not inexpensive are cheaper than SAN unit setup sold by manufacturers such as Dell. Basically a StoreVault Unit is a chaises that contains up to

12 bays for hard drives. Hard drives range in size from 750 GB per disk drive up to 1 TB per disk drive.

All-in-one storage options range in price from large-scale StoreVault units that can hold up to 36 TB of storage down to the new StoreVault S300 that costs $3,000 and hold up to 3 TB of information. The small-scale Buffalo TeraStation Pro II unit ranges in price from $1,200 to $3,000 and contains anywhere from 1 to 4 TB's of storage. The Buffalo TeraStation Pro II features drastically improved network performance over the older TeraStation unit and supports more RAID levels. The Buffalo TeraStation Pro II offers features targeted to businesses, such as Active Directory integration, DFS support, a heavy-duty power supply and cooling system, quick-swap SATA hard drives, and a status LCD. The Buffalo TeraStation Pro II is best suited for business and professional users who desire maximum reliability and robust features.

Appendix I — Project Documentation Template

Date:

Title:

Introduction
The introduction should abstract the main points of the project:
- What is the project about
- Key assumptions—audience, the significance of the project
- Subject keywords
- Overview of methods

Background
Provide background information relevant to the digital project
- Institutional background
- Collections background
- Predecessors of current project

Statistics
- Number of items digitized
- Staff numbers
- Project duration
- Equipment used

The Challenge
Briefly introduce what issues/problems the project addressed.

Key Stakeholders
Identify staff and/or institutional organizations/departments played a role in finding solutions to the issues address in the project.

Methods/Procedures Used
Outline the methods and procedures used to meet the challenge.

Resources

Include citations to resources used, vendors or additional resources used.

Key Contacts

Who would people contact for more information about the project who can answer questions not addressed within your project documentation form.

Findings

If you had to do it over what would you differently? What lessons were learned in the process? What would you recommend to new projects?