Preserving State Government Digital Information
Minnesota Historical Society

Retrospective Digitization of Government Records

Summary

Digitizing government records is a process that needs to be thought out and planned in detail before undertaking. The purpose of this paper is to summarize major concepts that need to be considered before starting a retrospective digitization project. The information below refers only to the scanning of textual documents; non-textual items such as photographs are not discussed here.

DISCLAIMER:
This white paper is a topical overview and nowise intended to offer legal advice. Consult an attorney for assistance with specific concerns or for advice.

Any comments, corrections, or recommendations may be sent to the project team, care of:

Carol Kussmann
Collections Assistant, State Archives
Minnesota Historical Society
carol.kussmann@mnhs.org / 651.259.3262

Introduction

It is very important to plan a retrospective digitization project. Below you will find a brief introduction to many of the issues surrounding digitization projects. Appendix A is an annotated list of selected resources to assist you with learning more about individual topics. Links to all resources and references were accurate as of March 2009.

Topics include:

Legal Requirements

Electronic records are subject to the same legal requirements as paper documents. To ensure your digital records are fully admissible in court, they must be trustworthy, complete, and accessible for as long as your retention schedule requires. Please review the *Trustworthy Information Systems Handbook*[1] produced by the Minnesota Historical Society for information on such systems. Following record keeping regulations will assist you in managing your records appropriately. Legislation varies from state to state, so it is important to research the requirements that may affect your particular digitization project, such as access to the records and disposition of the originals. Federal, state, local and organizational policies all apply.

Government records generally have more requirements to follow than business records. Laws relating to the "collection, creation, storage, maintenance, dissemination, and access"[2] of government records are common. Regulations specifically affecting government records usually address such issues as privacy and security, retention and disposition, and public access to information.

Some laws effecting electronic records in Minnesota include:

- Official Records Act (Minnesota Statutes, Chapter 15.17)[3]
- Records Management Act (Minnesota Statutes, Chapter 138.17)[4]
- Minnesota Government Data Practices Act (MGDPA) (Minnesota Statutes, Chapter 13)[5]
- Uniform Electronic Transactions Act (UETA) (Minnesota Statutes, Chapter 325L)[6]
- Electronic Signatures in Global and National Commerce (E-Sign), a federal law.[7]

---

[1] Minnesota Historical Society. *Trustworthy Information Systems Handbook*. July 2002. http://www.mnhs.org/preserve/records/tis/tis.html

[2] Minnesota Office of the Revisor of Statutes. *2008 Minnesota Statutes: 13.01 Government Data*. https://www.revisor.leg.state.mn.us/statutes/?id=13.01

[3] Minnesota Office of the Revisor of Statutes. *2008 Minnesota Statutes: 15.17 Official Records*. https://www.revisor.leg.state.mn.us/statutes/?id=15.17

[4] Minnesota Office of the Revisor of Statutes. *2008 Minnesota Statutes: 138.17 Government Records; Administration.* https://www.revisor.leg.state.mn.us/statutes/?id=138.17

[5] Minnesota Office of the Revisor of Statutes. *2008 Minnesota Statutes: Chapter 13. Government Data Practices*. https://www.revisor.leg.state.mn.us/statutes/?id=13

[6] Minnesota Office of the Revisor of Statutes. *2008 Minnesota Statutes: Chapter 325L. Uniform Electronic Transaction Act.* http://www.revisor.leg.state.mn.us/stats/325L

[7] Federal Trade Commission and Department of Commerce. *Electronic Signatures in Global and National Commerce Act*. June 2001. http://www.ftc.gov/os/2001/06/esign7.htm

Official Records Act[8]

The Official Records Act states that official records of government agencies will be preserved to document "a full and accurate knowledge of their official activities"; specifically stating that "government records may be produced in the form of computerized records" but "all government records shall be made on a physical medium of quality to insure permanent records." Records can be reproduced and will be admissible in court as long as it is in a "method that clearly and accurately reproduces the records"

Records Management Act[9]

The Records Management Act mandates the administration of government records. Destruction of records, retention schedules, and the preservation of historical records are covered under this act. It places the responsibility of records management on the head of each state agency and states that they must "maintain an active, continuing program for the economical and efficient management of the records of each agency, county, municipality, or other subdivision of government".[10] It also establishes the Records Disposition Panel which "reviews, evaluates, and then approves or disapproves requests to dispose of records, to transfer records, and to establish records retention schedules. Fundamentally, the panel provides oversight, but does not initiate any actions. If your agency wants to keep records forever, then you never have to work with the panel. However, if your agency wants to do anything else legally with your records, you must submit your proposal to the panel for approval."[11]

Minnesota Government Data Practices Act[12]

The Minnesota Government Data Practices Act addresses public access to government data, and the duties of record holders to ensure access to government information while protecting privacy. The Minnesota Government Data Practices Act "regulates the collection, creation, storage, maintenance, dissemination, and access to government data in government entities. It establishes a presumption that government data are public and are accessible by the public for both inspection and copying unless there is federal law, a state statute, or a temporary classification of data that provides that certain data are not public."[13]

---

[8] Minnesota Office of the Revisor of Statutes. *2008 Minnesota Statutes 15.17 Official Records*. https://www.revisor.leg.state.mn.us/statutes/?id=15.17

[9] Minnesota Office of the Revisor of Statutes. *2008 Minnesota Statutes: 138.17 Government Records; Administration.* https://www.revisor.leg.state.mn.us/statutes/?id=138.17

[10] *Ibid.* https://www.revisor.leg.state.mn.us/statutes/?id=138.17

[11] Minnesota Historical Society. *Electronic Records Management Guidelines: Introduction*. March 4, 2004. http://www.mnhs.org/preserve/records/electronicrecords/erintro.html

[12] Minnesota Office of the Revisor of Statutes. *2008 Minnesota Statutes: Chapter 13. Government Data Practices*. https://www.revisor.leg.state.mn.us/statutes/?id=13

[13] Minnesota Office of the Revisor of Statutes. *2008 Minnesota Statutes: 13.01 Government Data subd. 3*. 2008. https://www.revisor.leg.state.mn.us/statutes/?id=13.01

Uniform Electronic Transactions Act (UETA) [14]
The National Conference of Commissioners on Uniform State Laws (UCCUSL) created a law supporting interstate commerce in 1999; Minnesota put this law on its own books in 2000. The Act applies to electronic records and signatures related to business, commercial, and governmental matters, excluding only those transactions listed in the Act itself.

Electronic Signatures in Global and National Commerce Act (E-Sign)[15]
E-Sign was enacted in June 2000 and became effective in October of the same year. E-Sign "facilitates the use of electronic records and signatures in interstate and foreign commerce by ensuring the validity and legal effect of contracts entered into electronically."[16] This addresses the issues of electronic records being trustworthy and authentic.

Both "UETA and E-Sign were … intend to facilitate the use of information technology in government and business by addressing the legal obstacles that exist in a system oriented towards paper records and signatures. The primary message of the laws is that a court may not determine that an electronic record or signature is untrustworthy simply because it is in an electronic format. A court can, though, reject electronic records and signatures because a government agency is creating, using, or managing them in an untrustworthy system or manner. One indicator of untrustworthiness would be an agency's failure to respect the laws governing records."[17]

Cost Justification

Planning is critical when designing a digitization project or program. A cost analysis that answers the ever important questions of who, what, where, why, when, and for how long must be conducted. Digital imaging is an investment and, just like any other business decision, it should make financial sense for your organization or agency.

When looking to digitize retrospectively, it is important to know how the records will be used. Scanning documents just to eliminate an offsite storage space may not be cost effective. Bryan Apple, Director-Digital Document Management Solutions from Recall, a digital document management company, recently stated that "scanning could cost you

[14] Minnesota Office of the Revisor of Statutes. *2008 Minnesota Statutes: Chapter 325L. Uniform Electronic Transaction Act.* http://www.revisor.leg.state.mn.us/stats/325L; National Conference of Commissioners on Uniform State Laws. *Uniform Electronic Transactions Act (1999).* 1999. http://www.law.upenn.edu/bll/archives/ulc/fnact99/1990s/ueta99.htm
[15] Federal Trade Commission and Department of Commerce. *Electronic Signatures in Global and National Commerce Act*. June 2001. http://www.ftc.gov/os/2001/06/esign7.htm
[16] *Ibid.* http://www.ftc.gov/os/2001/06/esign7.htm
[17] Minnesota Historical Society. *Electronic Records Management Guidelines: Introduction*. March 4, 2004. http://www.mnhs.org/preserve/records/electronicrecords/erintro.html

ten times more than a year of physical storage".[18] He uses the example of one carton of materials or approximately 2000 pages stored for seven years would cost $46.40 as paper records compared to $181.40 in digital format.[19] The true value of digitizing comes from how the records will be used over time. There must be benefits of having access to a digital file before they become cost effective.

The Utah State Archives has created the *Guide to Digital Imaging* that lists advantages, disadvantages, and other factors to consider when digitizing records. A handful of their suggestions are listed below; for the complete list please review the *Guide to Digital Imaging*.[20]

Advantages:

- Multiple users and access levels possible
- Shorter retrieval time when the images are well indexed
- Ease of use of imaged copies of records in vital records and disaster recovery plans

Disadvantages:

- Digital images are not human-readable without computer equipment
- Significant equipment costs, including hardware and software
- Potential for hardware and software obsolescence. Generally, systems change every 18 months to 5 years, software every 2-3 years, and the life expectancy of media is relatively short.

Other Factors to Consider:

- Relationship to records on other media
- Records and information usage
- Legal acceptability
- Ease of maintenance
- Staffing requirements
- Document preparation

---

[18] The additional costs of paper documents include cost associated with activities during the life of the document, removal fees and destruction fees. In addition to scanning, the costs for digital documents include indexing, storage and access, necessary software, conversion needs, and destruction fees which can be up to 10 times more expensive than the costs for paper documents. Apple, Bryan. *Documents – Physical, Digital, or Both? A Framework for Evaluation*. Slide 14. November 11, 2008. http://www.twincities.arma.barr.com/Programs/Presentations/20081111BryanAppleARMATC.pdf

[19] A chart shows the associated costs for storing one carton of papers in both paper and digital format. The paper files need to be moved to the storage location and eventually destroyed. The digital files need to be scanned, indexed, and imaged all which have a per page cost which significantly increases the cost of storage. Access costs are not included. Apple, Bryan. *Documents – Physical, Digital, or Both? A Framework for Evaluation*. Slide 15. November 11, 2008. http://www.twincities.arma.barr.com/Programs/Presentations/20081111BryanAppleARMATC.pdf

[20] Utah State Archives. *Guide to Digital Imaging*. June 2005. http://www.archives.state.ut.us/recordsmanagement/ERM/guide-to-digital-imaging.html

- Quality control issues
- Condition of original records

Cost analysis or budget development must consider costs throughout the entire digitization process. Costs incur during project development, the digitization process, and continue as the digital collection is maintained over time. Below you will find a summary of project costs incurred at each stage. Often times many of the items listed will overlap into more than one phase of the project. The questions below are to get you thinking about your project goals and how to estimate costs.

Project Development
Project development includes the costs to select, prepare, and catalog the documents that are to be digitized. Cataloging includes creating or linking any necessary metadata to the original object. Selection of appropriate hardware and software is also part of this phase.

Questions to ask include:

- How many people are involved in the selection process? (cost of staff time)
- Does the size of the selected collection matter? Will a large collection that does not need as much preparation be a better use of time than a smaller collection that will need a lot of preparation?
- Are you going to pre-select materials? As a group, individual files? Or will you digitize on-demand?
- How organized are the files you want to digitize? Do they need to be sorted?
- Do the files already have attached metadata or does the metadata need to be created?
- What is the condition of the original? Is it strong enough to withstand the digitizing process?
- Will the digital object replace the original or simply become an access copy? How will this be handled?
- Do you have access to digital capture equipment? Will you need to purchase equipment?
- Do you have computers and storage drives with enough memory and space for the intended project? Will you need to purchase new hardware?
- Do you have the software for image capture and manipulation? Will you need to purchase new software?
- Are you able to partner with anyone to share experience and costs?

Digitization Process
The digitization process includes the scanning or capturing of a digital image, entering of the metadata, and developing a system that holds the images. Providing access to the images could be considered part of this process or part of the ongoing costs, as could

database creation. Answer the following questions to start thinking about costs and other possible issues related to the actual digitization process.

- What is the capture process?
- Who will be doing the capturing? Is the person trained? How much staff time will be used for this project?
- Do new people need to be hired? Trained? How long will this take?
- Will the project be outsourced? What has to be done in-house to ensure that documents are properly handled?
- How will the images and their metadata be linked?
- What system will be used to house the images? Their metadata? Will this be in-house or will you have to pay a service to assist you?
- Will the images have to be enhanced after creation? All? Some? How much time does this add to the process?
- Will there be a need for additional digital copies? Access copies? Various size files?
- What is the quality control policy?
- Who is in charge of the quality control?
- How often are the files checked for quality?

Ongoing Costs

Ongoing costs include the salary and benefits of current and new staff, money for external technical support if needed, additional training costs, maintenance of hardware and software, replacement costs for failed or obsolete equipment or software, and Internet connections.[21] These costs continue to occur after the collection has been digitized.

Questions to ask include:

- How much time will current staff devote to this project? What level staff are they? Include the time a manager must spend administrating the project.
- How many people will need to be hired for this project? Will these be temporary positions?
- How much will training cost? Hiring a trainer? Time of current staff?
- Will outside specialists need to be hired for the short- or long-term?
- What system will be used to store the digital images? How often will it need to be upgraded? What back-up measures are in place?
- What system will be used to disseminate the digital images? How often will it need to be upgraded? What back-up measures are in place?
- Who is in charge of the maintenance for the ongoing technological aspects of the project?

[21] Roderick, Elizabeth. More Than Just Pretty Pictures: A Cost/Benefit Analysis of Digital Library Holdings. CAUSE98 EDCAUSE Conference. December 9, 1998. http://net.educause.edu/ir/library/html/cnc9804/cnc9804.html

- What is the maintenance policy for ensuring continued availability of the images? For records with long-term or permanent retention, costs of file maintenance must be planned for.
- What is the policy for files that have a retention schedule? Who is in charge of following up?
- Will the original continue to be preserved upon creation of a digital surrogate?
- Where are files to be stored? Do you need to pay for offsite storage or do you have enough space in-house?
- Will Internet connections fees need to be included?

For a discussion on how to calculate 'actual' costs for a digitization project, review Steven Puglia's 1999 article, *The Costs of Digital Imaging Projects.* The summary states that 32% of the cost is digital conversion, 29% is metadata creation, and 39% is administrative activities such as quality control.[22]

Before the numbers scare you, go back to the plan and understand why you want to digitize a collection. What is the retention period of the records? Do the benefits outweigh the costs of creating files with short-term utility if retention is only a few years? Do you have the resources to maintain files with long-term or permanent retention? Are the records being considered for digitization in high demand or have they never even been requested before? Will the digital records become official records, reducing the amount of paper storage? What are the costs of storing paper vs. digital records? How much do you pay for storage of paper records? How much time is spent looking through old files? Whose time will be being used for digitizing the materials? Is there money available to sustain the project over the long term? Can a partnership keep costs down? Does it make sense to outsource some or the entire project? The costs will depend on the method used to create, store and disseminate information and the answers to these questions will only provide an estimate.

In-House vs. Outsourcing

Digitizing is a time-consuming process. Evaluation of your resources will help determine if your digitization process should be done in-house or outsourced to a vendor who specializes in digitizing.

[22] The associated costs listed in this study are almost ten years old and have probably changed with time. Now that digitization has become more common place, it would be useful for new studies to address this topic in more detail. Puglia, Steven. *The Costs of Digital Imaging Projects*. RLG DigiNews. October 15, 1999. Volume 3, Number 5.
http://worldcat.org/arcviewer/1/OCC/2007/08/08/0000070511/viewer/file422.html

The three main factors to consider are:

Time:

- How much time do you think the project will take?
- How much time do you have to spend on this project?
- How many people will take time away from other duties to work on this project? How much time will it take to hire new or temporary staff?
- How much time will it take to research and select a vendor?
- How much time will it take to work with a vendor?

Money:

- How much money do you think the project is going to cost?
- How much money is available to complete this project?
- Will additional staff need to be hired?
- Will you need to hire consultants or other experts in the field?
- If the project is outsourced, how much will it cost?
- What will the cost of housing the digital materials be?
- Will an outside vendor be storing the files after conversion?

Available Resources:

- Does your staff include people knowledgeable about the digitization process and technologies?
- Are you able to train staff members to do the process?
- What scanner capabilities do you need, keeping in mind possible future uses?
- Are there machines currently available for you to use or do you need to purchase new equipment?
- Do you have the software needed to complete a project or do you have to purchase additional software?
- Do you have storage media or space appropriate to your project's size and time frame?
- Do you have the physical space to set up a digitizing lab, space for prep work, scanning and storage during the process?

Other factors to consider include:

Content:

- Do the records include non-public materials that cannot be removed from the building?
- Do certain parts of documents fall under privacy laws?

Physical Condition:

- Are the materials fragile?
- Do they require special care?
- Do the materials require prep work before they are digitized (e.g., removing staples or disbanding)?
- Is it cost-effective to do the prep work as part of the digitizing process or as a separate process?
- Is it cost effective to do the prep work ahead of time if contacting a vendor?
- Can the vendor do the prep work?

Quality Checks:

- How will quality checks be performed?
- Do you lose control of quality checks if the work is outsourced?
- Does the vendor have their own quality control process, and does it meet your standards?

Multiple Copies:

- Is the goal of the project to create both master and access files?
- Can both be made during the initial scanning?
- Should both be done at this time?
- Will there be any differences between the two copies, such as adding a watermark to access images?

Long-Term Access:

- Does the digitization plan allow for long-term access to the digital images?
- Are the records covered by a retention schedule?
- Who will want access to the digital files?
- Who controls the access?
- Is there a plan in place for migration of digital files if needed?

Originals:

- What will you do with the originals?
- If you want to dispose of them, do you have the authority to do so?
- Does the digital version meet your requirements for an official record?
- If keeping the originals; is on-site or off-site storage more cost effective?

There are pros and cons for both digitizing in-house and using vendor services. Some of each are listed below as suggested by the Bibliographic Center for Research in their *Digital Imaging Best Practices* guide.[23]

In-House Advantages

- Staff gains the knowledge of how to complete a digitization project (useful for future projects)
- Control over imaging process including handling and security of originals
- Ability to make changes as needed during the entire process

In-House Disadvantages

- Large cost up front, with continuing costs to maintain files and access
- The amount of time it takes to develop a plan, train staff or find people with appropriate credentials
- Current staff will only spend part of their time on this project, which may extend the time line for the project
- Must enforce standards with multiple people doing the same tasks

Out-Sourcing Advantages

- Do not need to purchase hardware or software, pay only for what is scanned
- Existing staff time is not redirected for an extended time period
- Vendor expertise can equal high quantity and quality production

Out-Sourcing Disadvantages

- High cost of paying vendor to scan
- Contract negotiation for EVERY part of the process, including working out the problems before there are any problems
- Less control over entire process, making changes is difficult
- Original materials must be transported and then handled by vendor staff

In reviewing these points and answering the questions above about each factor you might find that some of the process is best done in-house, while outsourcing other portions is a better choice. As is often the case, the answer is not black and white. You might choose

[23] Bibliographic Center for Research (BCR). *BCR's CDP [Collaborative Digitization Program] Digital Imaging Best Practices Version 2.0* June 2008. http://www.bcr.org/cdp/best/digital-imaging-bp.pdf

to do most of the project yourself while outsourcing a few tasks or do a little prep work and outsource most of the project. The Northeast Document Conservation Center highlights this idea in the preservation leaflet *Outsourcing and Vendor Relations*.[24] The same resource goes into detail about how to find a vendor, how to interview vendors, and how to work with vendors.

Whether you choose to do the work in-house or outsource it, your decisions must be documented. You must be able to prove that the records, once digitized, are still trustworthy. Internal digitization processes and policies must be documented and, when using an outside vendor, documentation must include such details as background information on the vendor, discussion of how records were delivered to and handled by the vendor, and the digitization and quality assurance processes.

File Formats

When determining which file format/s to use in a digitization project, you must consider the project goals. Is your goal simply to provide increased access and searchability to current paper records, or will you be replacing your original paper files with digital records – making them the *only* version available? In Minnesota, the law states, "all government data collected, created, received, maintained or disseminated by a government entity shall be public unless classified by statute… The responsible authority in every government entity shall keep records containing government data in such an arrangement and condition as to make them easily accessible for convenient use."[25] Creating digital surrogates provides another avenue for the public to pursue when trying to access government records.

The file format chosen must be able to stand the test of time required by the retention schedule. If digital records are to permanently replace paper files, the digital copy becomes the legal copy of the document and the retention period now applies to it. Some records might only need to be saved for five years, others ten, while others will need to be retained permanently. It would be a costly mistake to choose a transitional or fringe format that might become obsolete in a few years; the records would then need to be migrated to a new format, sooner than if you had chosen a stable, standard one. Taking the time to understand your goals for digitizing and studying the trends and current guidelines is a good investment.

Choosing formats that provide access to the greatest number of people over a long period of time is ideal. Guidelines to follow include using open source or non-proprietary formats, choosing formats that are widely available and accepted, and using formats that

[24] Dale, Robin L. *Outsourcing and Vendor Relations*. Northeast Document Conservation Center (NEDCC). 2007. http://www.nedcc.org/resources/leaflets/6Reformatting/07OutsourcingAndVendorRelations.php

[25] Minnesota Office of the Revisor. *Minnesota Statutes: 13.03 Access to Government Data*. https://www.revisor.leg.state.mn.us/statutes/?id=13.03

have become standards in the industry. The formats must be stable, well-supported, and well-documented. Stable file types have technical specifications that do not change frequently and are not subject to major changes over time. Any new or current versions created must be backward-compatible with older versions, ensuring continued use over time. This compatibility factor increases format stability and shows support of the file type over time. Well-documented file types have their technical specifications publicly available and easily accessible. The specifications explain the file type, what it does, how it can be used and allows for the file type to be recreate-able if necessary, the more details the better.

Much has been written on the importance of choosing an open-source or non-proprietary file type over a proprietary file type. Open source or non-proprietary formats are created and improved though collaboration, the results are not controlled by a single entity, and their encoding standards are publically available. The final product is a public, open-access file type that is supported by a large variety of programs[26]. On the other hand, "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 the long-term access and preservation of images."[27]

Choosing an open-source file type may remove many elements of future uncertainty as there are more program choices available to read the file type. However, the most important concept to remember is overall readability and use[28]. People must be able to read and use the digital records over time. When choosing a file format, ask yourself, is the format stable, well documented, and well supported? Non-proprietary is not always the best solution. TIFF, for example, is considered an archival standard by many even though the specifications for the file type are copyrighted by Adobe[29]. However, Adobe has made the comprehensive specifications for TIFF 6.0 public and states that "the goal is that TIFF files should never become obsolete and that TIFF software should not have to be revised more frequently than absolutely necessary."[30] In this statement Adobe is addressing the availability of specifications as well as the support for and stability of the file type over time.

Before a decision is made, you must also determine if there are any enterprise/agency/state guidelines that must be followed. The Minnesota Office of Enterprise Technology has produced the *Enterprise Technical Architecture* for the state of Minnesota, a guide that discusses the ideas behind the practices and standards for the state of Minnesota. "Data architecture describes how the State's electronic data should

[26] Wikipedia. *File Formats*. October 29, 2008. http://en.wikipedia.org/wiki/File_format.
[27] Bibliographic Center for Research. *BCR's CDP Digital Imaging Best Practices Version 2.0.* June 2008. http://www.bcr.org/cdp/best/digital-imaging-bp.pdf
[28] JISC Digital Media. *File Formats and Compression: Open –v Proprietary- Formats and Compressions*. March 2009. http://www.jiscdigitalmedia.ac.uk/stillimages/advice/file-formats-and-compression/
[29] Joint Information Systems Committee. *JISC Standards Catalogue*. October 2006. http://standards.jisc.ac.uk/catalogue/TIFF.phtml
[30] Adobe Developers Association. *Adobe Specifications for TIFF Revision 6.0.* June 3, 1992. http://partners.adobe.com/public/developer/en/tiff/TIFF6.pdf

be defined, stored, maintained and retained to facilitate processing, accessing, sharing, and analyzing from any part of the enterprise for appropriate constituencies according to existing federal and state laws".[31] Some states have specific guidelines. The Government Records Branch in North Carolina has produced a best practice guide for digital preservation for government entities entitled *N.C. Department of Cultural Resources: Archival Process for Data and Image Preservation: The Management and Preservation of Digital Media*. This guide states that the preferred file formats for textual documents include rich text format (rtf), PDF-A, and PDF, while the preferred formats for image documents are TIFF, JPEG2000 and SQL database.[32]

Compression is another issue that must be considered. You must decide if compression of the files is acceptable or not. In general, for archive or master copies compression is not acceptable as the files lose information during the compression process. There are different types of compression, each with its own intended use. Lossy, loses information during the compression process, such as with a JPEG, while the lossless technique looks identical to the uncompressed file, as with a TIFF file. Other new methods of compression have recently been developed, including using fractal and wavelet compression. JPEG 2000 uses wavelet compression and is a method that may allow 'compressed' files to be used as archival masters in the future.[33]

As is the case with any choice, choosing a file format is just one piece of the puzzle and you must look at the entire project to see how all the pieces will fit together. The choice becomes an "attempt to balance the requirements for quality, stability, potential longevity and industry acceptance."[34]

File Format Standards

Current file format standards for basic textual documents include:

TIFF:

Use for: Master copy of scanned documents.

Details: As stated in the specifications:[35]

---

[31] Office of Enterprise Technology. *Enterprise Technical Architecture 2.02*. September 8, 2006. http://www.state.mn.us/portal/mn/jsp/content.do?subchannel=-536891918&programid=536911145&sc3=null&sc2=null&id=-536891917&agency=OETweb

[32] All of the recommended files types are all proprietary formats, however they are well supported and well documented. Government Records Branch of North Carolina. *Guidelines for Public Records*. April 1, 2008. http://www.records.ncdcr.gov/guidelines.htm

[33] JISC Digital Media. *File Formats and Compression*. March 2009. http://www.jiscdigitalmedia.ac.uk/stillimages/advice/file-formats-and-compression/

[34] InterPARES 2 Project. *General Study 11 Final Report: Selecting Digital File Formats for Long-Term Preservation*. March 2007. www.interpares.org/display_file.cfm?doc=ip2_file_formats(complete).pdf

[35] Adobe Developers Association. *Adobe Specifications for TIFF Revision 6.0*. June 3, 1992. http://partners.adobe.com/public/developer/en/tiff/TIFF6.pdf

- TIFF is capable of describing bi-level, grayscale, palette-color, and full-color image data in several color spaces.
- TIFF includes a number of compression schemes that allow developers to choose the best space or time tradeoff for their applications.
- TIFF is not tied to specific scanners, printers, or computer display hardware.
- TIFF is portable. It does not favor particular operating systems, file systems, compilers, or processors.
- TIFF is designed to be extensible—to evolve gracefully as new needs arise.
- TIFF allows the inclusion of an unlimited amount of private or special-purpose information.

JPEG:

Use for: Access images.

Details: JPEG is technically a type of compression, not a file type, created by the Joint Photographic Experts Group.[36] The file format commonly called JPEG is actually a JFIF file format with the file extension .jpg, which uses JPEG compression and is called JPEG.[37] The file size of a JPEG makes it quick and easy to download, but the image quality deteriorates with increased compression.[38]

JPEGs use compression to reduce the file size of an object, making them useful for providing online access to images, for example, thumbnail images. [39] JPEGs can produce high-resolution images, but because the compression algorithm used to reduce the file size is lossy, information about the image itself is removed to create a smaller file size; the overall quality of the image is compromised.[40] "The amount of compression given to the file is chosen at the time of saving the file and allows for variation in quality against file size, as a rule of thumb, it is normally considered that a file compressed with JPEG to 10% of its original size will be visually acceptable with no obvious compression artefects. However it is

[36] JISC Digital Media. *Choosing a File Format for Digital Still Images: File Formats for Capture.* May 2006. http://www.jiscdigitalmedia.ac.uk/stillimages/advice/choosing-a-file-format-for-digital-still-images/
[37] JISC Digital Media. *Choosing a File Format for Digital Still Images: File Formats for Capture.* May 2006. http://www.jiscdigitalmedia.ac.uk/stillimages/advice/choosing-a-file-format-for-digital-still-images/
[38] Strategic Web Ventures. *Definition of JPEG*. 2004. http://www.strategicwebventures.com/definitions/Glossary/JPEG/
[39] Bibliographic Center for Research. *BCR's CDP Digital Imaging Best Practices Version 2.0.* June 2008. http://www.bcr.org/cdp/best/digital-imaging-bp.pdf
[40]Sharpened Glossary. *Definitions of Computer Terms: JPEG.* http://www.sharpened.net/glossary/definition.php?jpeg

common if required, to compress right down to 2-4% if the lower quality is acceptable."[41]

JPEG is the most common format used online for storing and transmitting photographs and most digital cameras create JPEGs, both enforcing the popularity of the standard.[42]

PDF/A:

Used for: Newer format emerging as a long-term storage option.

Details: PDF/A is based on Adobe's PDF format and was designed with the purpose of creating "a standard format for electronically archived documents".[43] The ISO standard for PDF/A (ISO 19005-1) defines PDF/A as "a file format based on PDF … which provides a mechanism for representing electronic documents in a manner that preserves their visual appearance over time, independent of the tools and systems used for creating, storing or rendering the files."[44] The format is appropriate for documents that contain character, raster and vector data making it a more flexible option.[45] Features making PDF/A an attractive solution over a TIFF file include having a higher quality image with a smaller file size and the immediate ability to search within the document. Another advantage of PDF/A is the "metadata, such as the title, author, creation date, modification date, subject, keywords, etc. can be embedded in the PDF/A file which can then be used to classify the files without human intervention."[46]

JPEG2000:

Used for: Access images, but has potential for master archiving use in the future.[47]

---

[41] JISC Digital Media. *Choosing a File Format for Digital Still Images: File Formats for Capture.* May 2006. http://www.jiscdigitalmedia.ac.uk/stillimages/advice/choosing-a-file-format-for-digital-still-images/

[42] Mario Salexandrou. Web Strategist and Project Manager. *Definition JPEG*. http://www.mariosalexandrou.com/definition/jpeg.asp

[43] PDF/A Competence Center. *PDF/A – A New Standard for Long-Term Archiving*. 2008. http://www.pdfa.org/doku.php?id=pdfa:en:pdfa_whitepaper

[44] PDF/A Competence Center. *PDF/A – A New Standard for Long-Term Archiving*. 2008. http://www.pdfa.org/doku.php?id=pdfa:en:pdfa_whitepaper

[45] International Organization for Standardization. *ISO 19005-1:2005*. http://www.iso.org/iso/catalogue_detail?csnumber=38920

[46] PDF/A Competence Center. *PDF/A – A New Standard for Long-Term Archiving*. 2008. http://www.pdfa.org/doku.php?id=pdfa:en:pdfa_whitepaper

[47] Bibliographic Center for Research. *BCR's CDP Digital Imaging Best Practices Version 2.0.* June 2008. http://www.bcr.org/cdp/best/digital-imaging-bp.pdf

Details: "JPEG 2000 is a wavelet-based image compression method which has been designed as the successor of the popular JPEG image compression. One of the most important differences is that JPEG 2000 includes a lossless mode for image compression whereas JPEG always used lossy image compression algorithms."[48] JPEG 2000 also allows embedded metadata, which many others formats do not. "Its architecture should lend itself to a wide range of uses from portable digital cameras through to advanced pre-press, medical imaging and other key sectors."[49]

In the United States, both the Library of Congress [50] and the Collaborative Digitization Program of the Bibliographical Center for Research[51] have adopted TIFF (for masters) and JPEG (for web access) as their standards, while PDF-A is the preferred format for long-term preservation at the Bentley Historical Library at the University of Michigan.[52] Internationally, JISC Digital Media has adopted TIFF (for master) and JPEG (for web delivery) as standards[53] while the Digital Preservation Coalition (DPC) in the United Kingdom has recently endorsed PDF/A as the best file type to preserve electronic documents well into the future[54].

Resolution

The quality of images is most often determined by resolution. Resolution is commonly measured in PPI (pixels per inch) or DPI (dots per inch). Putting these pixels or tiny dots close together forms an image; increasing the number of pixels per inch captures more details of an image. As more details are captured with increasing resolution, the file size also increases.

When determining what resolution to use for scanning your documents, you need to analyze the contents of the materials to be digitized. Textual documents are generally text only, but they could also include graphs, photographs, or hand written annotations. Requirements for a high quality scan of a photograph differ from that of a textual document. Textual documents that include both text and photographs, or other details

[48] Dret.net. *Web and XML Glossary: JPEG2000*. March 13, 2007. http://dret.net/glossary/jpeg2000
[49] JPEG Committee. *JPEG 2000: Our New Standard*. 2007. http://www.jpeg.org/jpeg2000/
[50] Library of Congress. *JPEG Lossy (DCT) Compression Encoding*. March 7, 2007. http://www.digitalpreservation.gov/formats/fdd/fdd000017.shtml
[51] Bibliographic Center for Research. *BCR's CDP Digital Imaging Best Practices Version 2.0.* June 2008. http://www.bcr.org/cdp/best/digital-imaging-bp.pdf
[52] Bentley Historical Library, University of Michigan. *University Archives Best Practices*. May 14, 2008. http://bentley.umich.edu/uarphome/bestprac/word.php
[53] JISC Digital Media. *Choosing a File Format for Digital Still Images: File Formats for Master Archives.* May 2006. http://www.jiscdigitalmedia.ac.uk/stillimages/advice/choosing-a-file-format-for-digital-still-images/
[54] Digital Preservation Coalition. *Press Release: PDF should be used to preserve information for the future.* 2002. http://www.dpconline.org/graphics/advocacy/pdf-0804.html

such as hand written notes, may need to have a higher resolution than a basic textual document to capture page details.

Resolution is a complicated idea that is often not completely understood. The term resolution refers to image resolution (SPI), display resolution (PPI), printer resolution (DPI), and halftone resolution (LPI), all of which are measured slightly differently. To determine which type of resolution is most important, you will need to understand the goals of the digitization project. Are you working on a project that will place thumbnail images online? Will you have the need to print the scanned documents sometime in the future? Do you want to create digital master copies to keep permanently? Will you have to migrate the images from one storage media to another?

Below you will find simple definitions of each type of resolution to help explain why knowing and understanding your project goals are so important.

> **SPI (samples per inch):** is scanner and digital *image resolution*. To scan an image the scanner takes a sampling of portions of the image. The more samples it takes per inch, the closer the scan is to the original image. The higher the resolution, the higher the SPI.[55]
>
> **PPI (pixels per inch):** is the number of pixels displayed in an image on a screen. A digital image is composed of samples that your screen displays in pixels. The PPI is the *display resolution*.[56]
>
> **DPI (dots per inch):** is a measure of the resolution of the printer. It properly refers to the dots of ink or toner used by an image setter, laser printer, or other printing device to print your text and graphics. In general, the more dots, the better and sharper the image. DPI is *printer resolution*.[57]
>
> **LPI (lines per inch):** refers to the way printers reproduce images, simulating continuous tone images by printing lines of halftone spots. The number of lines per inch is the LPI, sometimes also called the frequency. You can think of LIP as the *halftone resolution*.[58]

Despite the differences outlined above, the terms SPI, PPI and DPI are often used interchangeably. Context is needed to decipher what type of resolution is truly being discussed. All further discussion below will use the term DPI to refer to resolution (unless text is quoted), as it is a term that is often used for image resolution.

---

[55] Jacci Howard Bear, About.com. *Measuring Resolution Inch by Inch*. http://desktoppub.about.com/cs/intermediate/a/meas_resolution.htm
[56] Jacci Howard Bear, About.com. *Measuring Resolution Inch by Inch*. http://desktoppub.about.com/cs/intermediate/a/meas_resolution.htm
[57] *Ibid.* http://desktoppub.about.com/cs/intermediate/a/meas_resolution.htm
[58] *Ibid.* http://desktoppub.about.com/cs/intermediate/a/meas_resolution.htm

It is important to realize that there is not one standard resolution that fits all types of documents. The size of the document, type of images or text on the document, and desired outcome of the digital copies all come into play when determining what resolution to use.[59] Keep in mind, the higher the resolution the larger the file size which may cause storage issues with large digitization projects. "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."[60] Also, it is not wise to always scan at the highest level the scanner allows, because at some point "adding more pixels per inch no longer adds content"[61] and all you end up doing is increasing the file size. The visual resolution is limited by the clarity and information contained in the original document, the resolution of the screen, as well as the resolution of the printer if it is to be printed.

Recommendations for Text Documents

Some institutions have set "standards" for typical document types. For example, many institutions use 400 to 600 DPI as a standard resolution for plain text documents, although some institutions go as low as 200 DPI in these cases.[62]

The Bibliographic Center for Research (BCR) provides more specific and technical guidelines in *Digital Imaging Best Practices*, that can be used to find the appropriate resolution for documents smaller or larger than the standard 8½ by 11 piece of paper or for documents that contain unusually small or large font sizes. For master images BCR recommends "adjusting 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 and 16 bit files."[63] The spatial dimensions should be "4000 to 6000 pixels across the long dimension."[64]

Pixels across the long dimension are the number of pixels found along the longest dimension of a document. To determine the number of pixels you need to know the resolution and document size. For example, an 8½ by 11 piece of paper scanned at 100 dpi resolution has 1100 pixels across the long dimension (100x11 = 1100) and the same piece of paper scanned at 400 dpi has 4400 pixels across the long dimension (400x11 = 4400); whereas a postcard 3½ by 5 would need to be scanned at 800 dpi to reach 4000 pixels across the long dimension (800x5 = 4000).

The number of pixels across the long dimension is used as a benchmark because if the smaller dimension was used, the resulting image would not have a high enough resolution

[59] Bibliographic Center for Research. *BCR's CDP Digital Imaging Best Practices Version 2.0.* June 2008. http://www.bcr.org/cdp/best/digital-imaging-bp.pdf
[60] *Ibid.* http://www.bcr.org/cdp/best/digital-imaging-bp.pdf
[61] *Ibid.* http://www.bcr.org/cdp/best/digital-imaging-bp.pdf
[62] U.S. Government Printing Office. *Report on the Meeting of Experts on Digital Preservation.* March 12, 2004. http://www.gpoaccess.gov/about/reports/preservation.pdf
[63] Bibliographic Center for Research. *BCR's CDP Digital Imaging Best Practices Version 2.0.* June 2008. http://www.bcr.org/cdp/best/digital-imaging-bp.pdf
[64] *Ibid.* http://www.bcr.org/cdp/best/digital-imaging-bp.pdf

to accurately represent the scanned document. If too high of a resolution is used, the image becomes very pixilated.

Other Considerations Affecting Resolution

Image Size

Image size is the physical size of scanned document. When using the standard settings on photo editing programs, resolution does not control the size of the image; changing the resolution changes the number of pixels in a file, but not the overall size of the image. Using photo editing software, the image size itself can be increased or decreased at any time, but doing so will affect the look of the image because the resolution (or number of dots per inch), does not change when the electronic image size is changed; the same number of dots will fill the selected image size, they just get closer or further apart. This can best be seen when you compare a 4x6 image with 72 dpi and the same 4x6 image with 36 dpi. The 36 dpi image will look very pixilated, as there are less dots available to fill the same amount of space. However, if you take the same 4x6 36 dpi image and reduce the image size to 2x3, the image will look almost the same as the 4x6 with 72 dpi resolution, because by decreasing the image size, the dots per inch increased. So, remember that image size matters when selecting a resolution.

File Size

At some point you may be restricted by file size. You may want to save files at 20 MB vs. 40 MB to conserve storage space. How would you do this? Higher resolution images have larger file sizes. If you have a particular file size that you need to maintain, you can keep the file size constant by using settings in photo editing software, but remember that "increasing the size of an image decreases its resolution and decreasing its size increases its resolution."[65] This will change your final product considerably, for better or worse. Make sure you know what the final image actually looks like in print and on screen.[66]

Bit-Depth

File size is also affected by the bit depth or the number of bits represented. The National Archives and Records Administration (NARA) explains that "bit-depth or signal resolution, sometimes called tonal resolution, defines the maximum number of shades and/or colors in a digital image file, but does not define or guarantee the quality of the

[65] IEEE. *A Word on Graphic Image Resolution.* April 29, 2002. http://www.ieee.org/organizations/pubs/magazines/imageres.htm

[66] We just stated above that changing the size of the electronic image does not change the resolution and it does not, however with a program like Adobe Photoshop you can make it do so.

image."[67] Digital image quality is dependent on the original image as well as the entire capture process. Bits determine how color is read by the imaging software, "each pixel or 'bit' contains color information for the image."[68]

> **1-bit file:** Each pixel is represented by a single binary digit (0 or 1), so the pixel can either be black or white.[69]
>
> **8-bit file:** Each pixel is represented by eight binary digits of data per channel which provides a maximum of 256 shades per channel ranging from black to white.[70]
>
> **16-bit file:** Each pixel is represented by sixteen binary digits of data per channel, which provides a maximum of 65,000 shades per channel. [71]

Why do bits matter? Both TIFF and JPEG are bitmap graphics. Once created "bitmap graphics have a fixed resolution which means that resizing a bitmap graphic can result in distortion and jagged edges."[72] Bitmap graphics "look best at the size they were created. Making bitmaps larger usually distorts them".[73] Fixed resolution means that once scanned at a certain resolution, changing the resolution affects the details and information of the image. It is not recommended to try to increase the resolution on master images, as the overall quality of the image will decrease because the added pixels are rendered based on the software's best guess of what color pixels should be there. Reducing resolution removes pixels (or information) from an image, reducing overall quality. Increasing resolution means that pixels need to be added to the image. To do this, the photo editing software has to take a best guess at what color pixels should be added, altering the digital image, making it not an exact copy of the original image.

Color Choice: Black and White, Grayscale, or Color

What 'color' should you capture the document in? Scanning in black and white, grayscale or color depends on the document in hand. A text document might only consist of black text or could contain multi colored text. A photograph contained within a text document could be in black and white, sepia, or color. What color scale should be used and when? Thinking back to bit-depth, scanning documents in black and white would

---

[67] National Archives and Records Administration (NARA). *Technical Guidelines for Digitizing Archival Masters for Electronic Access: Creation of Production Master Files – Raster Images*. June 2004. http://www.archives.gov/preservation/technical/guidelines.html
[68] Jacci Howard Bear. About.com. *Bitmap*. http://desktoppub.about.com/od/glossary/g/bitmap.htm
[69] National Archives and Records Administration (NARA). *Technical Guidelines for Digitizing Archival Masters for Electronic Access: Creation of Production Master Files – Raster Images*. June 2004. http://www.archives.gov/preservation/technical/guidelines.html
[70] *Ibid.* http://www.archives.gov/preservation/technical/guidelines.html
[71] National Archives and Records Administration (NARA). *Technical Guidelines for Digitizing Archival Masters for Electronic Access: Creation of Production Master Files – Raster Images*. June 2004. http://www.archives.gov/preservation/technical/guidelines.html
[72] Jacci Howard Bear. About.com. *Bitmap*. http://desktoppub.about.com/od/glossary/g/bitmap.htm
[73] *Ibid*. http://desktoppub.about.com/od/glossary/g/bitmap.htm

create 1-bit files. If the images are only rendered in black or white, what about shadowed text or hand-written annotations in the document? These might not come across in the 1-bit scanned document as hoped. To capture the finer details, it would be better to scan the document in grayscale, using 8- or 16-bits of code. Full-color documents or documents where the exact color makes a difference should be scanned with the largest bit depth to capture the widest range of color possible. The NARA explains the relationship between bit depth, grayscale and color in the following manner: "grayscale image files consist of a single channel, commonly either 8-bits or 16-bits per pixel with the tonal values ranging from black to white. Color images consist of three or more grayscale channels that represent color and brightness information, common color modes include RGB (red, green, blue), CMYK (cyan, magenta, yellow, black), and LAB (lightness, red-green, blue-yellow). The channels in color files may be either 8-bits or 16-bits. Display and output devices mathematically combine the numeric values from the multiple channels to form full color pixels, ranging from black to white to full colors."[74]

When working with color documents, true colors become very important. Colors must be managed throughout the entire digitization process – in the capture device (scanner or camera), on the viewing screen, in the photo editing software, and on the printer. Each step in the process has its own method for calibrating color. If color is not managed, the colors become misrepresented at some point in the process and the end product will not be a true representation of the original file. Please review BCR's summary of color management in their *Digital Imaging Best Practices Version 2.02* guide under 'Quality Control.'[75] The Getty also has a detailed guide written by Howard Besser covering issues of resolution, color management, and other imaging concerns.[76]

Photo Editing Software: Making Adjustments

Programs like Adobe Photoshop allow you to capture scanned images and manipulate them in a variety of ways. These tools offer great flexibility in what you can do with digital images. Adobe has many online tutorials and help guides to help you understand the features of any program of theirs being used.[77] For example, Photoshop allows the user to:

- Change the resolution of a file, by changing the number of pixels in the image, keeping the physical size of the image the same
- Change the resolution of a file, by reducing the physical size of the image, keeping the number of pixels the same (keeping the file size the same).

[74] National Archives and Records Administration (NARA). *Technical Guidelines for Digitizing Archival Masters for Electronic Access: Creation of Production Master Files – Raster Images*. June 2004. http://www.archives.gov/preservation/technical/guidelines.html

[75] Bibliographic Center for Research. *BCR's CDP Digital Imaging Best Practices Version 2.0.* June 2008. http://www.bcr.org/cdp/best/digital-imaging-bp.pdf

[76] Besser, Howard. Edited by Sally Hubbard with Deborah Lenert. *Introduction to Imaging – Revised Edition: The Image*. J. Paul Getty Trust. 2003. http://www.getty.edu/research/conducting_research/standards/introimages/image.html

[77] Adobe. *Photoshop Help and Support*. http://www.adobe.com/support/photoshop/

- Change the number of pixels across each dimension, by changing the document size, keeping the resolution the same
- Control the file size, number of pixels, and resolution independently or grouped in various fashions

This flexibility provides more options when creating digital images, but makes the issue more complex. Even though changes can be made to the resolution, file size and pixel dimensions, they should be done with great consideration for what the desired final result will be. In general, someone should spend time determining the best resolution based on the material, the document size, and desired file size. After this is determined, a standard process should be set and followed, creating the highest quality images possible for the desired output. As stated before, BCR recommends adjusting the scanning resolution based on the size of the original document, the size of the text, and the quality of the document being scanned. Their recommendations of 6000 and 4000 pixels across the long dimension will help set scanning parameters. If large number documents of the same size and type are to be scanned, take the time to find out what works and use the same rules for all of them.

Summary of Resolution

Terms to understand:

- **Raster/Bitmap:** An image created using individual pixels or dots.
- **Resolution:** measures the number of pixels or dots that are grouped together forming a raster or bitmap image.
- **Pixels:** the individual dots that store color information that when displayed together create an image.
- **Bit-Depth:** records the color information of a file based on the number of channels available – 1, 8, or 16.

Points to Consider:

- Increasing resolution increases file size.
- Higher resolution images are higher quality.
- Resolution is not the same as the total number of pixels in the image. Total number of pixels depends on the total file size.
- Photo editing software allows adjustments to be made, for better or worse.
- How to determine pixels across the long dimension – multiply the number of inches of the long dimension with the resolution of the file.
- Color management is important to research, and is dependent on the materials you choose to use.

The most important idea to take away from this discussion is that there is not a "one-size-fits-all" standard. Use industry guidelines as benchmarks on which to create your own standard. Standards must be set with the project goals in mind. The chosen standard must then be followed during the entire project for consistency sake. If variations are made, either intentional or unintentional, these should be documented.

File Naming

A strong file naming convention is the first step in facilitating access to digital records. This "set of agreed-upon rules used to assign identifiers to digital objects in a collection"[78] must be followed by all users for the system to be useful. "Naming records consistently, logically and in a predictable way will distinguish similar records from one another at a glance, and by doing so will facilitate the storage and retrieval of records, which will enable users to browse file names more effectively and efficiently."[79]

To assist with future records management, any decisions made about file naming conventions should be documented.[80] Documentation should include information about the collection the policy was made for, the names of the people involved in the decision making process, the date the policy was implemented, and the list of specific rules that define the specific convention.[81] Documenting the process and rules developed is part of a strong records management program. The *Electronic Records Management Guidelines*, produced by the Minnesota Historical Society (MHS), reminds us that "legally your records must be trustworthy, complete, accessible, legally admissible in court and durable for as long as your approved records retentions schedule requires. Records that are consistently and logically named are easier to manage to meet these requirements."[82]

Many organizations, institutions, and government agencies have implemented file naming policies for themselves by following industry standards. In *Naming Conventions for Electronic Documents*, the Alberta government has provided readers with an overview of important concepts to consider when developing a file naming convention. In general, they suggest that a file naming system should create a name that includes some or all of the following elements: title, version number, date [of creation], author/creator, business unit/program [associated with the record], type [such as a letter,

---

[78] Digital Projects Advisory Group, University Libraries at the University of Colorado at Boulder. *File Naming Conventions for Digital Collections*. March 4, 2008. http://ucblibraries.colorado.edu/systems/digitalinitiatives/docs/filenameguidelines.pdf

[79] University of Edinburgh, Records Management Section. *Standard Naming Conventions for Electronic Records*. July 6, 2007. http://www.recordsmanagement.ed.ac.uk/InfoStaff/RMstaff/RMprojects/PP/FileNameRules/FileNameRules.htm

[80] Digital Projects Advisory Group, University Libraries at the University of Colorado at Boulder. *File Naming Conventions for Digital Collections*. March 4, 2008. http://ucblibraries.colorado.edu/systems/digitalinitiatives/docs/filenameguidelines.pdf

[81] *Ibid.* http://ucblibraries.colorado.edu/systems/digitalinitiatives/docs/filenameguidelines.pdf

[82] Minnesota Historical Society. *Electronic Records Management Guide - File Naming*. March 2004. http://www.mnhs.org/preserve/records/electronicrecords/erfnaming.html

memo, report], and file extension.[83] The MHS points out a few more elements that may be important to your project, such as name the of intended audience, release date, publication date, project number, department number, and record series. A file name that includes such elements will support access and retrieval, file sorting, and version control.

File names can be descriptive or non-descriptive; the choice is yours and depends on the specific project and intended document usage. A descriptive file name uses the elements discussed above to show at a glance what the file represents before it is opened. For example, *2004-11_Annual_Meeting_Notes_JonesMJ_v02.doc* can be easily interpreted without knowing the context of the file. Descriptive file names work well for medium to small collections while non-descriptive file names generally work well for large collections; the larger the collection the harder it becomes to create unique file names.[84] Non-descriptive file names use alpha-numeric characters that when put together do not provide any identifying information about the file itself. For example, *XM00251.jpg* looks like a random grouping of letters and numbers that when taken out of context has no meaning. "The decision to use descriptive or non-descriptive file names should be based on the collection's characteristics and project specifications."[85]

Guidelines to Consider for File Naming[86]

- Create unique file names. Duplicate file names will cause problems.
- File names should be simple and easy to understand.

---

[83] Alberta Government. *Naming Conventions for Electronic Records*. August 2005. http://www.im.gov.ab.ca/publications/pdf/DocumentNamingConventions.pdf

[84] Digital Projects Advisory Group, University Libraries at the University of Colorado at Boulder. *File Naming Conventions for Digital Collections*. March 4, 2008. http://ucblibraries.colorado.edu/systems/digitalinitiatives/docs/filenameguidelines.pdf

[85] Digital Projects Advisory Group, University Libraries at the University of Colorado at Boulder. *File Naming Conventions for Digital Collections*. March 4, 2008. http://ucblibraries.colorado.edu/systems/digitalinitiatives/docs/filenameguidelines.pdf

[86] Information about file naming conventions gathered from the following resources: Alberta Government. *Naming Conventions for Electronic Records*. August 2005. http://www.im.gov.ab.ca/publications/pdf/DocumentNamingConventions.pdf; Bibliographic Center for Research's (BCR) *BCR's CDP [Collaborative Digitization Program] Digital Imaging Best Practices Version 2.0* June 2008. http://www.bcr.org/cdp/best/digital-imaging-bp.pdf; Brookhaven National Laboratory (BNL) Web Communication Standards. *File Naming Conventions and Directory Structure*. February 5, 2008. http://www.bnl.gov/webstandards/fileNaming.asp; Digital Projects Advisory Group, University Libraries at the University of Colorado at Boulder. *File Naming Conventions for Digital Collections*. March 4, 2008. http://ucblibraries.colorado.edu/systems/digitalinitiatives/docs/filenameguidelines.pdf; JISC Digital Media. *Choosing a File Name*. November 2008. http://www.jiscdigitalmedia.ac.uk/crossmedia/advice/choosing-a-file-name/; Minnesota Historical Society. *Electronic Records Management Guide - File Naming*. March 2004. http://www.mnhs.org/preserve/records/electronicrecords/erfnaming.html; North Carolina Department of Cultural Resources. *Best Practices for File-Naming*. May 7, 2008. http://www.records.ncdcr.gov/erecords/filenaming_20080508_final.pdf;

- Use only alpha-numeric characters. Avoid using special characters such as: ? / $ % & ^ # . \ : < >. Special characters are often reserved for use by the operating system.

- Use underscores (_) and dashes (-) to represent spaces. Spaces are often reserved for operating system functions and might be misread.

- Use leading zeros with the numbers 0-9 to facilitate proper sorting and file management.

- Dates should follow the ISO 8601 standard of YYYY_MM_DD or YYYYMMDD[87]. Variations include YYYY, YYYY-MM, YYYY-YYYY. This maintains chronological order. If dates of creation are used, these can make following retention schedules easier.

- Keep the file name as short as possible and always include the three character file extension preceded with a period (Ex: .jpg or .doc).

- Include the version number in the file name by using 'v' or 'V' and the version number at the end of the document. (Ex: 2004_Notes_v01.doc) Avoid using the word version or draft and the beginning of the file name for access purposes (Ex: Version1_2004_Notes.doc).

- Order the pieces of information or elements being used to create the file name in the most logical order based on retrieval methods. For example, use the date first on events that are time specific or reoccurring, and use the name of the event for events that are infrequent and will be easier to find by name rather than date.

Building on the above rules, the Bibliographic Center for Research (BCR) has suggested in the *Digital Imaging Best Practices Guide* specific ways to create file names. BCR recommends beginning each file name with a two or three character acronym representing the institution's name followed by another set of characters representing a department or unit name if necessary. These acronyms should be followed by an object ID number using the unique numbering scheme already used. In addition, if the objects are made of multiple parts, they should be identified by adding a letter (a, b, c…) after the ID number.[88]

[87] North Carolina Department of Cultural Resources. *Best Practices for File-Naming.* May 7, 2008. http://www.records.ncdcr.gov/erecords/filenaming_20080508_final.pdf

[88] Bibliographic Center for Research. *BCR's CDP Digital Imaging Best Practices Version 2.0.* June 2008. http://www.bcr.org/cdp/best/digital-imaging-bp.pdf

Variations on the Guidelines[89]

There are points of disagreement among file naming guidelines, including:

- Some guidelines suggest avoiding capital letters as some browsers may be case sensitive; other guidelines suggest using capital letters instead of underscores or dashes to delineate words (Ex: AnnualMeetingNotes rather than Annual_Meeting_Notes)

- Some guidelines suggest using dashes instead of underscores when working with online documents as the dashes allow for greater searchability. For example, if underscores are used, Google can only find the complete string of words during a search, but if dashes are used, each word in the string becomes searchable and will create more hits making the document more accessible.[90]

- Some guidelines suggest avoiding unnecessary repetition and redundancy in file names and file paths. For example, do not use part of the folder structure to name the file; it repeats information. However, depending on the naming convention, this could lead to files with the same name. Other guidelines state the reason not to use folder structure information as part of the name is because the structure itself can change overtime. Because structure can change over time, other guidelines highlight specifically that the name should always be descriptive and independent of file location. This allows a disassociated file to be returned to its proper location.

- Some guidelines suggest that names should be listed using first initial then last name (JSmith), others prefer using the last name first followed by the first initial (SmithJ). Using the last name first, sorts the files by last name, which is how paper files are commonly sorted.

---

[89] Information about file naming conventions gathered from the following resources: Alberta Government. *Naming Conventions for Electronic Records*. August 2005. http://www.im.gov.ab.ca/publications/pdf/DocumentNamingConventions.pdf; Bibliographic Center for Research's (BCR) *BCR's CDP [Collaborative Digitization Program] Digital Imaging Best Practices Version 2.0* June 2008. http://www.bcr.org/cdp/best/digital-imaging-bp.pdf; Brookhaven National Laboratory (BNL) Web Communication Standards. *File Naming Conventions and Directory Structure*. February 5, 2008. http://www.bnl.gov/webstandards/fileNaming.asp; Digital Projects Advisory Group, University Libraries at the University of Colorado at Boulder. *File Naming Conventions for Digital Collections*. March 4, 2008. http://ucblibraries.colorado.edu/systems/digitalinitiatives/docs/filenameguidelines.pdf; JISC Digital Media. *Choosing a File Name*. November 2008. http://www.jiscdigitalmedia.ac.uk/crossmedia/advice/choosing-a-file-name/; Minnesota Historical Society. *Electronic Records Management Guide - File Naming*. March 2004. http://www.mnhs.org/preserve/records/electronicrecords/erfnaming.html; North Carolina Department of Cultural Resources. *Best Practices for File-Naming*. May 7, 2008. http://www.records.ncdcr.gov/erecords/filenaming_20080508_final.pdf

[90] Matt Cutts. *Dashes vs. underscores.* Matt Cutts: Gadgets, Google, and SEO. August 25, 2005. http://www.mattcutts.com/blog/dashes-vs-underscores/

- Some guidelines suggest that file names should be no longer than eight characters, others suggest 11, 25 or 31. Most browsers now accept long file names, however, to ensure full backwards compatibility JISC Digital Media suggests that the file name be no longer than 8 characters as set by the ISO 9660 standard for CD-ROM to enable use across platforms.[91] The BCR and MNHS both suggest the length should be limited to 31 characters or less for files on a network or removable media (including the extension) but also recommend limiting the full name to 11 characters or less (including the extension) when transferring files to an external system to allow backward compatibility, also complying with the ISO 9660 standard.

Additional Considerations

The MHS suggests considering the following when deciding on a file naming convention:[92]

> *Access and ease of use*. The policy should be simple and straightforward. A simple policy will help staff members logically and easily name records and help ensure that records are accessible to staff members and/or to the public…. A simple policy will be more consistently used, resulting in records that are consistently named, and thus easier to organize and access.
>
> *Ease of administration*. The policy should work with your computer infrastructure, so that you can monitor policy compliance, manage records and records series, gather metadata, and perform other administrative tasks easily and in compliance with all legal requirements. For example, if all the records in a specific records series are easily identifiable by file name, they will be easier to gather and manage.
>
> *Scalability*. Consider how scalable your file naming policy needs to be. For example, if you want to include the project number, don't limit your project numbers to two digits, or you can only have ninety-nine projects.
>
> *Persistence over time*. File names should outlast the records creator who originally named the file. With good stakeholder and staff input, and training, you should be able to develop file names that make sense to staff members once the file creators are no longer available.

[91] JISC Digital Media. *Choosing a File Name*. November 2008. http://www.jiscdigitalmedia.ac.uk/crossmedia/advice/choosing-a-file-name/
[92] Minnesota Historical Society. *Electronic Records Management Guide - File Naming*. March 2004. http://www.mnhs.org/preserve/records/electronicrecords/erfnaming.html

Reviewing the guidelines implemented by other institutions is the first step to creating your own file naming convention. The next steps when developing a file naming convention should be to utilize the work of others and think about your project goals. Talk with staff and users about your ideas. Use forward thinking to ensure long-term solutions. Develop a plan, test it, and record the process. Did the convention work? If so, train others in the process stressing the importance of following the policy. Follow up during the digitization process to ensure consistency with file naming conventions. Understanding file naming conventions as well as good record management policies will help you create a strong naming convention of your own.

Metadata

Metadata is often defined as data about data. Metadata records information about objects, including documents or files in either paper or digital format. The recorded information describes the object itself, its relationships with objects, and how the object was and should be treated over time. As stated by the Bibliographical Center for Research (BCR) Collaborative Digitization Program (CDP) Metadata Working Group, metadata must provide information that:

- certifies the authenticity and degree of completeness of the content
- establishes and documents the context of the content
- identifies and exploits the structural relationships that exist between and within information objects
- provides a range of intellectual access points for an increasingly diverse range of users
- provides some of the information that an information professional might have provided in a physical reference or research setting.[93]

Metadata is important to a digitization project because it tells you what you have and how to use it. The main categories of metadata are referred to as descriptive metadata, technical metadata, structural metadata, and administrative metadata. The following table defines these four categories of metadata and is taken from the BCR CDP's *Digital Imaging Best Practices Version 2.0* guide.[94]

[93] Collaborative Digitization Program (CDP) Metadata Working Group. *Dublin Core Metadata Best Practices Version 2.1.1*. September 21, 2006. http://www.bcr.org/cdp/best/dublin-core-bp.pdf
[94] Bibliographic Center for Research. *BCR's CDP Digital Imaging Best Practices Version 2.0.* June 2008. http://www.bcr.org/cdp/best/digital-imaging-bp.pdf

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 dimension and hardware. The information is critical for migration and long-term sustainability of the digital resource

Metadata itself is very fluid and often fits into more than one of the categories listed above. The National Archives and Records Administration (NARA) describes the above categories in more detail and also utilizes additional metadata categories that address other issues, including metadata about rights management, behavior or presentation, image quality assessment, records management, workflow processes, and even metadata about the metadata (meta-metadata).[95]

Metadata is either captured automatically or collected and entered manually. Automatically generated metadata may include the date an image was created, the program used for capture or manipulation, resolution, file size and color scheme. Descriptions of the content including object type (memo, letter, bill, act, proposal), names of the people involved (author, intended audience, about who), and why the object was created must usually be entered manually. The more manually entered metadata the harder it becomes to enforce standards. Policies and procedures for metadata entry are a must. By standardizing the process it will be easier to manage, access, and preserve the files long-term. There are many technical standards available for metadata, including Dublin Core (DC).[96]

Dublin Core[97] is an ISO/ANSI standard and internationally recognized as a set of standard core metadata elements whose simplicity leads to great flexibility while at the same time facilitating access. Dublin Core mandates that fifteen standard elements be

[95] United States 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
[96] Spindler, Robert P. *Digital Preservation (New 2007).* Northeast Document Conservation Center. 2007. http://www.nedcc.org/resources/leaflets/6Reformatting/05DigitalPreservation.php
[97] Dublin Core Metadata Initiative. *Home Page*. http://dublincore.org/

used to describe an object; it is then your choice to add any other elements appropriate for your project. For example, if you complete twenty small digitization projects and used DC as the standard base-line for cataloging, even though the projects are very different and the additional metadata collected is different, the baseline metadata makes sharing information between the collections possible.

The purpose of preservation metadata is to facilitate management and access to digital records over time; therefore a preservation metadata schema inherently includes descriptive, administrative, structural, and technical metadata elements.[98] A report sponsored by the Digital Preservation Coalition, recommends that preservation metadata should specifically include information about:

> *Provenance:* Preservation metadata should record information bearing on the custodial history of the digital object, potentially stretching back to the time of the object's creation, and moving forward through successive changes in physical custody and/or ownership.[99]
>
> *Authenticity:* Preservation metadata should include information sufficient to validate that the archived digital object is in fact what it purports to be, and has not been altered, either intentionally or unintentionally, in an undocumented way.[100]
>
> *Preservation activity:* Preservation metadata should document the actions taken over time to preserve the digital object, and record any consequences of these actions that impact the look, feel, or functionality of the object.[101]
>
> *Technical environment:* Preservation metadata should describe the technical requirements, such as hardware, operating system, and software applications, needed to render and use the digital object in the state in which it is currently stored in the repository.[102]
>
> *Rights management:* Preservation metadata should record any binding intellectual property rights that limit the repository's powers to take action to preserve the digital object, and to disseminate the object to current and future users.[103]

With international recognition of the importance of preservation metadata, the Online Computer Library Center and the Research Library Group sponsored the PREMIS (PREservation Metadata: Implementation Strategies) working group. The main goals of

---

[98] Lavoie, Brian and Richard Gartner. *Preservation Metadata*. Digital Preservation Coalition Technology Watch Report. September 2005. http://www.dpconline.org/docs/reports/dpctw05-01.pdf
[99] *Ibid.* http://www.dpconline.org/docs/reports/dpctw05-01.pdf
[100] Lavoie, Brian and Richard Gartner. *Preservation Metadata*. Digital Preservation Coalition Technology Watch Report. September 2005. http://www.dpconline.org/docs/reports/dpctw05-01.pdf
[101] *Ibid.* http://www.dpconline.org/docs/reports/dpctw05-01.pdf
[102] *Ibid.* http://www.dpconline.org/docs/reports/dpctw05-01.pdf
[103] *Ibid.* http://www.dpconline.org/docs/reports/dpctw05-01.pdf

PREMIS were to "1) define a core set of implementable, broadly applicable preservation metadata elements, supported by a data dictionary; and 2) identify and evaluate alternative strategies for encoding, storing, managing, and exchanging preservation metadata in digital archiving systems."[104] In 2005 PREMIS released the *PREMIS Data Dictionary*; a comprehensive, practical resource for implementing preservation metadata in digital archiving systems which was updated in March 2008.

How do you keep track of all the metadata that accumulates over the life of a digital object? Without associated metadata the value of a digital file is greatly reduced. For example, NARA will not consider a file archival or high quality without its associated metadata.[105]

Metadata can be stored with or imbedded in the digital file or stored separately, such as in a relational database. There are advantages and disadvantages to both and, once again, it depends on your project for which would be more advantageous to you.[106]

To assist with keeping track of the various types of metadata, Metadata Encoding and Transmission Standard or METS was developed. "METS is an XML schema designed specifically as an overall framework within which all the metadata associated with a digital object can be stored. A METS file comprises four major constituent sections: a file inventory for all the files associated with the digital object; a section for administrative metadata; a section for descriptive metadata; and a structural map for the object. METS allows two approaches to the storage of the metadata and data associated with a digital object: both may be either stored internally within the METS file, or held externally and referenced from within METS. The content of each section is not prescribed by METS itself: any XML data or metadata may be used; however, METS does recommend a number of schemas. The flexibility of METS implies that its practical implementation can be very flexible as well: any system capable of handling XML documents can be used to create, store and deliver METS-based metadata. METS Profiles can be used to document a particular METS implementation within a project."[107]

As with every step in the digitization process, questions need to be asked and answered to determine the best way to use metadata. JISC Digital Media, formally the Technical Advisory Service for Images (TASI), suggests a few basic questions, listed below.

- What metadata do my users need or expect?

---

[104] *Ibid.* http://www.dpconline.org/docs/reports/dpctw05-01.pdf

[105] United States 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

[106] The Metadata Encoding Transmission Standard (METS) is an XML schema developed to provide a framework to store all metadata associated with a digital object, either internally or externally. Any system used for XML documents can be used to create, store, and deliver METS-based metadata. METS's home page is: http://www.loc.gov/standards/mets/

[107] Lavoie, Brian and Richard Gartner. *Preservation Metadata*. Digital Preservation Coalition Technology Watch Report. September 2005. http://www.dpconline.org/docs/reports/dpctw05-01.pdf

- What metadata do I need to manage the collection?
- What are others with similar collections doing?
- Are there any institutional, technological or legal demands I need to consider?
- What metadata do I already have?
- What expertise can I call on?
- How much time/money can I afford to spend on creating metadata?[108]

In addition, JISC Digital Media recommends creating metadata entry guidelines and developing a metadata workflow to assist in metadata capture or creation.[109]

The importance of metadata is summed up by the Northeast Document Conservation Center when they say:

> "At this point in time, extensive metadata is our best way of minimizing the risks of a digital object becoming inaccessible. Properly used metadata can:
>
> - Identify the name of the work, who created it, who reformatted it, and other descriptive information
> - Provide unique identification and links to organizations, files, or databases that have more extensive descriptive metadata about this work (this is particularly important in the likely event that the digital file and its external metadata become separated)
> - Explain the technical environment needed to view the work, including applications and versions numbers needed, decompression schemes, other files that need to be linked to it, and so forth.
>
> Various types of metadata that appear unimportant today may prove critical for properly viewing these files in the future. (For example, saved information about a particular scanner's color profile will be critical for future color management systems to account for display device differences and to properly display color on a particular device.) A good rule of thumb is to save any metadata that is cheap and easy to capture, or that someone has indicated might eventually be important."[110]

Storage

After you create your digital files, what are you going to do with them? Digital files do not take up shelf space, but they do take up storage space. Where and how do you store them? The answer will depend on the file size, number of files, and how you want to use

---

[108] JISC Digital Media. Metadata Standards and Interoperability. March 2009. http://www.jiscdigitalmedia.ac.uk/stillimages/advice/metadata-standards-and-interoperability/

[109] *Ibid.* http://www.jiscdigitalmedia.ac.uk/stillimages/advice/metadata-standards-and-interoperability/

[110] Sitts, Maxine K. *Handbook for Digital Projects: A Management Tool for Preservation and Access*. Northeast Document Conservation Center . 2000. http://www.nedcc.org/resources/digitalhandbook/dman.pdf

and access your files. As storage concerns are related to the data structure of an institution, please review the Bibliographic Center for Research's CDP Digital Imaging Best Practices guide for a basic review of data centers and network design.[111]

There are three main types of storage systems including online, nearline, and offline, each serving its own function. Frequency of access will help determine which system will work best for your project.

Online Storage

Defined by Altera Corporation, "Online storage is in constant use in the data center performing real-time data transactions for server applications. Online storage consists of disk drive-based storage that resides in or is attached (direct or fabric) to a server. Direct-attached storage allows only that server attached to the storage to access the storage. Fabric-attached storage enables all servers attached to the fabric to share the available storage resources. The storage is shared by multiple servers commonly referred to as a storage area network (SAN). SAN protocols are usually Fibre Channel, SCSI, or Ethernet in the case of a network attached storage (NAS) enclosure".[112]

Altera explains that "Online storage devices, such as just a bunch of disks (JBODs) and disk arrays, allow high-speed access to the storage, while at the same time providing data protection and security. The high-speed access to the storage is achieved with the high-speed I/O for the network, system bus, and disk drive interfaces. Data protection and security are provided with algorithms such as RAID for data protection and encryption."[113]

Just a bunch of disks (JBOD) is a collection of disks that are set up to store back-up files. If one of these disks goes bad, only the information on that particular disk is lost. Disks arranged in a particular manner are called Redundant Array of Inexpensive (or Independent) Disks (RAID) systems and are more commonly used for online storage. RAIDS often create mirrors of themselves to help mitigate loss of data if a single disk goes bad.[114] A RAID array uses a group of disks to help back up data on a daily basis. This does not replace full backups that must be done on a regular schedule. Please review BCR's CDP *Digital Imaging Best Practices Guide's Appendix G* for a more detailed look at the first five levels of RAID systems.[115] Storage.com also provides a

[111] Bibliographic Center for Research. *BCR's CDP Digital Imaging Best Practices Version 2.0.* June 2008. http://www.bcr.org/cdp/best/digital-imaging-bp.pdf
[112] ALTERA Corporation. *Storage*. http://www.altera.com/end-markets/computer-storage/storage/cmp-storage.html
[113] *Ibid.* http://www.altera.com/end-markets/computer-storage/storage/cmp-storage.html
[114] Everything 2. *Just a Bunch of Disks*. July 12, 2004. http://everything2.com/e2node/Just%2520a%2520Bunch%2520of%2520Disks
[115] Bibliographic Center for Research. *BCR's CDP Digital Imaging Best Practices Version 2.0.* June 2008. http://www.bcr.org/cdp/best/digital-imaging-bp.pdf

good overview of the RAID and RAID levels.[116] RAID is often used, but depending on the size of your organization, may not be cost effective.

Recently there has been another meaning to online storage — storing files 'online' or on the Web with a vendor or organization (also known as "cloud" storage). When outsourcing your file storage, remember to ask the appropriate questions to ensure that your records will remain trustworthy, including:

- How are files backed-up?
- What is the disaster recovery strategy?
- What are the security measures?
- What logging and audit data is available?
- What is the services downtime history?

Cloud storage may not be appropriate for confidential files or for files that need to be readily available around the clock. Two examples of this long-term preservation storage service on the Web include the Online Computer Library Center's (OCLC) Digital Archives and Amazon's Simple Storage Service (Amazon S3). Each stores files on the Web, but with different purposes. The goal of the Digital Archive is to preserve master digital files and the system accounts for metadata and preservation controls. Amazon S3's goal is to provide 24/7 access to files by using a web interface; long-term preservation is not a main concern. For a comparison of the two systems, please see the blog entry "Long-term Preservation Storage: OCLC Digital Archive versus Amazon S3" written by the Peter E. Murray, the Disruptive Library Technology Jester.[117] This type of online storage is not discussed any further in this paper; future mention of online storage refers to networks and computer access, not the Internet.

Nearline Storage

Altera defines nearline storage as follows: "Nearline storage has many of the same features, performance, and device requirements as online storage. However, nearline storage is deployed for backup support for online storage. Demand for nearline storage is growing rapidly because more information must be archived for regulatory reasons. Nearline storage is frequently used for data backup because large volumes of data must be quickly backed up, which sometimes cannot be achieved with slower bandwidth rates to tape-based solutions. Nearline storage is built using less expensive disk drives such as SATA drives to store information that must be accessed more quickly than is possible through tape or tape libraries."[118]

---

[116] Storage.com. *What is RAID?* September 9, 2008. http://searchstorage.techtarget.com/sDefinition/0,,sid5_gci214332,00.html

[117] Murray, Peter E. *Long-term Preservation Storage: OCLC Digital Archive versus Amazon S3*. Disruptive Library Technology Jester. May 16, 2008. http://dltj.org/article/oclc-digital-archive-vs-amazon-s3/

[118] ALTERA Corporation. *Storage*. http://www.altera.com/end-markets/computer-storage/storage/cmp-storage.html

Nearline storage reduces the amount of space needed in the online environment, reducing costs. Systems allow virtually seamless transition between moving your materials from an online environment to a nearline environment and offer great flexibility of what materials are stored where. This reduces the amount of space needed in the online environment, which also reduces the overall cost of an online RAID system, which might still be needed.

Offline Storage

Defined by Altera Corporation, "Offline storage is commonly referred to as "archive" or "back up" storage and is typically a tape drive or low-end disk drive (virtual tape). Offline storage is used to back up the data stored on both the online and nearline storage devices and is designed for storage of data for long periods of time. Because data is archived, offline storage appliances focus on data accuracy, protection, and security."[119]

Offline storage may be in-house or out-sourced. When held in-house, files are generally stored on removable magnetic or optical media. Magnetic formats include hard drives as well as magnetic tapes. Optical media includes CDs and DVDs, as well as the Erasable Optical (EO) disk and Write-Once Read Many (WORM) disks. For integrity purposes, read-only disks are preferred; you do not want to use Read-Write (RW) disks as the information on the disks can be changed at any time. As stated above, offline storage is generally reserved for archived materials, suggesting long-term preservation. However, file migration to new media and format conversion are practical concerns. The shelf life of removable media is finite, so you should not store the media on a disk and forget about it until the files are needed twenty years down the road, content will need to be moved to new media over time. Likewise, software and file formats change quickly, so a conversion plan is necessary. Digital media is fragile and proper handling procedures must be followed. Storage conditions are also important. Keep the materials in a clean, dry, and temperature controlled space will protect the media from unnecessary harm.[120]

Another type of offline storage includes flash drives which have become a common storage option. A USB flash drive is "a small, portable flash memory card that plugs into a computer's USB port and functions as a portable hard drive."[121] The amount of storage on a flash drive varies from 512 MB to 64 GB, providing great flexibility in a very small device.[122] The main advantage of these devices is that they contain no moving parts and

---

[119] ALTERA Corporation. *Storage*. http://www.altera.com/end-markets/computer-storage/storage/cmp-storage.html

[120] Optical Storage Technology Association (OSTA). *Technology: Q & A: Understanding CD-R & CD-RW Disk Handling, Storage and Disposal.* 2001. http://www.osta.org/technology/cdqa12.htm; Other guidelines can be found at http://www.itl.nist.gov/iad/894.05/digitalPhysicalMediaAndDevice.html.

[121] Webopedia Computer Dictionary. *USB Flash Drive*. http://www.webopedia.com/TERM/U/USB_flash_drive.html

[122] Kanguru Solutions. *Kanguru Flash Drive FB Series*. http://www.kanguru.com/flashblu.html ; CBR Storage. *SanDisk introduces backup button USB flash drive*. http://storage.cbronline.com/news/sandisk_introduces_button_backup_usb_flash_drive_120109

are fairly durable.[123] In January 2009, SanDisk, a maker of flash drives and memory cards, introduced a new line of USB flash drives that back-up files with a touch of a button.[124] This added feature saves a considerable amount of time when backing up files. The portability of a flash drive can be useful if there is a need to transport data frequently, however; because of the flash drive's relatively small size, it can easily be misplaced or stolen.

Larger than flash drives, external hard drives are hard drives that are connected to a computer with a USB or Firewire cable. These hard drives are portable, but store anywhere from forty gigabytes to two terabytes of information, drastically increasing the amount of information that can be readily transported. Uses for external hard drives include simple external storage, off-site back-up, transferring large amount of data between machines, and protecting sensitive data.[125]

If out-sourcing your offline storage needs, remember to ask the vendor all of the same questions of the vendor about security, preservation method, backup procedures, storage conditions and handling procedures.

Summary of Storage

When determining the best storage method, one must consider your institution's current storage capacity and digital file management system. Is there enough space for the project at hand, for future projects? Does your current system work for you? Are you able to produce, manage, and store back-up copies of the files or will you need outside help? Do you have a disaster recovery plan? Are all your files stored in one place or do you have backups offsite? How often are backups done? Who is in charge of them? How are they documented?

Looking at all of the available choices, their benefits and potential problems, it may be hard to determine best practices for long-term storage. You must study your digitization project, determine what is important to your institution, and understand your current and future resources before determining the best storage method for you.[126]

---

[123] Flash memory technology has begun to creep into computer hard-drives and is called Solid State Drives (SSD). There are many advantages to Solid State Drives including lower power usage, faster access, and greater reliability. A few laptop computers currently use this technology, but it is not yet main-stream, as the cost to storage ratio is high. Kyrnin, Mark. *SSD- Solid State Drives: A Hard Drive Alternative Based on Flash Memory*. About.com. http://compreviews.about.com/od/storage/a/SSD.htm

[124] CBR Storage. *SanDisk introduces backup button USB flash drive*. http://storage.cbronline.com/news/sandisk_introduces_button_backup_usb_flash_drive_120109

[125] SuperMediaStore.com. *External Hard Drive Enclosure 101*. http://www.supermediastore.com/external-hard-drive-enclosures-guide.html. In 2007, it was written that Seagate will have developed a 37.5 terabyte hard drive by 2010. As of 2009, two terabytes is the largest external hard drive available. Zaharov-Reutt, Alex. *Seagate to offer 300 TB hard drive by 2010*. ITWire. January 4, 2007. http://www.itwire.com/content/view/8350/532/

[126] To highlight the lack of standards, just in the past year, the South Carolina Department of Archives and History recommended using magnetic tape for long-term storage of electronic records while the North Carolina ECHO (Exploring Cultural Heritage Online) program specifically recommends NOT using

Preservation Strategies

Once you have decided on a file format and a storage plan, the challenge will be to keep that file accessible and viable. Digital files are not able to sit on a shelf for decades like paper files could in the proper environment. The storage medium, file type and software and hardware used to create and store the file all affects the file shelf life. Files must be preserved over time to ensure accessibility and use. The action of preserving digital files must be addressed in any digitization plan and should "involve a number of organized tasks associated with a variety of technical approaches or strategies that ensure digital resources are not only stored appropriately, but also adequately maintained and thus consistently useable over time."[127]

There are three common methods for preservation of digital files: migration/conversion, technology emulation, and technology preservation, of which the first one focuses on keeping the digital material immediately accessible and the last two focus on the technology used to create the digital file.

> **Data Migration/Conversion:** The Task Force on Archiving of Digital Information defines migration as "the periodic transfer of digital materials from one hardware/software configuration to another, or from one generation of computer technology to a subsequent generation."[128] Data migration and data conversion are often used interchangeably, although data migration may or may not involve data format changes and by definition, data conversions involve a change from one file format to another.[129]
>
> Migration is necessary to keep files up-to-date; files created in one version of software are moved (or migrated) to the most recent version of the software to keep them accessible. Migration is also essential when your chosen storage medium needs upgrading or is about to become unreliable. Storing files on a network server will greatly reduce the amount of data migration necessary over time, as servers are not in need of being constantly upgraded. However, if your files are stored on removable media, they will need to be migrated more often to remain viable. Data stored on floppy disks must be migrated to CDs before floppy

---

magnetic media for long-term storage. South Carolina Department of Archives and History. *Electronic Records Management Guidelines: Digital Media Storage - Facilities and Procedures Version 2.* March 2008. http://arm.scdah.sc.gov/NR/rdonlyres/E03AB5A2-2B90-490B-96BF-D3E838FABCF7/0/ermDMSFP.pdf and North Carolina ECHO (Exploring Cultural Heritage Online). *Digitization Guidelines: Chapter 6 – Digital Preservation*. 2007. http://www.ncecho.org/dig/guide_6preservation.shtml#6.4

[127] UKOLN at the University of Bath. *Good Practice Guide for Developers of Cultural Heritage Web Services: Digital Preservation*. April 11, 2006. http://www.ukoln.ac.uk/interop-focus/gpg/Preservation/

[128] The Commission on Preservation and Access and the Research Libraries Group. *Preserving Digital Information*. Report of the Task Force on Archiving Digital Information. May 1, 1996. http://www.ifla.org/documents/libraries/net/tfadi-fr.pdf

[129] Minnesota Historical Society. *Electronic Records Management Guidelines: Glossary*. March 2004. http://www.mnhs.org/preserve/records/electronicrecords/erglossary.html

disk drives are obsolete. Files stored on CD or DVD, for example, must be moved before the physical disk begins to degrade.

Data conversion, on the other, hand refers to format changes of the data itself. For example, if files were created in proprietary software and your institution chooses not to use that software any longer, the files will soon become unsupported and inaccessible. To ensure continued access, files should be converted to a well-supported format, minimizing the need to convert the files again in the future.

While this may sound easy, there are some drawbacks to migration. The process of migration can be time consuming, and it is unknown how often it will need to be done. Migration must occur to keep up with evolving technology. The larger the project, the more time and cost absorbed by the action of migrating the files. Another disadvantage to migration is the risk involved; some of the original data may be lost in translation. Associates at the Cornell University Library were asked by the Council on Library and Information Resources (CLIR) to assess the risk of migrating digital files. Their results include a list of eight categories of risk associated with file format migration including content fixity, security, context and integrity, cost, staffing, functionality, legal concerns, and searching capabilities.[130] Please review their results for an in-depth discussion on the risks as found in *Risk Management of Digital Information: A File Format Investigation*.[131]

Technology Emulation: Emulation preserves the technologies used to create a digital file by recreating the environment in which a file was created. Emulation "seeks to preserve that environment not through the preservation of original hardware/software but by using current technology to mimic the original environment".[132] Metadata that describes the environment in which the file was created in must be carefully recorded to ensure that emulation can occur. Emulation requires the data file, the software that created the file, the operating system used, and the "hardware environment emulated in software using detailed information about the attributes of that hardware"[133] which adds to the complexity of this method.

Some believe "that for truly long-term preservation emulation is the best solution. It accepts the necessary conundrum of preserving the original technical environment but it ensures that material is not held hostage to obsolete

---

[130] Lawrence, Gregory W. et al. *Risk Management of Digital Information: A File Format Investigation*. Council on Library and Information Resources. June 2000. http://www.clir.org/pubs/reports/pub93/contents.html

[131] *Ibid.* http://www.clir.org/pubs/reports/pub93/contents.html

[132] UKOLN at the University of Bath. *Good Practice Guide for Developers of Cultural Heritage Web Services: Digital Preservation*. http://www.ukoln.ac.uk/interop-focus/gpg/Preservation/

[133] Lawrence, Gregory W. et al. *Risk Management of Digital Information: A File Format Investigation*. Council on Library and Information Resources. June 2000. http://www.clir.org/pubs/reports/pub93/contents.html

technologies,"[134] while others recognize the possible problem of having to preserve and migrate the emulations themselves[135].

Technology Preservation: Technology preservation preserves the physical technology that created the file. This approach requires the saving and preserving of the hardware and software that was used during creation, which in itself creates a museum of physical components that are no longer in use. The physical space required for saving these materials, as well as the degradation of the machines themselves, makes this a very impractical short-term solution for digital preservation.

Establishing a plan for preserving your data is required for any digitization project. The procedures should address many of the issues included in this paper as well as periodic checks that can help to identify any data loss that may occur over time for quality control and authenticity purposes. One must also keep updated on new technologies and standards as they are developed, as they may become useful for future preservation activities.

Disposition of Originals

Depending on the purpose of your digitization project, you will need to consider what to do with the original files. Do you preserve them or destroy them? Was your purpose for digitizing to create greater access while preserving the originals or was your purpose to eliminate the paper records altogether? What are the retention requirements for the records? Can you legally dispose of them?

If you choose to keep the originals, you should evaluate the storage conditions. Follow best practices developed for storage environments, including temperature, humidity, and security. Resources in Appendix A provide information on storage best practices.

Before disposing of scanned materials, there are several things to first consider. If you are working with government records, for example, you must first determine if the records have a retention schedule. All records must be on an approved retention schedule, and you must have the authority to dispose of them before doing so. If your goal for digitizing was to make the digital files the official version of the records and the current retention schedule specifies the paper copy as the official version, you must amend the retention schedule and have it approved by the appropriate authority before the paper copy can be disposed of. Internal policies should also be updated to reflect this

[134] UKOLN at the University of Bath. *Good Practice Guide for Developers of Cultural Heritage Web Services: Digital Preservation*. http://www.ukoln.ac.uk/interop-focus/gpg/Preservation/
[135] Lawrence, Gregory W. et al. *Risk Management of Digital Information: A File Format Investigation*. Council on Library and Information Resources. June 2000. http://www.clir.org/pubs/reports/pub93/contents.html

change. If the records are not listed on a retention schedule, one must be created for them and approved before the records can be disposed of.

You must also determine the classification of the records you wish to discard. Is the content of the records public, or is there a level of privacy or security attached to them? This information should also be covered in the retention policy. If records contain non-public information, they must be disposed of properly. At a minimum, sensitive paper records must be shredded; using the cross-cut method is more secure than the strip method of shredding. For highly sensitive materials, methods of disposal include being "pulverized (rendered into a powder by grinding), macerated (rendered into a pulp by chemicals) or incinerated (burned)."[136] Secure disposal is usually required for any document that contains private information. The process of disposal should be written into a digitization plan or policy and should address record types, responsibilities (who is in charge), schedule (time frame) and location (in-house or outside vendor) when the records are to be disposed of. Risk prevention policies and a paper trail documenting the steps taken will assist you in the worst case-scenario of sensitive documents being improperly disposed of and mis-used.

If you do choose to dispose of the original files, do not be in a hurry to do so. You will want to be extremely confident that you have the legal authority to dispose of the records. In addition, you will want to make sure that you have no reason to go back to the paper records. Complete all of your cataloging, quality checks and indexing on the digital files. You may find that some of the records are unreadable, or were skipped, in which case you will need to access the paper records to correct this problem.

Providing Access

Now that you have taken the time, money, and energy to digitize your records, how do you plan on providing access to them? Who will you provide access to? What are your terms and conditions for use? How will you ensure or verify the authenticity or trustworthiness of the records? How will you secure any non-public content?

Gail Hodge and Evelyn Frangakis posed general accessibility questions in their report, *Digital Preservation and Permanent Access to Scientific Information*, many which also apply in this context. Some of the questions they asked include:

- Is the archive accessible on a routine basis?
- Through what interface/software?
- To whom? When? At what cost?
- What are the terms and conditions for use?
- Can the users download the objects to personal files?

[136] University of Miami Leonard M. Miller School of Medicine. *Secure Data Disposal Methods*. http://it.med.miami.edu/x677.xml

- If the material archived requires special software for re-use, how do you plan to maintain accessibility as the software changes?[137]

Before determining a method for providing access, it might be helpful to determine how users currently access your records. Analyze the current process, determine if it is a good one, and modify your new access procedures as necessary, especially if providing access to digital materials is new to your institution. There are many options available, including creating your own custom web interface or using a digital asset management system.

In some cases it might be necessary to restrict access to documents with private or confidential information. There are a number of ways to do this, including linking access privileges to log-in type (e.g., the public cannot see the records, but staff can), flagging confidential records and having the system filter them out of search results or browsing options, or segregating confidential records so that access is through a separate process.

How will people find the information they are looking for? Will you create an index of the records? Will only certain fields in your database be keyword searchable or will a search be performed over the entire database? Will the documents themselves be searched? If so, do you need to employ optical character recognition (OCR) on them? Again, there are a number of commercial search tools available, or you may wish to develop one specific to your needs and records. Look at what others with similar records have done to get ideas.[138]

After people find the information they are looking for from your database or online interface, how do they know that the information they are receiving is authentic, accurate, or trustworthy? A report written about information assurance issues and requirements for the National Archives and Records Administration states that in order to have an authentic and secure environment, policies must address the availability, integrity, authentication, and confidentially of data.[139] The archives staff at the Minnesota Historical Society Archives wrote a white paper that addresses such issues.[140] There are

---

[137] Hodge, Gail and Evelyn Frangakis. *Digital Preservation and Permanent Access to Scientific Information: The State of Practice*. The International Council for Scientific and Technical Information and CNDIS US Federal Information Managers Group. http://www.icsti.org/documents/preservationrpt.pdf

[138] The variety of interfaces range from simple point and click navigation to being able to search various fields within a record. In 2008, the Minnesota Office of the Revisor of Statutes Office began offering online digitized versions of the state's session laws dating back to 1849 (https://www.revisor.leg.state.mn.us/laws/). Users click through years and chapters to reach PDFs of images of individual pages. In contrast, the Minnesota Historical Society has digitized the state's birth records from 1900-1934 (http://people.mnhs.org/bci/). Users can search for full names, partial names, and even misspelled names by entering the information into the search fields. Entering information in additional fields, such as year of birth and county, narrows the search results.

[139] Nguyen, Binh Q. *Information Assurance Issues and Requirements for Distributed Electronic Records Archives*. Army Research Laboratory. April 2003. http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA413692&Location=U2&doc=GetTRDoc.pdf

[140] This website includes a white paper written on the authentication of online primary legal materials as well as a links to a few additional resources on the topic. Minnesota Historical Society. *Preserving State*

many tools available that help protect your information and computer systems including use of firewalls, intrusion detection systems, file integrity checks, and secure computing technologies such as HTTPS. Completing a risk assessment of your records and computing environment will allow you to make appropriate, cost-effective decisions.

After the necessary steps have been taken to authenticate the records and establish yourself as a trusted source, it is important to inform people about the terms and conditions of use. Conditions of use inform users of your policies, provenance of the collections, copyright holders, and permissions of use specific to the records of interest. Contact information should also be included if the user requires more assistance. Informing the user of such things shows that you are a responsible curator for the information entrusted to you. This in turn fosters further authenticity of the information and a greater trust in the repository.

Developing a Program

Often retrospective digitization projects are just that, projects. A selected amount of materials are scanned, indexed, and made accessible. Once a project is complete departments may tackle one digitization project after another without considering the overall digitization goals of the institution. Does the organization have institution-wide digitization goals? Should they? Have enough projects been completed that it is time to start thinking about the bigger picture and possibly implementing a digitization program for the whole institution? As Abby Smith says in *Developing Sustainable Digital Library Collections*, "Every library, regardless of size or mission, will need to determine for itself how and when digitization will move from being an experiment to becoming a collection-development strategy that is well integrated into its daily practice."[141]

Digitization programs take research and planning and buy-in from every level of the institution. The program must be monitored and controlled by policy and procedures. If the institution has already completed some digitization projects, it is assumed that they were documented. Use this documentation to evaluate past projects and to create institution-wide digitization policies. Save past documentation for future use; this will allow staff to go back and see what previously did or did not work. Documentation creates a foundation for a digitization program and makes it easier to collaborate with others in the future.

Digitization must be integrated into the daily workflow of the organization to ensure continued success. Studies show successful digitization projects and programs are the ones that digitize material that are "integrated into the fabric of library services; focused primarily on achieving mission related objectives; funded from predictable streams of

---

Government Digital Information: White Paper and Resources on Authentication. July 2008. http://www.mnhs.org/preserve/records/legislativerecords/authentic.htm

[141] Smith, Abby. *Developing Sustainable Digital Library Collections: Strategies for Digitization*. Digital Library Federation. September 2001. http://www.clir.org/pubs/reports/pub101/pub101.pdf

allocation, be they external or internal; and include a plan for the long-term maintenance of digital assets."[142]

Digitization projects are often funded on a per-project basis. As the investment in the digitization process grows the time, money and energy allocated to the program must also increase. Sustainability becomes a greater factor as digitization projects turn into a digitization program; resources must be sustained over time if the digitization program is to succeed.

To assist with the allocation of resources, there has been a growing trend to collaborate with similar institutions locally, regionally, and nationally. This collaboration has facilitated access to increasing numbers of digitized materials and limits the number of institutions that must create their own repository or access system. Using the strengths of others increases the strength of the digital collection.

As reiterated throughout this paper, planning your digitization project is imperative for success. Much thought should be put into developing a digitization program. The topics discussed in this paper will need to be considered on a program wide level, rather than on the project level. Remember that "decisions [made] when creating a digital information resource (e.g. about its contents, format, data model, level of description, etc.) impact directly upon how, at what cost, and by whom that resource can be used, integrated into collections, maintained, and supported."[143] As the field of digitization continues to grow and change, attending workshops, conferences and continuing education opportunities will assist you in the development of a digitization project or larger program.

[142] *Ibid.* http://www.clir.org/pubs/reports/pub101/pub101.pdf

[143] Greenstein, D. *Strategies for Developing Sustainable and Scalable Digital Library Collections*. Digital Library Federation. May 1, 2000. http://www.diglib.org/collections/collstrat.htm

Appendix A: Selected Resources

Below you will find a list of resources for many concerns related to digitization projects. This appendix includes resources on the following topics:

General

ARMA International. *Home Page.* http://www.arma.org/

> "ARMA International is the oldest and largest association for the records and information management profession." This website provides users with access to publications and information about electronic records management, standards and best practices, professional development, and upcoming conferences and seminars. Local chapters of ARMA have been developed in many metropolitan cities, of which a list can be found on this ARMA International site.

Bibliographic Center for Research. Bibliographic Center for Research's (BCR) Collaborative Digitization Program (CDP) *Digital Toolbox.* http://www.bcr.org/cdp/digitaltb/index.html

> The Digital Toolbox includes information on Digitization Training, Best Practices, Getting Started, Digital Imaging, Metadata, Digital Audio, EAD, Rights Management, Digital Preservation, and the Dublin Core Builder. Each section provides the user with valuable information when starting a digitization program. The *Digital Imaging Best Practices Guide* was updated in June 2008 and addresses issues to think about before starting a project, provides general information on upfront and ongoing costs, standard file formats, computer hardware and software requirements, the importance of quality control, metadata standards, and storage options. Appendixes provide more in-depth detail as well as resources for more information. Also includes basic discussion of descriptive, administrative, structural, and technical metadata. More attention is paid to preservation metadata including information on PREMIS (PREservation Metadata Implementation Strategies) and METS (Metadata Encoding and Transmission Standard).

California Digital Library (CDL). *CDL Guidelines for Digital Images (CDL GDI).* Version 2.0. September 2008. www.cdlib.org/inside/diglib/guidelines/bpgimages/cdl_gdi_v2.pdf

> Reviewed and updated annually, this resource provides information specifically on how to create digital images with the intent to deposit them in the California Digital Library (CDL), however the importance of the topics covered is useful for anyone contemplating a digital imaging project. The guide includes information on file formats, compression, watermarking, and definitions of terms to understand. Additional guidelines are provided for both digital masters and derivative files taken from various media types (textual documents, maps, graphics, photographs, slides or negatives, and 3D objects).

Cornell University. *Digital Preservation Management: Implementing Short-term Strategies for Long-term Problems*. http://www.icpsr.umich.edu/dpm/dpm-eng/eng_index.html

An online tutorial that addresses many preservation issues and preservation strategies in a step-by-step guide. Topics include terms and concepts, obsolescence and physical threats, digital repositories, metadata, legal issues, balancing stakeholders, access issues, financing the future, and organizational and technological infrastructure.

Cornell University, Department of Conservation and Preservation. *Moving Theory into Practice: Digital Imaging Tutorial.* www.library.cornell.edu/preservation/tutorial

Created in 2000, this tutorial provides hands on learning about the digitization process. Ten sections related to digital imaging walk the user through the basic concepts of digital imaging with self-checks along the way. Sections include terminology, selection, conversion, quality control, metadata, technical infrastructure, presentation, preservation, management and further resources.

Council of State Archivists (COSA). *Archives Resource Center (ARC).* http://www.statearchivists.org/arc/index.htm

Website of program resources developed by and for state archives and record centers. Topics include directories, legal statutes, digitization, training and education, records retention policies, disaster planning, preservation, collaborations and information on various records types. Digital Imaging guidelines and standards: www.statearchivists.org/arc/states/res_imag.htm . Collaboration and partnerships: www.statearchivists.org/issues/ocp/index.htm

Federal Agencies. *Federal Agencies Digitization Guidelines Initiative*. http://www.digitizationguidelines.gov/

A work group of federal agencies created to develop digitization guidelines and standards. Developing guidelines to support the long-term preservation efforts of the Library of Congress. Resource includes a glossary of terms to help standardize word usage.

JISC Digital Media (formally TASI) http://www.jiscdigitalmedia.ac.uk/

Home page for the UK's JISC Digital Media. Provides links to advice, training, and news about still images, moving images, and sound. Much of the information

provided is specific to digital media. Topics include but are not limited to deciding how to digitize, managing the workflow, legal issues, project management, risk assessment, choosing a file format, compression issues, developing a digital preservation strategy, file naming, and metadata overview and schemas.

Joint Information Systems Committee (JISC). *Digital Images Archiving Study*. March 2006.
www.jisc.ac.uk/whatwedo/programmes/programme_preservation/imagesarchiving.aspx

The results of a six long month study address the concerns of long term preservation of digital materials. User requirements, digital image properties, preservation methods, metadata requirements, organizational models, and projected preservation costs are evaluated.

Minnesota Historical Society. *Electronic Records Management Guidelines*. March 2004. http://www.mnhs.org/preserve/records/electronicrecords/erguidelines.html

This guide discusses many topics of concern in relation to electronic records management. Topics include legal issues, management strategy, long-term preservation, metadata, file naming, file formats, storage facilities and procedures, digital media, electronic document management systems, digital imaging, e-mail management, web-content management, electronic and digital signatures, and a glossary. The entire guide or specific sections of interest can be viewed online or downloaded for your convenience.

Minnesota Historical Society. *Trustworthy Information Systems Handbook*. July 2002.
http://www.mnhs.org/preserve/records/tis/tis.html

Trustworthy means that an information system, whether computer- or paper-based, is accountable and can produce reliable and authentic information and records. Government records and record keeping systems must be accountable to citizens and elected officials. The Minnesota State Archives developed a set of guidelines and criteria to establish the trustworthiness of information systems.

North Carolina Office of Archives and History. *Government Records Branch of North Carolina*. October 30, 2008. http://www.records.ncdcr.gov/default.htm

Home page that links user to guidelines for public records, state agency records, county and municipal records, and university records. These guidelines include legal information, retention schedules, best practices for digital preservation, digital imaging standards, and project planning. Guidelines are easily understood

and explained further with examples. Some specific topics include the importance of dates on files for following retention schedules, the idea that files will move from their original location, multiple versions, and that there will inevitably be exceptions in every case.

North Carolina ECHO (Exploring Cultural Heritage Online). *Digitization Guidelines*. 2007.
http://www.ncecho.org/dig/digguidelines.shtml

This guide covers collaboration, standards and best practices, project planning, selection, legal considerations, digital production, metadata, digital preservation, presenting your digital project, targeting the K-12 audience, project evaluation, project management, resources and glossary. The guide is available in HTML as well as in PDF.

Northeast Document Conservation Center. *Preservation Leaflets.*
http://www.nedcc.org/resources/leaflets.list.php

The Northeast Document Conservation Center is a good source of preservation information and materials on both digital and paper files and photographs. Preservation Leaflets have been written by 'experts' in the field about how to plan and prioritize digitization projects, the physical environment and how to protect resources, emergency management and disaster planning, storage and handling procedures, reformatting, and conservation. A variety of aspects of a digitization project are reflected in the various leaflets.

Porter-Roth, Bud. "Digital Conversion Projects: A Decision-Making Checklist." *The Information Management Journal*. March/April 2007: 58-61, 63. www.arma.org/imj

An article that asks the questions that must be answered before starting a digitization project. Questions address the issues about the types of files that could be digitized, about the conversion process, whether done in-house or out-sourced, as well as other important concepts.

Preserving Access to Digital Information (PADI). http://www.nla.gov.au/padi/

Administered by the National Library of Australia, PADI provides a topical gateway to international preservation resources. Major topics include general resources on digital information; issues of authenticity, archiving, selection, storage and technical obsolescence; digital preservation strategies including emulation, encapsulation, migration and universal virtual computer; data documentation and standards for metadata, file formats, and persistent identifiers;

intellectual property rights management; information on various formats and media types; management issues; information on various national approaches; digital libraries; other digitization topics; and digital preservation tools.

Sitts, Maxine K. *Handbook for Digital Projects: A Management Tool for Preservation and Access*. Andover, Massachusetts: Northeast Document Conservation Center, 2000. http://www.nedcc.org/resources/digitalhandbook/dman.pdf

A guide that discusses the complete digitization process, from the moment it is considered to the long term preservation of the records. Discusses the pros and cons for digitizing materials, selection of materials, project management, copyright issues, technical specifications, vendor relations and long term access.

South Carolina Archives and Records Management. *Electronic Records Management Guidelines*. February 2005, with updates done in 2008 on certain sections. http://arm.scdah.sc.gov/erp/ermanagement+guidelines.htm

The home page for the Electronic Records Management Guidelines links the user to short documents covering electronic records management issues. File naming, file formats, digital media storage, document management systems, digital imaging, e-mail management, web content management, electronic signatures, and the Trustworthy Information System Handbook for South Carolina are all discussed in more detail.

TAB Products. *Digital Imaging Needs Assessment Tool*. June 2008. www.tab.com/ecms.aspx/RM-HOMEus-DigitalImagingNeedsAssessment-june2008.htm

A worksheet that will help you understand your current records management system, assess if digital imaging is right for you, and help explain what might be involved in a digital imaging project. Questions focus around the topics of your imaging objectives, your current filing environment, collaboration and access, compliance, and your records management strategy.

UPDIG Working Group. *The Universal Photographic Digital Imaging Guidelines*. September 2008, Version 4. http://www.updig.org/index.html

Guidelines created to assist with accurate reproduction and management of digital files, including digital asset management, color profiling, metadata, and photography workflow. These guidelines are for photographers, designers, printers and image distributors alike.

Utah State Archives. *Guide to Digital Imaging*. June 2005.
http://www.archives.state.ut.us/recordsmanagement/ERM/guide-to-digital-imaging.html

A guide that highlights the possible advantages and disadvantages of a digitization project. Lists other factors to take into account before making any decisions. A cost-benefit analysis of time is provided. Long-term storage, retention and disposition of both paper and digital records are also taken into consideration.

Legal Requirements

Federal Trade Commission and Department of Congress. *Electronic Signatures in Global and National Commerce Act*. June 2001.
http://www.ftc.gov/os/2001/06/esign7.htm

Report on the Electronic Signatures (ESIGN) act. Includes background of e-commerce, summary of public comments, and an analysis of the benefits and burdens of the Act.

Library of Congress. *THOMAS*.
http://thomas.loc.gov/

An online resource for federal legislative information. Provides various search options to find bills, resolutions, acts, treaties and reports. Search by topic, creator, bill text, bill number. Track legislative sessions. Links to other government resources, such as the Federal House and Senate, are also linked from here.

Minnesota Legislative Reference Library. *Links to the World – Government Data Practices*. October 27, 2008.
http://www.leg.state.mn.us/lrl/links/data.asp

An online resource that includes links to government information for the state of Minnesota and the Federal government. Links to sites about privacy issues are also included.

Minnesota Office of the Revisor of Statutes. *Revisor of Statutes Home Page*. 2009.
https://www.revisor.leg.state.mn.us/index.php

The Minnesota Office of the Revisor of Statutes is responsible for compiling, editing and printing Minnesota Laws, Statutes and Rules for each legislative

session. The home page provides access to the Minnesota Laws, Statutes, and Rules through a search tool. Other links allow the user to search for Bills and follow their status through time. Information is also provided about the different branches of government and government resources.

National Conference of State Legislatures. *Uniform Electronic Transactions Act.* 2008. http://www.ncsl.org/programs/lis/CIP/ueta-statutes.htm

Website discussing the Uniform Electronic Transactions Act (UETA). A graphic of North America highlights states that have already enacted UETA. Links to each state law are included.

Office of Management and Budget, United States Government. *Memoranda: Guidance on Implementing the Electronic Signatures in Global and National Commerce Act.* September 25, 2000.
http://www.whitehouse.gov/omb/memoranda/esign-guidance.pdf

Explains the areas that the Electronic Signatures Act (ESIGN) covers, relates it to government activities, and discusses retention schedules, filing requirements, and effective dates.

Uniform Law Commission. *The National Conference of Commissioners on Uniform State Laws*. http://www.nccusl.org/Update/

The National Conference of Commissioners on Uniform State Laws works to create uniformity of state laws. Their site allows users to search or browse acts by title or state. With the evolution of e-commerce and the increasing need to do business across state lines, having uniform laws ensures compatibility.

USA.gov. *Laws and Regulations – General Reference Resources.* September 5, 2008. http://www.usa.gov/Topics/Reference_Shelf/Laws.shtml

Provides links to web resources related to federal, state, and local laws and regulations.

Cost Justification

Apple, Bryan. *Documents – Physical, Digital, or Both? A Framework for Evaluation*. November 11, 2008.
http://www.twincities.arma.barr.com/Programs/Presentations/20081111BryanAppleARMATC.pdf

> PowerPoint presentation given by the Director of Digital Document Management Solutions stresses the importance of knowing the overall business processes of an organization before making the decision to digitize files. Pluses and minuses of paper and digital files are suggested. Offsite storage and destruction policies are also considered important factors in the decision to digitize.

Avedon, Don M. "Electronic Imaging 101 part III – Application analysis, document preparation and digitizing". *ARMA Records Management Quarterly*. October 2004.
http://findarticles.com/p/articles/mi_qa3691/is_/ai_n8710569

> The content of this older report remains important as it breaks down cost analysis into hard costs and soft costs. Questions are asked about the system requirements as they pertain to customer service, competitive advantages, reducing processing time, and document quality. Other aspects of a digitization project are also touched upon.

North Carolina ECHO (Exploring Cultural Heritage Online). *Digitization Guidelines*. 2007.
http://www.ncecho.org/dig/digguidelines.shtml

> The chapter on Project Planning in the Digitization Guidelines created by North Carolina ECHO covers goals, audience, needs analysis, standard processes, documentation, and evaluation. A table separates the costs into the categories of hardware, software, staff wages, training costs, presentation and preservation costs, and material costs making estimated costs easy to view. Available in HTML and as a PDF file.

Puglia, Steven. "The Costs of Digital Imaging Projects". *RLG DigiNews*. October 15, 1999. Volume 3, Number 5.
http://worldcat.org/arcviewer/1/OCC/2007/08/08/0000070511/viewer/file422.html

> Discusses the many components that should be considered when developing a digitization plan. Many tables analyze the costs of per-item production including specifics on digitizing, metadata creation, and number of images completed per day. Each category is analyzed in more detail. Projected maintenance costs are also covered.

Roderick, Elizabeth. *More Than Just Pretty Pictures: A Cost/Benefit Analysis of Digital Library Holdings*. December 9, 1998.
http://net.educause.edu/ir/library/html/cnc9804/cnc9804.html

> A discussion of the Library of Virginia's Digital Library Project highlighting the cost benefits for digitizing parts of the collection. Project costs and ongoing costs are defined, as well as a comparison of the time and money needed to assist patrons online, via the mail, and in person. Time and money estimates are made for both the library and patron.

In-House vs. Outsourcing

ARMA International. *Buyer's Guide.*
http://www.arma.org/buyersguide/index.cfm

> Search vendors and companies that provide services by name, market type, product, or alphabetically. Government Market is an option.

ARMA Twin Cities Chapter. *Vendor Directory*. Updated November 2008.
http://www.twincities.arma.barr.com/Vendor%20Directory/venddir.html

> Includes links to vendor that address: commercial records/off-site storage, content/document management software, document destruction, imaging sales and services, records/retention management software, consulting services, disaster recovery, filing/storage sales and service, and micrographs.

Dale, Robin L. *Outsourcing and Vendor Relations*. Northeast Document Conservation Center (NEDCC). 2007.
http://www.nedcc.org/resources/leaflets/6Reformatting/07OutsourcingAndVendorRelations.php

> Leaflet is designed to address the pros and cons of both in-house and out-sourcing digitization options. Highlights that it is not usually a black and white decision, portions of the project can be done in-house and others outsourced. Provides information on setting up an in-house digitizing area. Request for Information, Request for Proposals, evaluating proposals, contract negotiation tips, and communicating with vendors is also highlighted.

Montana State Library. *Digitization Planning Power Point.* Montana State Library Training Materials Fall 2008.
http://msltraining.pbwiki.com/Fall%20Workshop%202008

> This Power Point presentation discusses the importance of planning a project, and knowing the pros and cons for each decision you make. Asks a lot of questions to get you thinking about your project purpose. Notes include more details about the pros and cons of both in-house and vendor digitization programs.

File Formats

Graves, Ralph. *File Formats Glossary*. Cruchfield. April 19, 2008.
http://www.crutchfield.com/S-zSaxMtp6oIW/learn/learningcenter/home/fileformats_glossary.html

> Definitions of file types as well as related terms that provide background information and understanding to the overall concepts surrounding file formats.

InterPARES 2 Project. *General Study 11 Final Report: Selecting Digital File Formats for Long-Term Preservation*. March 2007.
www.interpares.org/display_file.cfm?doc=ip2_file_formats(complete).pdf

> A study that discusses major concepts regarding digital file types, including the importance of standards and selection criteria. Stability, open file formats, standard formats, standard terms, selection criteria, and policy considerations are all covered as recommendations for developing and implementing policies related to long-term digital preservation discussed.

JISC Digital Media. *Choosing a File Format for Digital Still Images: File Formats for Master Archives.* May 2006.
http://www.jiscdigitalmedia.ac.uk/stillimages/advice/choosing-a-file-format-for-digital-still-images/

> Includes information on formats for capture, master archives, optimization, manipulation, and various forms delivery including printing, web delivery, and PowerPoint applications.

JISC Digital Media. *File Formats and Compression.* March 2009.
http://www.jiscdigitalmedia.ac.uk/stillimages/advice/file-formats-and-compression/

Explains file compression. Defines the different types of compression, lossy and lossless. Explains image format types, which effect compression.

Library of Congress. *Sustainability of Digital Formats Planning for Library of Congress Collections: Format Descriptions*.
http://www.digitalpreservation.gov/formats/fdd/descriptions.shtml

Includes detailed descriptions of various file types for still images, sound, textual images, and others. Encoding information included. Sustainability factors are also covered.

Minnesota Historical Society. *Electronic Records Management Guidelines – File Formats*. March 2004.
http://www.mnhs.org/preserve/records/electronicrecords/erfformats.html

Legal concepts, proprietary and non-proprietary, file format types, preservation, compression, and other key issues to consider as explained in this guide. Includes information for non-text based documents. Emphasizes the importance of planning the digitization project.

Office of Enterprise Technology State of Minnesota. *Enterprise Technical Architecture 2.02*. September 8. 2006
http://www.state.mn.us/portal/mn/jsp/content.do?agency=OETweb&id=-536891917&subchannel=-536891918&programid=536911145

Guidelines for the state of Minnesota to follow when managing records. Includes information on network architecture, storage, data exchange, application architecture, accessibility, workflow, security, and reliability. Home page also includes information on IT standards and securities, and IRM policies, standards and guidelines.

University of Michigan, Bentley Historical Library, University Archives. *PDF/A (ISO 19005-1) FAQ*. http://bentley.umich.edu/uarphome/bestprac/pdfafaqs.php

Provides general information about the PDF/A format, its uses, and instructions for creation. Answers questions about the format that help explain its features and benefits.

Wotsit.org. *The Programmer's File and Data Resource*.
http://www.wotsit.org/default.asp

Collection of information on file types and data formats. Includes specifications and technical information useful to programmers.

Resolution

Bibliographic Center for Research. *BCR's CDP [Collaborative Digitization Program] Digital Imaging Best Practices Version 2.0* June 2008.
http://www.bcr.org/cdp/best/digital-imaging-bp.pdf

Includes basic information on the overall concepts needed to understand resolution and its associated ideas including modes of capture, bit depth, color space, resolution, tonal dynamic range, compression, and file formats.

J. Paul Getty Trust. *Introduction to Imaging – Revised Edition: The Image*.
http://www.getty.edu/research/conducting_research/standards/introimages/image.html

Explanation of color management, bit-depth, dynamic range in relation to resolution. Discusses differences in on-screen color and print color. Definitions of screen resolution, monitor resolution, printer resolution, and capture resolution.

Library of Congress. *Technical Standards for Digital Conversion of Text and Graphic Materials*. December 2006. http://lcweb2.loc.gov/ammem/about/techStandards.pdf

Current Library of Congress standards for text and graphic images. Includes information on how to test processes and which ISO standards should be met when working with resolution, color, dynamic range, tonality, optical character recognition, and image capture. Chart summarizes document types with standards for various formats. Metadata and file management are also discussed.

National Archives Records Association (NARA). *Technical Guidelines for Digitizing Archival Masters for Electronic Access: Creation of Production Master Files – Raster Images*. June 2004. http://www.archives.gov/preservation/technical/guidelines.html

Technical specifications relating to resolution discussed including how to test for accuracy and quality control. Spatial resolution, bit depth, and color modes are explained. Color calibration techniques are discussed in detail.

United States Government Printing Office. *Report on the Meeting of Experts on Digital Preservation*. March 12, 2004. http://www.gpoaccess.gov/about/reports/preservation.pdf

Reviews specifications for resolution and metadata standards for seventeen institutions. Covers plain text, illustrated text, color documents, bit depth, compression, and master copy format.

File Naming

Alberta Government. *Naming Conventions for Electronic Records*. August 2005. http://www.im.gov.ab.ca/publications/pdf/DocumentNamingConventions.pdf

This resource provides a complete package of information relating to naming conventions. The introduction discusses the importance of a standard file naming convention. Covers file storage in general – personal vs. shared servers. Reviews elements of importance and provides examples of how each would be using in a file name. Engages the reader into further thought about what decisions will need to be made. Appendix provides four examples of different files names with explanation of what situation they would be good for.

Bibliographic Center for Research (BCR). *BCR's CDP [Collaborative Digitization Program] Digital Imaging Best Practices Version 2.0* June 2008. http://www.bcr.org/cdp/best/digital-imaging-bp.pdf

Lists eight important ideas to consider when developing a file naming convention. Includes more specific ways to track where the file originated from. Highlights the importance of unique file names following an established convention and being able to think long-term.

Brookhaven National Laboratory (BNL) Web Communication Standards. *File Naming Conventions and Directory Structure*. February 5, 2008. http://www.bnl.gov/webstandards/fileNaming.asp

Discusses file naming conventions for online directories.

CabinetOffice, e-Government Unit. (London). *Guidelines for UK Government Websites Illustrated Handbook for Web Management Teams: 3.1 File Storage and Servers*. http://archive.cabinetoffice.gov.uk/e-government/docs/resources/web_guideline_handbook/pdf/wgl_3-1.pdf

Resource for file organization and storage related to government websites in the UK. Provides a checklist of basic needs and explains the issues of case sensitivity, file name length, and spaces on various platforms.

Digital Projects Advisory Group, University Libraries at the University of Colorado at Boulder. *File Naming Conventions for Digital Collections*. March 4, 2008. http://ucblibraries.colorado.edu/systems/digitalinitiatives/docs/filenameguidelines.pdf

Provides background information on file naming. Explains importance of knowing what the project entails before naming files. Lists seven guidelines to follow when creating a file naming convention. Highlights importance of documenting file convention decisions and the decision making process. Includes example of needed documentation.

JISC Digital Media. *Choosing a File Name*. November 2008. http://www.jiscdigitalmedia.ac.uk/crossmedia/advice/choosing-a-file-name/

Recommends that file names only be eight characters long based on the ISO 9660 standard to ensure compatibility across platforms. Discusses the various ways of using folder structure and files names to name files with reasons to use descriptive or non-descriptive file names.

Minnesota Historical Society. *Electronic Records Management Guide - File Naming*. March 2004. http://www.mnhs.org/preserve/records/electronicrecords/erfnaming.html

Provides general concept knowledge needed to understand file naming conventions. Lists common file name elements that can be used, discusses challenges and issues related to creating a long term file naming convention. Asks questions that need to be answered before implementing a file naming system. Internet file names, domain names, and URL protocols are also discussed.

North Carolina Department of Cultural Resources. *Best Practices for File-Naming*. May 7, 2008. http://www.records.ncdcr.gov/erecords/filenaming_20080508_final.pdf

Guidelines for file naming are easily understood and explained further with examples. Addresses the importance of dates on files for following retention

schedules, the idea that files will move from their original location, multiple versions, and that there will inevitably be exceptions.

University of Edinburgh, Records Management Section. *Standard Naming Conventions for Electronic Records*.
http://www.recordsmanagement.ed.ac.uk/InfoStaff/RMstaff/RMprojects/PP/FileNameRules/FileNameRules.htm

Provides a list of thirteen guidelines to follow when creating a file naming convention. Rules have some specifics working with correspondence and emails. Each rule has its own page displaying the rule, an explanation, and a correct and incorrect example.

York University, Information and Privacy Office. *Information and Privacy Toolkit - Tip Sheet 6 – Naming Conventions for Electronic Files and Folders*. May 8, 2007
http://www.yorku.ca/secretariat/infoprivacy/infotoolkit/docs/TipSheet6NamingConventionsEFilesFolders.pdf

Provides an easy to read list of guidelines including examples and explanations for each. A scaled-down version of *Standard Naming Conventions for Electronic Records* produced by the University of Edinburgh Records Management Section is also offered here.

Metadata

American Library Association. *Update on Major Metadata Standards*. Library Technology Reports. November/December 2005. www.techsource.ala.org or
http://www.accessmylibrary.com/coms2/summary_0286-12329830_ITM

Resource explains numerous metadata schemas and provides information on where to go for more information. Schema discussed include: Metadata Authority Description Schema (MADS), Metadata Object Description Schema (MODS), Resource Description and Access (RDA), Metadata Encoding and Transmission Standard (METS), Dublin Core (DC) and many others.

Bibliographic Center for Research (BCR). *BCR's CDP [Collaborative Digitization Program] Digital Imaging Best Practices Version 2.0* June 2008.
http://www.bcr.org/cdp/best/digital-imaging-bp.pdf

Basic discussion of descriptive, administrative, structural, and technical metadata. More attention is paid to preservation metadata including information on PREMIS

(PREservation Metadata Implementation Strategies) and METS (Metadata Encoding and Transmission Standard).

Collaborative Digitization Program (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

The Collaborative Digitization Program (CDP) has adopted the Dublin Core metadata standard as a baseline standard for metadata to help ensure compatibility across various collections. The report defines Dublin Core itself and explains why it was chosen as the standard. Each Dublin Core metadata element is then defined, including specific rules to that element. Input guidelines and examples help the reader understand that function of each element in more detail.

Dublin Core Metadata Initiative. http://dublincore.org/

Provides an overview and history of the Dublin Core metadata set. Links to the metadata elements with a detailed description of how they are used. Includes general information on encoding guidelines and schemas. The main page also has a link for additional resources including papers, presentations, training, tools and projects.

Federal Agencies Digitization Guidelines Initiative. *Digitization Standards: A Resource List for Standards Related to Digital Imaging of Print, Graphic, and Pictorial Materials.* July 21, 2008. http://www.digitizationguidelines.gov/stillimages/digstandards.html

A resource list for digitization standards. Many resources are very technical in nature, and include information on the ISO standard.

JISC Digital Media. *Metadata Overview*. November 2006. http://www.tasi.ac.uk/advice/delivering/metadata.html

The metadata overview is very general but provides solid information. Also linked from this page are other metadata resources that go into more details including: Challenges of Describing Images, Getting Practical with Metadata, and Metadata Standards and Interoperability. A good resource for both general and more descriptive information.

JISC Digital Media. *Putting Things in Order: Links to Metadata Schemas and Related Standards*. http://www.jiscdigitalmedia.ac.uk/stillimages/advice/putting-things-in-order-links-to-metadata-schemas-and-related-standards/

Information about twenty-seven metadata schemas and related standards.

Kansas State Historical Society. *Digital Imaging Guidelines for State Agencies*. Kansas State Historical Society.
http://www.kshs.org/government/records/electronic/digitalimagingguidelines.htm

Guidelines for government records that adhere to the records preservation laws of Kansas. Seventeen recommendations to follow as best practice. Based on *Guidelines for the Use of Digital Imagining Technologies for Long-Term Government Records in Alabama*.

Kwon, Hyuckbin, Theresa A. Pardo and G. Brian Burke. *Building a State Government Digital Preservation Community: Lessons on Inter-organizational Collaboration*. Center for Technology in Government. March 2008.
http://www.ctg.albany.edu/publications/working/building_community/building_community.pdf

Discusses collaboration of projects and resources. Written specifically with government issues in mind. Partnerships should be within states, between states, between states and the private sector, and between states and the federal government. The Library of Congress and National Digital Information Infrastructure and Preservation Program (NDIIPP) is assisting with this collaboration.

Lavoie, Brian and Richard Gartner. *Preservation Metadata*. Digital Preservation Coalition Technology Watch Report. September 2005.
http://www.dpconline.org/docs/reports/dpctw05-01.pdf

Discusses the history and importance of preservation metadata. Elements of preservation metadata are defined and explained as well as the development of a preservation metadata schema. Discusses how the Open Archival Information System (OAIS), PREservation Metadata: Implementation Strategies) PREMIS, and the Metadata Encoding and Transmission Standard (METS) relate to each other and the development of preservation metadata.

Minnesota Historical Society. *Minnesota Recordkeeping Metadata Standard (IRM 20).* May 29, 2002. http://www.mnhs.org/preserve/records/metadatastandard.html

Explains the context for the Minnesota Recordkeeping Metadata Standard and its legal requirements. Both a summary and detailed list of the metadata elements are included.

Online Computer Library Center (OCLC). *PREMIS (PREservation Metadata: Implementation Strategies) Working Group*. 2008.
http://www.oclc.org/research/projects/pmwg/
http://www.loc.gov/standards/premis/v2/premis-2-0.pdf

Home page for the PREMIS Working Group. "The objectives of PREMIS were to develop a core preservation metadata set, supported by a data dictionary, with broad applicability across the digital preservation community and identify and evaluate alternative strategies for encoding, storing, and managing preservation metadata in digital preservation systems." This site links users to the Data Dictionary created by PREMIS. The Data Dictionary also includes background information on PREMIS itself, definitions of terms, descriptions of entities, and tables for each element with the associated rules. The website also keeps users updated with the latest activities of PREMIS.

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

Very detailed description of metadata types and technical aspects used by the National Archives and Records Administration (NARA). This detailed guide that includes descriptions of eleven types of metadata, the digitizing environment, necessary equipment, workflow and processing procedures, specification for various record types, file formats, naming conventions, storage recommendations and quality control issues. Provides guidance for accessing project metadata needs.

PREMIS Editorial Committee. *PREMIS Data Dictionary for Preservation Metadata version 2.0*. March 2008. http://www.loc.gov/standards/premis/v2/premis-2-0.pdf

A link to the Data Dictionary that includes background information on PREMIS, definitions of terms, descriptions of entities, and tables for each element with the associated rules.

Rieger, Oya Y. *Preservation in the Age of Large-Scale Digitization*. Council on Library and Information Resources (CLIR). http://www.clir.org/pubs/abstract/pub141abst.html

A white paper addressing recent large-scale digitization projects. Includes framework for addressing preservation issues and process recommendations.

Spindler, Robert P. *Digital Preservation*. Northeast Document Conservation Center. 2007. http://www.nedcc.org/resources/leaflets/6Reformatting/05DigitalPreservation.php

Written in 2007, this leaflet discusses challenges associated with digital records and highlights the use of metadata to help preserve documents. Some challenges are physical degradation of media, obsolescence of media, and migration failure. Discusses the human factor of digital preservation.

United States Government Printing Office (GPO). *Report on the Meeting of Experts of Digital Preservation: Metadata Specifications*. June 14, 2004. Revised June 2, 2005. http://www.gpoaccess.gov/about/reports/metadata_report_final.pdf

Discussion of what people want the GOP to do for preservation metadata. Highlights partnering with libraries to develop a no-fee public access database for government records. Stresses the importance of standards if multiple repositories will be creating metadata but understands that documents will come in various formats. Includes chart of what metadata schemas institutions are currently using for different purposes.

Storage

Bibliographic Center for Research (BCR). *BCR's CDP [Collaborative Digitization Program] Digital Imaging Best Practices Version 2.0* June 2008. http://www.bcr.org/cdp/best/digital-imaging-bp.pdf

This guide includes sections on storage systems, network concerns, and storage conditions. There is more detailed information on data center basics, network design for data protection, RAID arrays, and online storage.

North Carolina ECHO (Exploring Cultural Heritage Online). *Digitization Guidelines: Chapter 6 – Digital Preservation*. 2007.
http://www.ncecho.org/dig/guide_6preservation.shtml#6.4

Discusses challenges of digital preservation including storage media and storage concerns. Addresses preservation strategies. Guidelines for improving the lifespan of media formats.

Conway, Paul. *The Relevance of Preservation in a Digital World*. Northeast Document Conservation Center Preservation Leaflet.
http://www.nedcc.org/resources/leaflets/6Reformatting/04RelevanceOfPreservation.php

Overall discussion on digital preservation that defines preservation based on the purpose of the digitizing project which affects the storage goal for the project.

The PC Guide. *Redundant Arrays of Inexpensive Disks (RAID).*
http://www.pcguide.com/ref/hdd/perf/raid/index.htm

A good explanation of everything RAID including cost, benefit, limitations, concepts and issues, level definitions including a comparison chart, and configuration and implementation. The site itself, however, is ad heavy.

South Carolina Department of Archives and History. *Electronic Records Management Guidelines: Digital Imaging Version 2*. March 2008.
http://arm.scdah.sc.gov/NR/rdonlyres/33990F8B-D9E2-4303-A490-6CA6F6C14062/0/ermDI.pdf

General information on digital media. Defines magnetic and optical solutions.

South Carolina Department of Archives and History. *Electronic Records Management Guidelines: Digital Media Storage - Facilities and Procedures Version 2*. March 2008.
http://arm.scdah.sc.gov/NR/rdonlyres/E03AB5A2-2B90-490B-96BF-D3E838FABCF7/0/ermDMSFP.pdf

Planning and maintenance of storing digital media. Details storage facility requirements, components and conditions for digital media. Lists proper handling techniques. Includes legal requirements for South Carolina.

Storage.com. *What is RAID? Definitions*. September 9, 2008. http://searchstorage.techtarget.com/sDefinition/0,,sid5_gci214332,00.html

Explains how RAID works and defines the ten RAID levels. Links to other resources for more information.

Tom's Hardware Guide. *External RAID Storage: Direct Attached Storage is not Dead Yet.* http://www.tomshardware.com/reviews/external-raid-storage,1922.html

Discusses storage solutions. Reviews two specific RAID systems "aim to bridge the gap between inflexible high-speed internal storage and flexible, but expensive networked storage solutions."

Preservation Strategies

Besser, Howard. *Digital Longevity*. Handbook for Digital Projects: A Management Tool for Preservation and Access. 1999. http://www.gseis.ucla.edu/~howard/Papers/sfs-longevity.html

An older article, but it addresses the important aspects of digital files including life span, viewing concerns, and problems encountered such as scrambling, custodial, and translation problems. The article also discusses various approaches to improving digital longevity.

The Commission on Preservation and Access and the Research Libraries Group. *Preserving Digital Information*. Report of the Task Force on Archiving Digital Information. May 1, 1996. http://www.ifla.org/documents/libraries/net/tfadi-fr.pdf

An older report often cited when discussing digital preservation. The guide includes the challenges of archiving digital information (obsolescence, migration, and legal issues), the roles and responsibility of archives including operating environments, migration strategies, and finances.

Cornell University. *Digital Preservation Management: Implementing Short-term Strategies for Long-term Problems*. http://www.icpsr.umich.edu/dpm/dpm-eng/terminology/strategies.html

An online tutorial that addresses many preservation issues and preservation strategies in a step-by-step guide. Topics include terms and concepts, obsolescence and physical threats, digital repositories, metadata, legal issues, balancing stakeholders, access issues, financing the future, and organizational and technological infrastructure. This particular section defines the follow strategies: bit-stream copying, refreshing, durable/persistent media, technology preservation, digital archaeology, analog backups, migration, replication, reliance on standards, normalization, canonicalization, emulation, encapsulation, and the universal virtual computer.

Digital Preservation Testbed. *Migration: Context and Current Status*. December 5, 2001.
http://www.digitaleduurzaamheid.nl/bibliotheek/docs/Migration.pdf

A paper that discusses different methods of preservation including technology preservation, printing to paper, migration, emulation, encapsulation, virtual machine software, and XML. The method of migration is covered in depth and includes information on different migration strategies, risk management, issues and concerns, and the current research and knowledge surrounding migration.

JISC Digital Media. *Establishing a Digital Preservation Strategy*.
http://www.tasi.ac.uk/advice/delivering/digpres2.html

A report that discusses the importance of having an organizational strategy for a digital collection before you begin converting files. The processes of technological preservation, technological emulation and migration are explained.

Lawrence, Gregory W. et al. *Risk Management of Digital Information: A File Format Investigation*. Council on Library and Information Resources. June 2000.
http://www.clir.org/pubs/reports/pub93/contents.html

This report discusses both emulation and migration, with more details on the migration process. A risk management chart shows associated risks with the file formats and conversion process when using migration. States the importance of accessing risk, and explains how to complete a risk assessment for your own collection.

Markin, Andy. *CD and DVD Longevity: How Long Will They Last?* Audioholics Online A/V Magazine. August 25, 2004. http://www.audioholics.com/education/audio-formats-technology/cd-and-dvd-longevity-how-long-will-they-last

> How to test storage media to interpret a storage life. Explains the construction and components of optical media to help understand why they are fragile. Dos and don'ts for handling and storage of optical media.

Preserving Access to Digital Information (PADI). *Migration*. National Library of Australia. http://www.nla.gov.au/padi/topics/21.html

> Helps define different methods of migration including refreshing, and using standard formats. Addresses the issues around metadata, risk management, intellectual property rights, costs, and frequency of migration.

UKOLN at the University of Bath. *Good Practice Guide for Developers of Cultural Heritage Web Services: Digital Preservation*. http://www.ukoln.ac.uk/interop-focus/gpg/Preservation/

> The focus of this section of this guide book is digital preservation. Digital preservation is defined with details about each strategy; technology preservation, technology emulation, and data migration. Preservation metadata is also addressed

Disposition of Originals

Manitoba Government. *Managing Records in the Manitoba Government*. http://www.gov.mb.ca/chc/archives/gro/recordkeeping/managing1.html

> Discusses records schedules, transferring and use of files, destruction and records with archival value. Additional policies address how to transfer records for offsite disposal and destruction as well as the use of a paper shredder at the office level.

Mead, Kevin. *Get Serious About Paper Record Destruction*. American Health Information Management Association (AHIMA). http://library.ahima.org/xpedio/groups/public/documents/ahima/bok1_012344.hcsp?dDocName=bok1_012344

Discusses the importance of taking control of paper records and knowing the risks involved with various methods of destruction. Policies and procedures set the stage for the trustworthiness of records.

PALINET. *Digitization and Preservation.* http://www.palinet.org/ds_ps_pag_preservation.aspx

A resource slanted more towards preservation rather than storage, the PALINET website includes links to preservation policies that represent best practices. Disaster planning and preservation techniques that ultimately relate to best practices for storage of original and digital materials are also covered.

State of Wisconsin – Department of Administration. *Records Management Section Fact Sheet #4: January 2008 Topic: Options for Destruction of Paper Files*. http://www.doa.state.wi.us/facts_view.asp?factid=17&locid=2

Policies for the state of Wisconsin for both non-confidential and confidential materials including a witness of destruction service.

University of Miami Leonard M. Miller School of Medicine. *Secure Data Disposal Methods*. http://it.med.miami.edu/x677.xml

Defines secure data disposal and discusses methods of disposal for various types of media including paper, electronic, magnetic, and optical formats.

Providing Access

Adams, Margaret O. *Archival Reference Services for Digital Records: Experience with the Access to Archival Databases (AAD) Tool*. New Skills for the Digital Era Case Study 9. http://rpm.lib.az.us/NewSkills/CaseStudies/9_Adams.pdf

Discusses the National Archives and Records Administration's (NARA) Access to Archival Databases (AAD) tool as a method to providing access and assisting with reference services. Background information about records requests and how

the AAD has assisted with the digitization process. Also discusses the skills required of archivists when working with digital rather than analog records.

The Archivists' Toolkit. *Introduction to the Archivists' Toolkit*. 2007. http://www.archiviststoolkit.org/

The home page for the Archivist's Toolkit, the first open source archival data management system intended for a wide range of archival repositories. The application supports accessioning, description, establishing name and subjects, managing location information, and exporting finding aids and reports in various standard formats.

Archon. *Archon: The Simple Archival Information System*. University of Illinois. 2008. http://www.archon.org/

The home page for Archon, the web-based tool that automatically publishes archival descriptive information and digital archival objects to a user-friendly website.

Hodge, Gail and Evelyn Frangakis. *Digital Preservation and Permanent Access to Scientific Information: The State of Practice*. The International Council for Scientific and Technical Information and CNDIS US Federal Information Managers Group. http://www.icsti.org/documents/preservationrpt.pdf

Covers general topics of concern in the archival profession including access, preservation, standards, workflow, and trends. Discusses six different "off-the-shelf" systems for digital preservation and access including OCLC Digital Archives, PANDORA, LOCKSS, and Fedora.

Mengel, David J. *Access to United States Government Records at the U.S. National Archives and Records Administration*. Japan-US Archives Seminar. May 2007. http://www.archivists.org/publications/epubs/accesstoarchives/07_David_MENGEL.pdf

This resource provides a history of the legal requirements to provide access to government records. The role of the National Archives and Records Administration's (NARA) in holding the government accountable for their actions is discussed. Discusses access regulations and concerns particular to each branch of the US government including the Executive Records, Presidential Record, Legislative Records, and Judicial Records.

The National Archives and Records Administration. *Access to Archival Databases (AAD)* http://aad.archives.gov/aad/

This is the home page for the Access to Archival Databases website which is an example of a searchable database to find digital materials on various topics held by NARA. The Getting Started Guide helps answer questions that can assist a user as well as a developer of an online searchable system.

The National Archives and Records Administration. *Electronic Records Archives (ERA).* October 2, 2008. http://www.archives.gov/era/

"The Electronic Records Archive is the National Archives and Records Administration's strategic initiative to preserve and provide long-term access to uniquely valuable electronic records of the U.S. Government, and to transition government-wide management of the lifecycle of all records into the realm of e-government."

The National Archives. [UK] *Access to Archives (A2A).* http://www.nationalarchives.gov.uk/a2a/

An example of a database that gives users access to catalog records of about 400 repositories most of which are in England and Wales. This resource can be used to see holdings in many repositories at once.

Nguyen, Binh Q. *Information Assurance Issues and Requirements for Distributed Electronic Records Archives*. Army Research Laboratory. April 2003. http://aad.archives.gov/aad/help/getting-started-guide.html

A summary of the important issues and requirements in ensuring authenticity and accurate documents in an electronic records management system. Includes information on data availability, integrity, authentication, confidentiality, and nonrepudiation. Information on network setups, threats, and concerns. Appendix includes a list of other information assurance resources.

Pearce-Moses, Richard. *Caught in the Middle: Access to State Government Records in the United States.* Japan-US Archives Seminar. May 2007. http://www.archivists.org/publications/epubs/accesstoarchives/09_Richard_Pearce-MOSES.pdf

A discussion on how to balance providing access to government documents with the right to privacy of individuals. Tries to answer what is record or more

specifically what is a public record? Talks about the pressures put on archives to both allow and restrict access to various documents.

Schwartz, Scott W., Christopher Prom, Kyle Fox and Paul Sorensen. *Archon: Facilitating Global Access to Collections in Small Archives*. International Federation of Library Associations and Institutions 74th Annual Conference. August 2008. http://www.ifla.org/IV/ifla74/papers/159-Schwartz_Prom_Fox_Sorensen-en.pdf

Written by the developers of Archon, the web-based tool for publishing archival information on the web, this paper discuss the challenges small archives are faced in regards to providing access to their holdings and the importance of collaboration and global access. The features of Archon are described in detail including the collections management features and administrative features that were created to assist archives on every level of curation.

Spiro, Lisa. *Archival Management Software*. Council on Library and Information Resources (CLIR). January 2009. http://clir.org/pubs/reports/spiro/spiro_Jan13.pdf

A discussion of the different archival processing solutions available that help provides access to collections. The Archivists' Toolkit, Archon, Cuadra STAR, Eloquent, and CollectiveAccess systems are all discussed, including user feedback.

Developing a Program

Clareson, Thomas F.R. *Statewide Digitization Program Plan for the Delaware Division of Libraries.* PALINET. August 30, 2007. http://state.lib.de.us/For_Libraries/Planning/Digitization/documents/DEStatewideDigPlanFinal20070830.pdf

Addresses collaborations on the state level and steps that should be taken before a digitization program is created.

Greenstein, D. *Strategies for Developing Sustainable and Scalable Digital Library Collections*. Digital Library Federation. July 12, 2006. http://www.diglib.org/collections/collstrat.htm

Outlines steps needed for developing sustainable digital collections.

Government Printing Office (GPO). *Information Brief: Digitization and the FDLP.* United States GPO. April 2007. http://www.gpo.gov/su_docs/fdlp/pubs/proceedings/07spring/digitization-0407.pdf

An update to the GPO digitization project.

Hurst, Jill Ann. *Planning for Regional Digitization Projects: The SCRLC Approach. South Central Regional Library Council*. September 30, 2002. http://www.lakenet.org/Digitizing/2002digplan.pdf

This document provides an example of a digitization plan including the mission and scope of the program and a project time line. Questions are asked about project management, material selection, digitization, material access, long-term strategy, and funding. These questions can be used as a guide for any project or program.

Indiana University. *Indiana University Digital Library Program*. December 9, 2008. http://www.dlib.indiana.edu/index.shtml

Collaborating with multiple units from Indiana University they have created the Digital Library Program (DLP). The DLP offers various services relating to digital library development to University staff as well as external partners in the areas of project planning, digitization services, electronic text services, metadata services, and interface design and usability services. This homepage will lead the reader to many different areas of interest related to digital services.

Lahr, Thomas F., Robin Haun-Mohamed and Eleanor G. Frierson. *Developing the Digital World of Government Information and Official Publications: A View from the United States*. International Federation of Library Association 71st General Conference. August 2005. http://www.ifla.org/IV/ifla71/papers/037e-Lahr_Haun_Frierson.pdf

Discussion of the US government's digitization and preservation initiatives. Description of the Government Printing Office's project and the four institutions that will pilot the program.

Roth, Eric J. *Opportunities, Challenges, and Priorities: Developing a Collaborative Digitization Plan for the Mid-Hudson Valley*. Southeastern New York Library Resources Council. May 2001. http://www.hrvh.org/about/HRVHassessment.pdf

Discussion of the digitization project goals for the Hudson Valley area. Other local and national digitization efforts are summarized. After reviewing others

efforts, suggestions are made for developing a plan which also addresses possible issues of concern.

Smith, Abby. *Developing Sustainable Digital Library Collections: Strategies for Digitization*. Digital Library Federation. September 2001. http://www.ncsi.iisc.ernet.in/raja/is214/is214-2006-01-04/clir-pub101.pdf

Discusses the importance of sustainability and how they relate to policies, guidelines, and best practices. Using digitization as a method for preservation and access is discussed with the overall impacts on the institution.

Stanford University Digital Library Systems and Services. *Digitization Program, Scanning Labs, How to Start a Digital Project*. Stanford University Libraries and Academic Information Resources. January 17, 2006. http://www-sul.stanford.edu/depts/dlss/services/index.htm

Stanford University has an "ongoing program to produce and acquire digital library collections". This webpage details the scanning services available as well as what to do to start a digital project within the University. Provides background information on a developed digital program.

WebJunction. *Developing Digital Projects: Digitization Action Plan*. July 19, 2005. http://www.webjunction.org/216/articles/content/435617

A worksheet, originally created to use in conjunction with an online course on Digital Projects, asks important questions about all stages during the development of a digital project or program.