Digital Imaging

Summary

Increasingly businesses and government agencies are looking towards digital imaging to enhance productivity and to provide greater access to certain types of information. Digital imaging offers many advantages, including: improved distribution and publication, increased access, streamlined workflows, and a greatly reduced need for physical storage space. In addition, digital images can be used to create text-searchable files through the application of optical character recognition (OCR) software. Digital images can be made available over the web, allowing government agencies and businesses to provide information to business partners or the general public quickly and efficiently.

While digital imaging is becoming increasingly popular and commonplace, you must remember it is an investment with potentially very high up-front costs. Digital imaging should make financial sense for your business or agency. To assure your digitized records are fully admissible in court, they must be trustworthy, complete, and durable for as long as your approved records retention schedules require.

Legal Framework

Imaging is, by state law, a recognized and legitimate form of record reproduction. To assure that your imaged records are fully admissible and meet all evidentiary standards, you should review the requirements of the:

- Official Records Act [Minnesota Statutes, Chapter 15.17] (available at: <http://www.revisor.leg.state.mn.us/stats/15/17.html>), which mandates that government agencies must keep records to fulfill the obligations of accountability and specifies that the medium must enable the records to be permanent. It further stipulates that you can copy a record and that the copy, if trustworthy, will be legally admissible in court.

- Uniform Photographic Copies of Business and Public Records as Evidence Act [Minnesota Statutes, Chapter 600.135] (available at: <http://www.revisor.leg.state.mn.us/stats/600/135.html>), which establishes that an accurate reproduction of a record is as admissible in evidence as the original in any judicial or administrative proceeding. It further stipulates that if an accurate and durable reproduction is made, the original record may be destroyed in the regular course of business unless its preservation is required by law.

- Records Management Act [Minnesota Statutes, Chapter 138.17] (available at: <http://www.revisor.leg.state.mn.us/stats/138/17.html>), which establishes the Records Disposition Panel to oversee the disposition of records using approved records retention schedules.

- Minnesota Government Data Practices Act (MGDPA) [Minnesota Statutes, Chapter 13] (available at: <http://www.revisor.leg.state.mn.us/stats/13/>), which mandates that

government records should be accessible to the public unless categorized as not-public by the state legislature.

- Uniform Electronic Transactions Act (UETA) [Minnesota Statutes, Chapter 325L] (available at: <http://www.revisor.leg.state.mn.us/stats/325L>) and Electronic Signatures in Global and National Commerce (E-Sign), a federal law (available at: <http://thomas.loc.gov/cgi-bin/query/z?c106:S.761:>). Both UETA and E-Sign address the issues of the legal admissibility of electronic records created in a trustworthy manner and of the application of the paper-oriented legal system to electronic records.

To prove the legal admissibility of your digitized records, you will need to demonstrate the records were created in a trustworthy manner. You should be prepared to demonstrate the reliability of your system and show that it leaves no room for the manipulation of the stored record. For more information refer to the State Archives' *Trustworthy Information System Handbook*.

Key Concepts

Before you determine whether digital imaging will meet your long-term legal and operational needs, acquaint yourself with the following key concepts:

- Imaging terms
- File formats
- Image storage
- Metadata
- Justifying the cost
- Choosing a vendor
- Implementation strategy

Imaging Terms

Digital imaging is a process by which a document or photo is scanned by computer and converted from analog format to a computer-readable digital format. After scanning, the original document or photo is represented by a series of pixels arranged in a two-dimensional matrix called a bitmap or raster image. This image can then be transferred onto a variety of electronic storage media, such as CD-ROM, for storage and use.

For a better understanding of imaging you should be familiar with the following terms:

- Pixel Bit Depth: Defines the number of shades that can actually be represented by the amount of information saved for each pixel. These can range from 1 bit/pixel for binary (fax type) images to 24 bits per pixel or greater in high quality color images. The following are current standard bit depths for image files:

  Table 1: Standard pixel bit-depths

Bit-depth	Displays	Recommended for
1-bit or "bi-tonal"	black and white	Typewritten documents
8-bit grayscale	256 shades of gray	Black and white photographs, half-tone illustrations, handwriting
24-bit color	Approximately 16 million colors	Color graphics and text, color photographs, art, drawings, maps

  Information taken from: *Technical Recommendations for Digital Imaging Projects* (Columbia University Libraries, April 1997), pages 4-5 (available at: http://www.columbia.edu/acis/dl/imagespec.html).

- Resolution: The quality of a digital image is dependent on the initial scanning resolution. Resolution is expressed in the number of dots, or pixels, used to represent an image, expressed commonly as "dpi," dots per inch. You may also see "ppi" (pixels per inch) and "lpi" (lines per inch) used. As the dpi value increases, image quality increases but so does the file size.

  Unlike paper documents, the resolution of photographs is sometimes expressed in the number of pixels across the long-dimension of an image. When creating standard-sized images from photographs or negatives of differing sizes (e.g., 35mm, 4"x 5"), the scanning resolution in dpi varies. In such cases, it is often easier to measure resolution as the number of pixels across an image's long dimension. For example, each of the following files measures 3000 pixels in the long-dimension, although they have varying values of dpi.

Table 2: Resolution as the number of pixels across the long-dimension of an image

Original photo size	Digital image size	Scanning resolution
8”x10”	2400 x 3000 pixels	300dpi
4”x5”	2400 x 3000 pixels	600dpi
35mm negative	2400 x 3000 pixels	2100dpi

Information taken from: Maxine K. Sitts, *Handbook for Digital Projects: A Management Tool for Preservation and Access* (Andover, Massachusetts: Northeast Document Conservation Center, 2000), page 86 (available at: http://www.nedcc.org/oldnedccsite/digital/dman2.pdf).

To determine the scanning resolution you need, you first have to determine the desired quality of your images and the storage capacity of your computer system. You will also need to consider the desired speed of delivery of the images, especially if they will be accessed over the Internet. You may want to scan high-resolution masters of your images and then create lower resolution copies for web delivery. General recommendations for master files are as follows:

Table 3: Recommended scanning resolutions for master files

Material	Recommended resolution (8-bit grayscale and 24-bit color)
Prints, paintings, drawings, textual records	600 dpi
Maps and oversize	600 dpi (300 dpi minimum)
Photographs, negatives, slides	3000-6000 pixels in long dimension, or 600dpi.

Information taken from: *CDL Guidelines for Digital Images* (California Digital Library, April 2008) (available at: http://www.cdlib.org/inside/diglib/guidelines/bpgimages/reqs.html#guidelinesmaster).

- Compression: Data compression saves file space. There are two types of compression, lossless and lossy. Under lossless compression no data is lost (although the file is still compressed). Under lossy compression data is lost. Lossy compression attempts to eliminate redundant or unnecessary information. Depending upon the degree of compression, this information loss may be unnoticeable to the human eye. For example, it is possible for a JPEG file (a lossy compression) and a TIFF file (lossless) to appear exactly the same, although the JPEG file is missing data, making it significantly smaller. These file formats, and others, are discussed in the following section.

File Formats

In any digital imaging project, choosing the file formats you will use is important. Like scanning resolution, the file format directly affects the quality and file size of your images. Choosing the best file format for your needs requires knowing how your images will be used (e.g., archival or display functions), the type of materials you will be imaging (e.g., text, art, graphics, photos), and the desired speed of delivery and the necessary quality of your images.

For many digital imaging projects it is necessary to create master images. Master images are especially necessary when creating a digital archive. They will serve as archival copies and be

the basis from which derivative images are subsequently created. For this reason, high quality is crucial; master images must be in a high resolution and lossless format, insuring that the original document is captured as completely as possible. Master images are high quality images, and they facilitate such functions as implementing OCR, verifying textual information, or zooming into details in maps or photographs. The TIFF file format, which allows high resolution and utilizes lossless compression, is well suited for making master images.

Master images, in formats such as TIFF, have large file sizes, making their delivery cumbersome for some web and document management system applications. To enhance the speed of delivery, you can create copy images from the master images. Copy images have smaller file sizes, are of lower quality, and typically use a lossy compression. The JPEG file format is commonly used for copy images.

Common types of digital image file formats include:

- *Tagged Image File Format* (TIFF) files, which are widely usable in many different software programs. TIFF files utilize lossless compression and are commonly used for master copies. TIFF graphics can be any resolution, and they can be black and white, grayscale, or color. TIFF is a very extensible format, allowing variations to be created for specific applications. Variations include *GeoTIFF*, used in cartographic and GIS (geographic information system) applications; *TIFF Class F*, used in faxing applications; and *TIFF/IT*, used in the graphic arts industry. Files in TIFF format end with a .tif extension.

- *Graphics Interchange Format* (GIF) files. GIF supports color and grayscale. Limited to 256 colors, GIFs are more effective for images such as logos and graphics rather than color photos or art. It should be noted that although the GIF format is widely used, it is technically proprietary. A lossless compression, files in GIF format end with a .gif extension.

- *Joint Photographic Experts Group* (JPEG) files. JPEG is a lossy compression technique for color and grayscale images. Depending upon the degree of compression, the loss of detail may be visible to the human eye. Files in JPEG format end with a .jpg extension.

- *Bitmap* (BMP) files. BMP files are relatively low quality and used most often in word processing applications. BMP format creates a lossless compression. Files end with a .bmp extension.

- *Portable Network Graphics* (PNG) files. A lossless compression designed to replace GIF files, PNG files can be ten to thirty percent more compressed than GIFs. PNG is completely patent and license free and is of higher quality than GIF. Files in PNG format end with a .png extension.

- *Portable Document Format* (PDF) files. PDFs are useful for viewing and printing multiple documents and images. Commonly used to capture, distribute, and store electronic documents, PDF preserves the fonts, images, graphics, and overall “look” of the original digital files. As with the GIF format, the PDF format is proprietary, although widely used. Files in PDF often end with a .pdf extension.

For a more in-depth discussion of file formats, refer to the *File Formats* guidelines.

Image Storage

Digital images are stored on digital media. Digital media are divided into two types: magnetic and optical. Examples of magnetic media include:

- *Magnetic disk.* Magnetic disks include the hard disk found in your computer that stores the programs and files you work with daily. Magnetic disks provide random access. Also included are removable hard disks, floppy disks, zip disks, and removable cartridges.

- *Magnetic tape.* Magnetic tapes come in reel-to-reel as well as cartridge format (encased in a housing for ease of use). The two main advantages of magnetic tapes are their relatively low cost and their large storage capacities (up to several gigabytes). Magnetic tapes provide sequential access to stored information, which is slower than the random access of magnetic disks. Magnetic tapes are a common choice for long-term storage or the transport of large volumes of information.

- *Digital Audio Tape (DAT).* DATs are in a cartridge format a little larger than a credit card. The industry standard for DAT cartridge format is a digital data storage (DDS) cartridge. DDS cartridges provide sequential access.

Optical media options include:

- *Compact Disk (CD).* Compact disks come in a variety of formats. These formats include CD-ROMs that are read-only, CD-Rs that you can write to once and are then read-only, and CD-RWs that you can write to in multiple sessions.

- *Write-Once, Read-Many (WORM) disk.* WORM disks require a specific WORM disk drive to enable the user to write or read the disk. WORM disks function the same as CD-R disks.

- *Erasable Optical (EO) disk.* The user can write to, read from, and erase from EO disks as often as they can magnetic disks. EO disks require special hardware.

- *Digital Versatile Disk (DVD).* These disks are also called digital video disks, but do not necessarily include video. DVD disks are new types of optical disks with more storage capacity than CD-ROMs. Common types of DVDs include:

    - *DVD-ROM*. These DVDs are read-only disks that also have enough storage capacity for a full-length feature film. They are accessed using a special DVD drive attached to a personal computer. Most of these drives are backward-compatible with CD-ROMs and can play DVD video disks. DVD-Rs can be written to once and are then read-only.

    - *DVD-RAM*. These DVDs are rewritable disks with exceptional storage capacity. They come in one- or two-sided formats.

    - *DVD+RW*. DVD+RW is a direct competitor to DVD-RAM with similar functionality and slightly greater storage capacity.

Note: DVD-RAM and DVD+RW are not compatible. The two technologies are being developed by competing vendors and require different hardware.

It is highly recommended that you store digital images on WORM, CD-R, or DVD-R disks to assure that the stored records are tamper-proof, allowing the greatest security for the data. Rewritable disks, in contrast, are designed to be re-used, making data integrity uncertain.

Because of limited life expectancy, no digital storage medium is adequate for the long-term, archival preservation of records. The most generous estimate of physical obsolescence is thirty years. Technological obsolescence, though, will probably come within five to ten years. As a result, you should assume the need to migrate all your files to a new storage medium on a regular basis. In the meantime, you will need to protect your stored data with a comprehensive back-up system.

For more information on digital storage media, refer to the *Digital Media* guidelines.

Metadata

Metadata is crucial to any digital imaging project, enabling proper data creation, storage, retrieval, use, modification, and retention of your digitized records. In addition, proper metadata helps document the trustworthiness of your system, assuring the legal admissibility of your digitized records in court.

Metadata can be simply defined as "data about data." More specifically, metadata consists of a standardized structured format and controlled vocabulary which allow for the precise description of record content, location, and value. Metadata often includes items like file type, file name, creator name, date of creation, and the record's classification under the Minnesota Data Practices Act.

For digital images, metadata is especially important in facilitating retrieval. Unless you plan to use OCR, all of your records will be stored as graphic files. The only way to locate specific information will be through its metadata. Metadata makes it possible to locate, use, and evaluate information through standard search criteria such as subject heading, numerical identifier, or keyword. For this to work effectively, you will have to identify the metadata your employees or patrons use to search for records.

For more information concerning metadata, refer to the *Metadata* guidelines.

Justifying the Cost

Digital imaging should make financial sense for your business or agency. While digital imaging is becoming increasingly popular and commonplace, you must remember it is an investment with potentially very high up-front costs.

Is digital imaging financially right for you? A comprehensive cost-benefit analysis is a necessary step in determining the answer. Costs will include system hardware, system software, application software, long-term system maintenance, staff training, vendor costs, and other expenses. Benefits include higher office productivity, lower storage costs, and the option of using the Web to make digitized information easily accessible. Keep in mind that, because of the rapid pace of technological obsolescence, you will need to make continuing investments in all aspects of an imaging application on a routine and frequent basis.

Choosing a Vendor

Most agencies and businesses do not have the appropriate scanning equipment, software, or staff expertise to execute a large digitizing project. For this reason, vendors have become integral to the world of digital imaging. Quality varies among vendors, so selecting the right one is crucial to your project.

Vendors provide digitizing services, technical advice, and sometimes the long-term maintenance of the resulting electronic files. To better choose your vendor, you should become familiar with digitizing technology and the terms used by the industry. You must also have a clear idea of your project and its goals. Questions such as the following must be addressed:

- How much material will be digitized?
- What type of materials will be digitized? Textual documents? Photographs? Maps?
- Who is the intended audience? Staff members? Researchers? The general public?
- What is the required quality of the digital images? High or low resolution? Black and white or color?
- What is the desired end product? A document management system? A searchable online collection?

As technology products and vendors come and go, you should assume and plan for the possibility of business failure and the inevitability of product obsolescence. The best way to protect yourself is to insist on an open systems architecture, using non-proprietary hardware and software. Non-proprietary means that the chosen hardware and software is not specific to that vendor. If proprietary software is unavoidable, it should be licensed beyond the length of the contract. As there will inevitably be some bugs in the system, a contract should completely spell out the provisions for implementation, service, upgrades, and repair.

Implementation Strategy

To successfully implement a digitizing project in a timely manner, you must create an implementation strategy which manages workflow. A digitizing project incorporates a myriad of tasks, the successful management of which can save time and money. While a vendor may be contracted for the project, you will still need to manage an assortment of activities, including the:

- Selection of materials to be digitized
- Preparation of materials, including sorting files, removing staples and paperclips, weeding out unnecessary materials, and conservation of any deteriorating documents.
- Creation of standardized metadata
- Quality control of source materials and digital images

- Staff training on new hardware and/or software
- Advertising, promotion, and user evaluation
- Long-term maintenance of resulting electronic files

Your implementation strategy may include setting up a pilot project. A pilot project will allow you to test the technology, examine the effectiveness of your digital images in providing and managing information for patrons or employees, and help determine how you can better implement a digital imaging system. A pilot project is especially necessary to study the impact and effectiveness of imaging before undergoing a large digitizing project for a whole department or organization.

Phases are an effective approach to implementing large digitizing projects. Rolling out the system in phases enforces an organized and careful approach to implementation. This allows small errors to be caught and corrected before they snowball into large and costly issues. Phases can be applied in several ways depending upon the structure of your organization and scope of your project. For example, you may want to phase in the system by departments or by function. If your project will be implemented over a lengthy time period, you may want to phase in your system beginning with your organization's highest priorities.

Key Issues to Consider

Now that you are familiar with some of the basic concepts of digital imaging, you can use the questions below to discuss how those concepts relate to your agency. Pay special attention to the questions posed by the legal framework, including the need for public accessibility as appropriate, completeness, trustworthiness, and legal admissibility. Consider the resolution and delivery requirements of your digital images, and choose the file formats and digital storage media that will best fit your needs.

The goal is to determine the best option for your agency that meets your legal and operational needs, not merely to automatically upgrade technology. If you cannot justify the costs of digital imaging, keeping your records in their original form may be the best option.

Discussion Questions

- What are our goals for digital imaging?
- How is our agency affected by the legal requirements?
- What is the desired end product? A document management system? A searchable online collection?
- What type of materials will be digitized? Textual documents? Photographs? Maps?

- What is the required quality of the digital images? High or low resolution? Black and white or color?
- What file formats and digital storage media will best fit our needs?
- What are some strategies for implementing our digital imaging project?
- Can we justify the costs of digital imaging?

Annotated List of Resources

Primary Resources

California Digital Library. *CDL Guidelines for Digital Images.* April 2008.
<http://www.cdlib.org/inside/diglib/guidelines/bpgimages/index.html>

These standards, published by the California Digital Library at the University of California, provide recommendations for image quality, file formats, and storage media for digital image collections.

Columbia University Libraries. *Technical Recommendations for Digital Imaging Projects.* April, 1997.
<http://www.columbia.edu/acis/dl/imagespec.html>

These digital imaging recommendations were prepared by the Image Quality Working Group of ArchivesCom, a joint committee between Columbia University Libraries and AcIS (Academic Information Systems) at Columbia University. Provides recommendations for image quality, file formats, and other capture and storage issues.

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/oldnedccsite/digital/dman2.pdf>

This handbook, published by the Northeast Document Conservation Center, is geared towards librarians, archivists, and other cultural or natural resource managers. Provides a basic technical overview of digital imaging and emphasizes project management, cost justification, vendor relations, and related issues.

Additional Resources

Collaborative Digitization Program
<http://www.cdpheritage.org/>

The Collaborative Digitization Program began as the Colorado Digitization Project in 1998, a collaborative effort among Colorado's archives, historical societies, libraries, and museums. The CDP now includes institutional partners from other western states to provide access to cultural heritage collections. Its website features technical guidelines, digitizing standards, a digital imaging glossary, and links to many additional resources.

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

Produced by the Digital Imaging and Preservation Policy Research (DIPPR) team at Cornell University's Department of Conservation and Preservation, this web tutorial provides an overview of technical and project management issues regarding digital imaging. Tutorials in English, French and Spanish use examples of actual digital images to demonstrate variations in image quality.

Minnesota Historical Society, State Archives Department. *Trustworthy Information Systems Handbook.* Version 4, July 2002.
<http://www.mnhs.org/preserve/records/tis/tis.html>

This handbook provides an overview for all stakeholders involved in government electronic records management. Topics center around ensuring accountability to elected officials and citizens by developing systems that create reliable and authentic information and records. The handbook outlines the characteristics that define trustworthy information, offers a methodology for ensuring trustworthiness, and provides a series of worksheets and tools for evaluating and refining system design and documentation.

Newcombe, Tod. The Local Government Guide to Imaging Systems: Planning and Implementation. *United States: Public Technology, Inc., 1995.*

A publication by Public Technology, Inc. (PTI) and the International City/County Management Association (ICMA), this guide emphasizes planning and implementation issues associated with digital imaging projects. Also addressed are policy and legal issues including records retention, ownership and control of images, and public access.

Technical Advisory Service for Images (TASI)
<http://www.tasi.ac.uk>

This website by Technical Advisory Services for Images (TASI), based at the University of Bristol's Institute for Learning and Research Technology (ILRT), provides information for creating and using digital image archives. The site features technical and project management advice, and a glossary of digital imaging terms.