NORTH CAROLINA DEPARTMENT OF CULTURAL RESOURCES
OFFICE OF ARCHIVES AND HISTORY
DIVISION OF HISTORICAL RESOURCES
ARCHIVES AND RECORDS SECTION

Guidelines for Digital Imaging Systems

Phase II: Technology Assessment and Selection

August 1, 2003

ACKNOWLEDGMENT

The North Carolina Division of Historical Resources would like to acknowledge the assistance of the State Records Management of the Secretary of State's office of Missouri. These guidelines are based on their publication *Guidelines: Digital Imaging Systems*. In addition, we consulted publications from other states such as Alabama, Connecticut, and Mississippi.

PURPOSE:

We hope these guidelines will help you assess what imaging entails—the cost, the time, and the commitment. In addition, we hope they walk you through the process of determining whether or not the expense of imaging is in line with the cost and worth the commitment of resources and long-term planning. Phase II, Technology Assessment and Selection discusses hardware and software choices as well as file formats and media types.

Phase II: Technology Assessment and Selection

Today, managers and administrators face the tremendous challenges of continued viability of their systems and the preservation of their records. To meet these challenges, they must employ strategies that transcend rapid technological changes. The success of such strategies depends upon several factors, including open systems architecture, component upgrades, stable storage environment, and accurate data transfer. The goal of this section is to give state and local government agencies and officials the information they need to design systems to ensure long-term functionality for the records that will need it.

What is Imaging?

Imaging is the process of converting human readable media, such as paper or microfilm, into information that can be stored and retrieved electronically. Basically, an image is a digitized picture of a document, drawing, or photograph. Another way to think of it is that an image is like a photocopy that you view on your computer. An imaging system's components are interdependent. Hardware, software, compression techniques, file formats, and media all have to work together. Each component does this using interfaces and drivers. You need a basic understanding of what each does to make an informed decision.

Using open systems architecture as much as possible is the best method to ensure system viability over time. A system is a combination of components working together. The system's architecture is its design or configuration. An open architecture means that the system's components use standards or specifications that have been made public by their designers. The use of standards and publicly available specifications mitigates the impact of incompatibility. Technical staff can upgrade hardware and software with minimal impact on the overall system and without significant risk of data loss. Thus, the integrity of the system is more likely ensured by using open systems architecture.

Recommendation: Use open systems architecture. If you must use proprietary software and hardware, contractually require vendors of proprietary systems to provide a method to either migrate to non-proprietary configurations or provide the code in escrow should they go out of business or if you switch vendors. The code could be held in a mutually agreed upon place. This allows your information to be processible in the future.

Hardware

Imaging hardware performs three basic jobs. It captures, stores, and retrieves digital images. The results of the needs assessment and workflow analysis will guide your hardware selection process. When configuring the system, plan for

your peak workload. In other words, if time is a factor, configure the system's capacity to handle the highest volume you will need to push through it. Also, plan for predictable expansion. If you know your workload increases 10% per year, configure the system so that it can meet your needs for the near future (three to five years). The essential thing to remember is to purchase what is needed for current and anticipated needs, not the latest and greatest. Bear in mind how much funding is available for implementing a new system and how much will be available in the future. The more complex the system, the more costly it will be to implement and to maintain it.

Capture

Typically, when people think of a digital imaging system, they think of scanning. Within the scanning process are several components and considerations: scanner types, feed method and throughput, resolution and drop out inks. Also included in this section is a discussion on hardware interfaces and software drivers. Both are an integral part of the capture process.

1. **Scanner types.** There are two broad categories of scanners – document and graphics. As the name implies, document scanners are designed to capture typical black-and-white or grayscale documents. Graphics scanners, conversely, are designed to capture color.

2. **Feed method and throughput.** The feed method is literally the way paper or microfilm is moved across the scanner. When selecting the appropriate feed method, an important consideration is the condition of the documents. Frail documents should not be run through an automatic document feeder. Conversely, high-volume documents, such as checks, should not be placed individually on the flatbed. Throughput is the rated speed at which the scanner can process pages. For example, a scanner rated at 60 ppm should be able to process 60 pages per minute. Vendor claims of throughput are useful for comparison; however, throughput on real-world documents will vary somewhat. In fact, once you begin production, you are likely to find the scanner's throughput to be less than its rated speed. When you interview vendors, ask them to perform tests on samples of your documents of various types (e.g. maps, checks, memos etc.) to get an idea of how fast the scanner works versus what the vendor claims it will do. The difference can be as much as one-half to one-third the rate as quoted by the vendor.

Sheet -fed scanners feed a sheet at a time through the scanner and will not accommodate bound materials. Flatbed scanners consist of a platen glass on which you manually position books, magazines, and other documents that you want to scan. Some flatbed scanners also can accommodate automatic document feeders (ADF) that allow stacks of pages to be fed through the scanner.

Other comments on scanners:

- Simplex captures one side of a page at a time.
- Duplex captures both sides of a page in a single pass.
- High-speed can scan several hundred pages or more per hour.
- Film scanners have attachments to scan microfilm and microfiche.
- Fax machines can be connected to the imaging systems so that faxed images can be imported into your indexing and retrieval software.
- Photocopiers with memory can be connected to the imaging system. Scanned documents from these machines can be dumped into the imaging system.
- Rotary scanners, much like rotary cameras, can process large volumes of documents. However, you should have a good file management plan in place, as the file number assigned to the document by the scanner is different from the file numbering system you create. Without a file management plan, you will lose track of your information.

Recommendations:
- Choose a scanner appropriate for the condition and format of your documents.
- Plan the scanning process for your peak load and plan for expansion as needed.

3. **Scanning resolution.** Scanning resolution is measured in dots-per-inch (dpi). If you were to look at an image under a microscope, you would see that each character is made up of many, many dots. Scanners are capable of supporting resolutions from 72 to 8,800 dpi. For example, an image captured at 600 dpi, has 600x600 or 360,000 dots per square inch. According to the AIIM TR26 standard (1993), "Resolution as it Relates to the Photographic & Electronic Imaging," text documents generally can be accurately captured using a scanning resolution of 200 dpi. A higher resolution, generally 300 dpi, is necessary to produce a relatively legible text file that includes both handwritten and typed text. More detailed documents, such as maps or drawings, may require a higher resolution of 600 dpi. There is a tradeoff between image quality (legibility) and storage, cost and speed (of both scanning and retrieval). The higher the scanning resolution, the larger the image file. In addition, scanners generally support three types of scans:
- Bi-tonal—also known as line art or "black and white," it is best for printed text and high contrast graphics. While once a popular type of scan, it is not used as often today.
- Grayscale—provides a range of shades of gray in an image and delivers a better quality scan than black and white. It is best for continuous tone documents, documents with a virtually unlimited range of color or shades of gray, and black and white negatives.

- Color—duplicates the range of possible colors in an image with the higher the range the more accurate the scan at duplication. It is best for photographs and any document with color.

A note about resolution and file size: the greater the resolution and the more detailed the resolution, the greater the file size. The following chart illustrates this point:

Resolution (dpi)	400x400	300x300	200x200	100x100
2-bit black and white	20K	11K	5K	1K
8-bit grayscale	158K	89K	39K	9K
24-bit color	475K	267K	118K	29K

Review the type of document and its intended use before deciding on resolution. If it is a short-term record (less than five years), consider scanning it in a lower resolution. North Carolina ECHO, *Exploring Cultural Heritage Online*, is an evolving statewide, collaborative access to special collections and digitization project. Funded through a Library Services and Technology Act (LSTA) Grant and managed by the State Library of North Carolina, this innovative project was one of the first in the nation to build a statewide framework for digitization and addresses a full-range of digitization needs of the state's cultural collecting agencies (libraries, archives, and museums). Long-term records (ten years or longer) or archival records (permanent) may need to be scanned at 600 dpi, depending on the type of record. Consult *NC ECHO"s Guidelines for Digitization* at*, www.ncecho.org* for more detailed information regarding scanning for preservation and access of long-term, valuable records.

Please consult the Government Records Branch to discuss your system and your needs if you are planning to scan long-term or archival records.

Recommendations:
- Scan office documents no lower than 200 dpi. If you scan handwritten documents, use 600 dpi.
- Scan using grayscale or color, depending on the type of document you have. If it is simple text, bi-tonal scanning is fine.
- For data capture using OCR/ICR, scan no lower than 200 dpi. (More information on OCR and ICR may be found on page 16.)

4. **Drop-out inks.** For many offices, the implementation of an imaging system involves scanning forms that are pre-printed and then filled out by an individual. Utilizing scanners that detect and eliminate drop out inks from the image when the form is scanned can prevent invalid optical character recognition results. The goal is to use drop-out colors on the pre-printed forms to eliminate undesirable line intersections between the pre-printed forms and the data that was filled in by the person (originator, etc.) using the form. The OCR (optical character

recognition) package then "sees" only the data that was entered by the form originator.

The use of drop-out ink provides tremendous advantages when employing OCR packages. Drop-out ink is important to maintain the accuracy of the OCR translation. When people (originators, customers, etc.) fill out forms, they do not always stay within the lines provided on the form. Therefore, the data that the person fills in may intersect with form lines and create characters that cannot be recognized by the OCR packages, if the pre-printed form does not use drop out ink.

Recommendation:
- When scanning pre-printed forms, choose a scanner that eliminates "drop out" inks when imaging the form.

Storage

Storage systems can be divided into three categories: on-line, near-line, and off-line. Which type of storage system you select will depend upon how fast and how often you must access the images.
1. **On-line.** On-line storage means the image file is stored on a medium, typically a hard disk that is immediately available for retrieval. RAID is considered an on-line system. (*Redundant Array of Independent (or Inexpensive) Disks,* a category of *disk drives* that employ two or more drives in combination for *fault tolerance* and performance. RAID disk drives are used frequently on *servers* but aren't generally necessary for *personal computers*.)
2. **Near-line.** Near-line storage means the image file can be retrieved by the system and loaded into a drive to be read. Optical jukeboxes (including CD jukeboxes) are considered near-line systems.
3. **Off-line.** Off-line storage means the image file resides on media (such as optical disks, data tapes, etc.) that requires human intervention to load into a system drive.

Recommendation: Consider retrieval-time requirements when selecting storage options. Depending on your access requirements, you may want to consider a hybrid system – on-line access for active records, near-line for semi-active, and off-line for inactive.

Interfaces and Drivers

Interfaces connect devices, such as scanners and other peripherals, to each other and to the system software. In scanning, the interface controls how fast data is moved from the scanner to the software. The common interfaces are video, serial, parallel, and SCSI. (SCSI, pronounced *Scuzzy*, stands for Small

Computer Systems Interface.) Video interfaces tend to be used in proprietary systems. Through customization, video interfaces can enhance scanning performance. SCSI is an industry standard parallel interface. It can support very high transfer rates. Serial and non-SCSI parallel interfaces that cannot support high data transfer rates tend to be used with low-end scanners. These scanners typically have a lower throughput than those using video or SCSI interfaces.

Recommendation: Use the interface that is appropriate for your system. Fully document all customization. If proprietary interfaces are employed, budget for programming when the need to migrate occurs.

Software drivers are another means of communication between devices and software programs. Each device and each program has a set of commands that it understands. Drivers interpret the commands from the program to the device and from the device to the program. The standard software drivers that digital imaging systems use are TWAIN (Technology Without an Interesting Name) and ISIS (Images and Scanners Interface Standard). TWAIN drivers support basic functions. ISIS drivers can be customized to take advantage of all the scanner's features. Systems using TWAIN tend to experience slower throughput than those using ISIS.

Recommendation: Use the interface that is appropriate for your system. Fully document all customization. If proprietary interfaces are employed, budget for programming when the need to migrate occurs.

Retrieval

Once the image is stored, it must be retrieved for viewing and/or printing. The monitor and printer must be able to render the image as accurately as possible. The performance of these devices has seen dramatic improvements, and the costs have dropped considerably in recent years.

1. **Monitors.** The resolution at which a monitor can display an image is measured in picture elements or pixels. A pixel is a single point in a graphic image. The relationship between scanning dpi and display and print resolution is not one-to-one. Display quality largely depends on the monitor's resolution, that is, how many pixels it can display. The higher the display resolution, the more colors the monitor can display. Bear in mind, however, the display color may not be important for the task. A monochrome monitor may be appropriate for systems that do not use color. On the other hand, users tend to prefer color monitors to monochrome and user satisfaction is a critical consideration. Cost difference between monochrome and color monitors is minimal. A more important consideration is the viewable display size. Image viewing is not easily

done on small monitors. Eyestrain and frustration set in early. If users retrieve and view images often throughout the day, a large monitor is justifiable. Also, choose a large monitor for staff performing quality control since they will be viewing screen images frequently. If they will retrieve images sporadically, a mid-size monitor may be acceptable.

In addition, imaging can make it possible for people with disabilities to read and process documents that they would be unable to access otherwise. Assistive technology can enlarge images for people with limited vision, and manipulation of imaged documents may be easier than with paper for with certain kinds of motor disabilities. Conversely, people with other kinds of motor disabilities may find paper documents easier to manipulate. It is an issue the agency has to be prepared to accommodate on an individual basis.

Recommendations:
- Whether monochrome or color, select a monitor capable of high-resolution display.
- Select at least a 17–inch monitor for infrequent retrievals.
- Select at least a 19–inch monitor for frequent retrievals.
- Select a 21–inch monitor for continuous retrievals.

2. **Video/Graphics Adapter.** The video adapter is a card that plugs into the computer. Since monitors are analog devices, they require an adapter to translate the digital information into an analog signal. You can spend hundreds of dollars on a high quality monitor but it will only generate an image as good as the video adapter. Graphics adapters have graphic accelerators that have their own memory. As a result, the card uses its own memory to open graphic representation instead of relying on the central processing unit of the computer. This allows the machine to work more efficiently and utilize its capabilities effectively. Today's modern computers, including the most inexpensive PCs, include a graphics adapter that is adequate for most imaging tasks. However if you will be working with large images, or if you need to be able to display many images in a short amount of time, consider upgrading the computer to include a high-end graphics adapter.

Note: All computers have a graphics adapter, and do not know of any graphics adapter that can help off-load the task of decompressing a still image from the main processor. Moving the decompression tasks to the graphic card however is often done for moving pictures, MPEG etc.

Recommendations:
- Review monitor specifications and system compatibility carefully.
- If you will be compressing images or need to access compressed images, be sure to purchase equipment with a video/graphic adapter card. The card enables faster retrieval time and better image resolution.

3. **Printers.** Two broad categories of printers are laser and ink jet or non-impact printing ink. Laser printers fuse toner onto paper to produce images. Ink jet

printers use wet ink. The ink has to be placed in specified locations which means the jets have to pass over the page multiple times to produce the image. Laser printers generally produce higher resolution images faster than ink jet printers. No matter what type of printer is used, the print process can be extremely slow because uncompressed image files are so large. A print accelerator board can speed up the process dramatically. The accelerator board is installed in the printer and instructs the imaging system to send the image file to it in a compressed format. The board then decompresses the image and prepares it for printing. If any of your records are printed and are long-term records, use a laser print for longevity. The toner is fused onto the paper and can withstand moisture damage whereas inkjet printing cannot.

Recommendation: Consider print resolution and throughput requirements when choosing a printer. Use a laser printer for long-term records.

Software

Two broad software categories are operating systems (OS) and applications. The computer's OS performs basic input and output functions. Microsoft Windows™ and Apple Macintosh™ and UNIX™ systems are three well-known operating systems. The OS provides a platform on top of which applications, such as digital imaging systems, run. There are innumerable software application choices. Application programs must be written specifically to run on a particular OS. Which OS is used is dependent on the system hardware, which includes both network servers and user workstations. This can be an issue when future migrations occur. Please note that Apple software is limited to Macintosh machines while Windows and UNIX systems can run a variety of hardware.

Application software must perform certain tasks such as the following:
- **Scanning.** This function is required to operate the scanner. It translates the image captured by the scanner into a graphics file format capable of being stored.
- **Indexing.** This information is necessary to identify the image. Indexing information is stored in the database.
- **Database Management System (DBMS).** The DBMS catalogs the location of the stored image on the storage media and adds that information to the index information created by the indexing application. Whatever DBMS is used, it should understand Structured Query Language (SQL), as most DBMSs on the market today can do.
- **Image Retrieval.** This application must be capable of retrieving the image by querying the database for its location, calling the image to a monitor or a printer, and converting the file into a format that the device can use.

Recommendation: Choose the OS and applications that satisfy your business requirements. Bear in mind that your existing network infrastructure may influence the decision-making process.

Compression

Image files are quite a bit larger than text files. For instance, a two-paged document created and stored as a word processing file might be 20 kilobytes. The same document, printed, scanned, and stored uncompressed might be as large as one megabyte. Because image files are so large, they are usually compressed. Compressed files do not require as much storage space on the computer system, and they can be transmitted over the network faster. However, they then must be decompressed for viewing and/or printing. The compression technique you use depends on whether the image is going to capture color or not. Although proprietary techniques exist, their use is not recommended for retention or preservation, as they will severely limit your ability to migrate to newer technologies. Also, you will lose information by the very act of compressing the image.

1. **International Telecommunications Union (ITU, formerly the International Telegraph and Telephone Consultative Committee — CCITT) Group III and Group IV.** These techniques are used to compress images of black-and-white documents. Each serves a different purpose. Group III is specified for analog transmissions (standard facsimile machines) and Group IV is specified for digital transmissions (computer networks, digital facsimile machines, etc.). While Group III compression ratios are generally up to 10:1, Group IV typically ranges from 15:1 up to 20:1. Therefore, Group IV compression creates smaller files.

Recommendation: Use CCITT Group IV compression. Use Group III as needed.

2. **Joint Photographic Experts Group (JPEG).**
   JPEG is designed for compressing either full-color or grayscale photographs. Compression ratios can range up to 100:1 using JPEG. Within JPEG, there are two broad categories of compression— *lossy* and *lossless*.

   - *Lossy* compression determines whether a color is useful (different enough from other colors to be visually perceived). If the color is useful, it is kept. If it is not useful, it is dropped or lost. Therefore, the decompressed image is not an exact replica of the original. In addition, each time you open, close, and reopen a lossy JPEG file, you lose more pixels and you have no control regarding what pixels get lost.
   - *Lossless* compression does not drop colors; however, the ratios are not as high as with lossy compression.

Recommendation: Carefully consider how the images will be used and their retention. If computer analysis is anticipated, lossy compression will not render 100% accurate results. If the retention is long-term, use lossless compression or save as a TIFF image (see **File Formats** below) so as to render complete and accurate results.

3. **Proprietary formats.** Proprietary compression techniques may offer benefits such as higher compression ratios. Thus, the image files may require less storage space and can travel through the network more quickly.

Recommendation: Use proprietary compression techniques for daily use only. Require additional non-proprietary compression for retention, preservation, and back-up copies.

File Formats

The file format is a structured container for information about each digital image and the image data. Information about the digital image files includes, but is not limited to: its name, width, length, resolution, and compression techniques. The computer requires this information to interpret the digital images. It is essential to use a non-proprietary image file format (e.g. TIFF) to ensure the ability to transfer successfully digital images between different systems or when the system is upgraded or modified.

A number of other file formats exist, such as Graphics Interchange Format (GIF), Joint Photographic Experts Group (JPEG), and Bitmap (BMP). These file formats are commonly used in conjunction with hypertext markup language (HTML) for Internet and intranet applications. Many systems or third-party graphics packages will convert images from one to another, although often with unpredictable results.

1. **Tagged Image File Format.** TIFF is the most commonly used format for digital imaging systems. The file contains a tag or header that stores information (metadata) about the file. The header contains information such as how the image is compressed. It is important to understand that not all TIFFs are alike. A TIFF from one vendor's system may not be compatible with another vendor's system. Some systems will allow TIFFs to represent single-page or multi-page documents. For example, 12345.tif could be a one-page document or it could be a 120-page document. In such systems, the operator chooses the setting (single-page vs. multi-page) before scanning. File errors can be a major problem with multi-page TIFFs. If a multi-page TIFF file is corrupt, every page in the document is rendered inaccessible. However, if a document is made up of single-page TIFFs and one of the TIFF pages is corrupt, then only that page is inaccessible. Users can still access the other TIFF pages. Therefore, multi-page TIFFs are not recommended for retention, preservation, or back-up.

Recommendation: Use single-page TIFFs for retention, preservation, and back-up.

2. **JPEG.** JPEG technology employs compression technology. In most cases, the software determines whether a color is useful (different enough from other colors to be visually perceived). If the color is useful, it is kept. If it is not useful, it is dropped or lost. Therefore, the decompressed image is not an exact replica of the original. Files saved or created in JPEG have a .jpg extension.

1. **Text files.** Using Optical Character Recognition (OCR) or Intelligent Character Recognition (ICR) software creates text files from image files. OCR creates computer readable text from a digital image. ICR is often used for forms recognition. The system is set up to recognize key fields containing specific information. ICR converts that information into text that can be imported into a database. Optical character recognition (OCR) is the process of converting an image of text, such as a scanned paper document or electronic fax file, into computer-editable text. The text in an image is not editable: the letters are made of tiny dots (pixels) that together form a picture of text. During OCR, the software analyzes an image and converts the pictures of the characters to editable text based on the patterns of the pixels in the image. After OCR, you can export the converted text and use it with a variety of word-processing, page layout and spreadsheet applications. If a document has been OCR'd, it can be searched by keyword thus increasing the potential use of the document. Utilizing character recognition software can potentially make your documents much more valuable depending on your requirements and the intended use of the document.

Recommendations:
- Carefully review the system's file format. Ask the vendor to provide proof that its files can be transferred to another vendor's system.
- Budget for programming fees when migration is necessary.
- All permanent images should be tiff, minimally Group IV (Group V is now available).
- Carefully evaluate how your document will be used and determine if OCR/ICR will be useful to you. If so, make sure to budget for the purchase of OCR capabilities.

Media

When designing a storage solution as part of a digital imaging system, you must consider several factors including retrieval-time requirements, records retention and preservation, and costs over time.

Fixed Storage Media

Fixed storage is so named because the media stays within the computer or drive. The imaging system uses the computer's hard drive, or magnetic disk, throughout the conversion, storage, and retrieval processes. The hard drive's performance, capacity, and reliability are critical elements to consider when selecting a hard drive. From time to time, hard drives fail. Therefore, back-ups are critical. A back-up contains copies of the information on the hard drive.

Random access memory (RAM) is also important. Imaging programs use RAM to process information and data. Imaging programs tend to be more intense users of a computer's memory, therefore, the more RAM, the better. In the computing world RAM is expressed in terms of megabytes such as 64 MB or 128 MB.

Another fixed storage option is RAID, redundant array of independent disks (also discussed on p. 9.) There are several levels of RAID. Each has different options and benefits. In general, RAID can be used to increase on-line capacity and speed retrieval while providing a good back-up strategy.

Recommendations:
- Select a hard drive that meets or exceeds the vendor recommendations.
- Select RAM that meets or exceeds the vendor recommendations.
- Consider RAID to increase on-line capacity and increase retrieval speeds.
- Adhere to routine back-up schedules.

Removable Storage Media

Removable media, as its name indicates, can be removed from the imaging system. Three general categories of removable media are optical disks, magnetic tape, and microfilm. Factors to consider when selecting removable media include capacity, life expectancy, and cost over time. Two important cautions about media life expectancy or shelf life are in order. First, the term refers to how long the physical media will last in perfect environmental conditions. When you read that a medium has an expected shelf life of X-number of years, be aware that the prediction is based upon laboratory-controlled conditions. While microfilm has been in business use since the 1920s, optical media are still relatively new. The actual life expectancy of a given medium has not been proven yet. Second, shelf life does not refer to how long the information on the media will remain viable. As a result, you should error test and annually sample the media. In addition, you should have migration strategies in place to ensure the viability of records whose retention periods are longer than five years.

1. **Optical disks.** Lasers are used to write data onto an optical disk. Optical disks, typically, have high storage capacity. There are three major categories of optical disks used in imaging systems:

- WORM (Write Once, Read Many) optical disks have large storage capacities, generally measured in terms of gigabytes (thousands of megabytes). While the medium is intended to be nonerasable, it is possible to destroy data on the CD. Data is written to the disk sequentially. To improve efficiency, writing can be done intermittently or in batches. Because data is written sequentially, it can be read back from the disk faster than the same data on a CD-ROM. The indexing information the software uses to retrieve the image can be deleted. This makes it virtually impossible for the average user to retrieve the image. However, the image is still on the WORM disk. Sophisticated computer sleuths can retrieve images whose indexes have been deleted.
- Magneto Optical (MO), erasable, or rewritable optical disks also have large storage capacities; however, images can be deleted from the disks. Since MOs allow multiple write sessions and erasures, the media tends to degrade quickly if many write/rewrite sessions are performed.
- CD-Recordable (CD-R) and CD-Rewritable (CD-RW) are essentially the same thing. A CD-R becomes a CD-ROM once data is written to it. Therefore, like WORM disks, data cannot be erased from CD-R disks. Typically, CD-Rs hold about 640 to 700 megabytes. Data is written to the disk spirally and is done in one session. In other words, if you only write 16 megabytes to the CD with a capacity of 700 megabytes, then you have 684 megabytes of wasted space on that CD. CD-Rewritables are slightly different. You can do multiple writing sessions and erase data. Since CD-RWs allow multiple write sessions and erasures, the media tends to degrade quickly if many write/rewrite sessions are performed.
- Digital Versatile Disks (DVD) are much like CDs. They come in both recordable and rewritable formats. DVDs can store up to 12.4 gigabytes of data. That is about 20 times that of a CD. Shelf life is the same as CDs. However, as of this writing, we are unaware of any imaging systems that use DVDs.

2. **Tape.** Magnetic tape is useful for infrequent retrievals and for back-ups. Tape is generally slower than optical media. Tapes can hold anywhere from one to ten gigabytes of data.

2. **Microfilm.** Like tape, microfilm can be useful for infrequent retrieval and back-ups. Some imaging systems can digitize existing microfilm and others can create microfilm from images stored on the system or on optical disk. Shelf life for microfilm produced under standards and stored and handled properly, is rated from 100 to 200 years or more.

Cost

Another inverse relationship to consider exists between storage capacity and cost per megabyte. For example, the faster the medium (e.g. magnetic media) the fewer records it can hold, and the more it costs per megabyte to store those records. Conversely, the more records a medium can hold, the lower the cost per megabyte, and the slower your access to the records becomes.

Recommendations:
- For retention and preservation, permanent records should be stored on microfilm, an eye- readable format.
- Once the records have met their retention period, the media should be physically destroyed. Therefore, records with similar retention periods should be stored on the same disk, tape, or reel.

If a record or record series is identified to be a part of litigation and that record or record series resides on a disk, tape, or reel that is scheduled for destruction, you should cease destruction related to those records until the case is resolved.

Please see our section, Phase III, System Implementation, to review the next steps and learn about indexing, labeling, storage, and access.