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 III: System Implementation

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. Specifically, Phase III, System Implementation discusses the details of indexing, labeling, storage, and storage conditions as well as revisiting legal issues and risk management.

Phase III: SYSTEM IMPLEMENTATION

Implementing an imaging system requires careful planning for both use of your resources and for the imaging system itself. Have a plan and budget in place for updating and operating the imaging system, identify legal issues regarding access to and integrity of your records in an imaging system, and have a plan in place for records storage and migration.

Staffing

Successful implementation of an imaging system requires the integration of many different internal and external resources. Internally, staff must be hired to operate and maintain an imaging system or existing staff must be reallocated to the project and trained to use the system. The system administrator must have knowledge ranging from system design and maintenance to daily operations. Imaging operators must have a general knowledge of computers and of imaging systems. They must also have a good working knowledge of the organization and the records it creates. Processing staff will be needed to prepare the documents for scanning. Document preparation consists of organizing the documents in a logical order, removing staples or paper clips, repairing tears, flattening wrinkles and creases, etc. This process is necessary to ensure that pages do not get caught in the scanner. The imaging system staff must have a good working relationship with the organization's technical support staff. The support of the technical staff is essential to the successful operation and maintenance of your imaging system. Technical staff can advise on compatibility issues, data integrity, and system upgrades.

Externally, the system vendor must be able to provide training, maintenance, and ongoing support in operating your imaging system. Both the imaging staff and the technical support staff must have a good working relationship with the vendor's technical support staff to ensure the smooth operation of your system.

Legal Issues Revisited

The laws governing the creation and use of records will affect your decisions. Statutes and administrative regulations may define how records are created, the media on which they may be stored, and who has what type of access rights—read and write access. Develop strategies to meet your legal requirements including compliance with your approved records disposition schedule. We strongly recommend consultation with your organization's legal counsel before implementing an imaging system. Whether imaged or not, the ability to use records as evidence for legal, audit, and other purposes depends on establishing their authenticity and reliability. First, your organization must be able to prove that a record keeping system is used in the normal course of business. Documenting the specifications of the imaging system, training staff in the operation of the system, ensuring the integrity of the records, and conducting random audits can accomplish this. In addition, any modification of the system should be documented and include information about what was done, when it was done, and who did it. Last, state agencies and local governments must provide access to public records as mandated in General Statute 132, the Public Records Act.

1. Imaging system documentation. Maintain a record of information about your system. Documentation is necessary for providing audit trails, for establishing legal admissibility of images, and for use of the system by future system operators. The system administrator, working with the vendor, is responsible for such documentation. This documentation should become a part of an approved records retention and disposition program and should contain, at a minimum, the following information:

- Policies and procedures for all aspects of system operation and maintenance, including procurement, file and document preparation for scanning, data entry, quality control, indexing, corrections, expungement, redaction, back-ups, security, migration, application of safeguards to prevent tampering, and unauthorized access and printing.
- All system equipment specifications.
- A description of all hardware and software upgrades to the system, including date of maintenance and version of software.
- Contact information for manufacturers and vendors.
- Technical operation manuals.
- User operation manuals.
- All policies and procedures related to access to and security of the records.
- Any changes made to the system or the process should be documented.

2. **Training.** You must be able to show that the imaging staff has been trained to operate the system. Also, you must be able to prove that they follow the normal record keeping practices and quality controls procedures established by your organization. Your human resources department may be able to help you develop training programs and guidelines. Such records should be included on an approved records disposition schedule.

3. **Integrity.** Integrity of electronic records refers to both the physical and intellectual integrity of the information. Maintaining the physical integrity concerns two issues:

- *The actual condition of the media storage device*—questions such as: has the media deteriorated, been scratched, or has it been exposed to extreme temperatures need to be assessed.
- *The reliability of the record after compression or migration*— after such events, has the appearance of the document been altered?

4*.* The intellectual integrity of a record is based upon the authenticity or truthfulness of the information within the record. A system should be in place for electronic records that validates access procedures and documents modifications to the records over time. To establish the integrity of the records, at a minimum, the following standards should be in place:

- o The identity of a record's creator is verified.
- o Quality control on the records is conducted.
- o Permission to read, write, and delete files is appropriately restricted.
- o Periodic system audits are conducted.
- o Data transmission includes data error checking and correction.
- o Data are regularly backed up.
- o Updates and changes to the system are documented.
- o Complete the *Self-Warranty* (*www.ah.dcr.state.nc.us/e-records/manrecrd/swf.htm*) document found in the publication *North Carolina Guidelines for Managing Public Records Produced by Information Technology Systems_* (*www.ah.dcr.state.nc.us/e-records/manrecrd.htm*) issued by the Department of Cultural Resources.

5. **Auditing.** We strongly recommend that you conduct periodic and random audits of the imaging system to ensure that the system is operating within the established records management guidelines and that the data remains viable. Prior to beginning an imaging project, management should establish both acceptable error limits and procedures for correcting systems that do not meet those limits. Documentation of audits should be kept with the imaging system documentation and should become a part of an approved records disposition schedule and part of your overall project planning committee.

6. **Access.** Unless otherwise closed by statute, the N.C. Public Records Law requires the public records of state and local governments to be made available to the public in any reasonable format requested by the public. In addition, General Statute 132-6.1 (b) requires that all databases be indexed. You can find guidelines for indexing at *http://www.ah.dcr.state.nc.us/e-records/pubdata/default.htm.* To ensure the records are easily accessible throughout their retention period for internal, as well as public use, the record keeping system should:

- Provide for clear identification of the record.
- Permit easy and timely retrieval of individual records and record series.
- Retain the records in a usable format.

7. **Expungement, Redaction, Encryption Capabilities.** Government offices should have in place a strategy to guarantee that material exempted from disclosure is not made available to the public. Imaging systems should have the capability to expunge images and index entries and to redact confidential portions of images or indexes when required by law. You cannot refuse access to a file because of an inability to block certain records or parts of certain records. The potential need for expungement, redaction, and encryption capabilities must be assessed on the front end and discussed with vendors when planning for long-term usability of an imaging system. Explanation of procedures for expunging information on WORM optical systems may be found in **ANSI/AIIM TR28-1991**, *The Expungement of Information Recorded on Optical Write-Once-Read-Many (WORM) Systems*.

Indexing and Labeling

Records constitute a corporate resource. To assure that the resources remain accessible, an indexing database that facilitates efficient retrieval, ease of use, and up-to-date information about the images stored in the system should be developed. To ensure that the value of the information in the record is maintained and can be retrieved from an electronic record, three elements must be present:

- *Content* refers to the subject matter of the record, while
- *Structure* focuses on use of fonts, headings, spacing, etc., as part of the meaning of the content, and
- *Context* refers to the relation of one record to other records.

The content, structure, and context a system captures in the indexing and labeling components are commonly referred to as metadata. Metadata is essentially "data about data," which describes an information resource. It is important to capture the image metadata to allow future users to discover the context in which the record was produced and to permit the owner of the record to manage it. There are several national and international organizations working toward establishing indexing standards. There are many metadata schemes, including but not limited to, the Dublin Core (DC), Encoded Archival Description (EAD), Text Encoding and Interchange (TEI), and Machine Readable Cataloging (MARC), which are being used at various institutions today. If you choose not to adopt one of these standards, then at a minimum, the metadata you capture should include:

- Title – Of the document
- Creator/Author – Including other contributors
- Date – Date of document creation/creation of image
- Unique identifier – Such as a coding system for different document types
- Format –To include operating system, software configurations, and versions thereof
- Definition – A statement that clearly represents the concept and essential nature of the record

- Rights and Security – Indicate if special authority is needed to access the information, and who has that authority
- Data type –Indicates if it is a written document, photograph, etc.
- Keywords – use a controlled vocabulary dictionary or list if possible. Several industries have published their thesauri on the World Wide Web.
- Comment – For additional information

Whatever media you choose for your imaging system, make sure that you clearly label all disks, tapes, or other removable storage containers with enough information that the contents of the storage instrument can be determined years later, by different staff. At a minimum, the label should include:

- Identifiers— including creator, date created, division or agency where created
- Name of agency, unit, and division that is responsible for the records on the disk
- Hardware, operating system, and software required to access the index or information on the disk
- Encoding standard and version
- Level of security or restricted access
- Sequential number or other specific identifier that identifies the disk in the series of disks used by the system
- Identification of the disk as master or back-up storage copy
- Retention dates of the information on the media
- Data classification: If it is stored off-site, is the data confidential, who can access it, who can read the data, and are there different levels of confidentiality, e.g. are parts of the record public records while parts of it are confidential?

If the disk or other format is too small to include all of the information on the label, then establish a coding system that can be linked back to an index that holds all of the vital information. Documentation relating to the coding system and index must be maintained for as long as it relates to any labeled storage medium that utilizes that coding system, and should become a part of your organization's approved records retention and disposition schedule.

If you choose to use a marker on your CDs, use a water-based permanent marker. Do not use a Sharpie® pen, or one like it, on the CD. There is a modest amount of anecdotal evidence that the use of solvent-based ink markers (Sharpies® use an alcohol-based ink), particularly on CD-R/RWs without a protective coating and CD-R/RWs kept in a warm to hot environment can lead to long-term penetration of the ink to the data layer with resulting damage to the data. The standards organization for CD-R/RW media also advises that people use markers with permanent, water-based inks.

Batching

Once you have determined what types of records you want to scan, you need to address the issue of whether you will create a file for each image or for a file for each document. PDF files can contain multiple page as well as multi-page TIFF files. Multi-page files can be created as they are scanned, or you may choose to scan the images in batches, then assemble the images into multi-page files in a second step. After the images have been created, it is sometimes unclear where one document ends and the next starts. Inserting a "header" page in the stack of pages before the beginning of each document will help. This header page can simply be a blank page or can be prepared ahead of time to include metadata such as the document's title, etc. If each page of a document will be kept as a separate file, then some method should be used to record which images comprise which documents. In this case scanning should be done in small batches so that the images and document indexes can be checked for quality and accuracy and any rescans can be done without having to search through many files.

Quality Control

It is essential to the success of the imaging project that the image produced is the best image possible. Imaging operators should test the system at least at the beginning of each series of similar documents by scanning a test target. Test targets can be obtained from the Association for Information and Image Management, **AIIM TR 37-1996**, *Compilation of Test Targets for Document Imaging Systems*. The image operator should also monitor the scanned images. If the quality is poor, then the system operator must rescan the documents or pages and possibly adjust the scanning parameters. The frequency of quality control monitoring should be determined by the type of the documents, but should occur at least once during each session. The accuracy of the index must also be verified through visual inspection by a second staff member of each index entry following either entry of terms or created through optical or intelligent character recognition. The system should also include the ability to rescan and correct indexing errors before the image and/or index is written to optical media. Quality control issues must be raised with vendors during the selection process and be considered when planning for time and staff budgeting. Since original records are usually destroyed once reformatted, the importance of image and index quality control must not be underestimated. Information regarding the establishment and use of procedures for the ongoing control of quality within an electronic imaging system may be found in the ANSI/AIIM MS44-1988 (R1993), *Recommended Practice for the Requirements and Characteristics of Documents Intended for Optical Scanning* (**ANSI/AIIM MS52, 1991**)

In order for images to be easily retrieved, an index needs to be created from the scanned images. This can be done by manually by entering descriptive information into a database for each document, or automatically by first OCRing each document, then, with the aid of a full-text database software application, having the computer add each word to the index. The type of document you scan, will help you determine whether or not you will use a manually created index, an automatically created or both. In some cases such as forms, the documents can be designed so that a portion or the form when scanned and then OCRed can be used to supply specific manual-type indexing fields, while other parts are OCRed to supply full text databases. Some documents do not OCR well, particularly handwritten documents. Those images will have to be manually indexed. By keeping the batch numbers relatively small, e.g. 100 images or less, it will be easier to perform quality control and to add indexing terms. These issues should be discussed as earliest as possible, as they will determine in large part how documents will be accessed, and which software systems will be employed.

Once the images have been captured and placed on the storage medium, you will need to monitor the maintenance of the storage medium to ensure the records are accessible and their integrity intact. The type of storage medium you choose will determine how you will set up your quality control guidelines for storage. You should check with your vendor for recommendations for winding magnetic tape. A schedule and percentage base for random sampling of disks for signs of deterioration and corrupt files should be in place.

Error detection and correction is the ability to predict the point at which an optical storage disk is no longer readable. This ability is critical if the recopying of disks is to take place at the appropriate time. Refer to **ANSI/AIIM MS 59-1996,** *Media Error Monitoring and Reporting Techniques for Verification of Stored Data on Optical Digital Data Disks.*

Security Copies

Producing a security copy of the output serves to protect your organization in the event the working copy becomes damaged, lost, or destroyed. We recommend storing security copies in an off-site location, with strict control access. Security copies should be properly labeled with information to include date, system, and software used, and any existing restrictions on access. For more information on security copies, please see the *Security Back-up Files as Public Records In North Carolina: Guidelines for the Recycling, Destruction, Erasure, and Re-use of Security Backup Files* (*www.ah.dcr.state.nc.us/records/BackupsProcedsfinal020822.pdf*)

Environmental Conditions and Storage

You should adhere to the media manufacturer's vendor's recommendations for specific environmental conditions in which the media should be stored. These recommendations should include information relating to ideal temperature, humidity, and storage orientation. Both the working and security storage locations should adhere to the recommended environmental specifications. If the recommended storage conditions cannot be met, you should ensure that the environmental conditions of the storage location are stable and do not or rarely fluctuate. Technical experts recommend a stable environment with a temperature between 65 and 75 degrees Fahrenheit and a relative humidity between 30 and 50 percent. Adverse storage conditions, especially high humidity, can cause rapid deterioration of the media. All media should be kept free of condensation. In addition to the guidelines for the physical storage of the media, it is vital that guidelines be in place for the storage and maintenance of the records on the media. As a rule, electronic media are less stable than paper. Dust, debris, and fingerprints affect optical disks. To protect disks from warping they should not be subject to pressure and should be stored in an upright position when not in the disk drive.

In addition, unlike paper or microfilm, digital images are not readable without the assistance of computer hardware and software. Because digital storage media are not permanent, and rapid changes in computer technology are constant, a conversion strategy for retaining and retrieving stored information should include refresh and migration guidelines.

1. **Refresh**. Refreshing involves copying information from one storage medium onto a newer more stable storage medium. For example, transferring the information stored on a three-year-old WORM disk to a new WORM disk. It is recommended that the media should be refreshed every three to five years, depending on the media type.

2. **Migration**. Migration is the process of transferring digital information from one storage format to another, or from one generation of hardware and software to the next. For example, transferring data from 5¼-inch floppy disks to a CD-ROM, or upgrading from imaging software version 3.0 to version 4.0. Currently, migration is the best practical means for retaining and retrieving data over time. Migrations have to be carefully planned, executed, and audited to ensure against data loss. Though migration is a time consuming and expensive process, with proper guidelines in place, the costs can be minimized. You should also consider copyright laws when creating a migration strategy. Currently, the same copyright laws apply to both printed materials and electronic media. Since most software is copyrighted, make sure you negotiate with vendors for the rights to the data you have created and to have the ability to migrate necessary software components to be able to access your data. In some cases, you may lose the right to use an earlier version of software once you upgrade to a newer version. Another option for avoiding copyright infringement is to use open systems architecture and know the licensing agreement. Migration and long-term usability planning must also include consideration of continuing information retrieval requirements.

Retention Schedules

The fact that you are storing information in an electronic format does not by itself impact the decision as to how long you will retain that record. It is the information contained on the medium, whether paper or electronic, which must be considered. General Statute 132 requires governmental offices to retain and to make available certain types of records for a determined period of time. Each record series should have a defined retention schedule. If you are unfamiliar with the retention schedules for your agency, contact your records analyst. A listing of records analysts can be found at *http://www.ah.dcr.state.nc.us/sections/archives/rec/analysts.htm* or *http://www.ah.dcr.state.nc.us/sections/archives/rec/contact.htm* For the purpose of this document, retention periods will be broadly defined as:

- Short-Term – creation to five years
- Medium-Term – six to nine years
- Long-Term – ten years and longer

Guidelines for each period will need to be in place. You will need to determine how often you will refresh and/or migrate the records, if different quality control measures should be used for different retention periods, and if you will store all the records in one location or separate locations based on the retention periods.

After a record has met its retention period, then your agency should proceed with the disposition of the record. If you will be destroying the records, it is important that the plan for destruction is in place before you implement the imaging system. Depending on the type of storage system you choose, this may mean destroying disks or magnetic tapes. By placing like records on the same storage disk you will be able to destroy records with the same retention periods. Otherwise, you would have to migrate records that are not eligible for destruction to another medium first. You must develop standards that ensure the record is completely destroyed once the retention period has been met.

Preservation

Preservation in the digital world means ensuring continuing access to high quality, eye-readable original source documents. Permanently valuable records maintained in an image based record keeping system will require a well-developed migration strategy and the most diligent efforts to keep them accessible. As much contextual information as possible must be captured to ensure the survival of the historical meaning. At this time we recommend microfilm for the retention of permanent records. With the proper equipment, digital images can be scanned to and created from microfilm. Microfilm's longevity is proven, it does not require software to read, and it is a cost-effective alternative to refreshing optical or magnetic media.

Risk Management

As part of your initial planning phase, a comprehensive disaster management plan should be developed. The plan should include standard back-up and recovery procedures, as well as quality control and storage procedures such as those mentioned previously. In the case of a natural disaster, having off-site copies of records may be the only answer for recovering data. Tests of both the prevention and data recovery guidelines should be conducted on a regular basis.

Back-up procedures and disaster recovery plans should be in place with specified provisions of the imaging system. Detailed information on back-ups and disaster recovery should be obtained from vendors. Back-up expense and complexity can vary depending on the type of media and the amount of data to be stored and must be considered during the planning and selection process. A regular schedule of back-ups should be instituted for all data on the system, including indexes. Control access to the tapes. Security copies should be properly labeled with information to include date, system, and software used, and any existing restrictions on access. It is preferable that security copies be stored off-site, in an area with stable environmental conditions and with adherence to the manufacturer's specifications for the storage of the media, whether magnetic or optical. Information regarding optical media storage may be found in **ANSI/PIMA IT9.25-1998**, *Imaging Materials-Optical Disc Media-Storage*.

Final Recommendations

The Imaging Guidelines provide an introduction to some of the many issues that you will face when planning, selecting, and implementing an imaging system. The guidelines are intended to be used as a starting point for the decision making process. Because of the number of issues involved, no single set of guidelines will be able to answer all of your questions. Regardless of the issues that you will face, project planning is the key to the successful selection and implementation of an imaging system. Please contact your analyst to consult with them about your imaging projects and concerns. If you have additional questions, please contact the Office of Archives and History, Archives and Records Section at *records@ncmail.net* or visit our website at *http://www.ah.dcr.state.nc.us/records/*