

OHIO ELECTRONIC RECORDS COMMITTEE

http://www.ohiohistory.org/ohiojunction/erc/

GUIDELINES FOR DIGITAL IMAGING OF PUBLIC RECORDS

SCOPE	2
INTENT AND PURPOSE	2
Introduction	2
1.0 Project Planning	3
1.1 Prior to selecting a digital imaging system, conduct a records and workflow analysis to determine and document existing and planned agency information needs	3
1.2 PRIOR TO SELECTING A DIGITAL IMAGING SYSTEM OR SERVICE PROVIDER, CONDUCT A COST BENEFIT ANALYSIS TO DETERMINE THE COST JUSTIFICATION OF A SYSTEM PURCHASE OR PURCHASE OF OUTSOURCE SERVICES. COMPARE THE COSTS OF YOUR CURRENT OPERATION WITH THE COSTS OF THE NEW SYSTEM, INCLUDING ADDITIONAL BENEFITS OF ELECTRONIC RECORDS.	
2.0 Technical Specifications and Selection	5
2.1 Require open system architecture for digital imaging applications or require vendors to provide a bridge to systems with non-proprietary configurations	5
2.2 AUDITED PROCESSES SHOULD BE SPECIFIED BASED ON LEGAL REQUIREMENTS FOR RECORD INTEGRITY, WIT THE OPTIONAL USE OF NON-REWRITABLE MEDIA	
2.3 Use a non-proprietary digital image file format. If using a proprietary format, provide a bridge to a non-proprietary digital image file format.	7
2.4 USE INTERNATIONAL TELECOMMUNICATIONS UNION (ITU) GROUP 3 AND GROUP 4 COMPRESSION TECHNIQUES OR HAVE THE VENDOR PROVIDE A BRIDGE TO THESE TECHNIQUES	8
2.5 When determining document scanning resolution, consider data storage requirements, document scanning throughput rates, and the accurate reproduction of the image. Validate vendor claims using a sampling of the agency's documents	8
2.6 Select equipment that conforms to a standard methodology for media error detection and correction. The system should provide techniques for monitoring and reporting verification of the records stored on a digital media, and the system administrator should	
ACTIVELY FOLLOW THE STATUS OF THE MONITORS	
2.8 Use an indexing database that provides for efficient retrieval, ease of use, and up-to-date information about the digital images stored in the system. The indexing database should be selected after an analysis of agency operations and user needs.	<u> </u>
3.0 System Implementation	0
3.1 Assign a permanent staff member as systems administrator and require the vendor to provide a project director during the installation and training periods	
3.2 ESTABLISH OPERATIONAL PRACTICES AND PROVIDE TECHNICAL AND ADMINISTRATIVE DOCUMENTATION TO ENSURE THE FUTURE USABILITY OF THE SYSTEM, CONTINUED ACCESS TO LONG-TERM RECORDS, AND A SOUND FOUNDATION FOR ASSURING THE SYSTEM'S LEGAL INTEGRITY	
3.3 Institute procedures to ensure quality and integrity of scanned images	

OHIO ELECTRONIC RECORDS COMMITTEE

http://www.ohiohistory.org/ohiojunction/erc/

3.4 DESIGN BACKUP PROCEDURES TO CREATE SECURITY COPIES OF DIGITIZED IMAGES AND THEIR REINDEX RECORDS.	
3.5 Provide adequate environmental conditions for digital media	13
3.6 The retention and disposal of digital images and corresponding electronic records incorporated into the agency retention schedule	
3.7 A DISASTER PREPAREDNESS PLAN IS NECESSARY FOR A PROPERLY CREATED SYSTEM OF IMAGE AI BACKUP, STORAGE AND RECOVERY	
3.8 BUDGET ANNUALLY BETWEEN TWENTY AND TWENTY-FIVE PERCENT OF THE ORIGINAL SYSTEM A COST FOR UPGRADES, TRAINING, AND MAINTENANCE	
4.0 Archiving and Long-term Maintenance	15
4.1 Provide specific plans for creating and sustaining digital images that will be retain than 10 years.	
4.2 SCANNING SPECIFICATIONS AND FILE FORMATS	15
4.3 DIGITAL IMAGE METADATA	16
4.4 System Information and Maintenance	16
4.5 Copying and Refreshing Media	17
4.6 Sustainability	17
4.7 REFORMATTING AND MIGRATION	17
4.8 Scheduling	18
Definitions	18
IF YOU NEED ASSISTANCE	20
BIBLIOGRAPHY	20

SCOPE

These guidelines apply to State of Ohio executive agencies as well as local government entities.

INTENT AND PURPOSE

The intent of these guidelines is to provide and explain requirements, guidelines and best practices for digital document imaging projects that meet the criteria for records as defined by the Ohio Revised Code.

Introduction

Public officials are responsible by law for ensuring that their records are protected and accessible. This responsibility applies regardless of the records' form or storage media.

These guidelines are for public officials in the design of a digital imaging system. The guidelines are advisory and are not intended to be a requirement. National technical standards, established practices, and research gathered from literature form the basis for

OHIO ELECTRONIC RECORDS COMMITTEE

http://www.ohiohistory.org/ohiojunction/erc/

these guidelines. They are designed to identify critical issues for public officials to consider when designing, selecting, implementing, operating and maintaining digital imaging technology. These issues are especially important for systems used for mission critical records.

Digital document imaging is defined as the conversion, storage, and distribution of information displayed but not directly modified by a computer.

These guidelines have been divided into four sections:

- 1.0 Project Planning. It is recommended that everyone involved in a proposed imaging project read this section to determine if an imaging project is appropriate for your agency's records. Those involved in a proposed imaging project may include records managers/officers, records commissions, office administrators, and information technology staff.
- 2.0 Technical Specifications and Selection. This section is geared toward information technology staff and anyone who will be working directly with vendors to design a system that will meet your agency's needs. It is meant to highlight the recommended technical specifications that should be present in an imaging system.
- 3.0 System Implementation. When an imaging system has been selected, certain procedures need to be followed to implement and maintain the system and the information that it contains. This involves having the proper information technology staff and long-term budget to maintain the system as well as documentation of the system, back-up procedures and disaster recovery plans.
- 4.0 Archiving and Long-term Maintenance. This section discusses additional measures that should be taken into account if your agency is planning on maintaining digital images for longer than ten years without another copy of the records existing on an eye-readable media.

Within these sections, the recommendations are listed in order of their implementation.

These guidelines are based partly upon the work done by the Alabama Department of Archives and History and published in their technical leaflet "Guidelines for the Use of Digital Imaging Technologies for Long-Term Government Records in Alabama." Their recommendations formed the basis for this work and are included here with their permission. We appreciate their expertise and generosity.

1.0 PROJECT PLANNING

1.1 Prior to selecting a digital imaging system, conduct a records and workflow analysis to determine and document existing and planned agency information needs.

An analysis of an existing workflow is a crucial first step to determine the need for an imaging system. A <u>workflow</u> provides an opportunity to reengineer the business process for operational efficiency. It identifies areas within the process that are inefficient and redundant. Work processes can be significantly impacted by an imaging system because personnel create, use, distribute, and maintain records in different

OHIO ELECTRONIC RECORDS COMMITTEE

http://www.ohiohistory.org/ohiojunction/erc/

ways. The detail and complexity of workflow analysis and reengineering will affect your project schedule and cost justification since such analysis is time consuming.

Applications for a single process may be appropriate, negating the need for a workflow analysis. The volume of records, number of retrievals, length of time the records are active, and the need for multiple and/or simultaneous access will determine the need for such applications.

A <u>record analysis</u> determines which records are best suited for a digital imaging application. Such applications are primarily designed to enhance access to information that is frequently used, provide simultaneous access to information, and/or shorten the time needed to access information. Consequently, those records best suited for digital imaging applications are records that are used or viewed by multiple people, have multiple points of access or index fields, and require rapid access. Digital imaging applications are not designed to save storage space, even though this is often used as a selling point for such a system. Any cost-benefit analysis will demonstrate that it is less costly to rent storage space than it is to create and implement a digital imaging system. However, storage space costs may be a factor to consider when doing a cost-benefit analysis.

If the intent is to destroy the paper records and to maintain only the digital images, we strongly recommend imaging records with retention periods of less than ten years. This will avoid the costly and perilous task of trying to preserve digital images for long periods of time. Since digital images and the technology surrounding them are in a continuous state of change, any record in digital format cannot be considered stable and capable of remaining reliable, authentic and accessible over any long time period. Therefore, it is our recommendation that digital imaged records of permanent or long-term value (meaning greater than 10 years) be maintained in either paper or microfilm formats in addition to digital formats.

While the above recommendation and the reasoning behind it are valid, there will be times when digital images must be maintained permanently. The fast pace of technological change may make it necessary to migrate imaged records that have short-term value to avoid having to return to analog (paper or microfilm) records. While this method may retain the information, it is not feasible operationally. The recommendations in Section 4.0: "Archiving and Long-term Maintenance" address these issues.

The State Archives of the Ohio Historical Society can assist in analyzing an agency's record keeping system. Agencies that are considering imaging records with permanent retention periods should contact the State Archives for a system and records analysis to determine if maintaining the records in an eye-readable format is also necessary.

1.2 Prior to selecting a digital imaging system or service provider, conduct a cost benefit analysis to determine the cost justification of a system purchase or purchase of outsourced services. Compare the costs of your current operation with the costs of the new system, including additional benefits of electronic records.

Cost justifying a digital imaging system allows a financial comparison between the current and proposed record-keeping systems to help in making a procurement

OHIO ELECTRONIC RECORDS COMMITTEE

http://www.ohiohistory.org/ohiojunction/erc/

decision. The cost justification goal of a digital imaging system is to offset the cost of the equipment and software by reducing storage costs and increasing productivity through the improvement of work processes.

A typical cost justification includes the following:

- A study of current operations
- Potential changes/improvements to current operations
- Proposed system architecture
- Equipment pricing
- Financial measures such as payback period, rate of return or net present value

To determine the benefit derived from the new system, consider the following current costs:

- *File creation:* Includes file folder, labels, paper, file tracking system, labor to create files and add to system.
- File maintenance: Includes filing equipment, floor space for files and access, labor to retrieve/copy/refile documents, time waiting for information, cost of misfiles, cost of lost files.
- *File disposition:* Includes boxes for off-site records center storage, labor to move from active to inactive storage, destruction (recycling, pulping, shredding).

Below are links to two spreadsheets that will assist in the cost benefit analysis. The first allows the user to compare the costs of the current operation with the expected costs of the imaging operation. The second provides information to compare in-house vs. outsourcing costs. Certain formulas have been included in the spreadsheets. You should leave the formulas intact, but change the individual cost numbers to reflect your present experience.

Cost/Benefit Analysis Paper vs. Imaging (image analysis in-house vs. outsource 2003-15-03.xls) Cost/Benefit Analysis In-house vs. Outsourcing (ROI Analysis worksheet 03-15-03.xls)

2.0 Technical Specifications and Selection

2.1 Require open system architecture for digital imaging applications or require vendors to provide a bridge to systems with non-proprietary configurations.

Although the term <u>"open systems architecture"</u> may be defined in various ways, public officials should follow a system design approach, permitting flexibility in future upgrades without incurring extra costs. Open systems architecture permit future component upgrades with minimal degradation of system functions. An open system architecture allows the system to be upgraded over time without significant risk of losing records and unnecessary cost.

Open systems also support the importing and exporting of digital images to and from other sources. A key factor in achieving open systems architecture is the adoption of

OHIO ELECTRONIC RECORDS COMMITTEE

http://www.ohiohistory.org/ohiojunction/erc/

<u>non-proprietary</u> standards. The flexibility of open systems architecture helps enable long term-term records to be accessed and transferred from one hardware or software platform to another. With non-proprietary standards, an organization is able to examine multiple vendors when upgrading their system, with the price saving opportunities such competition brings.

2.2 Audited processes should be specified based on legal requirements for record integrity, with the optional use of non-rewritable media.

From a legal standpoint, all record solutions (paper or electronic) must be able to provide documentation certifying the integrity of the records. This means that the record is a true representation of the original referred to as "best available copy". In the case of a digital imaging system, it can be represented by the original scanned record if it exists, or a printed or displayed image from the system. Certification of record integrity can be achieved through the following, either separately or in concert:

- Implementation of media that allows the record to be written permanently and does not permit deletion or alteration. This is referred to as WORM (Write Once Read Many), and can take the form of optical disk (including CD-R, DVD-R, etc.), magnetic tape or magnetic disk. This can be enforced through the media itself, firmware "rules" at the hardware level or the software involved. If a record is no longer needed, software may allow the pointer to the data to be disabled preventing normal access. Because the data cannot truly be deleted it may remain accessible by other means and, therefore, presents a potential liability.
- Regardless of the media used, an audited process that ensures that the record represented has not been altered from its original form should be specified. This audit process is often called an <u>audit trail</u>. The rationale is that it is unlikely the media itself would be used in a legal proceeding, and so the evidence would be a printed copy, along with an affidavit certifying the authenticity of the copy. This means an audited process is required to provide the basis on which the affidavit is issued.

Features within the digital imaging system to provide evidence of record integrity include:

- Architecture to keep all annotations logically separate from the image itself, usually as a separate "layer". The record can then be printed or viewed with or without the annotations, while the record itself is not changed by the user.
- Workflow should be designed so that records cannot be deleted once entered to ensure a complete set of records within the system.
- Security to ensure the record cannot be directly accessed by the user.
 The digital imaging system acts as an abstraction layer.

System administrative personnel may require a higher level of administrative access to the system than an ordinary user. In order to assure system integrity, such high level system access should be limited to as few people as possible. A list of all current and past authorized users along with their privileges and responsibilities should be

OHIO ELECTRONIC RECORDS COMMITTEE

http://www.ohiohistory.org/ohiojunction/erc/

maintained as part of the system security procedures. This list should be reviewed regularly to ensure the timely removal of authorizations for former employees and the adjustment of clearances for workers with new job duties.

The process to request a record for a legal matter would involve printing or accessing the record directly from the system with the possible use of a witness to testify to the affidavit. The detail of these processes, and the possible use of non-rewritable media, is a discussion to hold with your legal representatives. Their decision will determine the technical requirements. For more information on record integrity, also known as creating reliable and authentic records, please see the "Ohio Trustworthy Information Systems Handbook."

2.3 Use a non-proprietary digital image file format. If using a proprietary format, provide a bridge to a non-proprietary digital image file format.

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

American National Standards Institute (ANSI)/Association for Information and Image Management (AIIM) MS53-1993, Standard Recommended Practice - File Format for Storage and Exchange of Images - Bi-Level Image File Format: Part I details a standard definition for file formats.

Despite the existence of a standard, there is not an agreed upon, industry-wide image format standard. Many digital imaging systems use the Tagged Image File Format, or TIFF. Because different versions of TIFF exist, there is still no absolute guarantee that images can be transported seamlessly from one system to another. Comprehensive documentation of the digital image file format, including TIFF, is recommended. Generally, TIFF files are very large and are not used on the web. Instead they are used as a Digital Master, from which other formats (derivatives) are created for use on the web.

Although TIFF has been considered the best practice Digital Master for several years, Adobe's Portable Document Format (PDF) is also sometimes utilized as a Digital Master file. There are concerns about the proprietary nature of PDF files; however, it can be argued that PDF has become a de facto standard for creating and maintaining digital images.

Current best practices would indicate that for images that are not going to be maintained more than ten years, PDF should be an adequate format. However, digitization best practice continues to advocate TIFF as the best option for long-term access and maintenance.

If you are creating images that will be maintained long term, creating a high quality Digital Master image with appropriate metadata is vital to the longevity of the image. For more information, please see <u>Section 4.0: "Archiving and Long-term Maintenance."</u>

OHIO ELECTRONIC RECORDS COMMITTEE

http://www.ohiohistory.org/ohiojunction/erc/

A number of other file formats exist, such as Progressive Network Graphics (PNG), Multi-resolution Seamless Image Database (MrSID), 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.

2.4 Use International Telecommunications Union (ITU) Group 3 and Group 4 compression techniques or have the vendor provide a bridge to these techniques.

The large file sizes of typical scanned documents require digital image compression to support data transmission and to promote storage efficiency. Most imaging systems today use some sort of compression algorithms to reduce image file sizes. Standard compression techniques are instrumental in ensuring a migration strategy for records needed for long-term use. The most commonly used file formats are TIFF with either Group 3 or 4 compression. Group 3 compression is most commonly found in fax machines and Group 4 in digital imaging scanning software.

There are two issues that are of greatest importance in deciding on image file format. The format should not be proprietary and it should provide a "lossless" compression algorithm. TIFF Group 4 meets those criteria.

There are other file formats that may meet this standard in the future (like PNG), but they are not as universally used or as readily portable as of January, 2003.

2.5 When determining document scanning resolution, consider data storage requirements, document scanning throughput rates, and the accurate reproduction of the image. Validate vendor claims using a sampling of the agency's documents.

A digitized image consists of black and white dots or picture elements called pixels and are measured in dots per inch (dpi). Generally, the higher the number of dpi, the higher the legibility of the reproduced image. Images scanned at higher dpi rates use more storage space on the disk and may require longer scanning times. The selection of scanning density involves a trade-off between image clarity, storage capacity, and speed. When selecting a scanner, ask the vendor to perform a quality test on a broad sampling of documents at various dpi settings so that an appropriate end-to-end throughput rate and resolution can be determined.

For good quality images use a scanning density of 300 dpi. A higher scanning density is appropriate for deteriorating documents and documents with a visual element such as engineering drawings, maps and documents with background detail. If Optical Character Recognition (OCR) is being considered, the resolution should be tested against the intended OCR engine to ensure optimum recognition levels. Higher resolutions can result in lower recognition rates, as extraneous data is added to the image. The display resolution of the inspection/verification monitor and printer should match the scanning density of the document scanner. When scanning continuous tone images, such as photographs, maps, and illustrations, use gray scale or color imaging technology.

2.6 Select equipment that conforms to a standard methodology for media error detection and correction. The system should provide techniques for monitoring

OHIO ELECTRONIC RECORDS COMMITTEE

http://www.ohiohistory.org/ohiojunction/erc/

and reporting verification of the records stored on a digital media, and the system administrator should actively follow the status of the monitors.

Digital imaging technology uses two methods within the <u>Error Detection and Correction (EDAC) system</u> to minimize digital image recording and retrieval errors. The first method uses error correction codes to detect and correct data read errors automatically. The second employs correction code software to determine if and when the utilization of error correction codes is approaching a critical point. Monitoring the error correction status information provides an audit trail to measure the progress and degree of disk degradation. Tracking error correction trends will indicate an appropriate timetable for recopying disks.

The Association for Information and Image Management's (AIIM) Standards Committee has developed a standardized methodology for reporting the error rate data to the operating system for user evaluations. ANSI/AIIM MS 59-1996, Media Error Monitoring and Reporting Techniques for Verification of Stored Data on Optical Digital Data Disks, describes these standards, assuming optical media is used.

2.7 Specify that data is verified at the point of writing to the storage media. If this is not feasible, your quality assurance procedures must be specified.

Due to the critical nature of the records involved, and the effort and difficulty involved in retrieving the original paper record in the event of incorrectly written data, it is recommended that the record data be verified at the write point. This is dependent on the capability of the media and hardware.

- Data written to storage media via the <u>Small Computer System Interface</u> (<u>SCSI</u>) can use the <u>"Write and Verify"</u> command, which confirms all data once written.
- If there are no hardware or firmware options to verify data, it can be done at the software level of the imaging system.
- If none of this is feasible, proper quality assurance procedures can minimize delays and additional steps to rescan documents in the event of incorrect data. This can take the form of prompt viewing of documents by personnel downstream in the workflow, or spot checks of records as required.
- 2.8 Use an indexing database that provides for efficient retrieval, ease of use, and upto-date information about the digital images stored in the system. The indexing database should be selected after an analysis of agency operations and user needs.

Reliable access to scanned images depends on an accurate, up-to-date index database. Indexing a digital image involves linking descriptive image information with the header file information. Normally, index data is manually key-entered using the original documents or the scanned images, either at the time of image capture or later in the production process. Index data verification, in which data base entries are compared with the original source documents for completeness and accuracy, is crucial because an erroneous index term may result in the inability to retrieve related images.

There are other options available to increase indexing efficiency. Bar coding and "Match and Merge" are both options that allow for increased indexing speed. Match and

OHIO ELECTRONIC RECORDS COMMITTEE

http://www.ohiohistory.org/ohiojunction/erc/

merge requires a unique data element that can be used for the "match". If there is no unique data element, match/merge is impossible. In most cases you will need to create a delimited ASCII text file to do the match/merge. Bar coding, for example, is helpful when scanning very large case files. Bar coding can provide single or multiple index fields. Most scan software today will accommodate adding index values from barcodes in the scan process. The system reads the barcode on the first page and then fills in the index fields for that document with the index values represented by the barcodes, until it reads a barcode that indicates a new document. Discuss your agency's business practices as well as the type and nature of your files with the vendors. Together you should be able to develop an efficient scanning and indexing workflow system.

3.0 System Implementation

3.1 Assign a permanent staff member as systems administrator and require the vendor to provide a project director during the installation and training periods.

The assignment of a qualified staff member, preferably with systems administration experience, is critical to the effective implementation and maintenance of a digital imaging system. The systems administrator should be responsible for overall project management, and the development and maintenance of written system documentation which describes the requirements, capabilities, limitations, design, operation and maintenance of the digital imaging system. Vendor requirements should include installing the equipment and training the systems administrator. Other appropriate agency staff can also help to ensure successful implementation of the system.

3.2 Establish operational practices and provide technical and administrative documentation to ensure the future usability of the system, continued access to long-term records, and a sound foundation for assuring the system's legal integrity. It is the responsibility of office administrators, rather than vendors and manufacturers, to maintain written documentation of system procedures, also called Standard Operating Procedures or SOPs, including access and security policies and procedures. Security and access policies should be developed to protect the system and the records from alteration or unauthorized use.

It is important to maintain a written record of procedures, operating systems, decisions, changes and updates made to the system. It should be complete, specific, and updated on a regular basis. Documentation of operating procedures should include a description of methods for scanning, entering data, revising, updating, and expunging records, hardware and software operating manuals, indexing techniques, and backup procedures for disks, tapes, microfilm, etc. It should also include procedures for testing the readability of records, security safeguards to prevent tampering and unauthorized access to protected information and the disposition of original records.

Technical Documentation should include the following:

OHIO ELECTRONIC RECORDS COMMITTEE

http://www.ohiohistory.org/ohiojunction/erc/

- Hardware type, brand name, model number, and date of installation of all hardware components of the system.
- Software version number, implementation date, and backup copies of all systems software and application programs.
- Maintenance equipment maintenance log to document the occurrence of regular maintenance
- Publishing and Redaction capabilities technical explanation of how electronic records for public distribution will be created. A public office must have the capability to redact confidential information permanently.

Confidential information must be redacted from files for public distribution, according to state and federal law. When creating databases with redacted layers, the file produced for distribution should be a combination of the original file and the redacted layer, so that they are one image with one layer. Once the new file is saved, it should not be possible to remove the redaction from the confidential fields.

The systems administrator should be familiar with the rules of evidence as they apply to the legal admissibility and trustworthiness of records. Records stored on a digital imaging system should comply with state and federal laws just as paper, film, and magnetic disk and tape. It is important that the systems administrator consider records retention requirements and legal admissibility of records at the beginning of the imaging project. Procedural controls must be established and followed to protect the integrity of the records.

These procedural controls should be documented and should reflect requirements for the legal acceptance of records as outlined in AIIM TR31-1992, Performance Guideline for the Admissibility of Records Produced by Information Technology Systems as Evidence. This AIIM performance guideline stresses the importance of specifying the processes used to create the records, demonstrating that records are produced and relied upon in the regular course of business, establishing quality control and audit programs, procedures, conducting formal training and providing documentation for each procedure. Case history indicates that system requirements for good archival maintenance are consistent with the requirements for the admission of records under the rules of evidence laws. Records Administrators should be familiar with how the rules of evidence apply to Ohio's public records. Policies and procedures should be followed to protect the integrity of long-term records. Integrity of records, also known as the reliability and authenticity of records, is discussed in depth in the "Ohio Trustworthy Information Systems Handbook."

3.3 Institute procedures to ensure quality and integrity of scanned images.

Scanning of Records: At the beginning of a shift, or every day, verify each image scanned until accuracy of the scanner settings is determined. Visual quality evaluations of one hundred percent should be performed at the time of new installation or upgrades in software or hardware, or when new operators or projects begin. These evaluations can be reduced as confidence is met.

Data Integrity: Data integrity of digital records should include procedures for regular inspection of the images and system components to insure both short and long term

OHIO ELECTRONIC RECORDS COMMITTEE

http://www.ohiohistory.org/ohiojunction/erc/

accessibility. Create a schedule of how often inspections of digital records and hardware and software should be made. These procedures should also include inspection of digital images to confirm that the storage media and other system components are working properly. An annual inspection of sample images from both primary and backup storage media to verify continued accessibility is recommended. Data maintained on electronic media should be copied onto new media at regular intervals.

Audits: Audit procedures should include schedules for regular, special, and annual inspections. Each audit should produce a written report detailing when it was performed, the name of the auditor, and the findings. On a regular basis, or at least every six months, a sampling of recorded images should be retrieved from the storage media and displayed for inspection to confirm that the storage media and other system components are working properly. Special audits should be performed at the time of new installations, when upgrades are made to the hardware or software, or when determined by staff or users. It is recommended a more thorough audit be conducted annually to ensure procedures are followed. Include an evaluation of the outputs of the digital records and an evaluation by the users.

A detailed discussion on auditing can be found in the "Ohio Trustworthy Information Systems Handbook."

3.4 Design backup procedures to create security copies of digitized images and their related index records.

Backup is the short-term copying or duplication of an information base, including the operating system, the applications, the active data and image sets.

The purpose of backup is to provide replacement of data or images lost due to system or user error, or in the event of a disaster. This is especially vital in terms of imaging, as the original paper-based information may no longer exist. Creating a duplicate copy of records in another format or another system is an effective method of ensuring access to information should an emergency or equipment failure occur. Backup copies also support system integrity and legal admissibility requirements.

Backup and, more importantly, restoration of data requires the following to be strictly enforced:

- Documented procedure for backups.
- Regular audits of the procedure to determine validity and completeness of data and images to be restored.

Documentation should include the following:

- Instructions for performing the backup.
- If the backup is automated, instructions for both setting up the automated sequence and manually bypassing if needed.
- Instructions for implementing backup solutions, including hardware and software.

A chart which includes:

OHIO ELECTRONIC RECORDS COMMITTEE http://www.ohiohistory.org/ohiojunction/erc/

- A schedule showing normal start-time and estimated timeframe for the backup.
- Types of backups performed (examples are <u>full</u>, <u>incremental</u>, differential).
- Examples of logs generated by the process and/or the application.
- Instructions for performing both partial and complete restore from any point in the process.

As part of the documentation, all of the above steps should be checked for accuracy.

Backup audits should be performed by either internal or external personnel on a regularly scheduled basis. They should be based solely on the documentation provided and can only be certified if the documentation is correct.

In general, backups should allow for disaster recovery. In other words, they must be constructed in such a way that data can be retrieved no matter what the circumstances. There are two ways of ensuring this.

- Duplicate backups done at more than one site. In the event of a disaster, backups should be available at the other site. This allows backups to be done remotely without human intervention at each site.
- Backups should be immediately moved to an offsite location so it is accessible in the event of a disaster.

No backup plan for disaster recovery is complete without ongoing testing to see if the data on the backup media can be restored successfully. The best backup procedures will not yield effective results unless periodic testing is done to ensure that:

- The data being backed up is the correct data.
- The data can be restored from the backup media.
- The restored data matches the data originally backed up.

As with any other aspect of a disaster recovery plan, failure to test for successful restores may invalidate the entire backup process. Be sure to test for full restore as well as restoring only selected files.

3.5 Provide adequate environmental conditions for digital media.

Digital media are susceptible to deterioration when storage conditions are inadequate. ARMA International suggests a temperature of 65 to 75 degrees Fahrenheit and a relative humidity between 20 - 35 percent. Media should never be in direct sunlight or near heat sources. They should be protected from dust, debris, fingerprints and a high static environment. Assuming exposed media are used, it is recommended each disk be stored upright in its own plastic case.

Agency officials should know and adhere to the documented storage specifications provided by the vendor.

3.6 The retention and disposal of digital images and corresponding electronic records should be incorporated into the agency retention schedule.

Records created in the imaging system should be scheduled for retention according to the records management procedures for your agency. It is important that the system documentation be scheduled as well.

OHIO ELECTRONIC RECORDS COMMITTEE

http://www.ohiohistory.org/ohiojunction/erc/

Document images should be stored in such a way that they can be identified and destroyed as retention periods expire. If feasible, items with like retention periods should be stored on the same media. If this is not done, a tool to locate and destroy specific records is required.

In any case, destruction of electronic images requires more than the normal file delete command. File deletion typically only erases the index pointer and not the image itself. If the image is stored on a server's magnetic storage, its location may eventually be overwritten, but until that occurs, the information is retrievable - if only by using forensic methods. Images stored on optical media (CD, DVD, MO Optical, etc.) will require that the physical disk be destroyed by crushing or pulverizing.

Once the record is destroyed, remember to destroy duplicate copies from system backups, secondary archive copies, etc.

3.7 A disaster preparedness plan is necessary for a properly created system of image and data backup, storage and recovery.

A record disaster is a sudden and unexpected event that results in the loss of records or information essential to an organization's continued operation. Disasters include fire, flood, and tornado. Less obvious, but often equally disastrous, events include human error, vandalism, unauthorized access, loss, theft, equipment failure, leaking pipes, insects, rodents, mold and terrorism.

The effects of a disaster on digital document imaging systems can be more easily controlled than those on paper records due to the ease of duplication and the portability of the media. A disaster preparedness plan should include preparation for any disaster which effects would render the organization's information stored in the digital document imaging system un-retrievable.

An effective plan should include provisions for the off-site storage of index data, digital images and system documentation. It should also provide for adequate security for back up media, a recovery plan, a plan for routine system audits and a plan for destruction at the end of the retention period.

The organization should identify an off-site facility for the storage of backup copies of the information. This facility should be geographically located far enough from the host site to minimize the likelihood of it being affected by the same disaster. The facility should be accessible 24 hours a day, 365 days a year. The facility should provide appropriate climate-control for the storage of digital media and should have adequate security measures in place to protect the information from access by unauthorized personnel.

The retention schedule for each record series should include the required retention for backups regardless of media. The backup should not be retained for a longer period than the record copy.

Any disaster recovery plan should be tested periodically to validate its effectiveness and identify areas that can be strengthened. Plans should typically be tested at least annually, with portions of the plan subject to more frequent testing if appropriate.

Organizations should establish a recovery site location that will be available in the

OHIO ELECTRONIC RECORDS COMMITTEE

http://www.ohiohistory.org/ohiojunction/erc/

event of a disaster. This may be a "hot", "warm", or "cold" site depending on the length of time the organization can survive without access to its information. Deciding on the level of redundancy required is a risk management issue. In any case, a recovery site will be required. Recovery of vital records should be a primary concern at this site.

Vital records are those without which the organization cannot function and would be impossible or costly to replace. They typically document the organization's charter, its legal obligations to customers, employees, and other stakeholders. If these records are maintained in electronic form, specific care must be taken to protect them from premature destruction.

The redundant site will provide for current system recovery. This means that redundant hardware and operating and application system software must be available. Recovery can be met by redundant copies of software stored off-site and a contract with a hardware vendor to provide system components in a specific length of time.

3.8 Budget annually between twenty and twenty-five percent of the original system acquisition cost for upgrades, training, and maintenance.

Administrative managers should understand the high cost of maintaining and upgrading digital imaging systems. Unless these costs are factored into the continuing support of system maintenance and improvement, the system can become obsolete, requiring a costly outlay to restore its effectiveness, if at all possible.

4.0 Archiving and Long-term Maintenance

4.1 Provide specific plans for creating and sustaining digital images that will be retained more than 10 years.

Public officials are responsible by law for ensuring that their records are protected and accessible for the time period stipulated in the records retention schedule. This responsibility applies regardless of the storage media of the records. It is the responsibility of the agency to decide on which medium to maintain their records. Should an agency decide to destroy original records once imaged, the agency must ensure that their imaged records are accessible for the time period stipulated on their records retention schedule.

If digital images need to be maintained for long periods of time, i.e., more than ten years, it will be necessary to take several steps to ensure accessibility. These steps include requirements for master image capture (scanning specifications, file formats, metadata); system information and maintenance (system documentation, copying and refreshing media); and sustainability.

4.2 Scanning Specifications and File Formats

For images that need to be retained ten years or more, it is recommended that originals be scanned at a minimum of 300 dpi and saved as a TIFF (Tagged Image File Format) Group 4 or higher. Currently this would include TIFF 4, TIFF 5, and TIFF 6. Best practice indicates that it is preferably to use the most recent version of the file format (currently TIFF 6, soon TIFF 7). (Other file formats are currently under

OHIO ELECTRONIC RECORDS COMMITTEE

http://www.ohiohistory.org/ohiojunction/erc/

development that may be appropriate in the future for the creation of a Digital Master. These may include PDF-A and JPEG2000.)

This TIFF will serve as a "master image" or "archival copy" that is similar to a microfilm master negative. The master image should capture as much information as possible from the original in order to serve as a long-term, high-quality digital version. Derivative images for use on the web or within the application itself should be made from the master image. Master images can be bitonal, grayscale or color. Quality control applied to master images should be intensive. This is especially true if an agency is retaining ONLY a digital image (no paper or microfilm copies) of a record with permanent retention.

The TIFF images should be stored uncompressed, or the compression used should be lossless. If it is absolutely necessary to compress the master image files, current industry standards recommend the following:

TIFF (Tagged Image File Format) with CCITT Fax 4 Compression: Better suited to bitonal text documents. This format can provide a high level of detail combined with a smaller file size. May be used as a master image file format.

TIFF (Tagged Image File Format) with LZW Compression: A 24-bit, lossless compression format, commonly used by Adobe Photoshop and other image editing software. Used to store color and grayscale files. May be used as a master image file format.

The TIFF images should be stored in a secure and stable environment, preferably offline. Other <u>derivative images</u> can be created from the TIFF files to enable web access.

4.3 Digital Image Metadata

Metadata is simply defined as data about data. More specifically, metadata consists of a standardized structured format and controlled vocabulary that allow for the precise description of record content, location, and value. Metadata often includes, but is not limited to, attributes like file type, file name, creator name, date of creation, and use restrictions. Metadata capture, whether automatic or manual, is a process built into the actual information system.

Note that TIFF file headers can contain system-generated metadata. You should be aware of which metadata elements your scanning software is able to use and which viewer you plan to use. Please refer to the <u>National Information Standards Organization's Draft Data Dictionary - Technical Metadata for Digital Still Images [Link updated 18 October 2010 - no longer draft/dwn]</u>. Although this is a draft standard, it may be considered the most current Best Practice for this topic.

4.4 System Information and Maintenance

On a system level, documentation is information about planning, development, specifications, implementation, modification, and maintenance of system components (hardware, software, networks, etc.). System documentation includes such things as policies, procedures, data models, user manuals, and program codes. Documentation capture is not a system process.

OHIO ELECTRONIC RECORDS COMMITTEE

http://www.ohiohistory.org/ohiojunction/erc/

Please refer to the Ohio Electronic Records Committee <u>Trustworthy Information</u> Systems Handbook Section 9 for more complete information on system documentation.

4.5 Copying and Refreshing Media

Media copying and <u>refreshing</u> are essential for all digital media to avoid degradation and to facilitate longer-term preservation strategies. This involves periodically copying data onto identical media to address media degradation and impermanence, and periodically reformatting the data from an obsolete storage device to a newly emerging one, in some cases bypassing the intermediate generation that is mature but at risk of becoming obsolete

Media copying and refreshing should take place as follows:

- Within the minimum time specified by the supplier for the media's viability under prevailing environmental conditions.
- When new storage devices are installed.
- When an audit discloses significant temporary or read errors in the resource.

4.6 Sustainability

Hardware, software, and file formats could be operational for ten years or more, but technology will often be superseded within two to three years. If the system stores records with retention periods exceeding the life span of the hardware and software, it is essential that the application or system administrator plan for data sustainability. A digital sustainability strategy documents how an organization will transfer long-term or permanent records from one generation of hardware and software to another generation while maintaining system functionality and data. The strategy should be written and available with current system documentation and should be updated when technology changes. It is important to budget for these costs.

4.7 Reformatting and Migration

Copying and refreshing will not ensure that the system remains accessible. New software, platforms, and file formats will need to be utilized in order to facilitate long-term accessibility and reliability of the records. Further steps will need to be taken that may include the following:

- <u>Reformatting</u> of existing file formats to appropriate newly emerging formats.
- <u>Migrating</u> one component of the system, such as the database that provides indexing information to the system, to a new hardware and/or software platform.
- Migrating the whole system from one hardware and/or software platform to another.
- Documenting the changes made to the hardware, software, and file formats. Include changes that could affect data viability such as moving to a new file format or moving indexing information to a new database system.

OHIO ELECTRONIC RECORDS COMMITTEE

http://www.ohiohistory.org/ohiojunction/erc/

Reformatting and migration should take place when the existing file formats, software, platforms or systems are no longer viable, usually due to obsolescence or the necessity of enhanced system performance.

Reformatting and migration will be much easier if the technical specifications of the system and the metadata relating to the digital images are created appropriately. For systems, this requires creating and maintaining System Documentation.

4.8 Scheduling

Agencies must submit a revised records retention schedule for records that they are reformatting, i.e., changing medium from paper to digital image. Agencies that are considering imaging records with permanent retention periods should contact the State Archives for a system and records analysis to determine if maintaining the records in an eye readable format (paper, microfilm, etc.) may also be necessary.

DEFINITIONS

Audit Trail: An electronic or paper log used to track computer activity.

Bitonal: One bit per pixel representing black and white. Bitonal scanning is best suited to high-contrast documents such as printed text.

Cold Site: A site which is stocked with equipment and ready to go. However, the machines are not operational, data is not copied on a live basis, and time is required to bring the site up live.

Color: Multiple bits per pixel representing color. Color scanning is suited to documents with color information.

Differential Backup: A backup of files that have changed since a full backup was performed.

Derivative Images: Images that are commonly used in place of master copy images for general web access, and include "thumbnail" images that might be only 100 pixels square and "reference" or "service" images that should fit completely within an average monitor. Images created for this purpose commonly have smaller file sizes and, therefore, do not require a fast network connection and are in a web viewable format such as JPEG or GIF.

Digital Master: A faithful digital reproduction of a document optimized for longevity and for production of a range of delivery versions.

Error Detection and Correction (EDAC) System: Allows data that is being read or transmitted to be checked for errors and, when necessary, corrected on the fly. Also known as Error Checking and Correcting (ECC).

Full Backup: All the files and folders on the drive are backed up every time you use that file set.

Grayscale: Multiple bits per pixel representing shades of gray. Grayscale is suited to continuous tone documents, such as black and white photographs.

Hot Site: A duplicate computer center is set up in a remote location, with communications lines set up and actively copying data at all times. The site has a

OHIO ELECTRONIC RECORDS COMMITTEE

http://www.ohiohistory.org/ohiojunction/erc/

duplicate of every critical server, with data that is up-to-date to within hours, minutes or even seconds.

Hypertext Markup Language (HTML): The set of markup symbols or codes inserted in a file intended for display on a World Wide Web browser page. The markup tells the Web browser how to display a Web page's words and images for the user.

Incremental Backup: A backup of files that have changed or are new since the last incremental backup.

Lossless Compression Algorithm: Every single bit of data that was originally in the file remains after the file is uncompressed. All of the information is completely restored.

Migration: The process of moving records from one hardware and/or software platform to another.

Non-proprietary: A format that is NOT owned by a private individual or corporation under a trademark or patent. It is in the "public domain" and is easily portable between various hardware and software systems.

Open Systems Architecture: Allows the system to be connected easily to devices and programs made by other manufacturers. Included are officially approved standards as well as privately designed architectures whose specifications are made public by the designers.

Optical Character Recognition (OCR): The recognition of printed or written text characters by a computer. This involves photoscanning of the text character-by-character, analysis of the scanned-in image, and then translation of the character image into character codes, such as ASCII, commonly used in data processing.

Record Analysis: The examination and evaluation of systems and procedures related to the creation, processing, storing, and disposition of records.

Reformatting: As data formats change, data streams will need to be moved to new formats. This process will change the actual configuration of the data, and some contextual information might be lost.

Refreshing: Periodically moving records from one storage medium to another. It is a preventive measure and, because of rapid media obsolescence, it will be a necessary strategy for some years to come.

Small Computer System Interface (SCSI): A set of ANSI standard electronic interfaces that allow personal computers to communicate with peripheral hardware such as disk drives, tape drives, CD-ROM drives, printers, and scanners faster and more flexibly than previous interfaces.

TIFF: A very high-resolution image that is best suited for archival preservation or print publications. Also, a TIFF is an uncompressed image. Compressing a file results in a loss of data and image quality. A TIFF image preserves more colors and details found in the original item.

Warm Site: A site which is pre-positioned with equipment, software and other necessities, all ready to go in the event of a disaster. The equipment is idle, often turned off, but can be quickly restored and brought online if needed.

OHIO ELECTRONIC RECORDS COMMITTEE

http://www.ohiohistory.org/ohiojunction/erc/

Write and Verify: A process during which information is written from one magnetic medium to another and then a check is made to verify that all of the data transferred properly.

Workflow: The tasks, procedural steps, organizations or people involved, required input and output information, and tools needed for each step in a business process.

WORM (Write Once Read Many): Data storage technology that allows information to be written to a disk a single time and prevents the drive from erasing the data. The disks are intentionally not rewritable, because they are especially intended to store data that the user does not want to erase accidentally.

IF YOU NEED ASSISTANCE

For more information about scheduling records, contact your agency's records manager. Local governments should contact the Ohio Historical Society-Local Government Records Program.

BIBLIOGRAPHY

- AIIM TR2-1992, Glossary of Imaging Technology. Silver Spring, MD: Association for Information and Image Management, 1992.
- AIIM TR25-1995, The Use of Optical Disks for Public Records. Silver Spring, MD: Association for Information and Image Management, 1995.
- AllM TR26-1993, Resolution as it Relates to Photographic and Electronic Imaging. Silver Spring, MD: Association for Information and Image Management, 1993.
- AIIM TR27-1996, Electronic Imaging Request for Proposal (RFP) Guidelines. Silver Spring, MD: Association for Information and Image Management, 1996.
- AIIM TR28-1991, The Expungement of Information Recorded on Optical Write-Once-Read-Many (WORM) Systems. Silver Spring, MD: Association for Information and Image Management, 1991.
- AIIM TR31-1992, Performance Guideline for Admissibility of Records Produced by Information Technology Systems as Evidence Part 1: Evidence. Silver Spring, MD: Association for Information and Image Management, 1992.
- AIIM TR31/2-1993, Performance Guideline for Acceptance of Records Produced by Information Technology Systems by Government Part 2: Acceptance by Federal or State Agencies. Silver Spring, MD: Association for Information and Image Management, 1993.
- AIIM TR31/3-1994, Performance Guideline for Admissibility of Records Produced by Information Technology Systems as Evidence Part 3: User Guidelines. Silver Spring, MD: Association for Information and Image Management, 1994.
- AIIM TR31/4-1994, Performance Guideline for Admissibility of Records Produced by Information Technology Systems as Evidence Part 4: Model Act and Rule. Silver Spring, MD: Association for Information and Image Management, 1994.

OHIO ELECTRONIC RECORDS COMMITTEE

http://www.ohiohistory.org/ohiojunction/erc/

- ANSI/AIIM MS44-1988 (R1993), Recommended Practice for Quality Control of Image Scanners. Silver Spring, MD: Association for Information and Image Management, 1993.
- ANSI/AIIM MS52-1991, Recommended Practice for the Requirements and Characteristics of Original Documents Intended for Optical Scanning. Silver Spring, MD: Association for Information and Image Management, 1991.
- ANSI/AIIM MS53-1993, Standard Recommended Practice File Format for Storage and Exchange of Images Bi-Level Image File Format: Part 1. Silver Spring, MD: Association for Information and Image Management, 1993.
- ANSI/AIIM MS59-1996, Media Error Monitoring and Reporting Techniques for Verification of Stored Data on Optical Digital Data Disks. Silver Spring, MD: Association for Information and Image Management, 1996.
- Arizona State Archives & Library. "Document Imaging Technology" 2002 [No longer available online 2 June 2010/dwn]
- Cinnamon, Barry and Richard Nees. *The Optical Disk-Gateway to 2000*. Silver Spring, MD: Association for Information and Image Management, 1991.
- "Cornell Digital Imaging Tutorial" 2000-2001 http://www.library.cornell.edu/preservation/tutorial_tutorial_English.pdf
- D'Alleyrand, Marc R., Ph.D. *Networks and Digital Imaging Systems in a Windowed Environment*. Boston, MA: Artech House, 1996.
- "Digital Imaging of Office Records: Archives and Records Management at the Harvard Medical School" 2002-2003 [No longer available online 2 June 2010/dwn]
- Elkington, Nancy E., ed. Digital Imaging Technology for Preservation: Proceedings from an RLG Symposium held March 17 and 18, 1994. Mountain View, CA: The Research Libraries Group, Inc., 1994.
- National Archives and Records Administration. "Digital Imaging and Optical Digital Data Disk Storage Systems: Long-Term Access Strategies for Federal Government Agencies" Washington, D.C. 1994.
- National Archives and Records Administration and National Association of Government Archives and Records Administrators. "Digital Imaging and Optical Media Storage Systems: Guidelines for State and Local Government Agencies" Washington, D.C. 1991.
- National Archives and Records Administration. "Frequently Asked Questions about Imaged Records" http://www.archives.gov/records-mgmt/faqs/imaged.htm [Link updated 2 June 2010/dwn]
- National Archives and Records Administration. "Frequently Asked Questions about Optical Media" March 28, 2003 http://www.archives.gov/records-mgmt/faqs/optical.html [Link updated 2 June 2010/dwn]
- Northeast Document Conservation Center. "Handbook for Digital Projects" 2000 http://www.nedcc.org/resources/digitalhandbook/dman.pdf [Link updated 2 June 2010/dwn]

OHIO ELECTRONIC RECORDS COMMITTEE

http://www.ohiohistory.org/ohiojunction/erc/

- Saffady, William. "Stability, Care and Handling of Microforms, Magnetic Media and Optical Disks" Library Technology Reports, Vol. 27, January/February 1991: 63-87.
- Technical Advisory Service for Images http://www.jiscdigitalmedia.ac.uk/blog/entry/tasi-re-brands-as-jisc-digital-media-9th-february-2009/ [Link updated 2 June 2010/dwn]
- Warner, Will. "Special Report: An Introduction to TIFF." Inform, Vol. 5, February 1991: 32-35.
- Western States Digital Standards Group. "Digital Imaging Best Practices." 2003 http://www.bcr.org/dps/cdp/best/wsdibp_v1.pdf [Link updated 2 June 2010/dwn]