State of Michigan
Records Management Services
Best Practices for the Capture of Digital Images from Paper or Microfilm

1.0 Introduction

The Records Reproduction Act (MCL 24.401-24.406) regulates the reproduction of public records by Michigan government agencies at all levels. This law requires the Records Management Services (RMS) to promulgate technical standards to ensure the continued accessibility and usability of records that are digitized throughout their retention period. This document supplements the "Technical Standards for Capturing Digital Images from Paper or Microfilm."

2.0 Scope

This best practice document applies to the conversion of public records by any Michigan public body from paper or microfilm to digital image format for the purpose of maintaining official records in a digital format. This document is not intended to apply to digital photography, publications, or convenience copies of records produced purely for the purpose of informational distribution (i.e. web publication or e-mail distribution). If digital images are produced by a third party, the state agency or local government is responsible for ensuring that the third party is in compliance with the "Standards for the Capture of Digital Images from Paper or Microfilm" and the stated goals of the state agency or local government.

3.0 Intent

The "Standards for the Capture of Digital Images from Paper or Microfilm" represent the minimum necessary requirements for capturing digital images. This document is intended to assist state and local governments with implementing the standards in a practical manner. They also define certain processes that should be used to convert records to digital formats. Following the best practices and standards will, in part, help state agencies and local governments ensure that converted records are authentic, reliable, have integrity, and are usable. Since digitized records are dependant upon hardware and software to function, appropriate system design and implementation is critical to successfully meeting all four principals. The selection of a system used to manage digitized records is dependent upon the business requirements of the records being managed, and therefore are not addressed within this document.

RMS is aware that there may be instances where a public body has an imaging application which, due to the nature of the application, might require technology and techniques that are not in compliance with the standards and best practices. It is not the intention of RMS to impose standards upon a public body that will reduce the intended benefits of an imaging application, provided the public body can be confident that steps have been taken to ensure the future migration of the images in accordance with retention requirements. In the event that deviations from the standards are warranted, the agency should contact RMS to develop technically acceptable alternatives that meet the needs of the agency without the risk of implementing a non-viable or non-compliant solution. In all other instances where deviations from the standards are not required or requested, it is in the best interest of the agency to follow the practices detailed in this standard.

Some electronic formats and techniques may not be suitable for long-term retention. Issues that affect the permanent preservation of the records include, but may not be limited to: organization,

indexing, format, resolution, compression algorithm and storage media. The Archives of Michigan will work with state agencies and local governments to identify those public records that are designated on an approved Retention and Disposal Schedule for transfer to the Archives for permanent preservation.

4.0 Basic Principles of Record Keeping

Selecting an appropriate record keeping system is like selecting the appropriate level of insurance for your home. Records with a greater value to the agency warrant a greater level of insurance. Records with lesser value "may" warrant a lesser level of insurance. Regardless of the value, all records and record keeping systems maintained by a state agency or local government must possess four basic characteristics to be considered trustworthy. Records must be authentic, reliable, have integrity, and be usable regardless of the format and the media they are contained on.

An **authentic** record is one that can be proven to be what it professes to be, to have been created or sent by the person claiming to have created or sent it, and to have been created or sent at that time. To ensure the authenticity of records, state agencies and local governments should implement and document policies and procedures which control the creation, receipt, transmission, maintenance and disposition of records. This will ensure that record creators are authorized and identified, and that records are protected against unauthorized addition, deletion, alteration, use and concealment.

A **reliable** record is one whose contents can be trusted to be a full and accurate representation of the transactions, activities or facts to which they attest and can be depended upon in the course of subsequent transactions or activities. Records should be created at the time of the transaction or incident to which they relate, or soon afterwards, by individuals who have direct knowledge of the facts or by instruments routinely used within the normal course of business to conduct the transaction.

The **integrity** of a record refers to its being complete and unaltered. It is necessary that a record be protected against unauthorized alteration. Records management policies and procedures should specify what additions or annotations may be made to a record after it is created, under what circumstances additions or annotations may be authorized, and who is authorized to make them. Any authorized annotation; addition or deletion to a record should be explicitly indicated and traceable.

A **useable** record is one that can be located, retrieved, presented and interpreted. It should be capable of subsequent presentation as directly connected to the business activity or transaction that produced it. The contextual linkages of records should carry the information needed for an understanding of the transactions that created and used them. It should be possible to identify a record within the context of broader business activities and functions. The links between records that document a sequence of activities should be maintained. The records must be accessible for the duration of the retention period.

5.0 Agency Responsibility

State agencies and local governments have responsibilities associated with the management of their records and information. These may include but may not be limited to the following:

- Develop and maintain a **Record Retention and Disposal Schedule**.
- Implement a retention and disposal policy.
- Select an appropriate record keeping system.
- Implement appropriate records management policies, procedures, and business practices.
- Develop a business analysis to determine whether imaging will be a cost effective, efficient and durable method of managing records.
- Develop a quality assurance level of acceptance that confirms the expectations of the agency are being met.
- Follow specific laws, rules and standards that govern records specific to the industry the records pertain to.

6.0 Compression Method

The lossless compression method T-4 (formerly known as group III) or T-6 (formerly known as group IV) developed by the International Telecommunication Union - Telecommunication Standardization Sector (ITU-T). These are widely used and commonly accepted defacto standards. It is for this reason they are identified as the standard compression method in the "Technical Standards for Capturing Digital Images from Paper or Microfilm" (herein referred to as "technical standards").

A compression method is used to reduce the amount of data needed to store or transmit a representation of a specific image. Ideally, compression should be lossless, where the data are compressed by efficient coding of the information in the image and where the reconstructed image contains the same amount of information.

Lossy compression algorithms compress images by selectively removing information from the image. These algorithms are designed to remove statistically redundant information, as well as perceptually irrelevant or unimportant information, leaving only useful information. The decompression process results in the production of the image with reduced information.

Selecting a compression scheme is application specific. Appropriate selection requires the state agency or local government to take into account the characteristics of the document, the tasks the system is designed to perform, and the user requirements to be satisfied.

7.0 File Format

The Technical Standards state that "a file format with lossless compression shall be used, so long as the format has the capability of conversion of its images to TIFF without a significant loss of information." Any file format with lossless compression may be used so long as there is the capability to easily convert images to TIFF without significant loss of information (such as PDF).

Formats with lossy compression may be justified only when the retention value of the records is short term or when it is determined through testing that the desired results of the capture process cannot meet the business needs of the organization using a recommended format with lossless compression.

A file format is necessary to transfer compressed images between systems. An image file format contains a header with a description of the image attributes and a body containing the compressed image. The most commonly used interchange format is TIFF. The TIFF format supports the most commonly used compression algorithms including T4, T6, JPEG and JBIG. The TIFF format for electronic interchange has been standardized and was designed for the conveyance of facsimile images but is also suitable for many other image interchange purposes.

8.0 Resolution

A 200 DPI/PPI minimum resolution level for standard business documents, a 300 DPI/PPI minimum resolution level for line art (engineering drawings) and those intended for Optical Character Recognition (OCR) processing is required by the technical standards. For records that are designated to be transferred to the State Archives of Michigan in digital form, a higher resolution may be necessary. In such cases, the State Archives will work with state agencies and local governments to select the appropriate resolution. The use of a lower resolution level may be justified when it is determined through testing that the desired results of the capture process cannot meet the agency's business needs using a recommended format with a higher resolution level. According to the technical standards, images shall exhibit a degree of legibility and readability comparable to the source document.

Resolution is often referred to as dots per inch or DPI, in common usage the terms DPI and pixels per inch PPI are used interchangeably. Since raster image files (formed by a set of pixels in a grid pattern) are composed of pixels, technically PPI is a more accurate term. DPI is the appropriate term for describing printer resolution (actual dots vs. pixels); however, DPI is used often in scanning and image processing software to refer to spatial resolution. Therefore, the usage is understandable.

In electronic imaging, resolution is the ability of a scanning device to reproduce the details of an image. Generally speaking, the higher the resolution, the better the quality of the image will be. Since the ability to capture a quality image is dependent upon the content and quality of the original document, different resolutions should be tested for a given collection of records to determine the optimum resolution necessary to satisfy the intended use.

9.0 Enhancement Techniques

Enhancement techniques common in scanning software such as deskew, despeckle, crop, rotate, etc. are acceptable, so long as the content that exists in the original document is not altered.

10.0 Documentation

Standard operating procedures should be developed that define, but may not be limited to, the following:

- Testing and cleaning of equipment
- Document preparation
- Image capture
- Data and image backup
- Access and security
- Administration and maintenance

- Audit trails

A statement of work (SOW) or job statement should be developed for each record collection or record type to be scanned. The SOW defines the necessary tasks, scanner settings, and product deliverables of a given job set. The SOW should include, but not be limited to:

- Definition of current environment
- Definition of the desired result
- Document preparation requirements and instructions
- Documented results of the quality control sample
- Index attributes
- Scanner settings
- File format and compression method
- Resolution
- Quality control processes

Depending upon how the capture process and the organization is structured, procedural documentation may be contained in several locations and/or be in several formats. It is not necessary for a state agency or local government to strictly follow the recommended SOP/SOW format. However, it is necessary that a state agency or local government document their processes to ensure continued success in capturing satisfactory images.

11.0 Document Preparation

Preparation of the records prior to scanning is critical to the success of the imaging solution. Care should be taken in the preparation, content and arrangement of original records for scanning to ensure that a true, accurate and complete reproduction is created. The functional steps in the preparation process include but may not be limited to:

1. Remove all staples, paper clips or other fastening devices
2. Repair all torn or damaged documents
3. Remove creases or folds for the pages so that no information is covered or lost
4. Identify any significant categories or subcategories of the collection prior to scanning
5. Identify and locate missing or misfiled documents
6. Arrange the documents in the order in which they are to be scanned
7. Insert barcodes, if necessary

12.0 Indexing

The retrieval of images is entirely dependent upon effective indexing, and is vital to the success of the application. A variety of indexing methods can be employed. Indexing for each application will depend upon the characteristics of the record, the system requirements, and end-user retrieval requirements. Indexes may take many forms, including databases, spreadsheets, full-text OCR, and file naming conventions that help locate and present an image or series of images.

13.0 Pre-production Sample for Quality

The technical standards require a sample set of source documents or documents equivalent in characteristics to the source documents, be assembled for the purposes of evaluating scanner

results against defined quality criteria. Documents in the sample set should include examples of source documents whose quality is poor, relative to the majority of the documents. The results of the production sample will dictate the necessary steps to be taken in the quality control process that is developed.

The purpose of the pre-production sample is to establish a quality reference. It will define what is an "acceptable image." This quality reference should be maintained for the duration of a project or until a new quality reference is defined. This process allows the end user to continuously evaluate and ensure that the scanning process is successful as it provides a base upon which to compare future scanned images. The technical standards require a new sample for quality be performed if the conditions or attributes of documents to be scanned change or if the equipment used to scan the documents change. This process is particularly important when using external service providers and when the content and quality of documents within a collection vary.

14.0 Quality Control

Quality control is defined as those steps incorporated into the production process that are designed specifically to reduce error. Quality criteria may include, but may not be limited to:

- Overall legibility
- Smallest detail legibility captured
- Completeness of detail
- Dimensional accuracy compared with the original
- Scanner generated speckle
- Completeness of overall image area
- Density of solid black areas
- Color fidelity
- Image skew
- Image rotation
- Image cropping
- Index data accuracy
- Image and index format compliance

Once the quality criteria for the various attributes have been defined for the production process, procedures should be established to ensure that these criteria are met. These procedures should be documented in the statement of work/job statement.

15.0 Quality Assurance

The technical standards require an agency to adopt written quality assurance procedures for inspection of digital images that are produced. Quality assurance shall be conducted before the original documents are destroyed. State agencies and local governments should have a quality assurance process in place for each project or record collection being scanned.

Quality assurance is the process by which the total product is examined to ensure that the quality criteria initially established in the pre-production test have been met. The purpose of this quality assurance process is to establish sampling plans and procedures to inspect the individual attributes of the created product. Upon receipt of an image product the state agency or local government

should verify that the imaged documents match the expectations defined by the sample set. Verification is needed to ensure that the attributes agreed upon at the onset of the project are successfully delivered.

Keep in mind that there is a significant difference between those quality control steps provided during the capture process that are designed to detect and correct errors and quality assurance which is designed to verify the validity and accuracy of the overall delivered product. While the capture process should provide quality control prior to product delivery, the end user must also perform their own quality assurance in order to verify that the delivered work product is acceptable.

In order to establish a meaningful sampling process there are three categories of information that must be established prior to the scanning process and can be derived from the pre-production test. The end-user must:

1. Identify the specific attributes of the work product that are critical to them.
2. Establish the acceptability level expressed as a percentage.
3. Establish a batch size expressed as a number of items that are contained within the batch.

From this information an inspection model can be developed that will, within the limits of the acceptability level, assure the state agency or local government that the delivered work product has met the established standard. The specific attributes that need to be defined are those elements of the image that are determined to be critical to the overall success of the conversion process. In a production environment, it is not sufficient to simply say, "This is a good image." Objective criteria that define what a good image is must be established so that the production process can routinely and reliably produce the defined "good image." Attributes must be defined objectively. Subjective attributes cannot be measured reliably. Be careful to select attributes that are critical to the desired output.

The end-user quality assurance process must be performed in a timely manner in order to conform within agreed upon acceptance terms. **Quality assurance should be performed before the original documents are destroyed.**

For more information regarding establishing a statistical sampling model refer to ANSI/AIIM TR34-1996 – Sampling Procedures for Inspection by Attributes of Images in Electronic Image Management (EIM) and Micrographics Systems.

16.0 Media

Any media suitable for storage of digital images as defined in the Records Reproduction Act may be used provided the images are managed in a system that provides an appropriate level of certainty for the recovery and security of the images and related index attributes.

17.0 Scanners and Scanning

Except for regular computer enhancement routines used to improve the legibility of a scanned record, the technical standards require an agency to use an imaging system that is not capable of altering a public record as scanned. Recorded text or characters in original records shall not be edited. The technical standards require an agency to periodically evaluate scanner quality. For

more information on performing scanner quality evaluations, an agency may refer to ANSI/AIIM MS44-1998 (R1993) - Recommended Practice for Quality Control of Image Scanners.

18.0 References

The following standards and recommended practices issued by the American National Standards Institute (ANSI), the Association for Information and Image Management (AIIM) and the International Association for Standards (ISO) may contain additional information that will assist state agencies and local government with compliance with Michigan law. These publications are available from the Association for Information and Image Management, 1100 Wayne Ave., Suite 1100, Silver Spring, MD 20910-5699, http://www.ansi.org/.

ANSI/AIIM MS44-1998 (R1993) - Recommended Practice for Quality Control of Image Scanners

ANSI/AIIM MS50-1994 – Recommended Practice for Monitoring Image Quality of Aperture Card Film Image Scanners

ANSI/AIIM MS52-1991 - Recommended Practice for the Requirements and Characteristics of Original Documents Intended for Optical Scanning

ANSI/AIIM MS53-1993 - Recommended Practice; File Format for Storage and Exchange of Image; Bi-Level Image File Format: Part 1 MS 53-1993.pdf

ANSI/AIIM MS55-1994 - Recommended Practice for the Identification and Indexing of Page Components (Zones) for Automated Processing in an Electronic Image Management (EIM) Environment

ANSI/AIIM MS58-1996 - Standard Recommended Practice for Implementation of Small Computer Systems Interface (SCSI-2) (X3.131-1994)

ANSI/AIIM MS59-1996 - Media Error Monitoring and Reporting Techniques for Verification of Stored Data on Optical Digital Data Disks

ANSI/AIIM MS60-1996 - Electronic Folder Interchange Datastream

ANSI/AIIM MS61-1996 - Application Programming Interface (API) for Scanners in Document Imaging Systems

ANSI/AIIM TR1-1988 (A1992) – Guidelines for Metrics

ANSI/AIIM TR2-1998 – Glossary of Document Technologies

ANSI/AIIM TR15-1997 – Planning Considerations, Addressing Preparation of Documents for Image Capture

ANSI/AIIM TR17-1989 – Facsimile and Its Role in Electronic Imaging

ANSI/AIIM TR19-1993 – Electronic Imaging Display Devices

ANSI/AIIM TR21-1991 – Recommendations for the Identifying Information to be Placed on Write-Once-Read-Many (WORM) and Rewritable Optical Disk (OD) Cartridge Label(s) and Optical Disk Cartridge Packaging (Shipping Containers)

ANSI/AIIM TR25-1995 – The Use of Optical Disks for Public Records

ANSI/AIIM TR26-1993 – Resolution as it Relates to Photographic and Electronic Imaging

ANSI/AIIM TR27-1996 – Electronic Imaging Request for Proposal (RFP) Guidelines

ANSI/AIIM TR28-1991 – The Expungement of Information Recorded on Optical Write-Once-Read-Many (WORM) Systems

ANSI/AIIM TR29-1993 – Electronic Imaging Output Printers

ANSI/AIIM TR31:1-1992 – Performance Guideline for the Legal Acceptance of Records Produced by Information Technology Systems Part 1: Evidence

ANSI/AIIM TR31:2-1993 – Performance Guideline for the Legal Acceptance of Records Produced by Information Technology Systems Part 2: Acceptance by Government Agencies

ANSI/AIIM TR31:3-1994 – Performance Guideline for the Legal Acceptance of Records Produced by Information Technology Systems Part 3: Implementation

ANSI/AIIM TR31:4-1994 – Performance Guideline for the Legal Acceptance of Records Produced by Information Technology Systems Part 4: Model Act and Rule

ANSI/AIIM TR32-1994 – Paper Forms Design Optimization for Electronic Image Management (EIM)

ANSI/AIIM TR33-1998 – Selecting an Appropriate Image Compression Method to Match User Requirements

ANSI/AIIM TR34-1996 – Sampling Procedures for Inspection by Attributes of Images in Electronic Image Management (EIM) and Micrographics Systems

ANSI/AIIM TR35-1995 – Human and Organizational Issues for Successful EIM System Implementation

ANSI/AIIM TR38-1996 – Compilation of Test Target for Document Imaging Systems

ANSI/AIIM TR39-1996 – Guidelines for the Use of Media Error Monitoring and Reporting Techniques for the Verification of Information Stored on Optical Digital Data Disks

ANSI/AIIM TR40-1995 – Suggested Index Fields for Documents in Electronic Image (EIM) Environments

ISO 12653-1:2000 – Electronic Imaging – Test target for the black-and-white scanning of office documents – Part 1 – Characteristics

ISO 12653-2:2000 – Electronic Imaging – Test target for the black-and-white scanning of office documents – Part 2 – Method of use

ISO 15489-1:2001 – Information and Documentation – Records Management – Part 1 – General

ISO 15489-2:2001 – Information and Documentation – Records Management – Part 2 – Guidelines

ISO/TR 15801:2004 – Electronic Imaging – Information stored electronically – Recommendations for trustworthiness and reliability