

Chapter 3: IMAGING STATE RECORDS

SUMMARY: These rules provide (a) standards to be used by State agencies for imaging records; (b) criteria for using digital and microfilm copies; (c) standards for storing digital and microfilm copies of permanent records; and (d) information concerning digital and microfilm services available from the Maine State Archives.

1. APPLICABILITY

These rules apply to all agencies of State government, including the Executive, Legislative, Judicial, Independent and Quasi-Independent State Entities housing material at BABLO warehouse or have been transferred to the physical custody of the Maine State Archives.

2. RESPONSIBILITY FOR ENFORCEMENT

The head of each agency shall be responsible for the application and enforcement of these rules.

3. DEFINITIONS

The following definitions are established for terms used in these rules:

- A. "Access images" (aka surrogate or derivative) are the user copies of images, those copies that users access in the system. The use copy may be identical to the master in all respects except that it is the copy used by internal and/or external users. If the access images will be retained long term, it is best if they are produced as common long-term formats with lossless compression (such as TIFF and PDF/A), even if their ppi and/or bit depth is reduced to allow for faster processing on an electronic system. However, access images may be produced in any format desired, so the JPEG and GIF formats are sometimes employed for access images. These guidelines do not set technical specifications for access images, except that such images must be in a universally accessible file format. Since user copies are not to be the copies of record, these images may be enhanced in any way desired to improve legibility or visibility of information.
- B. "Agency" means any Department, Bureau, Division or other branch of State Government.
- C. "Archives" means noncurrent government records that have been determined by the State Archivist, with advice from the Archives Advisory Board, to have sufficient value to warrant their continued preservation.
- D. A "backup image" is a copy of any version of an image (master, access,

- or thumbnail) that is stored off line and outside of the imaging system to protect the images from intentional or accidental destruction or tampering. Every version of every image on an imaging system must be backed up.
- E. "Imaging" for the purpose of these rules will encompass digital capture of documents; while microfilming is sometimes categorized as an imaging process. In some systems, imaging and micrographics technologies are combined to form a hybrid system.
 - F. "Digital imaging" is the process of converting records on paper or film into electronic images of the original records. The process typically requires a document scanner or digital camera, a computer, and software to capture the image, as well as labor needed for preparing the documents for imaging, indexing the digitized images, and performing quality control steps to ensure the proper conversion of the record.
 - G. "Duplicate Microfilm" whether positive or negative and regardless of base, emulsion or generation is any microfilm that has been produced by any means of exposure from another microfilm.
 - H. "Electronic Content Management System (ECMS)" means a system designed to properly manage digitized images and other electronic records, so the Archives recommends an ECMS to capture, store, retrieve, display, and transmit records electronically. An ECMS uses a database to manage descriptive information about the images that aids in the retrieval of records contained in the ECMS repository. Although storage outside of an ECMS is allowed, the Archives does not recommend storing digitized records outside of an ECMS due to the greater chance of accidental deletion of these records and lack of an audit trail to ensure the records' authenticity.
 - I. The "master image" is the one primarily governed by these guidelines, including file format, compression, imaging resolution, image pixel depth, and color mode. The master image will be the copy of record (unless the source paper or film copy is retained), and it will be the one of the highest quality. To ensure there is a satisfactory backup to all master images, master images must always be stored outside an imaging system, even if it is stored within the system as well.
 - J. "Nonpermanent record" is any record that has been determined by the State Archivist to have a retention period with a destruction date. Retention periods are stated in records retention schedules approved by the Maine State Archives.
 - K. "Original microfilm" is any microfilm, produced in camera or by digital means, that is to be considered a first generation production regardless of base or emulsion type. Any microfilm that has not been created by means of being copied from an existing roll of film.
 - L. "Records" means all documentary material, regardless of media or characteristics, made or received and maintained by an agency in accordance with law or rule or in the transaction of its official business. The term shall not include extra copies of printed or processed material of which official or record copies have been retained, stocks of publications and processed documents intended for distribution or use, or material relating to personal matters that may have been kept in an office for convenience.
 - M. A "thumbnail image" is a very small copy of an image that is created so that it can be displayed, usually in a tabular array, so that users can choose the particular images they desire

from a search return of images. Thumbnails are created in any format desired (often JPEG, GIF, or PNG) with ppi ranging from 72 to 100 ppi. These guidelines do not set technical specifications for access images nor require their production, except that such images must be in a universally accessible file format.

4. GUIDELINES FOR THE DIGITAL IMAGING OF RECORDS

State agencies are required to contact the Maine State Archives prior to imaging any state government record.

State agencies engaged in digitally imaging records shall comply with the guidelines set forth by the Federal Agencies Digitization Guidelines Initiative (FADGI). The guidelines are a collaborative effort by federal agencies to define common guidelines, methods, and practices for digitizing historical content. See <http://www.digitizationguidelines.gov/>. The Maine State Archives' recommendation is to follow the star system in these guidelines based on document type.

State agencies interested in digitally imaging archival records need to contact the Maine State Archives as archival records cannot be destroyed; the original document must be kept.

5. STATE AGENCY RESPONSIBILITIES

State agencies are responsible for managing their own records. Before starting an imaging project, the state agency must identify the appropriate retention period for the records involved. The retention period affects the determination of the file format and compression method to use in any imaging project. The agency must also carefully consider whether it will be able to manage the imaged records for the duration of the retention period – *different rules apply for archival records*. If the intent of the agency is to destroy the original documents after imaging the agency files must notify the State Archives at the start of the project. State agencies must follow the State General Schedules and their own agency-specific schedules, and they must file a records disposition authorization and obtain the Archives' approval prior to disposition of paper originals after imaging.

State agencies must implement the appropriate policies, procedures, and business practices to ensure the proper protection, authenticity, reliability, integrity, and usability of records, regardless of format.

If a third-party vendor or some other outside entity digitizes a record for a state agency they must ensure the third party is in compliance with these guidelines. In such cases, a properly written contract must be in place containing the basic requirements of the intended project as well as reference to these guidelines.

6. AUTHENTICITY

- A. These guidelines, when accompanied by adherence to industry best practices, are designed to ensure any digital images remain authentic and reliable throughout their required retention periods.
- B. Authentication is necessary for legal, fiscal or administrative purposes to ensure that imaged records are acceptable as evidence. To introduce imaged records into evidence in a court or to use them in other legal or adjudicatory situations, a state agency must demonstrate that the processes and system used to produce imaged records are designed and managed to ensure the accuracy and authenticity of the records.
- C. The authenticity and viability of the images are determined by sound systems and processes that include, but are not limited to, well documented procedures, proper indexes and finding aids, quality assurance controls, reliable storage and backup practices, and tested and audited processes that ensure imaged records have not been altered, destroyed, or replaced. Systems that produce records must be shown to do so in the normal course of business and in an accurate and timely manner. Policies, procedures, training and support programs, and controls must be documented to demonstrate that the systems that produce records are reliable.
- D. Legal acceptance of records requires proof that the process or system is reliable and hence capable of producing trustworthy records. State agencies must be able to demonstrate that imaged records have been produced in the normal course of business, that the system producing and managing the records (the technology, as well as the policies, procedures, training, and audit controls) is trustworthy and functioning properly, and that system documentation is maintained. System documentation must be up to date and show how the system creates, manages, and destroys records; define the roles and responsibilities of the individuals managing the system; demonstrate how the system provides for consistent quality control; document staff training, system and audit controls; and provide for adequate security.
- E. The trustworthiness of imaged records depends on how closely documented procedures are followed. Since courts may scrutinize deviations from established procedures, state agencies must ensure that procedures are followed and deviations are detected and remedied.

7. DOCUMENT PREPARATION

The integrity of the original records authorized for imaging shall be maintained by ensuring that the images are adequate substitutes for the original records and serve the purposes for which such records are created or maintained. The following measures and any others found necessary shall be observed in any State imaging project to insure preservation of the integrity of the records:

All documents will be appropriately prepared (through the removal of paper clips and staples, flattening, mending, etc.) prior to imaging, either by the state agency or by a qualified third-party. Any page with an attached note on its must have the attached note removed from the page prior to imaging and placed on a clean, blank page and digitized separately, unless the attached notes have been designated as non-records, in which case these may be discarded.

Obsolete records and non-records may be removed from files during document preparation. Staff of the state agency will conduct this work, unless the instructions for identifying obsolete records are completely objective and the ability of a third party to follow these instructions can be

verified by staff prior to imaging. If the percentage of obsolete records or non-records is low, the effort of weeding will likely not be worth the cost.

Multi-part forms and continuous-feed computer paper must be separated into single pages. If there are multi-part forms included within the file, then the parts must be separated. In cases where such forms repeat the information on each page, only the highest quality page will be digitized (such as the top or white copy of the form). Other pages of such multi-part forms will not be digitized.

Fragile items or items that are torn, otherwise damaged, or on delicate backing (such as onionskin paper or glass) must be handled with care to avoid further damage or even destruction. If these items are interspersed within a body of records, they must be separated from the other items within a folder that marks them as fragile, and the entity conducting the imaging must be provided with instructions on how to handle these documents carefully. If entire boxes of records are fragile, these boxes must be marked as fragile.

Before imaging documents, the state agency or alternatively the vendor chosen to conduct the imaging, will produce an inventory of the documents that records the type of documents (paper, volumes, photographic paper prints, diapositive slides, glass plate negatives, drawings, etc), general characteristics (size, paper color, number of pages, the presence or absence of staples or other fasteners, etc.), and general document quality (undamaged documents, torn documents, onionskin paper, blurry images, or other damage) for the set of records as a whole. This inventory will allow for an assessment of the difficulty of imaging the documents.

Test images of the digital documents intended to be digitized must be created prior to wholesale conversion to ensure quality images can be produced. The number of test images produced will be governed by variations in the source documents. If the records vary widely by age, type, color, or condition, or if they have been produced at various times by various entities in the case of microfilm, test digital images must be made of samples of each set of similar documents.

8. USE OF MICROFORMS

An agency having custody of a previously produced original microfilm (or, if the original microfilm is no longer in existence, the master copy) of archival records shall prepare a silver duplicate negative copy for its own use, if required, when it finds that such copies are deteriorating or in danger of deterioration. The agency shall also provide an identical copy to the Maine State Archives. Before imaging microforms, the state agency or alternatively the vendor chosen to conduct the imaging, will produce an inventory of the microforms that records the number of microforms (in reels or fiche), the type of microform (16mm, 35mm, microfiche, aperture cards, etc.), its length if applicable (100 or 215 feet), general film quality (resolution, density, image spacing, and the existence of targets, blips, splices, scratches, warping, separation of the emulsion from the base, or other damage), and general image quality (such as blurry images), the age of the film, and the creator of the film (if known).

(NOTE: The Maine State Archives recommends all microfilm cellulose nitrate or cellulose acetate based be duplicated onto polyester based film.)

9. INDEXING AND METADATA

Since digitized images do not have intelligence within them indicating their contents, appropriate index information or metadata is required to properly identify and later retrieve digitized images. For digital images, indexing and file naming schema are essential for locating and retrieving stored imaged records. Indexing typically consists of a structured format and controlled vocabulary that allows more precise description of a record's content.

The state agency must define and document specific indexing requirements needed to access the records efficiently prior to the performance of any imaging and indexing.

Indexing must comply with the specific requirements of the state agency but at minimum it must include the following:

Unique Identifier for Documents: Each document (including each multi-page document) must have a unique filename or other identifier, preferably sequential, which can be numeric, alphanumeric, or alphabetic as required by the government entity. Each filename must be unique across all records series and storage media, not merely within a single disc or other piece of removable media. If required, images will be filed in appropriate electronic folders on the designated storage media.

Indexing Fields/Descriptive Metadata: The index of documents must consist of a number of fields to ensure adequate access to the records. Whenever possible, the field data must consist of objective indexing terms (such as personal names, file numbers, and dates) or terms from a controlled vocabulary (such as subjects or geographical information), rather than subjective data. Index data often includes information such as record type, creation date, record creator, disposition date, among other information.

Indexing Structure: Although the structure of an electronic content management system (ECMS) database is outside the scope of these guidelines, the state agency must have a methodology in place to transfer all the images and corollary index data to the intended retrieval system. The indexing data must be stored in a non-proprietary format to allow its transfer to other systems and databases as needed through the conversion project and for the entire retention period of the records. Each record within the database must be associated with the respective digital image or document via its unique filename.

Optical Character Recognition: If required, optical character recognition (OCR) or intelligent character recognition (ICR) may be performed to convert digital images into electronic text. The government or its chosen vendor must certify the conversion to be at least 95% accurate as measured by character count, and the converted text must be associated with the respective digital image or document. Due to this error rate, OCR will not be used as the sole finding aid when retrieving digitized images. Some manual indexing is always required.

- a. **Correcting or Making Allowances for OCR Output:** Depending on the need for accuracy in the OCR'd text, the text may be reviewed and corrected or fuzzy

searching may be used to retrieve character strings. Post-OCR correction consists of review of the OCR output against the original text and hand-correction of the OCR output. Fuzzy searching works by searching for character strings that match or predominately match the character string being searched.

Directory Structure: Regardless of the image filename, files will be organized in a file directory or folder system that will link to metadata stored elsewhere in a database. Directories may have their own organization independent of the image files, such as folders arranged by date or records series number, or they may replicate the physical or logical organization of the originals being digitized.

Technical Metadata: During the imaging process, production metadata will be maintained either within the individual images or separate from but associated with each body of digitized images. For instance, these metadata may be created as part of a digital file during actual imaging, may be added to the file after imaging, may be associated with each file in an ECMS, or may be retained entirely separate from the files but associated to each file by their unique filenames. These metadata will include, at minimum, the following:

- Unique identifier
- Title of records series
- State Archives or other retention schedule name and item number (from the State General Schedule, Agency Specific Schedule or Records Disposition Authorization Number)
- State agency name
- Name of the imaging vendor or government staff person conducting the imaging
- Date of the imaging
- Pixels per inch (ppi)
- Equipment used to capture the images
- Software used to capture the images

D. The state agency will maintain these metadata for the life of the records.

10. QUALITY ASSURANCE

- A. Quality assurance procedures must be in place to ensure the creation of accurate and authentic images and accurate metadata (index terms) that follow these guidelines, as well as ensuring that the specific requirements of the state agency are met. Quality assurance must be conducted before the destruction of any original documents. Each image of every page of all digitized documents must be visually inspected, not verified by a machine, to ensure clarity, readability, and accurate representation of the original record. Similarly, each indexing field must be checked against the original or imaged record.
- B. If producing access copies and/or thumbnails of the images, conduct quality assurance for these as well.

- C. Digitizing must capture each digital image with the same level of clarity as each document page so that every legible line and character on the original document appears and is legible in the image.

A person or persons other than those digitizing or indexing a particular record must perform the final quality control procedures outlined within this document. In most situations, quality assurance is performed in a two-step process: the scanner or digital camera operator will perform an initial quality check during the imaging process, and then a different individual will perform a second review in a separate process. If a vendor is conducting the imaging, either the government agency using the services of a vendor or a third party must conduct the quality assurance.

The quality control process must be documented and maintained throughout the digitization conversion process. Information to document includes problem resolution procedures and reporting requirements for each step of a conversion project.

Quality control steps for digitized images must verify the following items:

- Correct image file naming convention, as agreed upon
- Correct file format (including verification of compliance with the PDF/A format for purported PDF/A files)
- Quality of image is the same as in the original
- Correct size and resolution
- Image digitized at appropriate ppi for each image type
- Proper reading orientation (landscape or portrait)
- Image is not skewed
- Image is neither too light nor too dark
- Curvature of the page does not obscure or distort the text
- Appropriate contrast within the image
- No distortion of the image
- No extraneous materials (attached notes, fasteners, etc.) obscure the image
- No additional information added to the image that is not part of the original document
- Appropriate indexing terms associated with the digitized image

For acceptable correction of digitized images refer to FADGI Guidelines.

- H. Upon inspection, any image deemed of unacceptable quality must be redigitized followed by a re-inspection of the new image.
- I. Index Accuracy: The index data, not including OCR'd text used to search against, must be verified with the goal of achieving 100% accuracy. Acceptable methods include verification of data by another individual other than the person performing the initial data entry; dual data entry where two operators independently index the same document and the results are compared to find any discrepancies (this is also known as double-blind indexing); or any other means as appropriate to ensure 100% index accuracy.

11. STORAGE

Storage Environment: Storage media must be kept in secure, dust-free area under the proper environmental conditions. Contact the Maine State Archives for proper consistent temperature control levels relative to humidity.

Confidential and sensitive information: Special care must be taken when physically transferring media from the imaging source to the customer if they contain confidential or sensitive information. Contracts with vendors must delineate the procedures vendors must use to ensure they provide adequate protection to sensitive information.

Back ups: A periodic backup process of all digitized records should be performed along with geographically remote offsite storage to ensure the accessibility of records in the event of a disaster.

Backups must be stored in areas geographically remote from the offices where the use copies of the records are stored. An appropriate backup location will be one where it is highly unlikely that the backup location will simultaneously suffer the same disaster as the offices.

Regardless of the media used, images must be accessible for the records' entire retention period.

12. DATA MAINTENANCE

Any stored record must be protected against file corruption, alteration, or deletion throughout its required retention period. Adequate processes and documented procedures to ensure the integrity of the digital image should be in place.

1. Reliability: Controls to ensure 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;
2. Authenticity: Controls to protect against unauthorized addition, deletion, alteration, use, and concealment;
3. Integrity: Controls, such as audit trails, to ensure records are complete and unaltered;
4. Usability: Mechanisms to ensure records can be located, retrieved, presented, and interpreted;
5. Content: Mechanisms to preserve the information contained within the record itself that was produced by the creator of the record;
6. Context: Mechanisms to implement cross-references to related records that show the organizational, functional, and operational circumstances about the record, which will vary depending upon the business, legal, and regulatory requirements of the business activity; and
7. Structure: controls to ensure the maintenance of the physical and logical format of the records and the relationships between the data elements.

Digital images and their associated index data must be effectively and efficiently managed over time. Regardless of format, records must be retained until their retention requirements have been met.

Agencies must design and implement migration strategies to counteract hardware and software dependencies of electronic records whenever the records must be maintained and used beyond the life of the information system in which the records are originally

created or captured. To successfully protect records against technological obsolescence.

- Carry out upgrades of hardware and software in such a way as to retain the functionality and integrity of the electronic records created in them;
- Any necessary conversion of storage media to provide compatibility with current hardware;
- Maintaining a link between records and their metadata through conversion or migration, including capture of all relevant associated metadata at the point of migration (for both the records and the migration process); and
- Ensure that migration strategies address non-active electronic records that are stored off-line.

13. PROCEDURES FOR ARRANGING FOR REIMBURSABLE SERVICES

State and non-state agencies desiring imaging services, whether in agency, private vendor or RFP bidder should contact the Maine State Archives. An agreement of services to be provided will be negotiated before work is begun. This agreement shall specify (a) the location of the records to be imaged; (b) a description of the records, including their volume, size, physical condition, filing arrangement, and the nature and frequency of additions, changes, and deletions, etc.; (c) the size and format of the film to be used (if applicable); (d) the type of camera to be used and the reduction ratio at which the documents are to be imaged; (e) the type of reader to be used (if applicable); and (f) the number and type (silver, diazo, vesicular) of reference copies if microfilmed.

14. FEES FOR IMAGING AND CONVERSION SERVICES

A fee schedule for imaging services is available from the Maine State Archives. Imaging and conversion services available are: microfilm processing, silver roll film duplication, digital to microfilm conversion, microfilm to digital conversion, imaging of oversized and fragile documents, and other services are available upon request.

15. VIOLATION OF RULES

Violation of any provision of these rules, except those violations for which specific penalties are provided, is a Class E crime.

STATUTORY AUTHORITY: 5 M.R.S.A. §95, sub-§ 3

RE-ADOPTED:

August 31, 1979

EFFECTIVE DATE (ELECTRONIC CONVERSION):

April 29, 1996

NON-SUBSTANTIVE CORRECTION:

May 29, 1996 - §10 (D)

CONVERTED TO MS WORD:

July 11, 2003

FULLY EDITED: