New York State Archives

Digital Imaging Guidelines

2022

The University of the State of New York
The State Education Department
New York State Archives
Albany, New York 12230
www.archives.nysed.gov

New York State Archives

Digital Imaging Guidelines

2022

Table of Contents

General Guidelines

1. Purpose

1.1. The purpose of these guidelines is to assist New York State local governments and state agencies that are considering imaging records on paper and film. The document provides recommendations that follow industry standards and best practices to ensure the long-term accessibility and authenticity of the electronic records involved. These guidelines must be used by governments imaging records in house, as well as any third-party vendor contracted to perform imaging services on their behalf. See Appendix A for guidelines specifically for vendors providing imaging services to state agencies and local governments. See Appendix B for a tabular summary of these guidelines.

1.2. This document is one of many resources created by the State Archives for state agencies and local governments considering the conversion of paper records into an electronic image format. Additional workshops, reference material and other resources can be found on the New York State Archives' website (www.archives.nysed.gov).

2. Use of These Guidelines When Working with a Vendor

2.1. State agencies and local governments must not use these guidelines alone as a specification. Instead, they must compile descriptions, special handling, and other procedures for each records series to be digitized. The following are essential pieces of information to add to the guidelines:

- Customer name
- Name of records series
- Years covered by the records series
- Total number of images
- Preparation requirements
- Condition of records series (sizes, type, whether single- or double-sided)
- Color and age of paper
- Type of fasteners
- Percentage of documents flat, rolled, or folded
- Type and color of imprint
- Preparation requirements
- Required retrieval aids (indexing, file naming, etc.)
- Required transfer media numbering
- Number of duplicates required
- Transfer media types required (CD-R, DVD-R, external hard drive, USB drive, etc.)
- Delivery information (date, destination, etc.)
- Technical enhancements to the guidelines needed to accommodate the needs of the records series

2.2. Packaged together, the appropriate imaging guidelines and the supplemental requirements outlined above will form an imaging project specification. A specification

describes the essential technical requirements for items, materials, services, and procedures. Incorporate the specification into the standard contractual documents used by your organization, and ensure that your legal counsel has reviewed these documents.

3. Digital Imaging Defined

3.1. For the purposes of these guidelines, 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. This document uses the terms "imaging" and "digitizing" interchangeably to refer to this process.

3.2. An Electronic Content Management System (ECMS) is 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.

3.3. While microfilming is sometimes categorized as an imaging process, for the purpose of these guidelines imaging will encompass digital capture of documents. In some systems, imaging and micrographics technologies are combined to form a hybrid system.

4. Scope

4.1. These guidelines apply to all state agencies and local governments, as defined in the Arts and Cultural Affairs Law and the Regulations of the Commissioner of Education.

4.2. This document is intended to provide guidance to state agencies and local governments interested in converting documents on paper and film to digital images, as well as to provide best practices to ensure their accessibility and protection throughout their retention periods. These guidelines list the minimum standards for producing and inspecting digital images of hardcopy records.

4.3. *Limitations:* These guidelines do not apply to New York State legislative or judicial records, even judicial records maintained within local government facilities. Local governments or courts managing court records may contact the New York State Office of Court Administration's records management department, which oversees the management of court records, for relevant imaging guidelines. No digital imaging guidelines have been prepared expressly for New York State's legislative records.

5. State Agency and Local Government Responsibilities

5.1. State agencies and local governments are responsible for managing their own records. Before starting an imaging project, the government agency must identify the appropriate

retention period for the records involved. As outlined later in this document, the retention period affects the determination of the file format and compression method to use in any imaging project. The government must also carefully consider whether it will be able to manage the imaged records for the duration of the retention period.

Local governments must follow the retention and disposition provisions outlined in the appropriate records retention schedule (LGS-1 and County Boards of Elections schedules), and may dispose of originals after imaging and confirming the accuracy of the digitized images. Authority to replace paper originals with digital surrogates is granted to local governments through the Regulations of the Commissioner of Education, § 185.8, Retention and preservation of electronic records.

State agencies must follow the State General Schedule and their own agency-specific schedules. The requirements for the retention and preservation of state agencies' electronic records are detailed in the Regulations of the Commissioner of Education, § 188.20, Retention and preservation of electronic records.

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

5.3. If a third-party vendor or some other outside entity digitizes a record for a state agency or local government, the government 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

6.1. 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.

6.2. Authentication is necessary for legal, audit, 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 or local government 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.

6.3. 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.

6.4. Legal acceptance of records requires proof that the process or system is reliable and hence capable of producing trustworthy records. State agencies and local governments 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.

6.5. The trustworthiness of imaged records depends on how closely documented procedures are followed. Since courts may scrutinize deviations from established procedures, state agencies and local governments must ensure that procedures are followed, and deviations are detected and remedied.

7. Document and Record Categories

7.1. For purposes of these guidelines, there are five broad categories of documents and records:

7.1.1. *Small Textual Documents:* Textual documents are defined as paper documents up to 11" x 17" in size. Computer printouts on continuous-form paper must be separated at the perforations for digitizing and are considered small textual documents for the purposes of these guidelines. (See section 15 for the guidelines relevant to such records.)

7.1.2. *Maps and Plans:* Maps and plans are defined as rolled, folded, or flat maps and engineering or architectural plans (on paper, fabric, or polyethylene, etc.) that are larger than 11 x 17 inches in size (ledger size). Large textual documents (those larger than 11" x 17") also fall under this category. (See section 16 for the guidelines relevant to such records.)

7.1.3. *Bound Volumes:* Bound documents include books and similar material that cannot easily have their pages removed for placement in a scanner's sheet feeder. Examples include minute books, older financial ledgers, and other documents with sewn or glued bindings. Post-bound records (records stored in binders held together with straight, removable posts) are not considered bound documents and must be dismantled and digitized following guidelines for small or large documents, depending on paper size. If volumes will be guillotined prior to imaging, to separate the pages of a volume from its binding, follow the guidelines for small textual documents or maps and plans, as appropriate. (See section 17 for the guidelines relevant to such records.)

7.1.4. *Photographic Records:* Photographic records cover all photographic reproduction formats (paper, plastic, glass; photographs, negatives, diapositive photo slides) except for microforms and motion-picture film. (See section 19 for the guidelines relevant to such records.)

7.1.5. *Microforms:* Microfilm, microfiche (including ultrafiche), and aperture cards are types of microforms. For the purposes of these guidelines, microforms consist of rolls or sheets of film (usually acetate or polyester) that store physically smaller copies of pages of records. Micro-opaques, microforms recording information on hard-copy media (cardstock, nickel alloy, or silicon), are not

covered by these guidelines. (See section 19 for the guidelines relevant to such records.)

8. Document Preparation

8.1. 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 local government, or by a qualified third-party. Any page with a sticky note on it must have it removed from the page prior to imaging, placed on a clean, blank page, and digitized separately unless sticky notes have been designated as non-records, in which case they may be discarded.

8.2. Obsolete records and non-records may be removed from files during document preparation. Staff of the state agency or local government 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.

8.3. 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.

8.4. *Fragile items:* 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.

9. Image Versions

9.1. An imaging system may contain up to three different versions of each image based on the images' use.

9.1.1. *Master Image:* 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 record copy of the image (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.

9.1.2. *Access Image:* Access images are the use 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 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 use copies must not also be the record copies of any records, these images may be enhanced in any way desired to improve legibility or visibility of information.

9.1.3. *Thumbnail:* 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.

9.2. *Backup Image:* 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.

10. Compression

10.1. Digitized image files are large and can require significant storage space if left in their native format. Compression is an efficient technique for reducing the size of an image file, thus saving storage space.

10.2. Any compression technique used by state agencies or local governments must be a non-proprietary, lossless compression method that does not remove data or otherwise alter the appearance of the original image, such as ITU-T (formerly CCITT Group IV) or JPEG 2000 encoding.

10.3. Lossy compression techniques are not acceptable for image compression for records. Lossy compression is a compression method that removes data from the image file to create a smaller file size.

10.4. For the best preservation practice, the State Archives recommends that the master digital images of archival records remain uncompressed.

11. File Formats for Master Digitized Images

11.1. For master digitized images with retention periods of 10 years or more, including those designated as archival or permanent, only two image file formats are allowed: Tagged Image File Format (TIFF), which is preferred for photographic records, and PDF/A, which is preferred for textual documents or hybrid documents with text and images. Lossless compression techniques must be used in either format. Both formats support black and white, grayscale, or color documents:

11.1.1. *Tagged Image File Format (TIFF):* TIFF-formatted files must be based on

International Telecommunication Union – Telecommunication Standardization Sector (ITU-T) encoding (formerly referred to as CCITT Group IV encoding).

11.1.2. *PDF/A:* The PDF/A format is the version of PDF (Portable Document Format) intended for the long-term preservation of electronic files. PDF/A is the preferred version of PDF for long-term retention since all of the elements to render the file are contained within the file itself in a PDF/A. A basic PDF file, on the other hand, may rely on linked fonts, non-standard header information, or other data stored outside of the file. PDF/A-formatted files must utilize lossless compression, such as ITU-T or JPEG 2000.

11.1.3. *PDF/A Conformance Levels:* Any version of PDF/A is acceptable, but versions using conformance level u (for Unicode) are preferred as these will save digitized text as images and also use optical character recognition to convert digitized text into Unicode, which is a text encoding standard covering text in all writing systems.

11.2. For records with retention periods of less than 10 years, there are three acceptable image formats for imaging: TIFF, PDF/A, and basic PDF. Lossless image compression must be used in any selected format. Note that many scanners and multi-function devices default to using lossy compression, the use of which is not permitted by these guidelines under any situation.

11.3. Digitized images used as surrogates of original records cannot have proprietary headers, which make the images inaccessible except in particular software environments.

12. Imaging Resolution

12.1. Imaging resolution refers to the number of pixels per inch (ppi) created during document capture. These guidelines set a minimum resolution of 300 pixels per inch as a base for small documents. However, higher resolution will be required in many cases, as noted in these guidelines.

12.2. The resolution of digitized images assumes a 100% imaging ratio (meaning a 1:1 ratio) and is intended to provide an accurate image of the original record when produced in the same size as the original.

12.3. Higher levels of resolution may be required for smaller, damaged, or low-contrast documents, or those where the creation of intelligent text (searchable text encoded in ASCII or Unicode) via optical character recognition is most essential to an imaging project.

13. Image Pixel Bit Depth

13.1. Digital imaging is a process by which the content of a document is digitized (through scanning or digital photography) and converted from an analog format to a computer-readable digital format represented by a series of pixels called a bitmap (or raster) image.

13.2. Pixel bit depth defines the number of shades that can actually be represented by the amount of information saved for each pixel. These can range from one bit per pixel for binary (or pure black and white) images to 24 bits per pixel or greater in high quality color images.

The pixel bit depth will vary depending on the characteristics of the document to be digitized as described below.

14. Color Mode

14.1. Color mode for digitizing color material may be either RGB (Red, Green, Blue) or CMYK (Cyan, Magenta, Yellow, Black).

14.2. CMYK is preferred for documents that may need to be printed in high quality, such as in books or magazines, in the future.

15. Digitizing Small Textual Documents

15.1. *Black and White Textual Documents:* Textual documents in black and white with good contrast between the printed text and the paper background will be digitized in bi-tonal mode (black and white, no gray), which equates to one bit per pixel. Digitize standard business documents at 300 pixels per inch to allow for sufficient clarity in the image and to support the use of optical character recognition (OCR) or intelligent character recognition (ICR) processing if desired.

15.2. *Low Quality Documents:* Documents with handwritten notes or markings, low contrast between the text and the paper background, or poor legibility will be digitized in grayscale mode (8 bits per pixel, representing 256 shades of gray) or in 24-bit RGB or CYMK color mode. Color mode may be necessary to capture details and information better in cases where the original paper record is yellowed, stained, or on onionskin paper.

15.3. *Documents with Illustrations:* Documents with half-tone illustrations included as part of the document will be digitized in grayscale mode (8 bits per pixel, representing 256 shades of gray). Documents with color illustrations (photographs, tables, charts, etc.) will be digitized in RGB or CYMK 24-bit color mode. These guidelines are necessary to capture the necessary detail of such information.

15.4. *Documents with Color:* Some documents do not include illustrations but still use color in such a way that the color does not add meaning to the record, and these may be digitized in bi-tonal mode. However, documents with meaningful color must be digitized in RGB or CYMK 24-bit color mode to capture the color accurately. Examples of records with meaningful color include those annotated with highlighter markers, those handwritten with more than one color of ink, or those in which the color designates the status of some value within a document (such as names in blue indicating those people chosen for a particular assignment).

15.5. *Document Inventory:* Before imaging textual documents, the state agency or local government, or alternatively, the vendor chosen to conduct the imaging, will produce an inventory of the documents that records the type of paper documents (letter-sized paper, legal-sized paper, ledger-sized paper, etc.), general characteristics (paper size, paper color, the presence or absence of staples or other fasteners, etc.), and general document quality (undamaged documents, torn documents, onionskin paper, 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.

15.6. *Test Images:* Test digital images of the 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 paper documents. If the paper records vary widely by paper type and condition, test digital images must be made of samples of each set of similar textual documents.

15.7. *Multi-page documents:* When a record contains multiple pages (e.g., a single invoice comprises two pages), then this document will be digitized as a single file, whether saving to a TIFF, PDF, or PDF/A file.

15.7.1. While the TIFF file format specification supports both single-page TIFF and multi-page TIFF, multiple-page documents will be digitized as a multi-page TIFF file to ensure document integrity.

15.7.2. An imaging system may, alternatively, combine individual images into a virtual multi-page document, if that is necessary to allow the system to display the document.

16. Digitizing Maps and Plans

16.1. *Resolution:* A resolution of 300 ppi will be adequate to scan many maps and plans. A higher resolution may be needed to properly capture smaller details within larger maps and plans. Resolutions up to 600 ppi may be necessary to capture the fine detail on some maps and plans. Test scans make it possible to test the level of resolution needed.

16.2. *Pixel Bit Depth:* Simple black and white maps and plans with good contrast will be digitized in bi-tonal mode. Documents with shades of gray but no color will be digitized in grayscale mode (8 bits per pixel, representing 256 shades of gray). Documents with color or in cases where the original carrier of the image is yellowed or stained will be digitized in RGB or CYMK 24-bit color mode.

16.3. *Document Inventory:* Before imaging text documents, the state agency or local government, or alternatively the vendor chosen to conduct the imaging, will produce an inventory of the documents that records the type of paper documents ("C" drawings, "D" drawings, "E" drawings, etc.), general characteristics (paper size, paper color, the presence or absence of staples or other fasteners, etc.), and general document quality (undamaged documents, torn documents, or other damage). This inventory will allow for an assessment of the difficulty of imaging the documents.

16.4. *Test Images:* Test digital images of the maps and plans 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 maps and plans. If the maps and plans vary widely in paper type and condition, test digital images must be made of samples of each set of similar maps and plans.

17. Digitizing Bound Volumes

17.1. When a bound volume cannot or will not be disbound, the pages of that volume must be digitized in such a way that the image of each page is not excessively warped and that all the information on each page, even handwritten additions, is captured. If the entire image of each page cannot be captured, the bound volume must be retained after imaging.

17.2. Bound volumes must be digitized by use of a book scanner or camera that holds the volume open at an angle which reduces the curvature of the pages, or a book cradle that holds the surface of the pages flat. Curvature correction of the document is allowed so long as the correction does not obscure or distort the original image, and all data in the record is captured.

17.3. *Resolution and Pixel Bit Depth:* The guidelines for resolution and pixel depth when digitizing of bound volumes will conform to the guidelines for small textual documents or maps and plans, as appropriate.

17.4. *Document Inventory:* Before imaging bound volumes, the state agency or local government, or alternatively, the vendor chosen to conduct the imaging, will produce an inventory of the volumes which records the number of volumes, estimated number of pages, and general quality (torn pages, wormholes, tight bindings, missing or damaged covers, or other damage). This inventory will allow for an assessment of the difficulty of imaging the volumes.

17.5. *Test Images:* Test digital images of the volumes 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 volumes. If volumes vary widely in age or condition, test digital images must be made of each set of similar volumes.

18. Digitizing Photographic Records

18.1. *General Guidelines:* Photographs require sufficient resolution to capture all the significant detail in the originals. Although 300 ppi will be adequate for many photographs, higher resolutions are necessary for smaller photographs that will be enlarged or ones with fine detail. Damaged photographs must be digitized at higher rate than those indicated below in order to capture sufficient detail.

18.2. *Black and White Photographic Prints:* Digitize in 8-bit grayscale mode, representing 256 shades of gray for continuous tone black-and-white photographs. Damaged or faded black-and-white photographs will be digitized in 16-bit grayscale or 24-bit color mode to capture sufficient detail. Color settings will be employed in situations where the photograph displays color outside the grayscale range. The minimum ppi for such materials will be 800 ppi for prints when the longest side is up to 5.5", 400 ppi for larger prints up to 11.5" long, and 300 ppi for prints larger than this.

18.3. *Color Photographs:* Digitize color photographs in 24-bit RGB or CMYK color mode, representing 16.8 million colors. The minimum ppi for such materials will be 800 ppi for prints when the longest side is up to 5.5", 400 ppi for larger prints up to 11.5" long, and 300 ppi for larger prints.

18.4. *Positive Film:* Digitize positive film (such as diapositive slides) at 2800 ppi for 35mm images and at 1600 ppi for 120 and 220 slides, in 16-bit grayscale mode for black and white images and in 24-bit RGB or CYMK color mode for color images. For higher quality images suitable for full color or grayscale printing, digitize at resolutions between 3000 and 6000 ppi.

18.5. *Photographic Negatives:* Digitize photographic negatives using the same guidelines for positive film, but reverse the polarity during imaging from negative images (light text on a black background) to positive ones (black text on white background) for better usability.

18.6. *Glass-Based Photographic Materials:* Glass plate negatives are pieces of glass with an emulsion layer displaying a negative image, and lantern slides are the same except that they are meant for projection, so they display a positive image. The minimum ppi for such materials will be 800 ppi for 3¼" X 4¼", 600 ppi for 4" X 5", and 400 ppi for 8" X 10" glass plate negatives and lantern slides. For greater image clarity and in cases when using the digital images as preservation copies, scan at much higher resolutions, approximately six times these minimums. When digitizing glass plate negatives, reverse the polarity during imaging from negative images (light text on a black background) to positive ones (black text on white background) for better usability.

18.7. *Document Inventory:* Before imaging photographic materials, the state agency or local government, or alternatively, the vendor chosen to conduct the imaging, will produce an inventory of the material that records the type of photographs (paper prints, diapositive slides, glass plate negatives, etc.), their size or sizes, and damage (tears, scratches, etc.), general image quality (such as blurry images), and the age of the photographs or film. This inventory will allow for an assessment of the difficulty of imaging the photographic materials as a whole.

18.8. *Test Images:* Test digital images of the photographs 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 photographs and film. If photograph and film types vary widely by type, color, or condition, test digital images must be made of each set of similar photographic materials.

19. Digitizing Microforms

19.1. *Limitations:* At times a state agency or local government may need to convert records stored on microform (microfilm or microfiche) to a digital raster image for speedier electronic access. Due to the photographic limitations of microfilm and the variable quality of older microfilm, it may not be possible to produce what would normally be considered reproduction quality digital image files from all microfilm. To produce images of adequate quality, the approach used to digitize microfilm may vary from the recommendations cited here for paper-based records.

19.2. *Resolution and Pixel Bit Depth:* Digitize microfilm at 5600 ppi for 16mm images and 2800 ppi for 35mm images, in 16-bit grayscale, and higher if necessary to offset the poor quality of micrographic images. If converting continuous tone (grayscale) microfilm, digitize in 16-bit grayscale mode. If converting color microfilm, digitize in 24-bit color mode (RGB or CYMK). Since images on microfiche are much smaller than those on microfilm, the digitization of them must raise the bit depth even higher to capture passable images.

19.3. *Polarity:* When digitizing negative microfilm, reverse the polarity during imaging from negative images (light text on a black background) to positive ones (black text on white background) for better usability.

19.4. *Microfilm Inventory:* Before imaging microforms, the state agency or local government, 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).

19.5. *Test Images:* A test digital image of the microfilm or microfiche intended to be digitized must be created prior to wholesale conversion to ensure quality image scan be produced. The number of test images produced will be governed by variations in the source microform. If microforms vary widely by type, color, or condition, or if they have been produced at various times by various entities, test digital images must be made of each set of microforms sharing similar characteristics.

20. Image Enhancement

20.1. In most cases, image enhancement cannot be employed through or after the imaging process because such processes may call into question the integrity of the digital surrogates produced. However, some modifications of the original digital image are acceptable. If image quality cannot be enhanced through acceptable techniques, the record must be re-digitized.

20.2. *Acceptable techniques:* Digital enhancement techniques commonly used in imaging software (de-skewing, cropping image data beyond the document, and rotating an image to the proper orientation) are allowed, so far as the information on the record image is not altered by such processes. Curvature correction is allowed so long as the correction does not obscure or distort the original image, and all data in the record is captured. Proper image orientation for optimal viewing must be maintained: portrait or landscape orientation, as appropriate. All documents must be digitized at their original size.

20.3. *Unacceptable techniques:* No image enhancement techniques may be used if these will alter existing content in an original record. Such techniques include sharpening, retouching, or otherwise adding or removing information from the images.

20.4. Annotations or "sticky notes" added during the imaging process must be separated from the image, and not burned onto the image file itself.

21. Indexing

21.1. Since digitized images do not have native 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 is essential for locating and retrieving stored imaged records.

21.2. Indexing typically consists of a structured format and controlled vocabulary which allows more precise description of a record's content. Index data often includes record type, creation date, record creator, disposition date, among other information. The state agency or local government must define specific indexing requirements needed to access the records efficiently prior to imaging and indexing.

21.3. Indexing must comply with the specific requirements of the state agency or local government, but at minimum it must include the following:

21.3.1. *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.

21.3.2. *Indexing Fields:* The index of documents must consist of a limited number of field names 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, rather than subjective data.

21.3.3. *Indexing Structure:* Although the structure of an electronic content management system (ECMS) database is outside the scope of these guidelines, the state agency or local government 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.

21.4. *Archival Indexing:* Indexing of potentially high-use archival or long-term records may include more metadata fields, but such metadata requirements must be implemented based on actual need for access.

21.5. *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.

21.5.1. *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.

21.6. *Directory Structure:* Regardless of the image filename, files will be organized in a file directory, or a 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.

22. Production Metadata

22.1. 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, added to the file after imaging, associated with each file in an ECMS, or retained entirely separate from the files, but associated to each file by their unique filenames.

22.2. 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, a specific local government schedule, or a special schedule) or Records Disposition Authorization Number (for state agencies only)
- State agency or local government 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

22.3. The state agency or local government will maintain these metadata for the life of the records.

23. Quality Assurance

23.1. 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 or local government 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.

23.2. If producing access copies and/or thumbnails of the images, conduct quality assurance for these as well.

23.3. 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.

23.4. 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.

23.5. 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.

23.6. 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 (sticky 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

23.7. Acceptable correction of digitized images is limited to the following:

- Correcting image filename
- De-skewing, rotating, or flipping the image to correct its orientation
- Adjusting brightness, contrast, or tone through re-digitizing only
- Curvature correction that does not obscure or distort the original image and that captures all data in the record
- Reversing polarity for photographic negatives or negative microfilm
- Re-digitizing, followed by a re-inspection of the new image
- Updating index database to correct errors

23.8. Upon inspection, any image deemed of unacceptable quality must be re-digitized followed by a re-inspection of the new image.

23.9. *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.

23.10. *Content Verification: If* the state agency or local government destroys the original paper or microform record after scanning, they must conduct a page-by-page verification of 100% of the records to ensure that each page has been imaged. If the state agency or local

government maintains the original records after scanning, they may use one of two sampling methods outlined in Appendix D, Sampling Methods for the Content Verification of Digital Images.

23.11. If conducting 100% verification, the state agency or local government must verify the images in batches during the scanning process and not at the completion of the scanning project.

23.12. For those contracting with an outside vendor or other third party to perform document imaging services, additional guidelines are included in Appendix A of this document.

24. Storage

24.1. Copying digital files onto CD-Rs or DVD-Rs and then placing them in storage will not ensure the long-term readability of the images. Any storage media used must comply with the applicable International Organization for Standardization (ISO) standards, which specify how this type of media disc must store information, and which allows the interchange of discs within different systems. Refer to http://www.iso.org for more information on ISO standards. Appendix C of this document includes selected ISO standards and technical reports.

24.2. For the purposes of these guidelines, there are three categories of storage media:

- Media used to transfer images and index data from the imaging source (such as a vendor hired to digitize documents) to the customer
- Media used to transfer images of archival records from the state agency to the State Archives
- The digital repository (preferably, an ECMS) used by the state agency or local government to store the images

24.2.1. *Transfer Media (from imaging source to customer):* Transfer media is only intended for short-term storage while moving images and index data from the imaging source (such as a vendor hired for imaging services) to the state agency's or local government's records repository. For the purposes of transferring images, CD-Rs, DVD-Rs, external hard drives, or USB-drive media are preferred. Note that an external hard drive will facilitate a faster transfer of content to the desired repository, whereas optical media (CD-Rs and DVD-Rs) will be slower. The Archives does not recommend the use of transfer media for the long-term storage of digital records or their associated metadata, because of media instability and fragility.

24.2.2. *Transfer Media (from state agency to State Archives):* When the retention period has been satisfied for state agency archival records and the records are ready for transfer to the Archives, contact the State Archives' Electronic Records Unit to determine the best method of transfer.

24.2.3. *Storage Media:* Final images and their associated metadata are best stored on server-class hard drives utilizing a RAID (Redundant Array of Inexpensive Discs) configuration. RAID 5 or higher is typically the preferred RAID configuration to ensure proper protection and availability in the event of a

disc failure.

24.2.4. If use of RAID 5 or other RAID level drive array is not available, storing images and their associated index data on server-class hard drives which are designed for greater tolerances and durability than standard desktop PC hard drives can be used, assuming that daily backup and offsite storage of the data is available. Note: Use of a non-RAID configured hard drive can result in the loss of data if a hardware failure occurs after the last backup process was performed.

24.3. *Media Storage Environment*: Storage media must be kept in secure, dust-free area under the proper environmental conditions: consistent temperature from 65 to 72 degrees Fahrenheit and 30% to 50% relative humidity with limited fluctuations.

24.4. *Confidential and Sensitive Information:* Special care must be taken when physically transferring media from the imaging source to the customer if records contain confidential or sensitive information. To prevent unauthorized access to sensitive information, the use of data encryption (such as the use of encoding that uses algorithms to ensure data cannot be read by unauthorized individuals), locked media storage containers, or other secure method of transfer must be employed. Contracts with vendors must delineate the procedures vendors must use to ensure they provide adequate protection to sensitive information.

24.5. A periodic backup process of all digitized records must be performed along with geographically remote offsite storage to ensure the accessibility of records in the event of a disaster. Annual testing of the backup media must be performed to ensure all files have been backed up and are readable.

24.5.1. Backups must be stored in areas geographically remote from the offices where the use copies of the records are stored. The necessary distance of backup storage from offices will vary based on the geological, meteorological, and human risks of the offices. An appropriate backup location will be one where it is highly unlikely that that backup location will simultaneously suffer the same disaster as the offices.

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

25. Data Maintenance

25.1. 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.

25.2. *Data Integrity:* 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 must be in place. Integrity indicates that the data is an exact copy of the original, and that the data has not been corrupted while either writing to or reading from the storage medium.

25.2.1. When data is written to a storage medium, an error-checking value called a checksum is computed and written along with the data. Any time the data is read, the checksum is recalculated and compared to the stored value to verify that the data on the disk was written and read correctly. All stored data files will be checked annually for integrity using a disk-error checking utility, which is built into most operating systems.

25.3. *Media Refreshing:* Due to technological advances and potential media obsolescence, local governments and state agencies will move records from one storage medium to another every three to five years, as a preventative measure to ensure future media readability. State agencies and local governments must prepare a plan for refreshing stored data to ensure its continued readability for the entirety of required records retention periods.

25.4. *Tape-based Media:* To ensure readability and proper performance of tape- based media, tapes will be rewound (also referred to as re-tensioning tapes) at least once every three years or less depending on the tape quality, type, storage conditions, and manufacturer's recommendations. Doing so relieves the stress normally placed on tape-based media through normal use when reading and writing data, and it can provide an early indication of any potential issues with the media. A random sample of 10% of all tapes must be read annually to identify any loss of data and to discover and correct the causes of data loss (including poor quality tape, high usage, improper handling, or a sub-standard storage environment). Data maintained on tape must be copied onto new tape a minimum

of once every five years.

25.5. *Migration:* Due to the technological advances and the potential obsolescence of technology currently in place, local governments and states agencies must plan for future migrations to new media and systems. Storage media often become obsolete and are replaced with new technology before the end of their life expectancy. Often, the device used to read the storage media is replaced with higher performing or higher capacity devices, and finding older hardware that can read older media may be difficult or impossible. If a system stores records with retention periods exceeding the lifespan of the hardware and software in use, it becomes essential to plan for future data migration. To ensure the contents of the media are readily accessible, migration of the stored images and associated metadata to a newer media platform every five years is recommended. Local governments and state agencies must establish a migration plan *before* imaging and review the plan annually. The reality of technological obsolescence requires that local governments and state agencies monitor technology trends and industry developments to ensure their records are accessible over the required retention periods of the systems on which they are stored.

26. Equipment Maintenance and Inspection

26.1. *General Guidelines:* If conducting its own imaging, a state agency or local government must regularly maintain and inspect its digital imaging equipment to ensure all of it is in good working order, and must maintain logs of such activities. These logs must note any problems identified with each piece of equipment and the steps taken to eliminate each problem. When applicable, the state agency or local government must follow manufacturer's guidelines for equipment maintenance, unless it determines that its differing processes are more reliable than those of the manufacturer.

26.2. *Additional Guidelines:* To ensure quality digital imaging, a state agency or local government should follow the detailed maintenance and inspection guidelines detailed for vendors in Section 31, Handling, Transfer, Storage, and Security, below.

27. System Trustworthiness

27.1. To ensure the trustworthiness of any imaging or electronic content management system, a state agency or local government must create policies and procedures that define the normal operations and use of such systems. These written policies and procedures must be kept current, be quickly accessible if needed for training and legal situations, and include the following:

27.1.1. An overview of the system that describes the purpose and uses of the system; the methods used to create, modify, duplicate, and destroy records; the roles and responsibilities of those individuals involved in records imaging or creation, maintenance, and destruction; and systems in place to ensure consistent quality control and problem resolution.

27.1.2. Policies and procedures for training and support that include instructions for imaging, indexing, quality assurance, and retrieval, and which document all staff training relating to the use of the system.

27.1.3. Auditing systems (human, machine, or both) which verify no unauthorized deletions, additions, or changes have entered the system, and that support the state agency or local government's ability to identify the source of any such unauthorized action.

27.1.4. System performance assurance processes which routinely test the hardware and software and document system testing and performance issues.

27.1.5. System security protocols which limit system access and update privileges to appropriate personnel, prevent unauthorized modification of records, divide staff responsibilities to ensure that individuals with an interest in the contents of records are not responsible for administering tasks related to those records, and which include disaster preparedness and security backup procedures.

Appendix A

Additional Vendor Requirements

When a state agency or local government (the Customer) contracts with a vendor (the Vendor) to perform imaging of government records, additional criteria must be met.

28. Facilities Inspection

28.1. The Customer reserves the right to inspect and approve the Vendor's worksite before and at any time during the performance of a contract, to ensure the Vendor's production and quality control capabilities. Inspecting a Vendor's facilities is highly recommended.

29. Quality Assurance Process

29.1. The Vendor must describe its quality assurance process used to verify high-quality images of digitized documents. This quality process must include reviewing every page digitized. However, this content review of the imaging does not eliminate the need for the state agency or local government to conduct its own review of the digitized images.

29.2. A Vendor must keep a log of the digitized images, to track and record any issues or concerns with an item being digitized. The Vendor must assign a quality grade to original documents that are deemed of marginal to low quality, and result in poor quality images. The quality grade must be recorded in the digitized images log. The quality grade must have, at minimum, three gradations denoting one of the following:

29.2.1. *Perfect capture:* No data loss or image deterioration. No notation is required in the digitized images log in such cases.

29.2.2. *Good capture:* No data loss but some image deterioration. Customer may consider maintaining the original document in such cases.

29.2.3. *Poor capture:* Data loss and/or significant image deterioration. Customer must retain the original document in cases where a better image cannot be produced from the original.

29.3. The Customer has the absolute right, after inspection, to reject any images determined not to meet the requirements of these guidelines. In such cases, the contractor must re-digitize at its own expense.

29.4. The Vendor must inspect each individual image, disc, tape, or other storage medium for compliance with the requirements herein, including resolution, image quality, accuracy of the index, and general workmanship. The Vendor must include an inspection report or certification covering each disc, tape, or other storage medium or transfer protocol used for each shipment.

29.5. Unless otherwise specified, the Vendor must maintain the original documents in their existing file order before, during, and after imaging. The Vendor must return the files to the original storage containers in the same order that existed before imaging, except that

the Vendor must maintain an account of any corrections to file order made during the preparation for imaging. The Vendor must not restore any fasteners (staples, clips, tape, etc.) removed during document preparation.

30. Handling, Transfer, Storage, and Security

30.1. All documents must be treated with care to ensure that the records are not re-arranged, damaged, lost, destroyed, or stolen.

30.2. The Customer will provide the Vendor with a transfer sheet that includes an accounting of every box of records and its contents and which notes any missing records, damage to the records, or other known problems with the records in their original paper form. Both parties will sign this document at the point of transfer of the records from the Customer to the Vendor.

30.3. Unless already performed by the Customer, the Vendor will pack all records on site, and provide personal transportation for the records in both directions to and from their facilities to the Customer.

30.4. Under no circumstances are any documents or their contents to be shared, copied, or transferred to another organization or individual outside of the expressed intent of the project, or without prior written permission from the Customer.

30.5. The Vendor must describe the entire chain custody for documents from the initial pick up to final delivery of the digitized documents back to the Customer.

30.6. The Vendor must maintain a tracking system located in its facility for instant tracking of the Customer's shipment. The Vendor must explain the cataloging or tracking process used to ensure the same files and the same numbers of files are returned as were shipped out.

30.7. The Vendor must physically apply a tracking number onto each box and individual large document (defined as those greater than 11" x 17" in size) for inventory and tracking purposes. This number must be unique for each box and large document, and applied in a clean and professional manner.

30.8. The Vendor must make all efforts to apply the best industry practices to eliminate the risk of documents' being lost during transfer.

30.9. All work must be performed at the contracted Vendor's facilities, unless prior agreement is made with Customer.

30.10. Any removable media used to transfer images between the Vendor and the Customer must be properly labeled. At a minimum, the following information must be included on a label on each piece of storage media used to transfer images and index data:

- Customer Name
- Records Series Title and Date
- Range of Records (if appropriate)
- Package or Media Number

- Security Level (Low, Medium, High)

The Vendor must maintain control at all times over the physical and electronic public records provided to them, or created as a result of the work performed under the scope of the project.

30.11. The Vendor must store all documents in a secure environment, with protection from damage due to the elements, human error, and other unforeseen threats.

30.12. The Vendor's records storage facility must have a working security system that includes protection from burglary, fire, and other events that may cause harm to the records provided.

30.13. The Vendor's records storage facility must have protection from water damage due to outside elements.

31. Vendor Equipment Maintenance and Inspection

31.1. *General Guidelines:* The Vendor must regularly maintain and inspect its digital imaging equipment to ensure all of it is in good working order, and must maintain logs of such activities. These equipment maintenance logs must note any problems identified with each piece of equipment, and the steps taken to eliminate each problem. When applicable, the Vendor must follow manufacturer's guidelines for equipment maintenance, or explain why its differing processes are more reliable than those of the manufacturer.

31.2. *Scanner and Camera Maintenance:* The Vendor will provide the Customer at the Customer's request a copy of the equipment maintenance logs demonstrating that the Vendor is regularly inspecting all scanners and cameras used for digital imaging at least monthly, to ensure they are clean and in good working order.

31.3. *Modulation Transfer Function:* The Vendor must test, at least annually, the Modulation Transfer Function (MTF) of the equipment it is using for digital imaging, to assess the imaging quality of its imaging system objectively. If the vendor discovers problems during this test, it must log the problem and its solution in its equipment maintenance logs and provide this information to the Customer at the Customer's request.

31.4. *Reference Targets:* To assess the functioning of digital imaging scanners and cameras, the Vendor must employ targets that include photographic reference standards (for assessing, at minimum, grayscale tone and color accuracy and lighting evenness).

31.5. *Lighting:* The Vendor must ensure the evenness of lighting for any scanner or camera used for digital imaging, and the Vendor must provide information to demonstrate how it ensures the evenness of lighting in any of its imaging equipment.

31.6. *Squaring Cameras to the Object:* The Vendor must ensure that any planetary digital camera used in imagining is kept square to the object being captured, and the Vendor must be able to demonstrate and explain the validity of its method for achieving squareness when asked to by the Customer.

31.7. *Calibration of Monitors:* Any monitor used by the Vendor to view the results of digital images must be equipped with a colorimeter or similar device and its associated software and use these at least monthly to calibrate the monitor. Vendors must also visually assess the functioning of monitors periodically to ensure they are in good working order.

32. Imaging Process

32.1. All procedures must be performed by the Vendor in accordance with generally accepted standards of conservation and conversion practices. Alterations, changes, or the insertion of any new material in any record is strictly forbidden.

32.2. The Vendor must provide bound volume imaging services that allow the bound volume to remain intact as a part of their service, unless directed otherwise by the Customer.

33. Access During Conversion Processing

33.1. The Vendor must complete the project in a manner that will maximize the Customer's access to records at all times. The Customer requires that any files in the possession of the Vendor as a part of this project which are needed prior to the completion of the conversion must be returned to the requesting party within forty-eight (48) hours from the initial request. An emailed copy of the image of the requested document will be acceptable, if the document is not legally restricted. A faxed copy will also be acceptable so long as the quality of the fax is deemed adequate by the Customer upon receipt.

33.2. The Vendor will allow access to the Customer's records only by individuals authorized in writing by the Customer prior to the onset of the project or authorized later, also in writing, if so required.

34. Post-Processing

34.1. All files and folders must be placed back into the original boxes and returned to a predetermined storage area.

34.2. All boxes of files must be returned in the same condition and document order as provided, minus staples, paper clips, and other fasteners. If the Customer has ensured the completion of the content and quality review of the digitized files, the Vendor may conduct or arrange for the shredding of paper files at the written direction of the Customer.

34.3. The Vendor must provide written verification that no files were damaged and/or lost and that all information on all records remained confidential while under its control.

34.4. A digitized images log must be maintained for each conversion project. The log may be created as a paper or electronic document, and it must be kept for at least six months to support problem resolution

Appendix B

Table of Technical Guidelines

This table provides a summary of the most important of the technical requirements of these guidelines, but it should not be used in place of them.

Topic	Guideline	Notes	Ref.
Image Versions	• A copy of the master image must always be stored outside of the imaging system • Access image must be in a universally acceptable format • Thumbnail must be in a universally acceptable format		9
Compression	• Lossless only o ITU-U o JPEG 2000		10
File Formats for Master Digitized Images			11
Records with retention periods of 10 years or more	• TIFF (preferred for photographic records) • PDF/A (preferred for textual records)	Conformance level u for PDF/A preferred	11.1
Records with retention periods of less than 10 years	• TIFF • PDF/A • PDF		11.2
Headers	• No proprietary headers allowed		
Imaging Resolution	• Minimum resolution of 300 ppi • Image ratio of 1:1		12
Color Mode	• RGB or CMYK	CMYK preferred for documents that may need to be printed in high quality	13
Small Textual Documents	• 300 ppi bi-tonal (1-bit) • 300 dpi, in 8-bit grayscale or 24-bit color mode, for damaged, stained, or poorly legible, or illustrated documents, or documents with meaningful color • Test images must be created before wholesale imaging	Multi-page documents as one file preferred	15
Maps and Plans (and other documents larger than 11" X 17")	• 300 ppi in bi-tonal, grayscale, or color mode • Test images must be created before wholesale imaging	Use up to 600 ppi to capture fine detail if necessary	16

Bound Volumes	• 300 ppi bi-tonal (1-bit) • 300 dpi, in 8-bit grayscale or 24-bit color mode, for damaged, stained, or poorly legible, or illustrated documents, or documents with meaningful color • Test images must be created before wholesale imaging	Curvature correction allowed post-digitizing	17

Topic	Guideline	Notes	Ref.

Photographic Records	• Black and white photos
	  o 8-bit grayscale
	  o 800 ppi for prints when the longest side is up to 5.5"
	  o 400 ppi for larger prints up to 11.5" long
	  o 300 ppi for prints larger than this
	• Damaged or faded black and white photographs
	  o 16-bit grayscale
	  o 24-bit color mode
	• Color photographs
	  o 24-bit color mode
	• Positive Film
	  o 2800 ppi for 35mm
	  o 1600 ppi for 120 and 220
	  o 16-bit grayscale for black and white
	  o 24-bit color more for color
	• Photographic negatives
	  o 2800 ppi for 35mm
	  o 1600 ppi for 120 and 220
	  o 16-bit grayscale for black and white
	  o 24-bit color more for color
	  o Reverse polarity during imaging
	• Glass plates, positive and negative
	  o 800 ppi for 3¼" X 4¼"
	  o 600 ppi for 4" X 5"
	  o 400 ppi for 8" X 10"
	• Test images must be created before wholesale imaging
			18

Microforms	• Resolution
	  o 5600 ppi for 16mm images
	  o 2800 ppi for 35mm images
	• Bit Depth
	  o 16-bit grayscale for black and white
	  o 16-bit grayscale for continuous tone
	  o 24-bit color for color
	• Test images must be created before wholesale imaging
		Use higher resolution and bit depth for microfiche images and micrographic images of poor quality	19

Image Enhancement	• Acceptable
	  o De-skewing
	  o Cropping out image data beyond the document
			20

Topic	Guideline	Notes	Ref.
Image Enhancement (continued)	○ Rotating ○ Correcting curvature • Unacceptable o Any technique that alters the existing content of the record • Annotations or sticky notes adding during imaging must be separate from the image of any record		20
Indexing	• Minimums ○ Unique identifier for each document across all storage media ○ Indexing fields chosen for each set of records ○ Indexing data stored in a non-proprietary format ○ Unique identifier ties each digital image to its indexing data • Archival Indexing ○ Include more metadata fields for high-use archival or long-term records • Optical Character Recognition ○ 95% accurate as measured by character count ○ OCR will not be used as the sole finding aid ○ Editing of OCR 'd text allowed • Directory Structure ○ Files will be organized in a file directory or folder system ○ This directory will link to metadata stored in the database ○ Folders may have an organization independent of the image files	• Use objective indexing fields whenever possible • Fuzzy searching encouraged	21

Topic	Guideline	Notes	Ref.
Production Metadata	• Minimums o Unique identifier o Title of records series o State Archives Retention Schedule Item Number (from the State General Schedule or a specific local government schedule) or Records Disposition Authorization Number o State agency or local government name o Name of the imaging vendor or government staff person conducting the imaging o Date of the imaging o Pixels per inch (ppi) o Equipment used to capture the images o Software used to capture the images		22
Quality Assurance	• Verify images for the following: o Correct image file naming convention, as agreed upon o Correct file format (including verification of compliance with the PDF/A format for purported PDF/A files) o Quality of image is the same as in the original o Correct size and resolution o Image digitized at appropriate ppi for each image type o Proper reading orientation (landscape or portrait) o Image is not skewed o Image is neither too light nor too dark o Curvature of the page does not obscure or distort the text o Appropriate contrast within the image o No distortion of the image o No extraneous materials (sticky notes, fasteners, etc.) obscure the image o No additional information added to the image that is not part of the		23

Topic	Guideline	Notes	Ref.
Quality Assurance (continued)	original document o Appropriate indexing terms associated with the digitized image ◎ Acceptable correction of digitized images o Correcting image filename o De-skewing, rotating, or flipping the image to correct its orientation o Adjusting brightness, contrast, or tone through re-digitizing o Curvature correction that does not obscure or distort the original image and that captures all data in the record o Reversing polarity for photographic negatives or negative microfilm o Re-digitizing, followed by a re- inspection of the new image o Updating index database to correct errors ◎ Index Accuracy o Verify index with the goal of achieving 100% accuracy. ◎ Content Verification o Page-by-page verification of 100%		23

Topic	Guideline	Notes	Ref.

Storage	• Transfer media (from imaging source to customer)
	• CD-Rs
	• DVD-Rs
	• External hard drives
	• USB drives
	• Transfer Media (from state agency to State Archives)
	  o Contact the State Archives' Electronic Records Unit
	• Storage Media
	  o Server-class hard drives preferred
	  o Using RAID configuration preferred
	  o RAID 5 or higher preferred
	• Media Storage Environment
	  o Secure
	  o Dust-free area
	  o 65 to 72 degrees Fahrenheit
	  o 30% to 50% relative humidity
	• Prevent unauthorized access to confidential data via use of
	  o Data encryption
	  o Locked media storage containers
	  o Other secure method of transfer
	• Backup
	  o Periodic
	  o Geographically remote offsite location
	  o Annual testing of the backup media
			24

Data Maintenance	• Records must be retained until their retention requirements have been met
	• Data Integrity
	  o Protect against file corruption, alteration, or deletion
	  o Integrity implies that the data remains an exact copy of the original over time
	• Checksums
	  o Conduct when data written to a storage medium
	  o Recalculate checksum anytime data is read
	  o Use a disk-error checking utility annually
	• Media Refreshing
			25

Topic	Guideline	Notes	Ref.
Data Maintenance (continued)	○ Move records from one storage medium to another every three to five years ◎ Tape-based Media ○ Rewind tapes at least once every three years or less depending on the tape quality, type, storage conditions, and manufacturer's recommendations. ○ Read a random sample of 10% of all tapes annually to identify loss of data and to correct the causes of data loss ○ Data maintained on tape must be copied onto new tape a minimum of once every three years Migration ○ Plan for future migrations to new media and systems ◎ ○ Migrate stored images and associated metadata to a newer media platform every five years		25

Topic	Guideline	Notes	Ref.
Equipment Maintenance and Inspection	• Regularly maintain and inspect imaging equipment	Recommend that guidelines for vendors, outlined in section 31, are followed	26
System Trustworthiness	• Policies and procedures that define the normal operations and use must be kept current o Overview of the system o Training and support o Auditing systems o System performance assurance processes o System security protocols		27
Quality Assurance Process	• Vendor must maintain a digital images log • Customer has absolute right to reject images based on quality • Vendor must inspect all images and storage media for compliance with guidelines • Vendors must retain original records in original order		29
Handling, Transfer, Storage, and Security	• Customer will provide vendor with tracking sheet accounting for every box of records • Vendor may not share contents of documents with anyone without express written permission • Vendor must describe entire chain of custody of documents • Vendor must maintain a tracking system • Vendor must apply a tracking number onto each box • Work must be performed at contracted vendor's facilities • Removable media from vendor to customer must be properly marked • Vendor must store documents in secure environment • Vendor's storage facility must have a working security system • Vendor's storage facility must have protection from water damage		30
Vendor Equipment Maintenance and Inspection	• Vendor must maintain and inspect its digital imaging equipment, including equipment logs		31

Topic	Guideline	Notes	Ref.
Vendor Equipment Maintenance and Inspection (continued)	• Vendor must test the Modulation Transfer Function of its equipment at least annually • Vendor must use reference targets to assess the functioning of its digital image capture equipment • Vendor must ensure evenness of lighting of any of its digital imaging equipment • Vendor must have means to ensure the squaring of cameras to any object being captured digitally • Vendor must ensure the calibration of its monitors		31
Access During Conversion Process	• Vendor must return to the customer any within 48 hours any documents requested by the customer		33
Post-Processing	• Vendor must place all files and folders back into their original boxes • Vendor must return boxes of files to the customer in their original condition • Vendor must verify in writing that no files were damaged or lost • Vendor must maintain its digital images logs for at least six months		34

Appendix C

Available Standards and Technical Reports

The following is a list of industry standards and technical reports available for purchase:

- AIIM ARP1-2009, *AIIM Recommended Practice—Analysis, Selection, and Implementation of Electronic Document Management Systems (EDMS).* Silver Spring, Md.: Association for Information and Image Management, 2009.

- 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.

- AIIM MS60-1996, *Electronic Folder Interchange Datastream.* Silver Spring, Md.: Association for Information and Image Management, 1996.

- 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 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.

- ANSI/AIIM TR39-1996, *Guidelines for the Use of Media Error Monitoring and Reporting Techniques for the Verification of Stored Data on Optical Digital Data Disks.* Silver Spring, Md.: Association for Information and Image Management, 1996.

- ANSI/AIIM 25: 2012, *Assessing Trusted Systems for Compliance with Industry Standards and Best Practices.* Silver Spring, Md.: Association for Information and Image Management, 2012.

- ANSI/AIIM/ARMA TR48-2004, *Framework for Integration of Electronic Document Management Systems and Electronic Records Management Systems.* Silver Spring, Md.: Association for Information and Image Management, 2004.ISO 12231:2012, *Photography—Electronic still picture imaging—Vocabulary.* Geneva, Switzerland: International Organization for Standardization, 2003.

- ISO 12233:2000, *Photography—Electronic still-picture cameras—Resolution measurements.* Geneva, Switzerland: International Organization for Standardization, 2000.

- ISO 14524:2009, *Photography—Electronic still-picture cameras—Methods for measuring opto- electronic conversion functions (OECFs).* Geneva, Switzerland: International Organization for Standardization, 2009

- ISO 16067-1:2003, *Photography—Spatial resolution measurements of electronic scanners for photographic images—Part 1: Scanners for reflective media.* Geneva, Switzerland: International Organization for Standardization, 2003

- ISO 16067-2:2004, *Photography—Electronic scanners for photographic images—Spatial resolution measurements—Part 2: Film scanners.* Geneva, Switzerland: International Organization for Standardization, 2004

- ISO 19005-1, *Document Management—Electronic Document File Format for Long-term Preservation—Part 1: Use of PDF 1.4 (PDF/A-1).* Geneva, Switzerland: International Organization for Standardization, 2005.

- ISO 19005-2:2011, *Document management—Electronic Document File Format for Long-term Preservation—Part 2: Use of ISO 32000-1 (PDF/A-2).* Geneva, Switzerland: International Organization for Standardization, 2011.

- ISO 19005-3:2012, *Document management—Electronic Document File Format for Long-term Preservation—Part 3: Use of ISO 32000-1 with support for embedded files (PDF/A-3).* Geneva, Switzerland: International Organization for Standardization, 2012.

- ISO 21550:2004, *Photography—Electronic scanners for photographic images—Dynamic range measurements.* Geneva, Switzerland: International Organization for Standardization, 2004

Additional Industry standards are available at the International Organization for Standardization (ISO): www.iso.org.

Appendix D

Sampling Methods for the Verification of Digital Images

35. Sampling Methods

35.1. Local governments and state agencies may use the ANSI/AIIM standard AIMM TR34: *Sampling Procedures for Inspection by Attributes of Images in Electronic Image Management and Micrographic Systems* (36 CFR 1237.28 (d)(2)) developed by the Association for Information and Image Management.

35.2. Or local governments and state agencies may use a simpler method developed by the New York State Archives

35.2.1. Review the first one hundred images in the batch.

35.2.2. If there is an error rate of 3% or less, then review every 25th image in the batch to determine if the whole batch appears to have a consistent quality. If you find an error while checking every 25th image be sure to check a few images immediately before and after the error to determine if the problem is more widespread or just a lone error.

35.2.2.1. If it appears the errors are more widespread then return to checking clusters of 100 images at the point where you found the error. If the error rate exceeds the accepted threshold for that cluster, then review another cluster of 100. If that one exceeds the accepted error threshold, then reject the whole batch. If it doesn't, return to checking every 25 images.

35.2.2.2. If you detect few, if any, errors on every 25th image, then you may accept the whole batch. If you find more errors while checking every 25th image, then reject the whole batch.

35.2.3. If the error rate is over 3%, then verify the next cluster of 100 images. If the error rate is still over 3% then review one more cluster of 100 images, and if it exceeds the error rate, then reject the whole batch. If the third cluster does not exceed the error rate, return to checking every 25th image and follow the instructions from 35.2 2 2 above.

35.2.4. If the error rate falls below 3% for the next cluster of 100 images, return to checking every 25th image and follow the instructions from 35.2 2 2 above.