Office of the Public Records Administrator
Connecticut State Library
State of Connecticut

Digital Imaging Standards, 2014

These standards shall be read together with *Public Records Policy 2: Digital Imaging* to ensure a full understanding of the Office of the Public Records Administrator (OPRA) policy regarding the use of digital imaging technology for the reformatting of analog public records; and regarding the retention and disposition of original and digitized records.

For the purposes of this document, the term “shall” indicates a requirement and the terms “should” and “may” indicate a recommendation or best practice.

1. Legal Issues

1.1. Any agency contemplating using digital imaging technology for the reproduction of public records shall be aware of all applicable statutes or regulations and any legal issues. Consultation with appropriate legal counsel regarding rules of evidence and any other legal issues is advisable.

1.2. References to electronic records or digital imaging can be found in many sections of the *Connecticut General Statutes*, including but not limited to, sections contained within Chapter 3, *Public Records: General Provisions;* Chapter 15, *Connecticut Uniform Electronic Transactions Act*; Chapter 92, *Town Clerks;* and Chapter 899, *Evidence*.

2. Digital Imaging Defined

2.1. For the purposes of this document, digital imaging is the process of converting original records on paper or film into electronic images. 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 of 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.

2.2. It is recommended that agencies implement and use an electronic content management system (ECMS) to properly manage digitized images and other electronic records. An ECMS provides the ability to capture, store, retrieve, display, and transmit records electronically. An ECMS uses a database to manage descriptive information about the images that aid in the retrieval of records contained in the ECMS repository. Storing digitized records outside of an ECMS is not recommended due to the greater chance of accidental deletion of these records and lack of an audit trail to ensure a record’s authenticity. State agencies shall consult with the Department of Administrative Services (DAS)/Bureau of Enterprise Systems and Technology (BEST) regarding appropriate ECMS technology.

2.3. 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. Public agencies should refer to the current microfilm policy issued by the Public Records Administrator for microfilm standards.

2.4. For the purpose of these standards, there are two versions of a digital image:

2.5. *Master Image:* The master image is captured at the highest practicable quality or resolution and maintains the essential features and information of the original. The master image is the fixed, unalterable image from which all others are derived. The master image shall be created and maintained in compliance with *Public Records Policy 2: Digital Imaging* and these standards.

2.6. *Access or Derivative Image:* The access or derivative image is a use copy of the master image. The access or derivative image can be smaller in file size, of lower quality, and faster to access (for use on websites, for example). The agency may create an access or derivative file in any universally accessible file format.

3. Authenticity

3.1. Authentication is necessary for legal, audit, or administrative purposes to ensure that imaged records are acceptable as evidence. The authenticity of digital images is based on a sound process that includes well documented procedures, proper indexes, quality assurance controls, reliable backup and storage practices, and tested and audited processes to ensure the imaged records have not been altered, destroyed, or replaced. Additionally, having sound processes in place includes ensuring that any hardware, software, or other equipment used as part of the imaging process is functioning properly and is not introducing errors into the system.

4. File Format

4.1. File format directly affects the quality and file size of the images. Choosing the best file format for the agency's needs requires knowing how the images will be used, the type of materials that will be imaged, the desired speed of delivery, and the necessary quality of the images.

4.2. Digitized images shall not have proprietary headers, which make the images inaccessible except in particular software environments.

4.3. The agency shall use one of the following lossless image file formats, which support black and white, grayscale, or color documents:

4.3.1. *TIFF:* TIFF-formatted files shall be based on TIFF 6.0 baseline specifications. Refer to http://partners.adobe.com/public/developer/tiff/index.html for additional information.

4.3.2. *PDF*: PDF will be an acceptable format but PDF/A is the preferred version of PDF intended for the long-term preservation of electronic files. Sharing similar characteristics to PDF format, PDF/A is the preferred version of PDF for long-term retention since all requirements to render the file are contained within the file itself, while a basic PDF file may rely on linked fonts, non-standard header information, or other data stored outside of the file. PDF/A-formatted files shall utilize lossless compression, either CCITT Group IV or JPEG 2000. Refer to http://www.iso.org for additional information on the most recent versions of PDF/A that will meet agency requirements.

4.3.3. Any other file format meeting the requirements of *Public Records Policy 2: Digital Imaging* provided the agency obtains written approval of the Public Records Administrator in advance.

4.3.4. Refer to the Federal Agencies Digitization Guidelines Initiative (FADGI) for a comparison of file formats, including cost and system implementation factors, and settings and capabilities: http://www.digitizationguidelines.gov/.

5. Compression

5.1. 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.

5.2. Any compression technique used by the agency shall be a non-proprietary, lossless compression method that does not remove data or otherwise alter the appearance of the original image.

5.3. 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. Note that many scanners and multi-function devices default to using lossy compression and the settings may need to be adjusted to meet this requirement.

6. Imaging Resolution

6.1. The quality of a digital image is dependent upon the initial scanning resolution. The agency shall confirm that the selected resolution is sufficient to permit the recording of an accurate image as legible as the original public record. 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.

6.2. The recommended minimum resolution level for standard business documents is 200 pixels per inch (ppi). The recommended minimum resolution for the use of optical character recognition (OCR) processing is 300 ppi. Higher levels of resolution ranging from 300 to 600 ppi may be required for some records, including smaller, damaged, or low-contrast documents. Test scans make it possible to test the level of resolution needed.

7. Image Pixel Bit Depth

7.1. 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.

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

8. Color Mode

8.1. Color mode for digitizing color material can be either RGB (Red, Green, Blue) or CMYK (Cyan, Magenta, Yellow, Black), but CMYK is preferred for documents that in the future may need to be printed in high quality, such as in books or magazines.

9. Digitizing Text Documents

9.1. *Black and White Text Documents:* Text documents with good contrast between the printed text and the paper background should be digitized in bitonal mode, which is one bit per pixel.

9.2. *Lower Quality Documents:* Documents with handwritten notes or markings, low contrast between the text and the paper background, poor legibility, or illustrations included as part of the document should be digitized in grayscale mode (8 bits per pixel, representing 256 shades of gray) or in RGB or CYMK 24-bit color mode. Color mode will capture details and information better in cases where the original paper record is yellowed, stained, annotated with highlighter markers, or on onionskin paper.

9.3. *Multi-page documents:* When a record contains multiple pages (e.g., a single invoice comprises two pages), the document should be digitized as a single file, whether saving to a TIFF, PDF, or PDF/A formatted file. If the digital images are imported into an electronic content management system (ECMS), the images may be stored as single page files depending upon the system.

10. Digitizing Microfilm or Microfiche

10.1. At times the agency may need to convert records stored on microform (microfilm or microfiche) to a digital raster image to allow for access. Due to the photographic limitations of microforms and the variable quality of older microforms, it may be difficult to produce reproduction quality digital image files from all microforms. To produce images of adequate quality, the approach used to digitize microforms may vary from the recommendations cited here for paper-based records. These variations can include digitizing at a higher resolution, digitizing in grayscale greater than 8 bits per pixel, converting a negative image (white text on a black background) to a positive image (black text on white background), and making additional adjustments in contrast and brightness to create legible digital images.

10.2. Before imaging microforms, the agency, or alternatively the vendor chosen to conduct the imaging, should produce an inventory that records the location of the master negatives, duplicate negatives, if any, and service copies, the type of microforms (16mm, 35mm, microfiche), film base (nitrate, cellulose acetate or polyester), the length if applicable (100 or 215 feet), general film quality (resolution, density, image spacing, and the existence of targets, blips, splices, and scratches or other damage), and general image quality (such as blurry images). This inventory will allow for an assessment of the difficulty of imaging the microforms. The agency should discuss with the vendor the advantages of scanning the master or duplicate negatives against the disadvantages.

10.3. Refer to *General Letter* 96-2 *Standards for Microfilming Public Records* for the current microfilm standards of the Office of the Public Records Administrator.

10.4. A test digital image of the microfilm or microfiche intended to be digitized should be performed prior to a wholesale conversion to ensure a quality image can be produced.

11. Digitizing Photographs and Still Images

11.1. Photographs require sufficient resolution to capture all the significant detail in the originals. Although 300 ppi will be adequate for many photographs, higher resolutions may be needed for smaller photographs that will be enlarged or ones with fine detail. Damaged photographs should be digitized at a minimum of 600 ppi to capture sufficient detail.

11.1.1. *Black and White Photographs:* Digitize in 8-bit grayscale mode, representing 256 shades of gray for continuous tone black-and-white photographs. Damaged black and white photographs should be digitized in 16-bit grayscale mode to capture sufficient detail.

11.1.2. *Color Photographs:* Digitize in RGB or CMYK 24-bit color mode, representing 16.8 million colors for color photographs.

11.2. Positive film (such as 35mm slides) and photographic negatives require digitizing at a higher resolution to capture all available detail given their smaller format. Although 2,000 ppi is adequate for some applications such as web publishing, it is recommended that film be digitized at resolutions between 3,000 and 6,000 ppi based on anticipated use of the resulting digital images. Best results are generally achieved by digitizing positive and negative film at the device's highest optical resolution within these parameters.

11.3. Refer to Federal Agencies Digitization Guidelines Initiative (FADGI) for the most up to date standards for digitizing photographs and still images: http://www.digitizationguidelines.gov/.

12. Digitizing Maps and Plans

12.1. A resolution of 300 ppi is 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.

12.2. Simple black and white maps and plans with good contrast can be digitized in bitonal mode. Documents with shades of gray but not color should 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 RBG or CYMK 24-bit color mode.

13. Digitizing Bound Volumes

13.1. When a bound volume cannot or will not be disbound, the page of that volume should 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 should be retained after imaging.

13.2. Bound volumes should be digitized by use of a book scanner or camera that holds the volume open at an angle that reduces the curvature of the pages of 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.

13.3. The resolution and pixel depth for digitizing bound volumes should follow the recommendations for text documents in Section 9 and maps and plans in Section 12 of this document.

14. Image Enhancement

14.1. Image enhancement may call into question the integrity of the digital images produced. The use of digital enhancement techniques commonly used in imaging software shall only be permitted so far as the information on the record is not altered or compromised by such processes.

14.2. If adequate image quality cannot be obtained without unacceptable enhancement techniques, then the paper or film shall be retained.

14.3. Proper image orientation for optimal viewing should be maintained: portrait or landscape orientation as appropriate.

14.4. Annotations or “sticky notes” added during the imaging process should be separate from the image and not burned onto the image file itself. Annotations can be text, redactions, or “stamps.” Stamps are preconfigured overlays, such as but not limited to, “Paid,” “Original,” or “Signed.” These are generally added post-scan as overlays to the scanned images. Stamps and text annotations can be moved so content beneath the stamp or text box can be viewed.

14.5. An ECMS may allow users to save a rendition of the image with a redaction. Although an agency may choose to create or save images with redactions (in order to fulfill FOIA requests, for example), the agency is responsible for maintaining the unredacted master image.

15. Indexing and Image Retrieval

15.1. Since digitized images do not have native intelligence within them indicating their contents, appropriate index information is required to properly identify and later retrieve digitized images. For digital images, indexing is essential for locating and retrieving stored imaged records.

15.2. Indexing typically consists of a structured format and controlled vocabulary that allows more precise description of a record's content. Index data often includes information such as record type, creation date, record creator, and disposition date, among other information.

15.3. The agency shall create and maintain accurate indexes to properly identify and retrieve digital images. The agency shall be responsible for defining the specific indexing requirements needed to access the records efficiently prior to the performance of any imaging.

15.4. Indexing shall comply with the specific requirements of the agency and include a unique identifier for each image.

15.4.1. *Unique Identifier for Images:* Each image should have a unique filename or other identifier, preferably sequential, which can be numeric, alphanumeric, or alphabetic as required by the agency. Each filename should be unique across all records series, separate storage media, and electronic folders.

15.5. *Indexing Fields:* The index of images should consist of a limited number of field names to ensure adequate access to the records. Whenever possible, the field data should consist of objective indexing terms (such as personal names, file numbers, retention schedule numbers, and dates) from a controlled vocabulary, rather than subjective data.

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

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

15.8. *Optical Character Recognition:* Optical Character Recognition (OCR) can be performed to convert digital images into electronic text. Due to error rates, OCR should not be used as the sole tool for the retrieval of digital images, and it is not a substitute for indexing and production metadata.

16. Production Metadata

16.1. During the imaging process, production metadata shall 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.

16.2. The agency shall maintain production metadata for the life of the records.

16.3. Production metadata shall include a unique identifier, and may include additional metadata, such as but not limited to: agency name, title of records, retention schedule item number, name of the imaging vendor or agency conducting the imaging, date of the imaging, and pixels per inch (ppi).

16.4. *Directory Structure:* Regardless of the image file name, files will likely be organized in some kind of 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 record group number, or they may replicate the physical or logical organization of the originals being digitized.

17. Quality Assurance and Control

17.1. For the purpose of this section, quality assurance refers to administrative and procedural activities implemented by the agency for the creation of accurate and authentic images and accurate indexes and production metadata. Quality control refers to the process by which the agency reviews the quality of all steps involved in the creation of imaged records.

17.2. Quality assurance procedures shall be in place to ensure the creation of accurate and authentic images and accurate indexes and production metadata that follow these standards, as well as ensuring that the specific requirements of the agency are met.

17.3. The quality control process shall be documented and maintained throughout the digitization conversion process. Quality control information to be documented includes, but is not limited to, problem resolution procedures and reporting requirements for each step of a conversion project.

17.4. Quality control shall be conducted by the agency before requesting authorization for the destruction of any original documents, including, but not limited to, visual inspection the digitized documents to ensure clarity, readability, and accurate representation of the original record and checking the indexing field against the original or imaged record.

17.5. Other quality control steps for digitized images may include, but are not limited to:

- 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)

- Correct size and resolution
- Image digitized at appropriate ppi for each image type
- Proper orientation (landscape or portrait)
- Image is not skewed
- Image is neither too light nor too dark
- 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
- All pages are present and in proper order
- Appropriate indexing terms associated with the digitized image

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

17.7. *Index Accuracy:* Index data shall be verified to ensure accuracy. Industry 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 other means as appropriate to ensure index accuracy. Best practices dictate that a person or persons other than those digitizing or indexing a particular record should perform the quality assurance procedures. 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 second review is performed in a separate process by another individual.

18. Transfer and Storage Media

18.1. For the purpose of these standards, there are two categories of media:

1. Transfer Media: Media used to transfer images and index data from the imaging source (for example, from the vendor hired to digitize documents) to the agency.

2. Storage Media: The digital repository (preferably, an ECMS) used by the agency to store the images.

18.1.1. *Transfer Media:* Transfer media is intended only for short-term storage while moving images and index data from the imaging source (such as a vendor hired for imaging services) to the agency’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.

18.1.2. The use of transfer media (such as CD-Rs, DVD-Rs, external hard drives, USB-drive media) shall not be permitted for the long-term storage of digital records because of media instability and fragility.

18.1.3. *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.

18.1.4. 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 since the last backup process was performed.

18.1.5. Any storage media used shall comply with the applicable International Standards Organization (ISO) standards, which specify how the media type should store information. Refer to http://www.iso.org for additional information on ISO standards.

18.2. Storage media should be kept in a secure, dust-free area under proper environmental conditions. For additional information regarding proper environmental conditions, refer to http://www.iso.org.

19. Cloud Storage

19.1. Agencies are required to manage their records through records retention schedules pursuant to CGS 11-8a. Regardless of which cloud computing service and deployment models are adopted, agencies are still required to comply with this statute. Variations among cloud computing services and deployment models, however, will affect how and by whom (agency/contractor) records management activities can be performed.

19.2. Any data sourced to a cloud shall remain legal property of the agency and this should be clearly articulated in a service level agreement with the provider. The agency should have an exit strategy to protect its data in the event the storage contract or relationship is terminated with the provider.

19.3. Cloud providers may not be able to easily meet the type of security and/or information access controls that satisfy state and/or federal regulation. Cloud providers may desire to charge additional costs to cover unique or specialized security requirements. Further, the agency may have little control over who among the cloud provider’s employees is authorized to agency data.

19.4. Storage of data outside of the physical and/or legal boundaries of the state may compromise the agency’s ability to manage and control its data.

19.5. State agencies shall obtain written approval of the state’s Chief Information Officer before committing to the services of a cloud vendor.

19.6. For guidance and best practices refer to the DAS/Bureau of Enterprise Systems Technology (BEST) publication *Use of Public Cloud Service by Executive Branch Agencies* and the National Institute of Standards and Technology (NIST) Special Publication 800-145 *The NIST Definition of Cloud Computing*.

20. Back Up Copies

20.1. The agency shall perform periodic backups of all digitized records, associated indexes and production metadata to ensure the continued accessibility of records in the event of a disaster. The agency shall perform and certify annual testing of the backup media to ensure all digitized records have been backed up and are readable.

20.2. A backup copy shall be stored in a location that is geographically remote from the location where the use copies of the records are stored. The necessary distance will vary based on an analysis of the geological, meteorological, and human risks present at the agency location where the records are used. An appropriate backup location is one where it is highly unlikely that the backup location will simultaneously suffer the same disaster as the agency offices. For example, if the agency is located in or near a flood plain, the backup location should be in an area that is away from that flood plain.

20.3. Regardless of the backup method used, the purpose of the backup is to ensure the digital images, associated indexes, and production metadata remain accessible through the records' entire retention period.

21. Maintenance

21.1. Digital images, associated index data, and production metadata shall be effectively and efficiently managed throughout the designated retention period.

21.2. *Data Integrity:* All stored records shall be protected against file corruption, alteration, or deletion throughout the designated retention period. Integrity means that the image is an exact copy of the original and that the data has not been altered through loss, tampering, or corruption. The agency shall have policies and procedures in place to ensure the integrity of the digital image.

21.3. All stored data files shall be checked regularly for integrity according to agency policy, such as using a disk-error checking utility which is built into most operating systems such as Microsoft Windows. 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.

21.4. *Migration*: Due to the technological advances and the potential obsolescence of technology currently in place, the agency shall 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 the older hardware is no longer in use to read the older media. 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 remain accessible, the agency shall migrate all digital images and their associated metadata to a newer media platform as needed.

21.5. The agency should establish a migration plan *before* imaging and review the plan annually. The reality of technological obsolescence requires that the agency 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.

21.6. The cost of migration should be factored into the agency's operating expense budget, as migration is an ongoing expense that may grow substantially with time depending on the storage medium.

22. System Trustworthiness

22.1. To ensure the trustworthiness of any imaging or electronic content management system, the agency shall create policies and procedures defining the normal operations and use of such systems. These written policies and procedures shall be kept up to date, be quickly accessible if needed for training and legal situations, and include the following:

22.1.1. 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.

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

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

22.1.4. System performance assurance processes that routinely test the hardware and software and document system testing and performance issues.

22.2. The system shall include security protocols that limit system access and update privileges to appropriate personnel, prevent unauthorized modification of records, and include disaster preparedness and security backup procedures.

22.3. All security controls required by regulation, policy, and/or law for paper records shall be addressed for the digital image unless those regulations, policies, and/or laws state otherwise. The agency shall ensure the protection of records that contain confidential or sensitive information.

23. Disposition of Digital Images

23.1. The disposition of digital images shall be in accordance with *Public Records Policy 2: Digital Imaging*.

23.2. The agency shall have documented policies and procedures that specifically address the destruction of digital images.

23.3. These practices shall be consistent with the agency's procedures for the lawful disposition of public records in other formats and should follow a regular and systematic disposition schedule.

23.4. *Confidential and Sensitive Information:* Digital images shall be destroyed in a manner that ensures that any information that is confidential or sensitive, including proprietary or security information, cannot practicably be read or reconstructed. Recorded media previously used for electronic records containing information that is confidential or sensitive, including proprietary or security information, shall not be reused.

23.5. Refer to National Institute of Standards and Technology (NIST) Special Publication 800-88 *Guidelines for Media Sanitization* for additional information.

Appendix A: Recommendations for Digital Imaging Planning and Implementation

24. Project Planning

24.1. The key to the successful design, integration, and implementation of a digital imaging project is thorough planning. Imaging projects require a combination of skills from a variety of staff with different areas of expertise, including those with information technology, records management, and legal expertise. Digital imaging is most efficient for large quantities of records that are frequently retrieved. Records that become inactive after a very short period of time may not be appropriate candidates for digital imaging.

24.2. Implementing digital imaging will significantly impact work processes by requiring staff to create, retrieve, share, use, and store documents in different ways. Existing operations should be analyzed to identify the type and volume of public records that are best suited for digital imaging, including a review of content, legal requirements, use levels, and work processes.

25. Cost Analysis

25.1. A cost-benefit analysis should be conducted to determine the cost justification and possible benefits to the agency with the implementation of digital imaging. The cost of implementing digital imaging depends on several key components of an information system such as hardware, software, and support, including the reengineering, training, maintenance, and management associated with the system.

25.2. The cost of converting short-term records may be greater than storing the records in a paper format until they are eligible for disposal. If the agency has a high volume of short-term records, off-site storage can be an economical alternative to keeping the records in the office or scanning them.

26. Implementation

26.1. An implementation plan may include setting up a pilot project, which will allow the agency to test the technology, examine the effectiveness of using digital images to provide and manage information for staff, and help determine how to better implement a digital imaging project. A pilot project demonstrates the impact and effectiveness of imaging before undergoing a large digitizing project for an entire agency.

26.2. The use of phases is an effective approach to implementing large scale digitization projects. Rolling out the project in phases enforces an organized and careful approach to implementation. It allows small errors to be caught and corrected before they escalate into large and costly issues. Phases can be applied in several ways depending upon the structure of the organization and scope of the digital imaging project. For example, the project could be phased in by department or by function. If the project will be implemented over an extended period of time, it is important that the agency begins with its highest priorities.

27. In-House vs. Outsourcing

27.1. Evaluation of the agency's resources will help determine if the digitization process should be conducted in-house or if elements of the project should be outsourced to a vendor specializing in digital imaging.

27.2. While a vendor may be hired to assist with the project, many tasks will nevertheless remain the responsibility of agency staff, as further described within these standards. Agency staff responsibilities may include, but are not limited to, the selection and preparation of records to be digitized, the creation of standardized indexing attributes, the performance of quality assurance on source material and digitized images, and the ongoing management of storage, maintenance, protection and documentation for the records and system, and using appropriate contract standards.

27.3. In order to maintain the integrity of the agency's records, the agency shall retain ownership and access to all information maintained by the vendor either at the location of the agency or at the vendor's location. If a vendor digitizes records for an agency, it is the responsibility of the agency to ensure the vendor is in compliance with *Public Records Policy 2: Digital Imaging* and these standards.

27.4. See Appendix B for additional recommendations regarding vendor contracts.

Appendix B: Recommendations Regarding Vendor Contracts

If an agency contracts with a third-party vendor for imaging services, the agency shall ensure that the vendor is in compliance with all applicable sections of *Public Records Policy 2: Digital Imaging* and these standards.

A properly written contract shall be in place specifying the requirements of the intended project, the requirements *of Public Records Policy 2: Digital Imaging* and all criteria that shall be met.

For statewide contracts administered by the Department of Administrative Services (DAS), the contract shall specify the digital imaging standards and policies and require the contractor to ensure that it is in compliance with all applicable sections of the *Public Records Policy 2: Digital Imaging*. When using statewide contracts, the state agency shall include the requirements of their intended project in their statement of work/purchase order to the contractor.

For contracts administered directly by a state agency or municipality, the contract shall specify the requirements of the intended project, the requirements of *Public Records Policy 2: Digital Imaging* and all criteria that shall be met.

In addition to the information contained within the policy and these standards, the agency should consider the following factors when contracting with a vendor:

28. Facilities Inspection

28.1. The agency should have 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 should describe its quality assurance procedures to verify high-quality images of digitized documents. The quality control process should include visual inspection of the digitized documents to ensure clarity, readability, and accurate representation of the original record. However, this content review of the imaging does not eliminate the need for the agency to conduct this review itself as well.

29.2. The vendor should maintain a log to track and record any issues or concerns with an item being digitized. The vendor should assign a quality grade to original documents that are deemed of marginal to low quality and result in poor quality images. A quality grade should be recorded in the log. Grading can be negotiated between the agency and the vendor.

29.3. The agency should have the right, after inspection, to reject any images determined not to meet its requirements. In such cases, the vendor should redigitize at its own expense.

29.4. The vendor should inspect images, discs, tapes, or other storage medium for resolution, image quality, accuracy of the index, and general workmanship. The vendor should include an inspection report or certification covering discs, tapes, or other storage medium or transfer protocol used for each shipment.

29.5. Unless otherwise specified, the vendor should maintain the original documents in their existing file order before, during, and after imaging. The vendor should return file material to the original storage containers in the same order that existed before imaging, except that the vendor should maintain an account of any corrections to file order made during the preparation for imaging. The vendor should not restore any fasteners (staples, clips, tape, etc.) removed during document preparation.

30. Handling, Transfer, Storage, and Security

30.1. All documents should be treated with respect and care, as they are public records. If the records to be digitized are confidential, the agency shall specify its requirements for the protection of confidentiality.

30.2. The agency should provide the vendor with a transfer sheet that includes an account of every box of records and its contents and that notes any missing records, damage to the records, or other known problems with the records in their original paper form. Both parties should sign off on this document at the point of transfer of the records from the agency to the vendor.

30.3. Unless already performed by the agency, the vendor should pack all records on site, and provide transportation for the records in both directions to and from their facilities to the agency. The transportation of confidential or sensitive records shall not be contracted to a third party provider.

30.4. Under no circumstances should any documents or their contents 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 agency.

30.5. The vendor should document the entire chain of custody for documents from the initial pick up to final delivery of the digitized documents back to the agency.

30.6. The vendor should have a tracking system located in its facility for instant tracking of the agency's shipment. The vendor should 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 should 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 should be unique for each box and large document and applied in a clean and professional manner. In most cases, the label should not be attached directly to the document.

30.8. The vendor should make all efforts to reduce the risk of documents being lost during the transfer.

30.9. All work should be performed at the contracted vendor's facilities or on site at the agency, unless prior agreement is made with agency.

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

- Customer Name
- Records Title and Date
- Range of Records (if appropriate)
- Package or Media Number
- Security Level (Low, Medium, High)

30.11. The vendor should maintain control over the records provided to them or created as a result of the work performed under the scope of the project at all times.

30.12. The vendor should store all documents in a secured environment, with protection from damage due to the elements, human error, and other unforeseen threats.

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

30.14. The vendor's record storage facility should have protection from water damage due to outside elements.

31. Imaging Process

31.1. All procedures should be performed by the vendor in accordance with generally accepted standards of conservation and restoration practice. Alterations, changes, or the insertion of any new material in any record should be prohibited.

31.2. The vendor should provide bound volume imaging services that allow the bound volume to remain intact as a part of their service, unless directed by the agency otherwise.

32. Access during Conversion Processing

32.1. The project should be completed in a manner that maximizes the agency's access to records at all times. The agency should require that files in the possession of the vendor as a part of the imaging project that are needed prior to the completion of the conversion should be returned to the requesting party. The agency should determine the appropriate file return period from the initial request based on its operational needs and this should be specified in the vendor contract. An emailed copy of the image of requested document is acceptable, if the document is not restricted. A faxed copy is also acceptable as long as the quality of the fax is deemed adequate by the agency upon receipt.

33. Post-Processing

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

33.2. All boxes of files should be returned in the same condition and document order as provided minus staples, paper clips, and other fasteners.

33.3. The vendor should provide written verification that no files were damaged and/or lost and that all security and confidentiality requirements have been maintained for all records.

34. Use of these Standards

34.1. Do not use these standards, alone, as a specification. An agency should write a specification that describes the essential technical requirements for the items, materials, services and procedures; and that includes appropriate policies and standards. Incorporate the specification into the standard contractual documents used by your agency, and ensure that your legal counsel reviews these documents.

34.2. The specification should include descriptions and special handling and other procedures for each records series to be digitized. The following are essential pieces of information to be added:

- 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

Appendix C: Definitions

Agency means a state agency or municipality or political subdivision as defined in *CGS* 1-200(1(A) and falling under the authority of the records management program administered by the Connecticut State Library as specified in *CGS* §11-8.

Analog means a static, hard copy, non-digital format that is not computer-readable, such as but not limited to: printed materials, photographs, and microfilm.

Archival means and pertains to records that have been or may be appraised by the State Archivist to have permanent value to the State of Connecticut and that may be scheduled for transfer to the State Archives or to an archival repository approved by the Public Records Administrator.

Audit means a process designed to evaluate a system's accuracy, timeliness, adequacy of procedures, training provided, and the existence of audit trails.

Audit trail means a record of transactions in an information system that provides verification of the past activity of the system.

Authenticity means the quality of being genuine, not a counterfeit, and free from tampering, and is typically inferred from internal and external evidence, including its physical characteristics, structure, content, and context.

Backfile conversion means a process of scanning, indexing and inspecting a large existing group of documents.

Born digital means information created in an electronic format. Born-digital information is distinguished from digitized, the latter describing information in an analog format that has been scanned.

Defensible disposition means the process by which content is systematically deleted with an audit trail that is legally admissible in court.

Derivative copy means an image that is used in place of a master image for general access.

Digitization means the process of transforming analog material into binary electronic (digital) form, especially for storage and use in a computer.

Electronic Content Management System (ECMS) provides the strategies, methods and tools used to capture, manage, store, preserve, and deliver content and documents related to organizational processes. An ECMS can include features such as document management, content taxonomies, auditing capabilities, check-in/check-out and other workflow controls and security mechanisms.

Human-readable storage medium means paper, a photograph, a photocopy, or a microform, including, but not limited to, microfilm, microfiche, computer output microfilm, and aperture cards.

Imaging system means a system used to create a digital reproduction of a scanned document or record.

Integrity means the quality of being whole and unaltered through loss, tampering, or corruption.

Metadata is commonly referred to as "data about data"; structured data that describes, explains, locates or otherwise makes it easier to retrieve, use or manage an information resource. Metadata is typically organized into distinct categories, such as administrative, descriptive, preservation or structural.

Migration is the act of moving records from one system to another, while maintaining the records' authenticity, integrity, reliability, and usability.

Non-proprietary means the format of the record is device independent and not tied to one piece of software.

Offsite location means a premises, building, or structure that is separate and apart from the premises, building, or structure that houses an agency's official records.

Official record copy means the specific copy of a public record, as provided in *CGS* §1-200(5), designated by the public agency as the legally recognized copy that must be maintained for records retention, preservation and authentication.

Proprietary means the format of the record is controlled and supported by just one software developer.

Public records or files, in accordance with *CGS* §1-200(5), means any recorded data or information relating to the conduct of the public's business prepared, owned, used, received or retained by a public agency, whether such data or information be handwritten, typed, tape-recorded, printed, Photostatted, photographed or recorded by any other method.

Records custodian means the staff member within the agency having possession of and responsibility for the care and control of material.

Records retention schedule identifies and describes groups of related records (also called record series) and provides instructions for their disposition when they have fulfilled all administrative, fiscal, legal, and historical / research requirements. It lists the schedule number, item number, record series title, description of records, approved retention period, and any legal citations for every record series.

Redundant array of independent disks (RAID) is a method of storing the same data in different places on multiple hard disks. By placing data on multiple disks, input/output operations can overlap in a balanced way, improving performance. The use of multiple disks increases the mean time between failures (MTBF), and storing data redundantly also increases fault tolerance. Currently there are at least nine types of RAID.

Reliability means the quality of being dependable and trustworthy. The record contains accurate information at the time of creation.

Security copy means a reproduction of a record created and managed to preserve information in case the original is damaged.

State Archives means the department within the Connecticut State Library having custody of the records appraised for permanent retention by the State Archivist.

Sustainable format means the format will be accessible throughout the lifecycle of the record and as technology evolves, regardless of the technology that was used when the record was originally created.

Usability means the quality of a record that allows it to be accessed, processed, and understood over time.

Definitions of other technical terms used in this document are defined by Society of American Archivists (SAA), Association of Records Managers and Administrators (ARMA), and the Association for Information Management (AIIM).

Appendix D: List of References and Standards

References

Avedon, Don M. *Introduction to Electronic Imaging.* 2nd edition. Silver Spring, MD: Association for Information and Image Management, 1994.

Bureau of Enterprise Systems Technology (BEST). *Use of Public Cloud Services*. 2012. http://www.ct.gov/doitservices/lib/doitservices/EP&A_Agency_Communication_Bulletin_%282012-06%29_-_Use_of_Public_Cloud_Services.pdf

Federal Agencies Digitization Guidelines Initiative (FADGI), Still Image Working Group, http://www.digitizationguidelines.gov/.

Federal Agencies Digitization Guidelines Initiative (FADGI). *Raster Still Images for Digitization: a Comparison of File Formats. Part 3. Narrative and Summary Table*, April 2014. http://www.digitizationguidelines.gov/guidelines/FADGI_RasterFormatCompare_p3_20140417.pdf

Kentucky. Kentucky Department for Libraries and Archives, Public Records Division. *Ensuring Long-term Accessibility and Usability of Textual Records Stored as Digital Images: Guidelines for State and Local Government Officials*, January 2010.

Michigan. State of Michigan. Records Management Services. *Best Practices for Reproducing Public Records,* October 2009.

Minnesota. Minnesota State Archives, Minnesota Historical Society. *Electronic Records Management Guidelines: Digital Imaging*, March 2012.

National Archives and Records Administration and National Association of Government Archives and Records Administrators. "Digital Imaging and Optical Media Storage Systems: Guidelines for State and Local Government Agencies." Washington, D.C. 1991.

New York. New York State Archives. *Digital Imaging Guidelines*, 2014.

Newcombe, Tod. *The Local Government Guide to Imaging Systems: Planning and Implementation.* International City/County Management Association and Public Technology, Inc., 1995.

North Carolina. State Archives of North Carolina. *File Format Guidelines for Management and Long-Term Retention of Electronic Records*, September 2012.

Sitts, Maxine K. *Handbook for Digital Projects: A Management Tool for Preservation and Access* (Andover, Massachusetts: Northeast Document Conservation Center, 2000).

South Carolina. South Carolina Department of Archives and History. *Electronic Records Management Guidelines: Digital Imaging Policy*, March 2008.

United States National Archives and Records Administration (NARA). *Technical Guidelines for Digitizing Archival Materials for Electronic Access: Creation of Production Master Files – Raster Images for the*

Following Records-Types – Textual, Graphic Illustrations/Artwork/Original, Maps, Plans, Oversized, Photographs, Aerial Photographs, and Objects/Artifacts, June 2004.

Virginia. Library of Virginia. *Digital Imaging Guidelines*, September 2008.

Standards and Technical Reports

ANSI/AIIM MS44-1988 (R1993), *Recommended Practice for Quality Control of Image Scanners*. Silver Spring, Md.: Association for Information and Image Management, 1993.

ANSI/AIIM MS52-1991, *Recommended Practice for the Requirements and Characteristics of Original Documents Intended for Optical Scanning*. Silver Spring, Md.: Association for Information and Image Management, 1991.

ANSI/AIIM MS53-1993, *Standard Recommended Practice—File Format for Storage and Exchange of Images - Bi-Level Image File Format: Part 1*. Silver Spring, Md.: Association for Information and Image Management, 1993.

ANSI/AIIM MS59-1996, *Media Error Monitoring and Reporting Techniques for Verification of Stored Data on Optical Digital Data Disks*. Silver Spring, Md.: Association for Information and Image Management, 1996.

ANSI/AIIM/ISO 19005-1, *Document Management—Electronic Document File Format for Long-term Preservation—Part 1: Use of PDF 1.4 (PDF/A-1)*. Silver Spring, Md.: Association for Information and Image Management, 2005.

ANSI/AIIM/ISO 19005-2:2011, *Document management—Electronic Document File Format for Long-term Preservation—Part 2: Use of ISO 32000-1 (PDF/A-2)*. Silver Spring, Md.: Association for Information and Image Management, 2011.

ANSI/AIIM/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)*. Silver Spring, Md.: Association for Information and Image Management, 2012.

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.

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 TR34-1996, *Sampling Procedures for Inspection by Attributes of Images in Electronic Image Management (EIM) and Micrographics Systems*, 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.

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

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.

Mell, Peter and Timothy Grance. National Institute of Standards and Technology (NIST). Special Publication 800-145, *The NIST Definition of Cloud Computing,* 2011.

National Institute of Standards and Technology (NIST). Special Publication 800-88 Revision 1, *Guidelines for Media Sanitization,* 2013.

Additional industry standards can be found at the International Organization for Standardization (ISO): www.iso.org.