Public Records Standards 04-2: Digital Imaging

Standards effective December 1, 2022 and supersede *Digital Imaging Standards, 2014.*

Approved scanning projects prior to this date shall meet the previous standards document except records designated as permanent, archival or life of structure on a records retention schedule.

These standards shall be read together with *Public Records Policy 04: Electronic Records Management* and *Public Records Standards 04-1: Electronic Records* to ensure a full understanding of the Office of the Public Records Administrator (OPRA) and State Archives (SA) joint policy regarding (1) the preservation and authentication of electronic records, and (2) the retention and disposition of electronic records.

These standards apply to records of public agencies (hereinafter "public records/records") that are (a) born digital or (b) digitized images of analog records (both of which are hereinafter "electronic records"). Public agencies are defined as executive branch state agencies, as defined in C.G.S. § 4-5; certain quasi-public agencies; towns, cities, boroughs, and districts; and other political subdivisions of the state (hereinafter "public agency/agencies").

Public agencies should work in conjunction with appropriate IT staff, either individually or through central IT (where applicable), to implement digital imaging processes that are compliant with the below standards.

Public agencies must establish a clear and sustainable plan for maintaining long-term electronic records and dedicate sufficient resources to this plan. Electronic records require proactive attention as they are more fragile and complex to preserve than paper and microform records. Without preservation actions, electronic records can be overwritten in databases, lost in media migrations, or become inaccessible due to incompatible legacy systems. Public agencies must be aware of the new skill sets, training, considerable significant resources, and ongoing management that will be required over many decades to ensure that the electronic records remain available to future generations.

Public agencies should weigh the pros and cons of maintaining analog records as protection against electronic records loss. Agencies should also consider access needs and ensure that public access would not be negatively impacted by the disposal of the hard copy records. Any decision to dispose of analog permanent records after scanning should be clearly outlined in the agency's records management policies and procedures.

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

I. **Legal Issues**

   A. Any public 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.

   B. References to electronic records 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 14, *Freedom of Information Act*; Chapter 15, *Connecticut Uniform Electronic Transactions Act*; Chapter 15b, *Uniform Electronic Legal Material Act;* Chapter 92, *Town Clerks;* and Chapter 899, *Evidence.*

II. **Scope**

   A. This section covers the standards and procedures a public agency must apply when digitizing paper records using reflective digitization techniques. Such records include most paper-based documents regardless of size, such as modern office paper, maps, posters, manuscripts, graphic-arts prints (lithographs, intaglio, etc.),

drawings, bound volumes, and photographic prints. This section also covers any records that may be incorporated into mixed-media records.

B. This section does not cover standards and procedures public agencies must apply when digitizing records using transmissive digitization techniques. Such records include photographic negatives, transparencies, aerial film, roll film, and micrographic and radiographic materials. In addition, this does not cover records on dynamic media, such as motion picture and audio-visual records, videotapes, and audio cassette tapes.

C. These standards apply regardless of who performs the digitizing activities for the public agency, whether the agency itself, a department/division/unit of the agency, a vendor or other similar entity acting on the agency's behalf. This document uses the terms "imaging" and "digitizing" interchangeably to refer to this process.

D. For guidance on digitizing out of scope media types or non-paper-based portions of mixed-media records, such as dynamic media, x-rays, negative or positive film, or other special media types, contact the Office of the Public Records Administrator.

E. This section also does not cover standards and procedures for optical character recognition (OCR) technology. Public agencies may perform OCR during digitization to meet agency business needs and transfer the resulting files to the State Archives, but this section does not require OCR.

F. This section does not address other applicable laws and regulations governing documents and electronic files, including, but not limited to, proper handling of confidential, restricted, or exempt from disclosure information. Public agencies should work with their legal counsel and other officials to ensure compliance with these and other applicable requirements.

G. This section does not address other business needs or legal constraints that may make it necessary for an agency to retain original source records for a period of time after digitizing. Public agencies should work with legal counsel to determine whether such retention might be necessary because it relates to rights and interests, appeal rights, benefits, litigation holds, or other similar reasons.

H. Provided there are no statutory, regulatory, or policy requirements to retain the record in a physical eye-readable format, and that the record is not designated on a records retention schedule as (1) permanent, (2) archival, (3) archival review required, (4) may have historical value, or (5) life of structure, public agencies may destroy the paper or analog record after scanning, completion of quality control, and designate the digital image as the official record copy. There is no requirement to request permission for such destruction from the Office of the Public Records Administrator.

I. If a record is designated on a records retention schedule as (1) permanent, (2) archival, (3) archival review required, (4) may have historical value, or (5) life of structure, the public agency may digitize the records. After scanning and quality control is completed, prior to disposal of analog records, public agencies must contact the State Archives. The State Archivist may request transfer of the analog records and/or digitized records to the State Archives. Analog records or digitized records may also be transferred to another approved repository in accordance with *State Archives Policy 01: Transfer of Historical Records to the State Archives or Other Approved Archival Repository* and *Procedures for the Transfer of Historical Public Records to the State Archives*.

J. If there are statutory, regulatory, or policy requirements to retain the record in a physical eye-readable format, the agency may digitize the records but also must continue to maintain the required physical format.

III. **General Requirements**

A. Purpose and objectives. This section establishes processes and requirements to ensure that public agencies:

1. Identify the scope of each digitization project and efficient use of resources.

2. Account for all records included in the scope of the project regardless of their media type.

3. Produce complete and accurate digitized records that can be used for all the same purposes as the original source records.

4. Where applicable follow specific requirements for digitizing permanent, archival, or life of structure records.

B. Records management requirements. Public agencies must comply with existing records management requirements identified in *Public Records Policy 05: Disposition of Public Records*. Before starting a digitization project, public agencies must have intellectual and physical control over the original source records that will be included in the project. Having and maintaining an appropriate level of intellectual and physical control over source records is critical to the success of a project, regardless of whether the agency, or an agent acting on the agency's behalf (such as a vendor), performs the digitization activities.

1. Establish and document all the elements of intellectual control.

2. Establish and document all the elements of physical control.

   a. Understanding the physical properties of original source records is necessary to properly identify a project's scope and acquire appropriate equipment and/or vendor services.

   b. Non-standard media, such as sticky notes, envelopes, or onion-skin paper, may require special handling and equipment. Using improper equipment may result in damage to original records.

   c. Public agencies must also document any records that cannot be digitized according to the standards in this section.

   d. For more information about records that need special handling, contact the Office of the Public Records Administrator.

3. Public agencies must create an inventory of records to be digitized, ensure that the proposed series are complete, document any missing records or gaps in records, and document any restrictions relating to the source records that will also apply to digitized records and note them in the metadata. Public agencies will need to maintain intellectual and physical control over the records throughout the project.

4. Public agencies must also document the contents of any electronic or analog storage media, such as CDs, DVDs, or magnetic tapes, discovered when preparing records for digitization.

   Determine whether any files on the storage media are records. If the files are non-records, the public agency may dispose of them.

5. Vendors must provide digitized records in a format as required by the public agency.

6. Public agencies must place digitized records in an information system that can successfully search, retrieve, and manage the records over time.

IV. Preparing records for digitization.

A. A successful digitization project relies on maintaining source records in their original order throughout the process, capturing all the information and characteristics of the source material, and performing visual and automated quality control inspections at multiple stages during a project to ensure the resulting digital record is complete.

B. Quality control is a necessary component of digitizing as a records management activity. Public agencies must:

1. Account for all records included in the project's scope prior to digitization. Note any missing records or records being retained in their original form in the metadata and include scans of any charge-out documentation so that skipped or missing records can be inter-filed if they are located at a later date.

2. Survey source records for items that require special handling and select equipment that safely digitizes the originals without damaging them during the scanning process.

3. Capture all information in records or files, regardless of the original media type.

4. Establish a consistent image file naming convention.

5. Select an appropriate file format to capture source records, including verification of compliance with the PDF/A format for purported PDF/A files.

6. Accurately capture administrative metadata including access and use restrictions, descriptive metadata, and technical metadata.

7. Determine and apply an appropriate method for associating digitized records with each other, when relevant (such as when digitizing each page of a paper document separately, or each document in a paper file folder separately). Acceptable methods include associating individual image files in a folder structure matching the original paper folder structure or utilizing file formats with support for multi-page files such as PDF or TIFF.

8. Ensure that each individual file is usable including but not limited to:

   a. proper orientation (landscape or portrait),

   b. appropriate contrast in the images (neither too light nor too dark),

   c. no distortion of images,

   d. no extraneous materials (sticky notes, fasteners, etc.) obscure the images,

   e. and no skewed images.

   f. correct size and resolution

   g. image digitized at appropriate ppi for the source record's original format.

   h. no additional information added to the image that is not part of the original source record

   i. all pages are present and in proper order

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

10. Ensure that each individual file can be located, retrieved, and accessed over time.

11. Index data shall be verified to ensure accuracy. Industry acceptable methods include:

   a. 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)

   b. Verification of data by another individual other than the person performing the initial data entry

C. Public agencies must also take steps to maintain intellectual and physical control of source records during the digitization process. Public agencies doing digitization in-house must:

   1. Document the age, media types (CDs, microfilm, paper), dimensions (height and width), required level of detail, and condition of source records prior to digitization; and

   2. Implement procedures and controls that:

      a. Safeguard records against loss and damage,

      b. Restrict and log access to records while they are being digitized to minimize the risk of unauthorized additions, deletions, or alterations, and

      c. Ensure that staff appropriately digitize all records or, if the agency keeps some records in their original format (for example, if they are too fragile to scan), maintain the association between the digitized and original records relationship in the metadata. Public agencies should note in the metadata any records that are not digitized and include scans of any charge-out documentation so that skipped or missing records can be inter-filed if they are located or transferred at a later date.

D. Public agencies using vendors should ensure that the contract includes these same safeguards.

E. Vendors shall exercise all reasonable care to avoid damage to the public agency's property or to property being made ready for the public agency's use, and to all property adjacent to any work site. All forms, documents, data and/or reproduction are the property of the public agency. The vendor shall promptly report any damage, regardless of cause, to the public agency.

F. Background Checks. Federal and state criminal history background checks may be required by the State, the State of Connecticut Department of Emergency Services and Public Protection (DESPP), or as provided for in any State or Local document that governs procedures for background checks.

V. Project management and documentation requirements.

A. Public agencies must ensure that any project to digitize records meets the parameters in this section, and the records are complete, unaltered, and meet all quality control criteria.

B. Accordingly, public agencies must have the following documents when digitizing records and retain them in association with the digitized records

1. A defined project plan that identifies the:

   a. Record series or file units the public agency will digitize and note in the metadata any missing records and provide scans of check-out documentation.

   b. Estimated volume and media types of the original source records.

   c. Image quality parameters the public agency must meet to capture the appropriate level of detail present in the original in order to interpret the information in the records including resolution, color, and tonal fidelity. See section VI.C and VI.D for the minimum requirements for image quality parameters. The color mode must be either RGB or grayscale; bi-tonal mode is only acceptable for modern office documents that are black-and-white with good contrast.

   d. Public agencies must digitize in color when the original source records have color present.

   e. Estimated date range of the source records.

   f. Estimated storage requirements (number of bytes, gigabytes, etc.) for the records once digitized (which may affect project decisions, such as compression and file format).

2. Applicable Office of the Public Records Administrator approved records retention schedule(s).

3. Any related finding aids, indexes, inventories, logs, registers, or metadata the public agency uses to manage the records.

4. Quality control procedures to identify and correct errors during digitization.

   a. Quality control 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 public agency are met.

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

5. Identify detected errors and remediation steps.

VI. Digitization requirements for paper and photographic print records.

A. *Equipment requirements.*

   The equipment used to digitize public records must be appropriate for the media type, capable of achieving documented project objectives, and meet the parameters specified in paragraph C of this section for paper records in good physical condition that are suitable for mass digitization or paragraph D of this section for photographic print records and paper records that require higher resolution or color accuracy or that cannot physically be digitized by mass digitization.

1. The specifications in paragraph C of this section are applicable for paper records that are suitable for mass digitization using high-volume scanners. To be suitable for this set of standards, the records must be in good physical condition, with well-defined printed type (such as typeset, typed, laser-printed, etc.), and have moderate to high contrast between the ink of the text and the paper background.

2. The specifications in paragraph D of this section are applicable for photographic prints and paper records that are old, brittle, or folded, or that could be damaged by high-speed equipment. For records in poor physical condition, public agencies must use equipment that does not result in further damage. For records with poor legibility or diffuse characters (such as carbon copies, Thermofax/Verifax, etc.), handwritten annotations or other markings, low inherent contrast, staining, fading, halftone illustrations, or photographs, digitization equipment or record staging must be capable of capturing record content, including all text, any embossed seals, or other details that can't be digitized by mass digitization.

3. For records where the smallest significant detail in a record is 1.0 mm or smaller, such as aerial photographs and topographic maps (which require a high degree of enlargement and precision regarding the dimensional accuracy of the scans when compared to textual documents or other types of photographs), public agencies should follow the Federal Agencies Digital Guidelines Initiative (FADGI) Technical Guidelines for Digitizing Cultural Heritage Materials. For many imaging devices, increasing the ppi settings may not increase the actual level of resolution or capture the desired detail. The equipment selected for digitizing records with fine detail must be capable of meeting higher quality parameters.

B. *Implementation requirements.*

1. Public agencies must:

   a. Implement a quality control process and use device-level reference targets to verify that digitization devices conform to imaging parameters in this section.

   b. Replace reference targets as they fade, or accumulate dirt, scratches, and other surface marks that reduce their usability.

   c. Test equipment to ensure scanners and digital cameras/copy systems are performing optimally.

      i. If applicable scan a reference target containing a grayscale, color chart, and accurate dimensional scale at the beginning of each workday.

      ii. Perform additional tests when problems are detected.

   d. Test equipment with the specific software/device driver combination(s) in use, and re-test after every software update.

   e. Ensure that equipment operation, settings, and image processing actions remain consistent for the entire batch and are applied to all images in the batch.

   f. Encode original image files using a lossless compression type, and in a format, specified in Section X, and with the resolution, color mode, bit depth, and color space specified in table 1 to paragraph C of this section.

   g. Not reformat, use a lossy compression codec, or interpolate (up-sample) files to meet the standards in this section.

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

      ii. Permanent records shall not use lossy compression.

2. If a public agency creates or saves images with redactions (in order to fulfill FOIA requests, for example), the public agency must also maintain the unredacted master image.

3. 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.

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

4. Any page with a sticky note on it must have the sticky note removed from the page prior to imaging and placed on a clean, blank page and digitized separately, unless the sticky notes have been designated as non-records, in which case these may be discarded.

C. *Digitizing requirements for mass digitization of paper records in good physical condition.*

1. Records suitable for the specifications in table 1 are paper records with well-defined printed type (such as typeset, typed, laser-printed, etc.), and with moderate to high contrast between the ink of the text and the paper background.

2. The specifications in table 1 are not appropriate for records that include fine detail, require a high degree of color accuracy, or have other unique characteristics that cannot be captured using the specifications in this table, or that cannot safely undergo high-volume digitization because they are fragile, would be damaged, or have other physical conditions that do not lend themselves to high-volume or mass digitization.

3. For these records, public agencies shall produce image files at a minimum of 200 ppi sized to the original document.

   The required minimum resolution level for standard business documents is 200 pixels per inch (ppi). The required minimum resolution for the use of optical character recognition (OCR) processing is 300 ppi.

4. Digitize in an RGB color mode when the original source paper records have color present. Non-photographic print paper records may be digitized in grayscale mode if there is no color present.

5. At a minimum, public agencies shall digitize paper records covered by this paragraph to the following parameters:

TABLE 1 TO PARAGRAPH C

Digital file specifications[1]	Attributes
Color mode[2]	RGB color or grayscale
Bit depth	8- or 16-bit

[1] Count values are expressed as 8-bit equivalents.

[2] Must digitize in color when the original source paper records have color present.

D. *Digitizing requirements for photographic prints, and paper records not suitable for mass digitization.*

1. The photographic print specifications also apply to maps and plans, manuscripts, illustrations, or graphics, as well as documents with poor legibility or diffuse characters, such as carbon copies, Thermofax, etc.

2. For these records, public agencies shall produce image files (as described in table 2 to paragraph D) at a minimum of 300 ppi sized to the original document. A higher resolution may be necessary for some photographic prints to capture fine detail. The required 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.

3. Digitize photographic prints (and items outlined in paragraph D.1 of this section), including monochrome and black and white originals, using RGB color mode (which captures nuances in black, gray, sepia, etc, as

well as color contained in the original). Paper records may be digitized in grayscale mode if there is no color present; if color is present, digitize them using RGB color mode.

4. At a minimum, public agencies shall digitize the records covered by this paragraph to the following parameters:

TABLE 2 TO PARAGRAPH D

Digital file specifications[1]	Attributes
Color mode[2]	RGB color or grayscale
Bit depth	24-bit

[1] Count values are expressed as 8-bit equivalents.

[2] Must digitize photographic prints, manuscripts, etc., in color, even when originals are in black and white or monochrome. Must digitize other paper documents in color when the original source paper records have color present; otherwise, may digitize such paper records in grayscale.

VII. Digitization requirements for mixed-media files.

A. Related records may be managed together but stored on more than one media type. For example, a ''case file'' may include paper records, on-line electronic records, and electronic records on storage media such as magnetic tapes or other optical media. This reflects the way public agencies create, maintain, and use these records; these are mixed-media files.

B. When digitizing files that fall within the scope of this section but are part of a mixed-media file, public agencies must:

1. Assess any electronic records in the mixed-media file to determine if they are digitized copies of paper records.

   a. If they are not digitized versions of paper records, ensure the electronic records remain associated with the rest of the records in the original mixed media file.

   b. If they are digitized versions of paper records, determine whether they meet the digitization standards in this section. If so, ensure they remain associated with the rest of the records in the original mixed-media file. If not, re-digitize the original paper records to the standards of this section.

2. Digitize any paper records and photographic prints in the mixed-media file according to standards in Section VI.

VIII.   Digitization requirements for microfilm or microfiche.

A. To produce images of adequate quality, the approach used to digitize microforms may vary from the requirements for paper-based records.

B. Before imaging microforms, the public agency or vendor should produce an inventory that will allow for an assessment of the difficulty of imaging the microforms. The inventory should record the:

1. location of the master negatives

2. location of duplicate negatives

3. location of service copies

4. the type of microforms (16mm, 35mm, microfiche)

5. film base (nitrate, cellulose acetate, or polyester)

6. the length if applicable (100 or 215 feet)

7. general film quality (resolution, density, image spacing, and the existence of targets, blips, splices, and scratches or other damage)

8. general image quality (such as blurry images)

C. The public agency should discuss with the vendor the advantages of scanning the master or duplicate negatives against the disadvantages.

D. 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.

IX. **Digitization requirements for bound volumes.**

A. The resolution and pixel depth for digitizing bound volumes must follow the recommendations for paper records.

B. Unbind volumes for digitizing where possible without damaging the contents.

C. If a bound volume cannot be unbound, 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.

D. If the entire image of each page cannot be captured, the bound volume must be retained after imaging.

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

F. 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.

X. **Requirements for disposal of hard copies**

A. Quality control shall be conducted by the public agency before the destruction of any original source records, including, but not limited to, visual inspection of the digitized documents to ensure clarity, readability, and accurate representation of the original record and checking the indexing fields against the original or imaged record.

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

C. Prior to disposal of analog records designated on a records retention schedule as (1) permanent, (2) archival, (3) archival review required, (4) may have historical value, or (5) life of structure the public agencies must contact the State Archives. The State Archivist may request transfer of analog records to the State Archives or a scheduled transfer of the digitized records.

D. When a public agency has validated that the digitized versions meet the required standards, the agency may destroy the original source records, subject to any pending legal constraint on the agency, such as a litigation hold.

E. The public agency must treat the digitized versions, now the official record copies, in the same way it would have treated the original source records. The agency must retain the digitized versions for the remaining portion of any retention period established by the applicable records schedule.

F. Public agencies do not need to obtain Public Records Office approval to destroy scheduled less-than permanent records they have digitized according to these standards.

XI. **File Naming**

Creating unique, consistent, logical, and predictable file names distinguishes similar records from one another in the file hierarchy and facilitates the storage and retrieval of records. Well-named files allow users to browse file names more effectively and efficiently. In general, file names should be fewer than 20 characters, be short but descriptive, avoid special characters or spaces, include dates in the format YYYY-MM-DD, and include a version number or designate which is the draft or final version.

XII. Metadata Requirements

A. *General.* Whether embedded into image files or captured in an information system, metadata provides information explaining what each record contains, when and why it was created, what media it was recorded on, original dimensions, and whether any restrictions govern its access and use. Metadata also describes the digitization process and the technical attributes of the resulting electronic records. It is important to capture this information about original source records and about the intervening digitization steps because the original source records or other project documentation will not be available when maintaining the digitized versions as records in the future.

1. Public agencies should consider business and legal needs when developing a digitization project plan and how the public agency will capture metadata (indexing, elements, properties, etc.).

2. Depending on the public agency's existing record-keeping practices and level of intellectual control, the public agency may use information from the record series, file, or project as the source for administrative and descriptive metadata fields. If the components of a record have not been individually indexed with unique descriptions, the public agency may apply the series or file level descriptions to all records within that grouping. If the components of the record do not have individual titles, the public agency must apply unique identifier(s) instead.

3. Appropriate and accurate metadata (index) information is required to properly identify and later retrieve electronic records.

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

5. The public agency shall be responsible for defining the specific metadata requirements needed to access the records efficiently.

6. Indexing shall comply with the specific requirements of the public agency and include a unique identifier for each electronic record. Unique filenames or other identifiers are preferably sequential and can be numeric, alphanumeric, or alphabetic as required by the public agency. They should be unique across all record series and storage media.

7. The index of electronic records should consist of a limited but sufficient 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.

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

9. **Permanent records only:** If the public agency provides other metadata elements in addition to the metadata requirements in this section, the Connecticut State Library will accept that metadata as part of the transfer process.

10. ''Mandatory if applicable'' instructions in the tables in this section mean that public agencies must provide the metadata if the public agency captures the metadata as part of its business processes. Public agencies do not have to create ''mandatory if applicable'' metadata as an extra step to transfer records to the State Archives.

11. "Strongly Encouraged" instructions in the tables in this section mean that public agencies are strongly encouraged but not required to capture or create this metadata.

12. "Suggested" instructions in the tables in this section mean that it is only a recommendation that the public agency consider capturing or creating this metadata but it is not required.

   B. *Overall requirements*

      1. For all electronic records public agencies must:

         a. Capture the metadata specified by paragraphs C, D, and E of this section at the record series, file unit or project level.

         b. When public agencies determine that records are no longer in active use and no longer subject to changes that would alter a checksum or audit trail, public agencies must generate checksums or capture an audit trail and record them as technical metadata in an information system for each electronic record, and use them to monitor records for corruption or alteration.

         c. Create file names and unique identifier(s) for each file (although public agencies must capture other metadata at the file or item level, some might be common to multiple files or items, but not these two elements).

         d. Transfer metadata specified by paragraphs C and D of this section to the State Archives in CSV or other appropriate format as agreed upon between the public agency and the Connecticut State Library.

      2. For digitized records public agencies must also:

         a. Embed the metadata specified by paragraph C of this section in each image file, capture and maintain it in an information system, associate it with the records it describes, and keep it consistent and accurate in both places.

         b. Ensure that scanning equipment or camera embeds the system-generated technical metadata specified by paragraph E of this section in each image file and that image processing does not alter or delete it.

   C. *Administrative Metadata.*

      1. Capture in an information system the following administrative metadata:

TABLE 1 TO PARAGRAPH C.1

Metadata label	Description	Requirement level
File Name	The complete name of the computer file, including its extension.	Mandatory.
Unique Identifier(s)	The unique identifier(s) is assigned by a public agency or a information system.	Mandatory.
Records Retention Schedule Record Series Number	The records retention schedule series number assigned to the records.	Strongly Encouraged.
Relation Has Part	A related record that is either physically or logically required in order to form a complete record. Mixed-media files that contain records on multiple media types should use this element to identify all components.	Strongly Encouraged if a record includes multiple parts, such as the component parts of a case file or mixed-media file.
Relation Is Part Of	A related record or file in which the described record is physically or logically included. Records that are components of mixed media files should use this element to indicate their status.	Strongly Encouraged if file is a component of a multi-part record.

2. Capture in an information system the following access and use restrictions metadata inherited from the original source records:

TABLE 2 TO PARAGRAPH C.2

Metadata label	Required fields	Description	Requirement level
Access Restrictions	Access Restriction Status. Specific Access Restriction.	Indicate whether there are access restrictions on the record (i.e. not public). Specific access restrictions on the record, based on Freedom of Information Act (FOIA) exemptions, donor restrictions, court orders, and other federal and state statutory or regulatory provisions.	Mandatory if applicable. Mandatory if access restriction exists.
Use Restrictions	Use Restriction Status. Specific Use Restriction.	Indicate whether there are use restrictions on the record. The type of use restrictions on the record, based on copyright, trademark, service mark, donor, or statutory provisions, including Freedom of Information Act (FOIA) exemptions.	Mandatory if applicable. Mandatory if use restriction exists.
Rights Holder		A person or organization owning or managing intellectual property rights relating to the record.	Mandatory if there is a rights holder.

D. *Descriptive Metadata*
Capture in an information system the following descriptive metadata from source records at the lowest level needed to support access and preservation and to maintain contextual information.

TABLE 3 TO PARAGRAPH D

Metadata label	Description	Requirement level
Title	A name given to the original record. If a name does not exist, the mandatory metadata element File Name and/or Unique Identifier(s) serves as the title for the record.	Mandatory.
Description	A narrative description of the content of the record, including abstracts for document.	Suggested.
Creator	The agent (person, agency, other organization, etc.) primarily responsible for creating the original record.	Mandatory.
Creation Date	The date or date range of the original record.	Mandatory.
Last Modified By	The user to last modify the record.	Suggested
Last Modified	The date or date range the record was modified.	Suggested
Source Type	The medium of the original source record scanned to create a digital image.	Mandatory.
Source Dimensions	The dimensions of the original source record (including unit of measure).	Suggested.

E. *Technical Metadata*
   Capture in an information system the following technical metadata describing the electronic records:

TABLE 4 TO PARAGRAPH E

Metadata label	Definition	Requirement level
File Size	The size in bytes of the image file.	Mandatory.
Format Name and Version	The format name or description of the file format.	Mandatory.
Image Width	The width of the digital image, i.e., horizontal or X dimension, in pixels.	Mandatory.
Image Height	The height of the digital image, i.e., vertical or Y dimension, in pixels.	Mandatory.
Date and Time Created	The Date or Date Time the digital image was created.	Mandatory.
Scanner Make and Model	The manufacturer and model of the scanner used to create the image.	Mandatory if using a scanner.
Scanning Software Name and Version	The name and version of the software the scanner uses to create the image.	Mandatory if using scanning software.
Digital Camera Make and Model	The manufacturer and model of the digital camera used to create the image.	Mandatory if using a digital camera.

TABLE 5 TO PARAGRAPH E

Fixity metadata label	Description	Requirement level
Message Digest Algorithm	The specific algorithm used to construct the message digest for the digital object or bitstream.	Mandatory if using checksum.
Message Digest (checksum)	The output of Message Digest Algorithm.	Mandatory if using checksum.
Audit Trail	The output of the audit trail in the information system.	Mandatory if using audit trail.

F. *Transfer metadata for permanent records to the Connecticut State Library*

1. When a public agency transfers legal and physical custody of electronic records to the Connecticut State Library, it must also transfer the associated metadata specified by paragraphs C, D, and E of this section.

2. In addition, the public agency must follow *State Archives Policy 01: Transfer of Historical Records to the State Archives or Other Approved Archival Repository*; *Procedures for the Transfer of Historical Public Records to the State Archives*; and complete a *Memorandum of Transfer* form.

Definitions

The definitions below are from the National Archives and Records Administration (NARA) and the Society of American Archivists (SAA) *Dictionary of Archives Terminology*, except where noted. See also *OPRA Records Management Terms;* policy and standards resources, and additional information are available on the Office of the Public Records Administrator website (https://ctstatelibrary.org/publicrecords/).

Accessible is information arranged, identified, indexed, or maintained in a manner that permits the custodian of the public record to locate and retrieve the information in a readable format within a reasonable time. (Wisconsin Public Records Board (PRB), *Guidance for Managing Web Records for State Agencies and Local Units of Government Appendix A*)

Authentic/authenticity means that the record is unaltered from the original; that it is what it purports to be; and/or that its representation is transparent.

Approved archival repository is a repository that meets professionally accepted archival facility and infrastructure requirements including but not limited to the care, management, security, preservation, and accessibility of public records. Contact the Office of the Public Records Administrator and the State Archives prior to any archival records transfer to an archival repository.

Audit Trails link to specific records in an information system and track such information as the user, date and time of event, and type of event (data added, modified, deleted, etc.). Since audit trails may play an integral part in prosecution, disciplinary actions, or audits or other reviews, public agencies are responsible for ensuring that internal management policies are in place for retaining audit trails as long as necessary for these purposes following the minimum retention period as listed on the Connecticut State Library's Records Retention/Disposition Schedules. Audit trails help prove a record's authenticity.

Batch is a group of files that are created under the same conditions or are related intellectually or physically. During digitization, batches represent groups of records that are digitized and undergo Quality Control inspection processes together.

Chain of Custody is the complete, documented, chronological history of the possession and handling of a piece of information or a record from the time of its creation through its authorized destruction. The ability to demonstrate an unbroken chain of custody is an important test of the authenticity of records. This includes all information on a file's travels from its original creation version to its final complete version or a detailed account of the location of each document/file from the beginning of a project until the end. A sound chain of custody verifies that the agency or vendor has not altered information either in the copying process or during analysis.

Checksum is a function that takes an input string, which can be of any length, and generates an output of fixed length. The output, or hash, is used to authenticate information, such as whether a file has been corrupted or modified. The values returned by a hash function are called hash values, hash codes, digests, or simply hashes.

Cloud or Cloud Computing consists of three parts: 1) delivery of hosted services over the internet, or an organization's intranet, instead of on a user's local computer; 2) storing, accessing, sharing, and using data with those hosted services; and 3) the hardware, networks, and staffing required to maintain the data and services. When the Cloud resources are owned and operated by an organization itself, it is known as a "private cloud." Most commonly, cloud resources are offered as a service from a third-party provider and are known as a "public cloud.""
Color management is using software, hardware, and procedures to measure and control color in an imaging system, including capture and display devices.

Content means the basic data or information carried in a record. (Wisconsin PRB, *Guidance for Managing Web Records for State Agencies and Local Units of Government Appendix A*)

Context is the relationship of the information to the business and technical environment in which it arises. "Context" can include, but is not limited to, such elements as: the origin of the record; date and time the record was created; identification of the record series to which the information belongs. (Wisconsin PRB, *Guidance for Managing Web Records for State Agencies and Local Units of Government Appendix A*).

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

Digitization project is any action an agency (including an agent acting on the public agency's behalf, such as a contractor) takes to digitize records. For example, a digitization project can range from a one-time digitization effort to a multi-year digitization process; can involve digitizing a single document into an electronic records management system or digitizing boxes of records from storage facilities; or can include digitizing active records as part of an ongoing business process or digitizing inactive records for better access.

Digitized record is an electronic record created by converting paper or other media formats to a digital form that is of sufficient authenticity, reliability, usability, and integrity to serve in place of the original source record.

Disposition is "a final administrative action taken with regard to records, including destruction, transfer to another entity, or permanent preservation." (ARMA)

Electronic Content Management System (ECMS) is a software system that 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.

Enterprise content management (ECM) is used to create, store, distribute, discover, and manage unstructured content (such as scanned documents, email, reports, medical images and office documents) and ultimately analyze usage to enable organizations to deliver relevant content to users where and when they need it. (Gartner Information Technology Glossary)

Extranet is a computer network that allows controlled access from the outside, for specific business or educational purposes. In a business-to-business context, an extranet can be viewed as an extension of an organization's intranet that is extended to users outside the organization, usually partners, vendors, and suppliers, in isolation from all other Internet users. (Wisconsin PRB, *Guidance for Managing Web Records for State Agencies and Local Units of Government Appendix A*)

File (noun) is a document or group of documents related by use or topic, typically housed in a folder (or a group of folders for a large file).

File (verb) is the action of placing a record or document in a folder.

Folder is a container used to group records.

Geographically remote means storing backups or duplicate copies outside of the building in which the server resides.

Image quality is a measurement of a digital image's overall accuracy in faithfully reproducing an original. A digital image created to a high degree of accuracy meets or exceeds objective performance attributes (such as level of detail, tonal and color fidelity, and correct exposure), and has minimal defects (such as noise, compression artifacts, or distortion).

Information system is an organized set of procedures, tools, and techniques designed to store, retrieve, manipulate, analyze, and display information. Note: 'Information system' usually connotes the use of computers. If automated, 'information system' also refers to the hardware and software. Automated information systems are generally distinguished from real-time control systems, message-switching systems, and software engineering environments.

Intellectual control is having the information necessary to identify and understand the content and context of the records. This includes knowing the disposition schedule under which the records fall, the date range when the records were created, and any access or use restrictions that apply to the records.

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. This is verified using an audit trait or checksum.

Intranet is a private network inside a company or organization, which is for internal use only and not accessible to the public or outside the organization's network. (Wisconsin PRB, *Guidance for Managing Web Records for State Agencies and Local Units of Government Appendix A*)

Lifecycle means all phases of a record's existence: creation, active use, preservation and management through to disposition. "Disposition" includes permanent preservation as well as designation for destruction. (Wisconsin PRB, *Guidance for Managing Web Records for State Agencies and Local Units of Government Appendix A*)

Mass digitization is the large-scale scanning of source records using scanners capable of high-volume throughput. Mass digitization approaches are appropriate for paper records of uniform size and type that can be digitized without being damaged by the equipment, and in which there is no information requiring higher specifications to ensure accurate capture (such as fine detail or precise color accuracy).

Media are the physical forms on which records are stored, such as paper, photographs, compact discs (CDs), digital video discs (DVDs), analog tapes, flash drives, local hard drives, or servers.

Metadata is the characterization or description documenting the identification, management, nature, use, or location of information resources (data). Note: Metadata is commonly defined as "data about data." Metadata is frequently used to locate or manage information resources by abstracting or classifying those resources or by capturing information not inherent in the resource. Typically, metadata is organized into distinct categories and relies on conventions to establish the values for each category.

> **Administrative metadata** is necessary to manage and use information resources and that is typically external to informational content of resources. Note: Administrative metadata often captures the context necessary to understand information resources, such as creation or acquisition of the data, rights management, and disposition.

> **Descriptive metadata** is information that refers to the intellectual content of material and aids discovery of such materials. Note: Descriptive metadata allows users to locate, distinguish, and select materials on the basis of the material's subjects or 'aboutness.' It is distinguished from information about the form of the material, or its administration.

Embedded metadata are textual components that exist alongside the content (usually binary data) within the file. Embedded metadata may be used to make self-describing digital files that contain specified administrative, rights, and technical metadata and can be appropriately managed outside of a recordkeeping system.

Preservation metadata is technical information that can help support the longer-term sustainability of digitized content. information about an object used to protect the object from harm, injury, deterioration, or destruction.

Structural metadata is information about the relationship between the parts that make up a compound object.

Technical metadata are elements of information that describe processes used to create electronic files, and parameters that aid a system in rendering the files properly. Technical metadata may include elements such as a file's byte size, file format and version, color encoding, and the type of equipment used to make the file (camera name, scanner manufacturer, etc.).

Mixed-media files are records in different forms of media. A file, when used in the phrase ''mixed-media file,'' is a group of records—regardless of location and type of media—that belong together or relate to a topic, such as a case file. For example, a mixed-media case file could be a box with paper notes, audio recordings of interviews, and a CD of photographs, along with physical evidence stored separately in an evidence locker. Records in a file may be in more than one media type due to changes in how agencies create, maintain, and use records, shifts in technology, and the topic or activity involved.

Official Record Copy is the specific copy of a public record, as provided in C.G.S. § 1-200(5), designated by the public agency as the legally recognized copy that must be maintained for records retention, preservation and authentication. For example, if records are kept in both electronic and hard copy format, the agency must identify the official record copy.

Physical control is having the information necessary to physically manage the records. This includes knowing where the records are housed, whether any records that fall within the project's scope are missing or stored separately, and the records' physical form (such as media types, the records' dimensions, and the smallest level of detail used to convey information).

Project plan establishes the vision and goals for the project, summarizes key points of historical or referential context, identifies stakeholders, addresses any areas of concern or risk for the long-term preservation of and access to digitized materials, and communicates in broad strokes the overall plan for the project.

Public Record as defined by C.G.S. § 1-200(5), is any recorded data or information relating to the conduct of the public's business prepared, owned, used, received or retained by a public agency, or to which a public agency is entitled to receive a copy by law or contract under section 1-218, whether such data or information be handwritten, typed, tape-recorded, printed, photostated, photographed or recorded by any other method.''

Quality control (QC) is the process by which a public agency reviews the quality of all steps in the creation and maintenance of electronic records through inspection or testing to determine if they meet their specifications. The purpose is to detect defects (deviations from predetermined requirements) in records or processes.

Reference target is a chart of test patterns with known values used to evaluate the performance of an imaging system.

Reflective digitization is a process in which an imaging system captures reflected light off of scanned objects such as bound volumes, loose pages, cartographic materials, illustrations, posters, photographic prints, or newsprint.
Reliable means the electronic record produced accurately reflects the initial record each time the system is requested to produce that record. (Wisconsin PRB, *Guidance for Managing Web Records for State Agencies and Local Units of Government Appendix A*)

Reproduction scale accuracy measures the relationship between the physical size of the original object and the size in pixels per inch (ppi) of that object in the digital image.

Resolution is the level of spatial detail an imaging system can resolve in an image.

Rollback is the operation of restoring information to a previous state by canceling a specific transaction or transaction set. Rollbacks are either performed automatically by database systems or manually by users. (Wisconsin PRB, *Guidance for Managing Web Records for State Agencies and Local Units of Government Appendix A*)

Sharpening is used to artificially enhance details to create the illusion of greater definition.

Source record/original source record is the record from which a digitized version or digitized record is created.

Structure is the appearance or arrangement of the information in the record. "Structure" can include, but is not limited to, such elements as heading, body and form. (Wisconsin PRB, *Guidance for Managing Web Records for State Agencies and Local Units of Government Appendix A*)

System trustworthiness means a system that is believed to be capable of operating within defined levels of risk despite the environmental disruptions, human errors, structural failures, and purposeful attacks that are expected to occur in its environment of operation.

Transaction logs is a system generated history of actions for a specific business purpose. (Wisconsin PRB, *Guidance for Managing Web Records for State Agencies and Local Units of Government Appendix A*)

Transmissive digitization is a process in which the system transmits light through a photographic slide or negative.

User is any person who creates, modifies, deletes, or accesses electronic records. In the present context, users include, but are not limited to public agency employees, contractors, individuals on a PSA, interns, volunteers, or the public.

Versioning is creating updated versions of content records. (Wisconsin PRB, Guidance for Managing Web Records for State Agencies and Local Units of Government Appendix A).

Web Archiving is the process of collecting, preserving, and providing enduring access to web content.