

## DIGITIZATION GUIDE for RECORDS WITH LONG-TERM RETENTION AT NEW YORK CITY AGENCIES

In accordance with the policy on “**Converting Hard-Copy Documents to Digital Documents**”, the Department of Records and Information Services (DORIS) created this guide as a resource to assist agencies in creating digitization processes. The procedures necessary for the successful implementation of a digitization process are included as are other standards and items for consideration. All agencies are required to submit documentation to DORIS pre- and post- implementation in order to confirm that all digitization processes are aligned with the appropriate industry standards and records management practices.

Agencies that converted hard copy records to digital formats prior to the issuance of the policy/guide should submit to DORIS a completed checklist and documentation showing the record series converted and relevant disposal quantities and dates.

### Standards:

All digital documents must conform to the minimum technical, security and quality standards as determined by DORIS:

#### *Technical*

- All images must have unique identifiers and have the appropriate metadata attached to ensure they can be retrieved.
- All documents to be digitized must be converted to one of the approved formats (see “Approved Formats” section).
- Digitized documents should have a **minimum** resolution of **300 dpi** (dots per inch). Additional guidance is provided in the “**Preferred and Acceptable Formats**” section.
- Backup images of all digitized records must be stored off-line and outside of the imaging system in a different geographical location for disaster recovery / business continuity purposes.

#### *Quality Assurance*

- Digitized records must capture each page so every legible line, character, and mark on the original is readable on the image.
- The agency must conduct quality assurance and file a certification with the Department of Records and Information Services (DORIS) that the agency made a good-faith effort to ensure the images are accurate and have been converted to a format that does not allow the deletion or alteration of the records or their metadata.
- If Optical Character Recognition (OCR) is used, the converted text must be associated with the corresponding digital image. The OCR conversion must be in accordance with industry standards.
- The agency must develop and disseminate agency-wide policies and procedures on how to manage digitalized documents.

## Information Security / Privacy

- The agency digitization process must ensure that all documents containing confidential/ sensitive information (e.g. Personal Identifiable Information (PII), HIPAA, Financial Information, Personal Health Information) are protected throughout the imaging process.
- Confidential documents that are eligible for encryption should be tagged at the time of processing.
- Security measures should be placed on documents eligible for redaction. Access and change logs must be established to document employee access and changes to digitized records.
- A permission and access control protocol should be established that identifies security roles and levels of access to digitized records.

## Legal

- The agency records retention schedule is media-neutral and should be reviewed to ensure it is up-to-date prior to starting the conversion process. After the records have been digitized, the record retention information should be attached to the electronic records immediately.
- Barring a legal or statutory reason to preserve the original documents, the hard-copy file source files shall be destroyed after completing quality control and ensuring the records are accessible and readable with their form and content intact in order to authenticate their trustworthiness as records.
- Legal acceptance of records requires proof that the process or system is reliable and hence capable of producing trustworthy records. State agencies and local governments must be able to demonstrate:
  - imaged records have been produced in the normal course of business,
  - the system producing and managing the records (the technology, as well as the policies, procedures, training, and audit controls) are trustworthy and functioning properly, and
  - the system documentation is maintained.
- System documentation must be up-to-date and show how the system creates, manages, and destroys records (workflow); define the roles and responsibilities of the individuals managing the system; demonstrate how the system provides for consistent quality control; document staff training, system and audit controls; and provide for adequate security. **(See the New York State Archives Digital Imaging Guidelines Sec 6.4. for additional information.)**

## Preferred and Acceptable Formats\*

The following file formats are recommended for conversion, storage, and archiving of New York City digitized files:

<b>Scanned Text</b>	<b>Preferred</b>	Tagged Image File Format ( <b>TIFF</b> ) Portable Network Graphics ( <b>PNG</b> ) Portable Document Format/Archival ( <b>PDF/A</b> )
	<b>Acceptable</b>	Joint Photographic Experts Group ( <b>JPEG</b> ) Graphics Interchange Format ( <b>GIF</b> ) Portable Document Format ( <b>PDF</b> )
<b>Digital Photographs / Images</b>	<b>Preferred</b>	Tagged Image File Format ( <b>TIFF</b> )
	<b>Acceptable</b>	Joint Photographic Experts Group ( <b>JPEG</b> ) Digital Negative ( <b>DNG</b> )

For digital images, agencies should digitize analog originals at a minimum resolution of **3,000 pixels** across the long dimension; and the resulting digital images should be uncompressed or make use of lossless compression.

For scanned text, the following resolutions are recommended:

- **Bitonal (1-bit black and white)** images must be scanned at **300-600 ppi**. Scanning at **600 ppi** is recommended. This is appropriate for documents that consist exclusively of clean printed type possessing high inherent contrast (e.g., laser printed or typeset on a white background);
- **Gray scale (8-bit)** must be scanned at **300-400 ppi**. Scanning at **400 ppi** is recommended. This is appropriate for textual documents of poor legibility because of low inherent contrast, staining or fading (e.g., carbon copies, thermofax, documents with handwritten annotations or other markings), or that contain halftone illustrations or photographs; and
- **Color (24-bit RGB [Red, Green, Blue])** must be scanned at **300-400 ppi**. Scanning at **400 ppi** is recommended. Color mode (if technically available) is appropriate for text containing color information important to interpretation or content.

Please contact DORIS prior to using any other file formats for additional guidance.

\*Additional guidance can be found in the New York State Archives Digital Imaging Guidelines and the U.S. National Archives and Records Administration (NARA) recommendations regarding digitized file formats.

## Process for Implementing a Digitization Program

Agencies should perform the following steps prior to and after implementing a digitization program:

### Pre- Implementation

#### THE AGENCY SHOULD

- review and update the retention schedule prior to starting the conversion process. After the documents have been digitized, the record retention information should be attached to the records immediately.
- utilize the **Records Inventory Form** to identify the records to be converted and all pertinent information.
- conduct cleanup (e.g., destroy any duplicate documents, non-records and records eligible for disposal as per their retention schedules,) prior to starting the digitizing process.
- determine storage requirements and where the digitized records are to be maintained and backed up (both short-term and long-term).
- create a naming index, or uniform file names, so that the documents being converted can be properly stored and easily located. 'A naming index' is the name given to file directories so that documents can be located effectively and efficiently. The naming conventions should be consistent and allow for ease of locating and retrieving.
- determine who will perform the digitizing:
  - **Centralized Scanning** – Documents are sent to a central location in the agency for processing. High-volume production scanners are required with a high scan speed.
  - **Decentralized Scanning** – Documents are scanned through various locations in the agency.
  - **Third-Party Vendor** – Documents are sent to an outside contractor for processing
- verify through a proof of concept process (POC) that the selected electronic content management (ECM) system can successfully produce and manage the records according to the documented agency requirements.
- Optical Character Recognition (OCR) may be performed to convert digital images into text that accompanies the digital record.

All agencies must submit a completed **Pre-Implementation Checklist** to DORIS prior to beginning a digitization process. Once the checklist has been reviewed and approved by DORIS, the agency may begin implementation.

## Post- Implementation

- After the digitization process has been implemented, quality control must be conducted to ensure the trustworthiness of the records. Each image and index must be visually inspected for clarity, readability, and accuracy according to industry standards. The quality control check must verify the following:
  - Correct image filing naming convention.
  - Correct file format (including verification of compliance with the PDF/A format for such files).
  - Quality of the image is an accurate reproduction of the original and includes:
    - The document dimensions, size and resolution are correct.
    - Orientation is proper—landscape or portrait
    - Image is not skewed
    - Image is neither too dark nor too light
    - Curvature of the page does not obscure or distort the text
    - There is appropriate contrast (e.g., not too light or dark) within the image
    - There is no distortion of the image
    - Nothing obscures the image
    - Information that was not contained in the original has not been added to the image
  - Appropriate indexing terms have been entered and associated with the image

**NOTE:** Any image that is not acceptable must be re-digitized to meet the standards.

- Consult your agency Quality Assurance standards and retention schedule to determine what documents, if any, must be kept in hard-copy format and stored accordingly.

All agencies must provide a certification to DORIS of the accuracy and authenticity of the images and metadata. After the certification has been reviewed and approved by DORIS, the agency may destroy the original source hardcopy.