

# Guideline

## Digitization of Physical City Records

*This guideline falls under A1477 Data and Information Management Administrative Policy.*

<b>Program Impacted</b>	Strategy & Business <i>The City of Edmonton's corporate processes are robust and helpful for integrated service delivery.</i>  Technology & Data <i>The City of Edmonton's technology and data are leveraged to enable quality decision-making and enhance innovative service delivery.</i>
<b>Approved By</b>	City Clerk
<b>Date of Approval</b>	March 18, 2025
<b>Approval History</b>	N/A
<b>Next Scheduled Review</b>	March 2027

### Guideline Statement

This guideline provides direction for scanned or digitized records to make sure the City of Edmonton meets legal obligations and can access records long-term.

This guideline falls under the Lifecycle Management (LM) functional domain, as defined in *A1477 Data and Information Management Administrative Policy*.

### Scope

This guideline provides recommendations for all City of Edmonton departments completing digitization processes or processing paper documents using scan and capture technologies.

Before undertaking any large-scale scanning project, business areas must discuss the legal requirements of various document formats with their lawyer, as per the *Electronic Transactions Act, SA 2001, c E-5.5*. Some records will need to be kept as physical originals with wet ink signatures.

## **Guideline Description**

This guideline provides direction for:

- digitization equipment and software
- file formats, metadata and compression
- image resolution and optical character recognition (OCR)
- evidence of digitized document authenticity
- digitized document storage and retention

## **Definitions**

*Authenticity* refers to the quality of being genuine, trustworthy and free from tampering or corruption.

*City of Edmonton Classification and Retention Schedule* refers to a living policy document that categorizes records based on similar use and purpose to facilitate lifecycle management. The schedule provides rules that identify the length of time a record needs to be kept (retention period), what to do with the record when it is no longer needed (disposition method) and supplementary information.

*City record* refers to recorded information in any form that is acquired or developed during the course of City business.

*Digital records* refers to City information that is entered, created, manipulated and/or stored on digital media or storage devices. This includes:

- records that are born digital (that is, records that are created and saved digitally, like online forms or digital photos)
- digitized records (for example, records that have been converted from a non-digital format, such as scanned paper records)
- unstructured data (for example, Google documents)
- structured data maintained within digital systems (for example, tables within a database)

*Digitize* means to convert a document, image or other physical record into a digital format. City records may be digitized by scanning, a process which assigns a binary representation for each visual element sampled.

*Metadata* means data about data. Metadata documents the context, content and structure of information and supports the effective management of the City's information holdings. The term covers a range of structured tools including records profiles maintained electronically (e.g., the data in word processing properties fields), online or hardcopy manuals, thesauri, indexes and library catalogues. Metadata is also essential in establishing a record's authenticity. Alternatively, metadata (and even a lack of metadata) can be used to invalidate authenticity.

## **Guideline Specification**

### *Document Digitization Process*

The following steps should be completed when digitizing a hardcopy document:

1. Prepare the document for scanning, as per the Document Digitization Checklist.
2. Scan all pages of the document, including any attachments or related documents like illustrations, maps and release forms, using an appropriate scanning device.
3. Combine pages into one PDF unless necessary.
4. Confirm that all pages were scanned. Remove any blank pages.
5. Confirm page quality and orientation. Make corrections or rescan where necessary.
6. Add a signed copy of the Scanner Technician Certificate of Authenticity (see Supporting Resources section, p.9).
7. Save the PDF. Name the file as per the corporate naming convention.
8. Save the file in the correct location.

### *Equipment and Software*

The equipment and software used to scan or digitize records should support all recommendations presented in this guideline. The output format, Dots Per Inch (DPI), compression and optical character recognition (OCR) requirements may be automated using scanner settings.

### *File Formats*

The appropriate file format for a specific type of digitized record is determined by

- retention requirements — i.e. long-term readability and compatibility with different document readers
- media type — appropriate for the type of media to be stored
- record function — based on record purpose. Editable formats are available for digitized records that require further manipulation or editing. Lockable formats are available for final approved records.
- source — the scanning system may create digitized records in one of the formats listed in Table 1. Born-digital records may be saved in many different formats.

Table 1: Acceptable Formats for Short-Term Storage: 10 years or less

	Born Digital		Scanned	
	Modification Enabled	Modification Disabled	Modification Enabled	Modification Disabled
<b>Spreadsheets</b>	ACCDB, CSV, DIF, MDB, ODS, TSV, XLM, XLS, XLSB, XLSM, XLSX, Google Sheets	PDF/A (locked archivable format)	XLSX, PDF	PDF/A (locked archivable format)
<b>Text Documents</b>	DOC, DOCM, DOCX, MD, ODF, ODT, OTT, PPTX, RTF, TXT, Google Docs	PDF/A (locked archivable format)	RTF, DOCX, PDF	PDF/A (locked archivable format)
<b>Emails</b>	PDF	PDF/A (locked archivable format)	PDF	PDF/A (locked archivable format)
<b>Architectural and Engineering Drawings</b>	DWG, DXF, SVG, PDF/E	PDF/A (locked archivable format)	PDF/E (engineering format)	PDF/A (locked archivable format)
<b>Photographs</b>	TIFF, JPEG, PNG	PDF/A (locked archivable format)	TIFF, JPEG, JPEG2000, PNG	PDF/A (locked archivable format)
<b>Audio Recordings</b>	FLAC, WAV, MP3, AIFF, BWF	BWF Embedded Metadata	N/A	N/A
<b>Video Recordings</b>	MP2, MP4, MXP container with MJPEG2000, AVI, MOV	MOV	N/A	N/A

Table 2: Acceptable Formats for Long-Term Storage: 10 years or longer

	Born Digital		Scanned	
	Modification Enabled	Modification Disabled	Modification Enabled	Modification Disabled
<b>Spreadsheets</b>	CSV, TXT	PDF/A	N/A	PDF/A
<b>Text Documents</b>	RTF, ODF, PDF	PDF/A	RTF, PDF/A (relock)	PDF/A

	Born Digital		Scanned	
	Modification Enabled	Modification Disabled	Modification Enabled	Modification Disabled
<b>Emails</b>	PDF	PDF/A	N/A	N/A
<b>Architectural and Engineering Drawings</b>	PDF/E	PDF/A	PDF/E	PDF/A
<b>Photographs</b>	TIFF, JPEG, PNG	PDF/A	TIFF, JPEG, JPEG2000	PDF/A
<b>Audio Recordings</b>	FLAC (Lossless Compression), BWF	AIFF (PCM), WAV (PCM), BWF Embedded Metadata	N/A	N/A
<b>Video Recordings</b>	MP2, MP4, MXP container with MJPEG2000	AVI, MOV	N/A	N/A

*Metadata*

Metadata allows users to find and group records using search engines or query reports. It can be generated automatically or entered manually when a record is created. Auto-generated metadata includes information about the author, creation date and editing date. Digitized or scanned records have this kind of data, but staff may also need to manually enter metadata to reflect the original information about the record. See Tables 3 and 4 for recommended and required metadata elements. Business units can and should supplement these lists as required.

*Table 3: Metadata captured when a record is scanned*

Metadata Element	Required or Recommended	Metadata Location	Automated or Manual Entry
Date of scanning	Required	Document Properties	Automated
Scanning Technician Name	Required	Storage environment, generally description field	Manual or automated depending on the system
Creator of source record	Required	Description Field	Manual

<b>Metadata Element</b>	<b>Required or Recommended</b>	<b>Metadata Location</b>	<b>Automated or Manual Entry</b>
Creation date of source record	Required	Description Field	Manual
Type of source record (e.g. physical photograph, paper record)	Required	Properties	Automated
Official title of record	Required	Properties	Manual
Data security classification	Required	Description Field	Manual
Access controls	Required	Properties	Manual or automated, depending on storage environment
Retention Rules	Required	Document Properties or Description Field	Manual or automated, depending on storage environment
Copyright information	Required (if available)	Description Field	Manual
Number of pages scanned	Recommended	Properties	Automated
Description	Recommended	Description Field	Manual
Cross-referencing information	Recommended	Description Field	Manual
Last modified date (if saved as an editable PDF)	Recommended	Properties	Automated
Last modified by (if saved as an editable PDF)	Recommended	Properties or Description Field	Manual or automated, depending on storage environment

*Table 4: Metadata to add during the quality assurance check*

<b>Metadata Element</b>	<b>Required or Recommended</b>
Quality assurance auditor name	Required
Quality assurance approval date	Recommended

### *Compression*

Compression reduces file size for easier sharing and storage. Compression types include:

- Lossless compression — this type of compression makes the file smaller without any loss to image clarity or quality. Lossless compression formats include PNG and GIF.
- Lossy compression — this type of compression makes a very small file, but results in a poorer quality image. Lossy compression formats include JPEG AND MPEG.

Ideally, compression should eliminate only duplicate data in a digitized record; in practice, compression can create a visible difference. For this reason, lossless compression is better.

### *Image Resolution*

Scan resolution is measured by DPI. The recommended resolution is 600 dpi for black/white and 300 dpi for colour/greyscale. Higher scan resolution creates larger file sizes and takes up more storage space. Lower scan resolution results in smaller but lower-quality files.

### *Image Enhancement*

Some software comes with optimization settings. Optimization corrects image skew, reduces noise or speckle, and eliminates blank pages. Employees may use image enhancement if it helps to create a more accurate image of the original physical record.

### *Optical Character Recognition (OCR)*

OCR is the procedure that makes the content of a scanned document searchable. It is helpful for

- finding words or characters in single PDF file;
- running scripts to categorize documents; or
- removing skew, speckle, binarization and lines.

Optical character recognition adds extra time to the digitization or scanning process, but it does make the resulting scan easier to read and more useful. It is recommended that the OCR setting on the scanner be turned on.

### *Authenticity*

If a person is digitizing documents as a part of a large-scale batch scanning project outside of their regular workflow, they must complete a Scanner Technician Certificate of Authenticity. For example, if a person is digitizing a backlog of invoices spanning several years, they would need to complete a certificate. This certificate confirms that what was scanned is an exact copy of the paper document, which helps the City meet potential legal demands. The certificate is generally scanned and saved in the same place as the scanned record.

### *Digitized Document Storage and Security*

Records should be stored in method that allows them:

- to be kept long-term, as per the City of Edmonton Classification and Retention Schedule;
- to be disposed of or transferred at the right time;
- to be locked down using permissions, such as disabling the user's ability to delete records;
- to be backed up regularly; and
- to be monitored using error logs.

### *Digitized Document Retention Management*

The City of Edmonton's Classification and Retention Schedule outlines the rules for disposing of digitized paper records.

### **Quality Assurance for Digitized City Records**

Good records management practices require regular quality control checks to make sure that the digitized images are clear and have all the same information included in the original document. Business areas can check scan quality using an audit process; a scanning *audit* is when a person other than the scanning technician reviews a sample of the scanned documents for accuracy. A scanning audit should not replace the scanning technician's own quality control processes. The audit determines whether the digitized records pass or fail the quality control requirements.

This process should establish a minimum acceptable percentage of digitized records to audit; the percentage may increase if you are noticing a large number of poor-quality images and need to find and correct more errors.

#### *Example:*

Rose is asked to make sure a group of 100 scanned permits are complete and have good quality images. She begins with choosing 10 of the permits at random. She checks that all pages are included and that the scan quality is good. She's happy with all of the scans, so she does not review any additional scans. The group of permits has passed the audit and is good to go.

Rose is also asked to audit a group of 100 scanned applications. She chooses 10 scans at random and notices that two of them have missing pages. She chooses five more scans and finds another with missing pages. This group fails the audit. Rose sends the applications back to the scanning technician and asks them to redo them.

Business areas must develop an audit process to ensure quality. See ANSI/AIIM TR34-1996 for direction on developing an appropriate audit procedure for your business area.

## Acceptable Image Quality Elements

Digital records should be audited for acceptable image quality, including:

### *Format Resolution*

Scan resolution is measured by DPI. The recommended resolution is 600 dpi for black/white and 300 dpi for colour/greyscale.

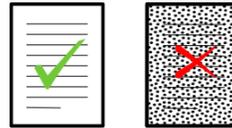


### *Page Orientation and Skew*

Pages should be oriented right-way-up with no skew.

### *Noise and Speckle*

Pages should be clear, with no noise or speckle.



### *Colour Dropout*

Scans should have minimal colour removal, which can occur when using OCR.

### *Missing Pages or Parts*

Scans must be a complete copy of the physical record.

### *Legibility*

There should be good contrast between the text and background. Words and images should be clearly recognizable.

## References and Supporting Resources

### *Legislation*

- [Municipal Government Act, RSA 2000, c M-26](#)
- [Alberta Evidence Act, RSA 2000, c A-18](#)
- [Electronic Transactions Act, SA 2001, c E-5.5](#)

### *Supporting Resources*

- [Scanner Technician Certificate of Authenticity](#)
- [Corporate Information Management Glossary of Terms](#)
- [Corporate Records and Information Management Accountability Matrix \(RASCI\)](#)
- [City of Edmonton's Classification and Retention Schedule](#)
- [Document Digitization Checklist](#)

- [ANSI/AIIM TR34-1996 Technical Report for Information and Image Management - Sampling Procedures for Inspection by Attributes of Images in Electronic Image Management \(EIM\) and Micrographics Systems](#)

## Data Governance Working Committee Sign-off: Subject Matter Experts

Only one person from each area needs to sign off.

Subject Matter Expert Area	Branch / Department (or Division)	Sign Off (Name)
Access and Privacy	Office of the City Clerk / Office of the City Manager	Jessica Fortin
Archives	Community and Recreation Facilities / Community Services	Kathryn Ivany
Business Solutions / Google	Open City and Technology / Financial and Corporate Services	Stephen Mundy
Data Analytics	Service Innovation and Performance / Financial and Corporate Services	Robin Song
Data Governance	Service Innovation and Performance / Financial and Corporate Services	Sean Clovechok
Data Ethics	Service Innovation and Performance / Financial and Corporate Services	Jared Bielby
Enterprise Commons	Open City and Technology / Financial and Corporate Services	Roddi Potter
Enterprise Performance Measurement	Service Innovation and Performance / Financial and Corporate Services	Sajid Majeed
Enterprise Architecture/ Technology Planning	Open City and Technology / Financial and Corporate Services	Mangulal Halai
GIS	Service Innovation and Performance / Financial and Corporate Services	Nicole Gaskell
Information Security	Open City and Technology / Financial and Corporate Services	Romel Alnazer
Information Management	Office of the City Clerk / Office of the City Manager	Saleem Akhtar
Legal Services	Legal Services / Employee and Legal Services	Michelle Bohn

## Data Governance Working Group Sign-off: Department or Division Representatives

Department or Division Representative	Sign Off (Name)
City Operations	Non-responsive
Community Services	Brendan Abbott Richard Williams
Employee and Legal Services	Karim Malik, Mostafa Nejati Ajibisheh
Financial and Corporate Services	Mike Chong
Integrated Infrastructure Services	Stephen Christensen

## Data Governance Working Group Sign-off: Department or Division Representatives

Department or Division Representative	Sign Off (Name)
Office of the City Manager	Mike Chong Mike Caverhill
Urban Planning and Economy	Peter Contessa