Guideline - Digitization

Issued By*:* The Office of the Chief Information Officer

1. Effective Date
This guideline takes effect on July 1, 2018.

2. Application
This guideline applies to all government bodies as defined in the *Archives Act*.

3. Context
This guideline is part of the GNWT's Information and Communications Technology (ICT) Policy Framework. They support and provide advice on implementing the *Directive – Digitizing* and *Standard – Digitizing*. These documents establish requirements for government bodies that wish to convert master analogue records to a digital format.

4. Resource Requirements
Section 4.5 of the *Standard – Digitizing* requires government bodies to create a business case in support of digitization programs. This business case should include the identification of staffing, equipment, software, file size, and storage requirements. This guideline provides additional information on two of these resource requirements: staffing and equipment.

4.1. Staffing competencies
Digitization of records must be carried out by employees or contractors who have the appropriate skills and training.

The basic competencies of digitization staff include:
- Basic computer skills;
- Basic/legible writing skills;
- Basic keyboarding skills;
- Reading comprehension;
- The ability to follow directions and procedures; and
- The ability to collect consistent and accurate information.

Digitization staff should have training on and be able to perform the following:
- Use and maintenance of the scanning hardware and software;
- Document preparation;
- Scanning of records;
- Use of naming conventions and metadata;
- Quality control; and
- Use of the trusted repository, including how to save or upload digitized records to

the repository.

4.2. Digitization Equipment

The digitization equipment required will depend on the project. Multifunction printer/scanners may not have the capabilities to meet a digitization project's requirements, particularly when digitizing large volumes of records or when digitizing records other than letter or legal size office paper.

The following should be considered before selecting digitizing equipment:
- Goal of the project;
- Format of the originals;
- Size and condition of the originals;
- Size of the scan area;
- Scanner speed and connectivity; and
- Compatibility with the government body's computer and network infrastructure.

There are many different types of scanning equipment and each type is optimized for a particular purpose. Some scanners may combine several of the features listed below. The following is not an exhaustive list and new equipment may be available as manufacturers identify new markets:
- **Large format scanners**: Used for scanning maps, blueprints, architectural drawings, site plans, posters and other large scale documents. The documents are placed directly on the bed of the scanner.
- **Flatbed scanners**: Used for scanning papers, photographs, negatives/film, slides, printed material. The documents are placed directly on the bed of the scanner.
- **Automatic feeder scanners**: Used for scanning large volumes of high contrast documents with printed type that is uniform in size and shape. Automatic feeder scanners should not be used for fragile documents because the feeder mechanism may tear them. A visual inspection of the documents to identify impediments such as staples, paperclips, tears or folds is required before running them through the automatic feeder to ensure the records will feed through without jamming or risk of being crumpled or torn.
- **Duplex scanners**: Used to scan both sides of a page at the same time.
- **Overhead scanners**: Used for books or other records that cannot be laid flat or may be damaged by being spread out and pressed down on a flatbed scanner. Overhead scanners sit on an arm or frame that is positioned above the document being scanned.
- **Slide and film scanners**: Slide and film scanners produce a better quality reproduction of slides and film than a flatbed scanner.
- **Digital cameras**: Used for books, documents and artwork that cannot be laid flat or may be damaged by being spread out and pressed down on a flatbed scanner. Digital SLR cameras mounted on a tripod or overhead mount are preferred because they produce a better quality image than point and shoot cameras. It may be necessary to

purchase additional lenses to achieve the best results.

- **Microfilm/Microfiche scanner**: Used for microfilm, microfiche, microfilm jackets, and aperture cards.
- **Digital recorders:** Used to capture an analogue audio signal from a cassette, reel-to-reel, or record player and convert it to a digital format. Digital recorders may be a separate unit attached to the output jacks on the player or it may be integrated into a player.
- **Video capture devices:** Used to capture analogue video and audio signals from a VCR player and convert it to a digital format. Video capture devices may be a separate unit attached to the output jacks on the player or it may be integrated into a player.
- **Motion picture film scanners:** Used to capture moving images and soundtracks from negative or positive original motion picture film, such as 35mm, 16mm, and 8mm film, and convert it to a digital format.

5.   Digitization methodology and quality control requirements

Section 4.5 of the *Standard – Digitizing* requires government bodies to create a business case in support of digitization programs. This business case should include the identification of the methods used for quality assurance and quality control. This guideline provides information on recommended approaches to performing quality control.

This guideline also provides information to support planning and decision making by government bodies when they are developing their digitization programs. It provides information on recommended approaches to performing digitization activities.

5.1. Considerations for digitizing textual records

Government bodies should use methods for digitizing records that meet their business and legal requirements. There are three different methods to digitize textual records. The different methods allow for different ways of accessing the records.

- **Page images**: The digitized record is static and cannot be changed or manipulated. Records cannot be searched unless appropriate metadata is added to the record and made suitable for browsing or online navigation.

- **Full text (OCR):** This method translates the digitized record into machine-readable text, which can then be changed or manipulated. Either the text can be keyed in from the original or an OCR program can be used to convert page images to ASCII. OCR only works with typewritten pages. For handwritten originals, the text must be keyed in.

  Full text is only permissible if used in conjunction with page images. The resulting file must contain both an unaltered digitized image layer as well as the text layer so the image of the original unaltered document is available. Full text alone must not be used.

- **Encoded Text, or Full Text with Mark-up**: This option is similar to full text, but adds additional search functionality by placing encoded text or mark-up language tags around selected text. Applications are able to search for captions created using the encoded text or markup language.

  Like full text, encoded text is only permissible if used in conjunction with page images. The resulting file must contain both an unaltered digitized image layer as well as the text layer so the image of the original unaltered document is available. Encoded text alone must not be used.

5.2. Considerations for digitizing non-textual records

Government bodies may digitize non-textual records as part of business process digitization or legacy digitization. The information presented in sections 5.4, 5.5 and 5.6 should be consulted prior to digitization.

However, digitization of non-textual records is a specialized skill that should be outsourced to a contractor who has the expertise and equipment to address potential problems and ensure quality control procedures are met.

5.3. Creating accurate digitized records

Digitization processes should be designed to create high quality master digitized records. Loss of information should be minimized and the resulting digitized records should not alter the original content or appearance of the master analogue records.

It is recommended that master digitized records be left unaltered in order to protect their authenticity, integrity, and reliability.

Corrections to the digitized record that are made to improve the readability or usability of the record, such as deskewing, despeckling, colour correction, OCR, or file size reduction, should be saved as a separate copy and clearly named to distinguish them from the unaltered master digitized record. Corrections may introduce unintended changes to the records, so the corrected copies should not replace the unaltered master digitized record.

5.4. Business process digitizing of textual records

Business process digitizing refers to digitizing performed as part of a regular business process. Business process digitizing typically involves new, current, or active records. Documents are usually scanned individually or in small batches.

5.4.1. Prepare the documents

Document preparation is usually minimal and done intuitively. Some document preparation steps should be taken:
- Remove staples, paper clips and other fasteners.
- Flatten folded or curled documents.

- Check to see if the document is printed on one or both sides.
- Check to see if the document contains any colour printing or handwriting

5.4.2. Check the scanner
Scanners should be checked regularly to make sure they are in good working order and have been set to the correct specifications.

5.4.3. Scan the documents
Documents should be scanned in compliance with the government body's digitization procedures manual and the scanner's instruction manual. The resulting preservation master file should be saved into the file format selected by the government body and named according to the naming conventions identified in the digitization procedures.

5.4.4. Quality control – scanner operator
Scanners should be checked regularly to make sure they are in good working order and have been set to the correct specifications.

The person operating the scanner should check the quality of the digitized record immediately after it is scanned and before it is uploaded to the trusted repository. The person should check and correct for the following:
- All pages were scanned and both sides of the page were scanned if necessary.
- All pages are in the correct orientation.
- All pages are legible.

It may be necessary to re-scan the documents in order to obtain an acceptable image.

5.4.5. Quality control – validation of digitized records
If the records contain confidential or personal information, the digitized records should be double-checked by a second employee to confirm the digitized image is useable.

If the digitizing is part of a workflow or business process, the person who receives the digitized record should open it and verify it is complete and legible.

5.4.6. Quality control – validation of metadata
Metadata is used to categorize and index digitized records for faster and more efficient retrieval. Metadata entered manually or captured automatically should be checked as part of the quality control and validation procedures.

Metadata should be checked for the following:
- Consistent use of naming conventions;

- Spelling, grammar, and punctuation; and
- All mandatory elements are complete.

5.5. Legacy records digitizing of textual records

Legacy records digitizing refers to digitizing performed in large scale on older records. Documents are frequently scanned in batches based on file classification number or record type.

5.5.1. Prepare the documents

Document preparation should be completed before the scanning project starts. Some document preparation steps should be taken:

- List each folder on the GNWT's Box Inventory Forms as it is being prepared.
- Check every document in a folder to ensure all documents are in the folder and they are in their proper sequence. For very large scanning projects, it may be practical to only check a representative sample of the files.
- Check page numbers to make sure multi-page documents are in the right order.
- Remove staples, paper clips, and other fasteners.
- Flatten folded or curled documents.
- Check to see if the documents are printed on one or both sides.
- Check to see if the documents contain any colour printing or handwriting.
- Tape torn records using non-yellowing, matte-finished tape. For single sided text records, place the tape on the back of the document so it does not cover the text.
- Remove sticky notes and flags. If they contain important information, tape them to the document or to a separate sheet of paper. Make sure they do not cover the text of the document.
- Tape small text records onto 8½x11" sheets of white paper. Tape the top and bottom, keeping them at least 1.5 cm away from the edge of the larger sheet of paper. This may not be necessary if all of the documents in the batch are the same small size.
- Transcribe hard to read or hand-written records and place the transcription behind the original. The transcription should indicate it is a transcript of the preceding document.

If scanning in batches, some additional preparation steps may be required, including:

- Place a batch header at the beginning of the batch to identify it. The batch header may include the following information: department name, batch identifier, file name, file number, volume number, and any special settings or handling requirements.
- Insert dividers or separator pages between each document or a group of documents. The dividers are used to indicate where each digitized record

starts and ends.

Personnel who prepare the documents should identify potential issues with the documents so the person operating the scanner can make the appropriate adjustments. Potential issues include:
- Non-standard paper sizes.
- Non-standard text orientation.
- Coloured paper.
- Coloured ink.
- Rough or textured papers.
- Show-through.
- Unusually thick or thin paper.
- Image contrast.
- Dot matrix printing.
- Character size, style and weight.
- Negative images.
- Margin settings.
- Barcodes.
- Photographs and images.
- Backgrounds and watermarks.

5.5.2. Check the scanner
Scanners should be checked regularly to make sure they are in good working order and have been set to the correct specifications. The scanner should be checked at predetermined intervals. At minimum, check the scanner before and after the settings are changed. It is best to check it daily.

To test the scanner:
- Create a sample set of test documents. These are not actual records, but a representation of the types of documents that might be scanned. The test documents should include a test chart.
- Scan the test set and save the scan as a quality reference. Print the quality reference and save the printouts for comparison later.
- Set up the scanner with the settings that will be used to scan the next batch of records. Scan the test documents and save the test scan.
- Examine the test scans against the quality reference on screen. It may be necessary to print the test scans and compare them against the printed quality reference. If the test scans are not satisfactory, print the quality reference documents again to rule out problems with the printer. If there are no problems with the printer, then the scanner should be checked for problems or the settings adjusted to improve image quality.
- Each test scan should be entered in a log. Test scans and printouts should be

dated and saved with the log. They should be available to show to a scanner repair technician if repairs are required.

5.5.3. Scan the documents

Documents should be scanned in compliance with the government body's digitization procedures manual and the scanner's instruction manual. The resulting preservation master file should be saved into the file format selected by the government body and named according to the naming conventions identified in the digitization procedures.

When scanning is complete, return each document (or batch) to the original folder and box in the same order in which they were originally packed. It is important to retain the original order in case it is necessary to rescan the records. Mark on the box it has been scanned and the date of scanning.

5.5.4. Quality control – scanner operator

The person operating the scanner should check the quality of the digitized record immediately after it is scanned and before it is uploaded to the trusted repository. The person should check and correct for the following:

- All pages were scanned and both sides of the page were scanned (if necessary).
- All pages are in the correct orientation.
- All pages are legible.

It may be necessary to re-scan the documents in order to obtain an acceptable image.

5.5.5. Quality control – validation of digitized records

It is not usually practical or feasible to perform quality control on each document or batch in a legacy digitization program. Instead, quality control is performed on a representative sample of the records. This practice does introduce a risk there will be errors in the digitized records even after quality control has been performed.

Validation of the quality of digitized records should be performed by someone other than the person who digitized the records. It should be performed at regular intervals (e.g. daily or weekly) while the records are being digitized.

Validation of the quality of digitized records should be performed in accordance with ANSI/AIIM TR34-1996, *Sampling Procedures for Inspection by Attributes of Images in Electronic Image Management (EIM) and Micrographics Systems*. The

following procedures provide an overview of the procedures in ANSI/AIIM TR34-1996.[1]

The sample size is based on the total size of the group of records that were digitized during the period being evaluated. For example, if validation is being performed once a week, then the sample size is based on the total number of records that were digitized during the previous seven days.

Government bodies must decide on a risk tolerance for errors. For some groups of records, it may be acceptable to have errors in up to 10 per cent of the documents. For other groups of records, the acceptable rate of errors may be lower.

The following chart is an example only. It identifies the total number of documents being inspected, the sample size to be pulled at random from the larger group of documents, and the number of acceptable errors.

For example, if the group size is between 281 and 500 records, then the sample size is 50 randomly selected records (e.g. pick every 6th file until you have 50). If the acceptable error rate is 1.0 percent, then each sample is permitted to have one error. If the sample has 2 or more errors, then the entire group of records must be rejected and re-scanned.

See ANSI/AIIM TR34-1996 for complete sampling and quality control procedures.

Group size	Sample Size	Acceptable Percentage of Errors If the sample has more than the acceptable number of errors listed below, it must be rejected									
		≤0.025	0.040	0.065	0.10	0.15	0.25	0.40	0.65	1.0	1.5 %
2 to 8	2	0	0	0	0	0	0	0	0	0	0
9 to 15	3	0	0	0	0	0	0	0	0	0	0
16 to 25	5	0	0	0	0	0	0	0	0	0	0
26 to 50	8	0	0	0	0	0	0	0	0	0	0
51 to 90	13	0	0	0	0	0	0	0	0	0	0
91 to 150	20	0	0	0	0	0	0	0	0	0	1
151 to 280	32	0	0	0	0	0	0	0	0	1	1
281 to 500	50	0	0	0	0	0	0	0	1	1	2
501 to 1,200	80	0	0	0	0	0	0	1	1	2	3
1,201 to 3,200	125	0	0	0	0	0	1	1	2	3	5
3,201 to 10,000	200	0	0	0	0	1	1	2	3	5	7
10,001 to 35,000	315	0	0	0	1	1	2	3	5	7	10

---

[1] ANSI/AIIM TR34-1996, *Sampling Procedures for Inspection by Attributes of the Images in Electronic Image Management (EIM) and Micrographics Systems* is a standard developed by the Association for Intelligent Information Management (AIIM) and issued by the American National Standards Institute (ANSI). There is no equivalent Canadian standard or international standard.

Group size	Sample Size	Acceptable Percentage of Errors If the sample has more than the acceptable number of errors listed below, it must be rejected									
		≤0.025	0.040	0.065	0.10	0.15	0.25	0.40	0.65	1.0	1.5 %
35,001 to 150,000	500	0	0	1	1	2	3	5	7	10	14
150,001 to 500,000	800	0	1	1	2	3	5	7	10	14	21
500,001 and over	1250	1	1	2	3	5	7	10	14	21	31

The following table provides a list of errors to look for. They are divided into Class A errors and Class B errors. Class A errors are major concerns. Class B errors are minor concerns. Government bodies may decide they have a higher risk tolerance for Class B errors than Class A errors. This means they may accept a higher percentage of Class B errors before they reject the group of digitized records than Class A errors.

Electronic Image Attributes	Class A	Class B
Size of digitized record relative to the original (measured horizontally and vertically)	X	
Adequate Contrast • For text there should be a high contrast ratio between the text and the background. • For greyscale images the contrast should represent the original image.		X X
Adequate brightness (more of a monitor problem)		X
Colour dropout (dropout of specific colours from the test target)		X
Poor thresholding (dropped low contrast features, dark backgrounds obscuring foreground)		X
Image skew on page (image is at an angle)		X
Incorrect orientation on page		X
Speckle or noise in the background		X
Character dropout (a missing character)	X	
Legibility of small characters or features (poor focus)	X	
Separation of black from white features (bar charts may be required on test target)	X	
Scan line drop out (missing scan lines, single or multiple, across the digitized record, either white or black)	X	
Repeated scan lines on successive digitized images	X	
Repeated pixel on successive digitized image (consistent noise or spot on each digitized record at the same location)		X
Borders not cropped		X
Missing portion of edge of digitized record	X	

Sampling plans may need to be adjusted from time to time. If large quantities of records are being rejected, a tightened sampling plan should be initiated. If few records are being rejected a reduced sampling plan may be adopted. The government body's digitization procedures manual should identify the rules for

changing sampling plans.

For example:
- If two out of five consecutive groups of digitized records are rejected on original inspection, switch to a tightened inspection plan.
- When the tightened inspection plan results in five consecutive groups being accepted, switch back to the normal inspection plan.
- When the 15 preceding groups have been accepted, switch to a reduced sampling plan.
- When on a reduced sampling plan, switch back to the normal sampling plan when a group of records is rejected.

When operating on a tightened or reduced sampling plan, the sample size remains the same, but the number of allowable errors changes. See ANSI/AIIM TR34-1996 for more information.

When a group of records is rejected it must be returned to the person doing the scanning and all of the source records must be re-digitized. The new set of digitized records should be saved separately from the original, rejected set. They must be resubmitted for verification. A group of records may have to be re-digitized several times before it passes verification.

5.5.6. Quality control – validation of metadata
Metadata is used to categorize and index digitized records for faster and more efficient retrieval. Metadata may be entered manually or captured automatically. It should be checked as part of the quality control and validation procedures.

Metadata should be checked for the following:
- Consistent use of naming conventions;
- Spelling, grammar, and punctuation; and
- All mandatory elements are complete.

5.6. Digitization of non-textual records (other media types)
This section discusses digitization of records other than textual records, including but not limited to photographs, microfilm and microfiche, blueprints, and audiovisual materials. These media types require specialized skill sets for digitization. Potential outsourcing to specialized companies can ensure quality control procedures are met.

Non-textual records may be digitized as part of business process digitization or legacy digitization. The information presented here should be consulted in addition to sections 5.4 and 5.5.

5.6.1. Digitizing photographs
Photographs and negatives should be inspected before they are digitized.

Photographs should be inspected for:
- Curling or distortion;
- Silver mirroring;
- Emulsion cracking, crazing, softening, or flaking;
- Mould or pest damage;
- Rips and tears; and
- Images separating from their supports.

Negatives should be inspected for:
- Physical damage from fingerprints or scratches; and
- Deterioration, including brittleness, disintegration, and discolouration of colour negatives.

Do not try to repair or clean damaged photographs or negatives. If photographs are attached to a page in a photo album or scrapbook, do not remove them from the page.

Damaged photographs and negatives may require professional conservation before they can be digitized.

5.6.2. Digitizing blueprints, maps, and architectural plans
Records should be inspected before they are digitized to identify any records that may require special care and handling.

Records should be inspected for:
- Damage;
- Brittleness; and
- Fading.

Do not try to unroll brittle or fragile records. Do not roll, fold or bend records. It may be necessary to bring in a professional conservator to handle records due to their fragility.

5.6.3. Digitizing microfilm and microfiche
Microfilm and microfiche should be inspected before being digitized to identify any issues with the film or with the images on the film that could affect the quality of the digitized images.

Microfilm should be inspected for the following:
- Contaminants, such as mould, damage from pests, dust, dirt, or debris;
- Physical damage to the film, such as scratches, tears, and reels without

headers;
- Damage to the reel, jackets, or aperture cards; and
- Quality of the images on the film, including legibility and completeness of the records on the film.

It may not be possible to create good quality digitized records from poor quality microfilm images.

Heavily contaminated or damaged microfilm may require professional cleaning and restoration before being digitized.

5.6.4. Digitizing audio reel tapes and cassettes

Audio tapes, such as reel to reel tapes and cassettes, should be inspected before they are played to identify tapes that could be damaged when played or could damage the tape player.

Audio tapes should be inspected for the following:
- Contaminants, such as mould, damage from pests, dust, dirt, debris, crystaline residue, and adhesive residue;
- Chemical degradation of the tape;
- Physical damage to the tape;
- Damage to the cassette housing or the reel;
- Check for labels that are falling off or not placed properly; and
- Check the pack of the tape to ensure there are no strands popping out, the wind is neither too loose nor too tight, or if the tape has been wound into a shallow cone

5.6.5. Digitizing video tapes

Video tapes should be inspected before they are played to identify tapes that could be damaged when played or that could damage video playback machines.

Video tapes should be inspected for the following:
- Damage to the cassette housing;
- Contaminants, such as dust, dirt, liquids, debris, mould, and damage from pests;
- Chemical degradation of the tape; and
- Physical damage to the tape.

5.6.6. Digitizing motion picture film

Motion picture film should be inspected prior to digitization to identify film that could be damaged when played.

Motion picture film should be inspected for the following:

- Physical damage, such as brittleness that causes the film to flake or break, mould, vinegar syndrome; and
- Optical damage, such as colour fading leaving the film appearing magenta, silvering, and scratches, tears, or broken perforations.

Damaged film requires professional restoration before it can be digitized.

6. Technical Requirements

Section 4.5 of the *Standard – Digitizing* requires government bodies to create a business case in support of digitization programs. This business case should include the identification of the technical specifications that will be followed when digitizing the records. This guideline provides recommended file formats and settings for digitizing analogue records in various media.

These specifications represent the minimum recommended requirements for the digitization of government bodies' records where digitization aims to replace the master analogue records with a master digitized record.

The following specifications are for creating master digitized records.

The file formats recommended in these specifications are based on the following principles:

- File formats should be open source or widely used and accepted with the GNWT; and
- Master digitized records should be created at the highest quality possible.

6.1. Black and white textual documents

Use this specification for textual documents that are clean, high contrast and where colour is not present or is not essential for understanding the record, and any graphics present are line art. Do not use for photographs.

Format	- PDF for non-archival records retained for less than 10 years - PDF/A for archival records or records that will be retained for more than 10 years - TIFF - PNG
Resolution	300 DPI sufficient for most applications 400 DPI recommended for text that is less than 2mm high 600 DPI recommended for text that is less than 1mm high
Scanning ratio	1:1
Colour profile	Greyscale
Bit depth	8 bit
Compression	Lossless preferred
Duplex	Recommended. Blank pages may be removed during post-processing.

Searchability	OCR is optional. OCR may be used only in conjunction with a non-OCR scan. If OCR is used, the file must contain both a digitized image of the source record and an OCR text layer to ensure an image of the original unaltered document is always available. OCR scans alone may not be used.

6.2. Black and white photographs

Use this specification for black and white photographs.

Format	• TIFF
Resolution	• 35 mm film: 2700 DPI • 2.75x4" print: 1000 DPI • 3x5" or 4x6" print: 800 DPI • 5x7" print: 600 DPI • 8x10" print or larger: 300 to 400 DPI
Scanning ratio	1:1
Colour profile	Greyscale
Bit depth	8 bit
Compression	Lossless preferred

6.3. Greyscale textual documents

Use this specification for records containing text or graphics that feature high contrast and low contrast (e.g. watermarks, grey shading) but where colour is not present or is not essential to understanding the record. Do not use for photographs.

Format	• PDF for non-archival records retained for less than 10 years • PDF/A for archival records or records that will be retained for more than 10 years • TIFF • PNG
Resolution	• 300 DPI sufficient for most applications • 400 DPI recommended for text that is less than 2mm high • 600 DPI recommended for text that is less than 1mm high
Scanning ratio	1:1
Colour profile	Greyscale
Bit depth	8 bit
Compression	Lossless preferred
Duplex	Recommended. Blank pages may be removed during post-processing.
Searchability	OCR is optional. OCR may be used only in conjunction with a non-OCR scan. If OCR is used, the file must contain both a digitized image of the source record and an OCR text layer to ensure an image of the original unaltered document is always available. OCR scans alone may not be used.

6.4. Colour textual documents

Use this specification for records where colour is present and important to understanding the

record, or for faded, low contrast records. Do not use for photographs.

Format	• PDF for non-archival records retained for less than 10 years • PDF/A for archival records or records that will be retained for more than 10 years • TIFF • PNG
Resolution	• 300 DPI sufficient for most applications • 400 DPI recommended for text that is less than 2mm high • 600 DPI recommended for text that is less than 1mm high
Scanning ratio	1:1
Colour profile	Colour
Bit depth	24 bit
Compression	Lossless preferred
Duplex	Recommended. Blank pages may be removed during post-processing.
Searchability	OCR is optional. OCR may be used only in conjunction with a non-OCR scan. If OCR is used, the file must contain both a digitized image of the source record and an OCR text layer to ensure an image of the original unaltered document is always available. OCR scans alone may not be used.

6.5. Colour photographs
Use this specification for colour photographs.

Format	• TIFF
Resolution	• 35 mm film: 2700 DPI • 2.75x4" print: 1000 DPI • 3x5" or 4x6" print: 800 DPI • 5x7" print: 600 DPI • 8x10" print or larger: 300 to 400 DPI
Scanning ratio	1:1
Colour profile	Colour
Bit depth	8 bit per channel RGB or 24 bit
Compression	Lossless preferred

6.6. Maps, architectural plans, and blueprints
Use this specification for maps, architectural plans and blueprints. Use greyscale where there is no colour present in the source records.

Format	• PDF for non-archival records that will be retained for less than 10 years • PDF/A for archival records or records that will be retained for more than 10 years • TIFF
Resolution	Adjust scan resolution to have a minimum resolution of 300 DPI to avoid loss of quality on large maps. Large analogue documents may require a higher resolution than 300 DPI to capture fine detail at 1:1 ratio.
Scanning ratio	1:1

Colour profile	Greyscale or colour
Bit depth	16 (greyscale) or 24 bit (colour)
Compression	Lossless preferred

6.7. Microfilm and microfiche

Use this specification for microfilm, microfiche, microfilm jackets, and aperture cards.

Format	- PDF for non-archival records that will be retained for less than 10 years - PDF/A for archival records or records that will be retained for more than 10 years - TIFF - PNG
Resolution	300 to 400 DPI is suitable if based on the size of the original. If scanning on a flatbed scanner, the size of the microform will have to be significantly higher – multiplied by the reduction ratio.
Scanning ratio	1:1
Colour profile	Greyscale or bitonal
Bit depth	8 bit
Compression	Lossless preferred

6.8. Audio recordings

Use this specification for audio recordings, such as vinyl records, reel-to-reel tapes, and cassettes.

Format	WAV
Sample rate	44.1 kHz, 48 kHz, or 96 kHz
Bit depth	24 bit
Compression	Uncompressed

6.9. Video and moving images

Use this specification for audiovisual recordings, such as those on video tapes (e.g. VHS, BETA) and motion picture film.

Format	MOV
Aspect ratio	Maintain original ratio (probably 1.375:1 (4:3))
Frames per second	Maintain original frames per second (probably between 18 and 24)
Timing, grading	One-light transfer, best settings determined prior to transfer
Image "tonal" settings	Per International Telecommunication Union, ITU-R BT.709-6 (June 2015): Scanning parameters must be kept within 5% of the ceiling or floor of Rec. 709
Picture encoding bit stream	Uncompressed 10 bit (V210), progressive, 4:2:2 chroma sampling

Sound encoding bit depth	48kHz, 24 bit

7. Documentation requirements

Section 4.7 of the *Standard – Digitizing* requires government bodies to create a procedures manual. This procedures manual is an essential tool for supporting the authenticity and reliability of the master digitized records in case of litigation. The procedures should be created before the digitization program begins. Procedures are required for both business process digitization and legacy digitization programs. Government bodies must document any changes to the procedures along with appropriate authorizations for the changes.

The procedures should address the following:

- **Planning:** the methods, people, equipment, and other resources needed to execute a successful digitization program.
- **Testing:** the tests needed to verify the planned procedures and methods which will yield a successful digitization of records.
- **Scope:** identification of the current and/or master records that have been approved for digitization. This listing should include information about the legal and business rationales for authorized destruction of any source records.
- **Digitization procedures:** the procedural steps to be performed in carrying out actual digitization of source records.
- **Metadata:** the information about the digitized record that will be captured and stored at the time of creation. This includes system generated metadata and metadata keyed in by the operator of the scanning equipment. This also covers naming conventions.
- **Quality control and quality assurance:** the procedural steps and methods to be used to verify the source records have been successfully digitized as well as the evaluations used to confirm quality control procedures are being carried out correctly and effectively.
- **Preservation and disposition:** the strategies required for the ongoing management of master digitized and converted source analogue records, for as long as they are required to be retained. This includes strategies for addressing integrity and security of the master digitized records, as well as strategies for addressing hardware, software, and file format obsolescence. It also includes provisions for storage, retrieval, and disposal of digitized records and converted source analogue records.
- **Training:** the strategies to be used to train employees or contractors who create, manage, or work with the digitized records.
- **Sign-off:** the authorizations needed to verify the digitization process has been successfully performed in compliance with approved policies and procedures. Sign-off may be built-in to workflow and business processes or it may occur at various

stages in a digitization project.

8. References

Digitizing is subject to a number of provisions, as established by the acts, policies, directives, and standards outlined in Appendix B.

9. Monitoring and Reporting

There are no monitoring or reporting requirements.

10. Enquiries

All enquiries regarding this guideline should be directed to the Department of Infrastructure, Corporate Information Management Division.

11. Approval

This guideline is effective from the date approved below.

Corporate Chief Information Officer	Signature	Date
Dave Heffernan		2019-02-01

Appendix A

Definitions

Analogue records refers to physical records of various media types (text, photographic, film, microfilm, blueprints, maps, audio, et cetera) that does not require a computer to view embedded information.

ASCII (American Standard Code for Information Interchange) is a character encoding standard for electronic communication. ASCII codes represent text in computers, telecommunications equipment, and other devices.

Bit depth is measure of the number of colours (or brightness in greyscale images) available to represent the colours (or shades of grey) in the original record. An image with a 1 bit depth has two possible values, black and white. An image with a bit depth of 8 has 256 possible values. RGB images are made of three colour channels, red, green, and blue. An 8 bit depth RGB image has 256 possible values for each channel. 8 bit RGB images are sometimes called 24 bit images.

Business process digitization refers to routine digitization of records and the incorporation of the digitized records into business information systems where future actions take place on the master digitized records, rather than on the converted source analogue record.

CMYK (Cyan-Magenta-Yellow-Black) refers to an image made up of pixels made up of these four process colours. CMYK is often used in offset printing for full colour records.

Compression relates to algorithms designed to reduce file size. There are two forms, lossy and lossless. In lossy compression, information is removed from the stored file during the compression process, resulting in smaller file size. In lossless compression, no information is lost and records appear the same as the original. Lossless file formats are larger.

Colour profile refers to the method by which a device will capture a digital image. The four common modes of capture are black and white (bitonal), greyscale, RGB, and CMYK.

Converted source analogue record is a record that has been digitized and is no longer the master record.

Deskewing is the process of straightening skewed (tilted) images to improve accuracy.

Despeckling is the process of removing imperfections from an image.

Digital is any data or recorded information that exists as binary code (zeros and ones).

Digital Quality Index (QI Formula for Text Records) is the recommended scanning resolution for textual records as determined using a Digital Quality Index (QI) formula. This formula is based on the Quality Index formula used in the micrographics industry and is based on three variables: the height of the smallest significant character, the desired quality to be obtained in the reformatted

version, and the resolution of the recording device. See AIIM TR26-1993 Resolution as it Relates to Photographic and Electronic Imaging.

Digitization is the process of converting records from analogue (physical) formats to digital formats.

Digitized record is a record that has been converted from an analogue record to a digital record format.

DPI (Dots per inch) is a measure of output of the device resolution and quality (e.g. number of pixels per inch used when printing with ink or toner). DPI measures the number of dots horizontally and vertically.

Duplex is the printing, filming, or scanning of both sides of a page.

Greyscale refers to an image made up of pixels that are black and white and a range of intermediate greys. Greyscale is suited to black and white, sepia, and other greyscale continuous tonal originals.

Integrity refers to the record being complete and unaltered in all essential respects.

Legacy records refers to an existing set of analogue record that are no longer being added to or modified. They may have been created using filing systems that are either no longer used, or have no apparent organization. They are commonly referred to as backlog records.

Lossy compression see Compression.

Lossless compression see Compression.

Master record is a record that is considered the official record and is considered a true and valid record by both the creator and for legal purposes. Also referred to as a substantive or authoritative record.

Master analogue record is a master record created and manipulated in an analogue (physical) state.

Master digitized record is a record that has been converted from an analogue record to a digital record format, and has met the qualifications (quality control, Records Disposition Authority approval) to be deemed the master record.

Metadata refers to data describing content, structure, and context of records and their management through time. Metadata can be divided into one of three categories:

- Descriptive metadata describes a resource for purposes such as discovery and identification. Metadata in this area can include such elements as author, title, and description.
- Structural metadata indicates how compound objects are put together. It identifies data

format, media format, or the type of data representation and file types, hardware and software needed to render the data, and the compression method and encryption algorithms used, if any.

- Administrative metadata provides information to help manage a resource, such as when and how it was created, and who can access it.

MOV (Apple Quicktime Movie) is a file format used to store video, audio, time code, and text tracks.

PDF (Portable Document Format)**:** A file format that captures formatting information from a variety of applications and makes it possible to transmit and display records in an identical way, independent of the platform. PDF is an ISO standard, ISO 32000-1:2008 and ISO 32000-02:2017.

PDF/A (Portable Document Format/Archival)**:** A long-term preservation formation for electronic records. Everything needed to render or print a PDF/A file must be contained within the file. This includes all visible content like text, raster images, vector graphics, fonts, colour information, etc. PDF/A is an ISO standard, ISO 19005-1:2005, ISO 19005-2:2011 and ISO 19005-3:2012.

Pixel: Pixel is short for picture (Pix) elements (el). It is the smallest element of a picture represented on a screen. Each pixel can represent a number of different shades or colours. It takes a number of pixels to form an image.

PNG (Portable Network Graphics)**:** A raster graphics file that supports lossless data compression.

PPI (Pixels per inch)**:** A measurement of resolution for computer display. It is the preferred term when referring to digital capture.

Quality assurance refers to procedures for monitoring and accessing the records system, aiming to maintain a desired level of quality.

Quality control refers to a system of maintaining predetermined standards in a digitized record by testing/reviewing a sample of the output against the specifications within the standard.

QI (Quality Index) see Digital Quality Index.

Raster image in computer graphics, a raster graphics or bitmap image is a dot matrix data structure, representing a generally rectangular grid of pixels, or points of colour, viewable via a monitor, paper, or other display medium. Raster images are stored in image files in varying formats.

Record is a record of information, regardless of its form and characteristics, the means by which it was created and the media on which it may be stored and, without limited the generality of the foregoing, include (a) a document, book, ledger, photograph, image, audio-visual recording, x-ray, map and drawing, and (b) a record created or stored in digital or other intangible form by electronic means, but does not include software or a mechanism that produces records.

Resolution is a measure of the ability to capture detail in the original work. The optimum resolution depends on the nature of the records being digitized. Photographs, for example, require

much greater resolution than text-based records. Resolution also refers to a measure by pixel dimensions (height and width) for on-screen use. It is often quantified in pixels per inch (PPI) or dots per inch (DPI). When digitizing textual records, resolution is determined using the Digital Quality Index (QI). In general, images scanned at a higher resolution will produce a better or clearer digitized record, but this also increases the file size. Depending on the record, the resolution will need to be adjusted based on the size, quality, and condition of the master analogue record, and the use of the digitized record. When digitizing records, the preferred term to use is pixels per inch, but most scanners define resolution in dots per inch.

RGB (Red-Green-Blue) is a colour profile made up of multiple bits per pixel representing colour. RGB images are made up of three colour channels, red, green, and blue. Colour/RGB capture is suited for use with colour continuous tonal originals.

Scanning refers to the act of moving a beam of light or electrons in a systematic pattern over a surface in order to reproduce or transmit an image.

Scanning ratio is the ratio of scanning resolution to printing resolution. The ratio of the image's width and height dimensions. A ratio of 100 per cent will produce a printed copy the same size as the master analogue record.

TIFF (Tagged Image File) is an open standard image format for storing raster graphic images. TIFF files are lossless, meaning they are in an uncompressed format. TIFF is an ISO standard, ISO 12639:2004.

Trusted Repository a trusted repository (secure storage location) provides reliable, long-term access to managed resources to its designated community, now and in the future. A trusted repository for digital assets must protect the authenticity and reliability of the digital assets stored and managed within it.

Vector graphics in computer graphics, a vector graphic uses paths, which are defined by a start and end point, along with other points, curves, and angles in between the start and end point. Unlike raster images, vector graphics can be scaled indefinitely without loss of quality because they are not made up of a grid of pixels.

WAV (Waveform Audio File) is an audio file format standard used for storing an audio bitstream. It is usually uncompressed.

Appendix B

References

Archives Act sets the legal framework for disposing, transfer, custody and access to records;

Electronic Transactions Act establishes the legal authority of digital records in regards to transactions;

Evidence Act establishes the power and authority of evidence admissible in court, including digital records;

Recorded Information Management Policy (6003.00.18) guides government bodies in the management of their recorded information (regardless of format) and defines the authority and accountability framework;

Records Scheduling Policy (6003.00.24) guides government bodies in the classification, retention, and final disposition of government records;

Management of Electronic Information Policy (6003.00.20) guides government bodies in the management of electronic information;

Electronic Information Security Policy (6003.00.26) guides government bodies in the security of electronic information;

Directive – Digitizing guides government bodies in the development of digitization programs;

Standard – Digitizing establishes the standards which government bodies must meet when implementing a digitization program;

Standard – Administrative Records (6003.00.19) establishes a common records disposition authority for administrative records;

Standard – Operational Records (6003.00.32) establishes the standard format for records disposition authorities for operational records;

Guidelines – ORCS Development (6003.00.25) provides guidance on the development and formatting of an Operational Records Classification System;

RDA 2018-02 Converted Source Analogue Records provides for the scheduling and disposal of converted source analogue records, when approved by the Territorial Archivist and the Deputy Head;

CAN/CGSB-72.34-2017, Electronic Records as Documentary Evidence is a Canadian national standard for the management of electronic information;

ISO 15489-1:2016, Information and documentation – Records management – Part 1: Concepts and Principles is an international standard for records management programs;

ISO/TR 15801:2009, Document management – Information stored electronically – Recommendations for trustworthiness and reliability is an international standard for the management of electronic information;

ISO 23081-1:2006, Information and documentation – Records management processes – Metadata for records – Part 1: Principles is an international standard for the capture of metadata about records;

ISO 23081-2:2009, Information and documentation – Records management processes – Metadata for records – Part 2: Conceptual and implementation issues is an international standard for the capture of metadata about records,

ANSI/AIIM TR26-1993 Resolution as it Relates to Photographic and Electronic Imaging is a technical report that provides information on image resolution for micrographics and electronic images.

ANSI/AIIM TR34-1996 Sampling Procedures for Inspection by Attributes of Images in Electronic Image Management (EIM) and Micrographics Systems is a quality control procedure for microfilming and digitizing of records.

Appendix C

Revision History

Version	Author	Description
Issue 1 Draft	Digitization Working Group	April 2017-April 2018: Draft prepared for endorsement by the Recorded Information Management Committee and approval by the Informatics Policy Council.
Issue 1 Final	Corporate Information Management, Department of Infrastructure	Final version submitted for approval by Chief Information Officer.