Government of New Brunswick Digitization Standard

Table of Contents

Table of Contents

1. INTRODUCTION

The records of government are a valuable resource and an important asset that document its business activities. Their effective management enables government to support future action and decision making, reduce costs, meet business, legal and accountability requirements, and preserve New Brunswick's documentary heritage.

Digitization programs form a part of a public body's overall records and information management program. Digitizing (also known as imaging or scanning) is defined as the process of converting any hard-copy, or non-digital record into digital form. This includes digitizing text, photographs, maps, microfilm, and other works. Meeting the requirements established in this standard will ensure the record's integrity, reliability, authenticity, and accessibility over the long-term.

Public bodies are responsible for creating full and accurate records of their activities and for maintaining them over time. This standard outlines the minimum requirements public bodies must meet if they are seeking to destroy original source records after they have been digitized.

The Archives' Corporate Information Management (CIM) unit, Digital Preservation Unit (DPU), and Photograph Archivist are available to assist public bodies in their digitization projects.

1.1 Scope

This standard applies to all public bodies as defined in the *Archives Act*. It identifies the functional and technical requirements for the implementation of a digitization program, and provides direction in the following areas:

- The creation of accurate, reliable, and authentic images in order to meet legal admissibility requirements;
- Ensuring images remain accessible for as long as required;
- Criteria for the disposal of original source records after digitization;

Born-digital records are outside the scope of this standard. For information on the management and long-term preservation of born-digital records consult the Provincial Archives *Guidelines for Transferring Digital Records* and *Managing and Exporting Web Resources for Long-Term Preservation*.

The digital conversion of audio and video records is also outside the scope of this standard. Please contact the Sound and Moving Images unit at the Provincial Archives of New Brunswick for recommendations and advice.

The Provincial Archives of New Brunswick invites your feedback on this standard. Please send questions, comments, and/or suggestions to records.centre@gnb.ca and/or Digitalpreservation.numerique@gnb.ca

1.2 Legal Accountability

Digitized records are subject to the same provisions as other record formats, as established by the following Acts and policies:

- *Archives Act* sets the legal framework for disposition, transfer, custody and access to records.

- *Copyright Act* sets the statutes governing copyright in Canada.

- **Electronic Transactions Act** establishes the legal authority of digital records in regard to transactions.

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

- **Government Information Technology Systems Security Policy** outlines principles and directives for managing all government information assets.

- **Public Records Act** establishes the Province's custody of records created and kept by any officer of the Province, municipality or rural community.

- **Records Management Policy (AD1508)** establishes the responsibilities of all New Brunswick Government bodies in the management of records, including digital records.

- **Right to Information and Protection of Privacy Act** provides public access to records and protects against unauthorized use or disclosure of personal information by public bodies.

2. PLANNING A DIGITIZATION PROJECT

A public body beginning a digitization project must consider the functional and technical requirements outlined in this standard to ensure that:

- Original paper records are not disposed of without authorization;
- The authenticity and integrity of the digitized records are able to be demonstrated;
- The images are accessible over the long term.

Public bodies should consider how the digitization project will integrate into existing workflows. Public bodies should work closely with their Records Managers who have a good understanding of existing records management practices and processes. Process decisions will need to be made and documented in relation to:

- The reasons why records should be digitized;
- When and where the records will be digitized, and who will digitize them;
- How the record will be digitized, what metadata will be captured, and how quality will be assured.

2.1 Benefits and Challenges of Digitizing

It is important to consider the benefits and challenges of digitizing records before starting a digitization project. Some benefits and challenges may be unique to the particular public body and digitization project.

Benefits:

- Reduction of paper records (saving space): Digitized records can be stored on electronic storage devices and thus reduce the need for physical space;
- Improved workflow and increased business efficiencies: Digitized records provide improved and efficient access to information and sharing by integrating the records with current business information management systems. This also removes the hybrid (dual) system of managing digital and paper records;
- Improved access: Digitized records can potentially provide improved access to the records internally by integrating with existing information systems, and externally by posting the images on the web;
- Protection: Digitized records reduce handling of source records, which may have deteriorated and become fragile, but need to be preserved because of their intrinsic value. It also reduces the risks associated with having a single physical copy of the record. Records can be integrated into existing back-up systems and disaster recovery processes.
- Security: Access rights can be used to control who can view the record;

Challenges

- The costs associated with the long-term preservation and accessibility of digital records may be more than required to physically store the original source records;
- Technical standards implemented by the organization will have an effect on the long-term preservation of the records, and the ability to re-use the records;
- The requirements to process the records may require additional resources, such as hardware and software, people, and storage space;
- Legal requirements to maintain authentic and reliable records may limit the ability to manipulate the digitized records;
- Legal requirements may require public bodies to keep the hard-copy version, even after the records have been digitized;
- Without careful planning, increased or unforeseen resources may be required.

2.2 Types of Digitization (Access Model)

There are three different ways to digitize records for purposes of access:

- 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 on-line navigation.

- Full Text (Optical Character Recognition) – This translates the digitized record into machine-readable text, which can then be changed or manipulated. Public bodies can either type from the original document (known as keying) or use an optical character recognition (OCR) program to convert page images to ASCII. OCR only works with pages that are typed. For handwritten originals, public bodies must re-type (key) the record.

- Encoded Text, or Full Text with Mark-up - This option requires the institution to provide the same choices and options as are noted for full text, but adds additional search functionality by placing Standard Generalized Markup Language (SGML) (ISO 8879:1986), Extensible Markup Language (XML), Hyper Text Markup Language (HTML) and others around selected text. Applications are able to search for possible captions created using encoded text or mark-up language.

If the public body chooses to use full-text or full-text with mark-up for web publication, it may want to consider using diplomatic rendition. Diplomatic rendition attempts to present the text as it appears in the original source document. Editorial comments are added to the document describing errors or issues such as illegible or unclear text. Diplomatic rendition however requires more resources in terms of finances, staff, and time.

2.3 Digitization Equipment

The equipment required depends on the project. It is important to consider the goals of the project, record format, size, condition of the records, and the costs of equipment. Other technical requirements may also influence your choice of device(s), such as the size of scan area, speed, and connectivity.

The physical format and size of the records will dictate which type of capture device to use.

Type of Record	Image Capture Deice
Oversized Textual Records, (e.g.) Newspapers, Maps, Plans	Large Format Scanner or Wide-format Scanner
Textual Records and Photographs	Flatbed Scanner
	Automatic Document Scanner
Books	Overhead Scanner
Film	Slide/Film Scanner
Art	Digital Camera

Large Format Scanner – These scanners are used for scanning maps, blueprints, architectural drawings, site plans, posters, etc. They are similar to a Flatbed Scanner, but are considerably larger.

Flatbed Scanner – These scanners are used for scanning papers, photographs, negatives/film, slides, printed materials, etc.

Automatic Document Scanner – These scanners are similar to a flatbed scanner, but have an automatic document feeder. These should only be used when scanning numerous clean, high-contrast records with printed type (e.g. laser printed or typeset). A visual inspection of the records will need to be done prior to using an automatic document scanner to ensure that the records will be fed without risk of being crumpled or torn by the automatic document feeder.

Overhead Scanner – These scanners are used to scan various sized books or other records that cannot be laid flat. Unlike flatbed scanners they allow for overhead copying, which reduces the damage to books and other records (particularly rare books) when pressing down. High end overhead scanners also include software to adjust and compensate for the distortion resulting from the curve of the page.

Slide/Film Scanner – These scanners are used to scan slides and film. Although flatbed scanners are capable of scanning these records, a dedicated slide/film scanner produces a higher quality image.

Digital Camera – For digitization projects, "point and shoot" cameras are not recommended. Ideally, public bodies should use a DSLR camera. These are capable of making high-quality images with appropriate image resolution, suitable for archiving and publication. Another advantage of using a digital camera is the ability to "tether" the camera to a computer. This process enables you to bypass the small screen on the camera and instantly view the image on the computer monitor. When using a digital camera, photographers will need to consider other factors, such as digital scan backs, camera stands, and lighting.

Before purchasing hardware (computer, scanner) read reviews and ask about the experience of other public bodies who own similar hardware about its reliability, warranty, and replacement costs. When choosing a scanner test the equipment and check spatial resolution, tonal reproduction, colour reproduction, noise, and artifact detection.

2.4 Software Considerations

The ability to select mode of capture, bit depth, resolution, original format etc., are important considerations when purchasing software compatible with the hardware device. In some cases, when purchasing a high quality scanner, the software is included. A review of the software capabilities should be performed along with the hardware device.

2.5 Planning for Quality Assurance

Quality assurance procedures should be established before beginning a digitization project. For recommendations on quality assurance see Section 7: Quality Assurance.

2.6 Preparing a Business Case

When the public body has a thorough understanding of the business requirement and the records, it is recommended that they prepare a business case. The business case should describe the records and their relationship to other records, the benefits, and costs of digitization. It should also address:

- The purpose of the project and benefits;
- Project management considerations, such as quality controls, and assurance procedures;
- Records management issues associated with the source records and the digitized records;
- Requirements established for the disposal of source records after digitization;
- Requirements established by the Records Retention and Disposition Schedule;
- Appropriate technical specifications.

2.7 Preparing an Internal Digitization Policy

Before a public body begins a digitization project, it should first develop a broad internal digitization policy applicable to all digitization projects. Such a policy should be supported by senior management, and outline the following:

- The criteria for identifying and selecting records to be digitized;
- Information on how the digitization process will be managed, including staff responsibilities;
- Information on how the public body plans to capture and manage the records metadata;
- A technical standard with approved image formats, compression algorithms, resolution and bit depth;
- Information on how the digitization process will be documented and monitored for risks;
- Quality assurance procedures that include:
    - The timing of equipment tests and equipment calibration;
    - Procedures for checking output, such as what proportion of the digital reproductions will be subject to visual inspection and how long the original records need to be retained after digitization to ensure that quality checking processes can be undertaken;
    - Procedures for re-imaging if quality standards are not met;
    - Roles and responsibilities for checking and approving output.

3. Requirements for Disposal of Original Source Records

If a public body plans to destroy the original source records after digitization, and declare the digitized record as the official records of business activity, it must create, review, and/or change the Records Retention and Disposition Schedule. **Records are only eligible for disposal after digitization under an approved Records Retention and Disposition Schedule. Only source records considered to be at low risk are eligible for disposal after digitization (See Section 4.1).**

Original source records that have a disposal freeze relating to litigation (including records that are scheduled for destruction), must not be destroyed while the freeze is in place.

3.1 Records Retention and Disposition Schedules

A Records Retention and Disposition Schedule specifies the length of time the records must be kept in the office (active), how long they are stored off-site (semi-active), if applicable, and how the records will be handled at the end of their life (final disposition). The final disposition of the records is based on the administrative, legal, fiscal, historical, or other uses of the records.

The final disposition of records is undertaken in one of three ways:

- Destroy – requires records to be shredded, recycled, deleted etc. Departments and agencies are responsible for making the appropriate arrangements for the destruction of records in their care, custody, and control.

- Selective Retention – requires records to be transferred to the Provincial Archives for appraisal. An archivist at the Provincial Archives will assess the records and permanently preserve some, none, or all of them.

- Archival Retention – requires the complete series of records to be transferred to the Provincial Archives for preservation.

3.2 Records Retention and Disposition Schedules and Digitized Records

Traditionally, Records Analysts and Archivists have identified three stages in the lifecycle of a paper record: Active, Semi-Active, and Final Disposition. In the case of digital records, there are only two stages in the lifecycle: actions performed by the creator (Active and Semi-Active) and actions performed by the preserver (Final Disposition).

- In the first stage, the public body has the responsibility to create a retention and disposition schedule that is approved by the Provincial Archives of New Brunswick. While in the care, custody, and control of the public body, digital records must be stored in a well-managed system with appropriate recordkeeping functionality for as long as they need to be retained in accordance with the approved schedule.

- At the time of disposition, if the records are scheduled for destruction, the public body is responsible for carrying that out. If the records are scheduled for retention, the public body transfers care, custody, and control of the digital records to the Provincial Archives of New Brunswick.

The Records Managers of each public body will need to work with the Archives to determine the number of years the public body should maintain their digital records before transferring them to the Archives for Final Disposition. It is recommended that related electronic and paper records be transferred at the same time.

3.3 Disposal of Original Source Records after Digitization

For source records to be eligible for disposal after digitization, public bodies must meet the minimum requirements outlined below.

Requirement 1: Original source records intended for disposal after digitization must be scheduled under an approved Records Retention and Disposition Schedule approved by the Provincial Archivist and must not be subject to a temporary disposal freeze.

To be compliant with this principle, at a minimum the public body must:

- Ensure all source records intended for disposal after digitization are covered under a current Records Retention and Disposition Schedule approved by the Provincial Archivist;
- Ensure a *Records Destruction Form* has been filed relating to the disposal of the source records, once approved by the Records Retention and Disposition Schedule.
- Follow any disposal freeze while the disposal freeze remains in place.

Requirement 2: A public body must have assessed the need to retain the original source records based on any ongoing legal, business or other requirements as low risk. (See Section 4.1)

To be compliant with this principle, at a minimum the public body must:

- Undertake a risk assessment and business needs analysis to determine which source records are eligible for disposal after the records have been digitized based on documented criteria;
- Seek legal advice if necessary, regarding the legal admissibility of the digitized record;
- Develop an internal policy statement outlining the handling and disposal of original source records after digitization based on the disposal requirements outlined by the public body's Security Information Officer.
- Monitor changes which may have an impact on the public bodies' need to retain the original source records and take any necessary steps to remain compliant with the *Government of New Brunswick Digitization Standard*.

Requirement 3: Trusted systems and processes to capture and manage digitized images as records must be in place.

Digitized records, once created, must be stored in a well-managed system with appropriate recordkeeping functionality for as long as they need to be retained in accordance with an approved Records Retention and Disposition Schedule. Trusted systems and processes in this context are those which operate reliably and consistently meet minimum recordkeeping standards. Trusted systems support the admissibility of digitized records in legal proceedings.

To be compliant with this principle, at a minimum the public body must have:

- Digitization plans, policies, and procedures covering the identification of records selected for digitization, how the digitization process will be managed including staff responsibilities, and the capture and management of images and technical metadata;
- A technical standard for reproduction, including approved image formats, compression algorithms, resolution and bit depth, and the use of any image processing procedures;
- Quality assurance procedures that include:
  - The timing of equipment tests and equipment calibration;
  - Procedures of checking output, such as what proportion of the digital reproductions will be subject to visual inspection and how long the original

records need to be retained after digitization to ensure that quality checking processes can be undertaken;
- Procedures for re-imaging if quality standards are not met;
- Roles and responsibilities for checking and approving output.

- A well-managed system with appropriate recordkeeping functionality to ensure that the digitized images can be managed as records for as long as they are required to be retained in accordance with an approved Records Retention and Disposition Schedule. This includes:
    - Adequate physical and other security safeguards to ensure the digitized records remain secure and can only be changed in an authorized manner;
    - Ability to capture and maintain metadata elements;
    - Capability to generate and capture appropriate technical metadata at the point of digitization;
    - An approved business continuity and disaster recovery plan and migration strategy to ensure that the digitized records are not placed at risk because of technological obsolescence.

4. RECORDKEEPING REQUIREMENTS

Public bodies are responsible for creating full and accurate records of their activities and for documenting the processes they used during the digitization process. The following recordkeeping requirements are designed to support the principles outlined in section 3.3.

4.1 Conducting a Risk Assessment

A public body must have assessed the need to retain the original source records based on any ongoing legal, business or other requirements as low risk. For the purpose of this standard, low risk refers to the improbable likelihood of risk occurring. To determine the risk level of the record, a risk assessment must be completed on the records before the records are digitized. Public bodies must use the following Risk Assessment Matrix to identify and assess the risk levels of their records.[1]

Risk Criteria	Questions to consider when assessing the risk level of the record	Comments
There is a risk that a full and accurate digitized image of the original source record will not be created.	Has the public body reviewed and identified the necessary technical and metadata specifications? Has the public body put in place the necessary quality assurance processes to ensure the full and accurate digitized images of the originals?	The public body should consider whether it can accurately digitize the original source record, or whether issues such as missing metadata, or the lack of quality standards will affect the digitization process.
There is a risk that access to the source record during its minimum records retention and disposition	Does the public body have a strategy in place for the possible long-term preservation of digital records?	It is the responsibility of the public body to care for records during the records active and semi-active period. Public bodies need to ensure that digital preservation strategies are in place to reduce the risks to digital records, such as technological obsolescence.

---

[1] The following Risk Assessment Matrix has been adapted from the Queensland State Archives.

schedule cannot be assured.		
There is a risk that the authenticity of the digitized record could be challenged (whereas the authenticity can be proven from the original source record).	Does the public body have a trusted digital recordkeeping strategy and system in place for managing the digitized records? Can the authenticity of the record be demonstrated?	The public body must have appropriate recordkeeping strategies and systems in place to ensure their reliability and security. The system must be able to store and provide access to records for evidential purposes in a trusted way. Without trusted systems and processes, the authenticity of the records and the admissibility of digitized images in legal proceedings may be compromised.
	Are legal processes relating to the records underway, pending, or to be reasonably expected?	Public bodies should assess a range of legal risks, including: • Whether the class of records is likely to be needed for any future legal action, including litigation and Right to Information requests, and • Whether the original paper record will be required to be produced. Depending on the rules and procedures under which the relevant judicial or review body operates, a preference for paper or electronic forms of evidence may apply. If a public body has operations in other provinces or countries, a legal risk assessment of that jurisdiction's evidence requirements should be considered. Public bodies should attach a high risk rating to any record classes that have been contested or part of litigation in the past.
	Will subsequent business action be taken on the digitized records?	If the original source record is still the record of business action, it must not be destroyed. If subsequent business action is taken on the digitized copy of the original source record, it can be considered as the official record of business and is eligible for destruction.
	Do the records contain physical elements which attest to their authenticity or evidential value?	For example: • Records with corporate seals or watermarks; • Hand signed documents that are important to retain in their original format; These elements can serve to demonstrate authenticity of authorship. There may be situations that require an original record to be produced. Public authorities should seek legal advice in this regard where relevant, to assist with determining the risk of disposing of the original source record after digitization.

There is a risk that Government, or community expectations will not be met through access to a digitized image of the source record.	Do the original records have value as physical objects?	For example, are they: • Original works of art; • Records of personal significance, e.g. handwritten letters. While they may not have ongoing archival value, these records may provide an emotional connection to the creator or have display or exhibition qualities. Therefore there may be high stakeholder expectations that the original records will be accessible for as long as they are required to be retained.
	Are there any requirements that stipulate the records should be retained in the original format?	Each public body must determine whether: • Any records needs to be kept in a particular format to satisfy particular requirements; • Whether such requirements impact on the destruction of the original source record for a limited period of time or for the life of the record; Public bodies should consider legislation, policies, standards, directives, etc. relevant to the operating environment.

4.2 Documenting and Monitoring Risks

Once risks have been identified and the assessment documented, these must be included in the internal policy created by the public body. The policy should clearly outline:

- The Records Retention and Disposition Schedule;
- The minimum period the original source records must be retained after digitization to allow for their quality assurance;
- A documented risk assessment, using the Risk Assessment Matrix;
- The trigger for, and frequency of, reviews of the risk assessment;
- A Compliance Declaration, attesting that the minimum requirements outlined in this standard have been undertaken (Appendix A).

The public body is not required to submit the compliance declaration and risk assessment to the Provincial Archives of New Brunswick, but should make it available upon request. If a change to the digitization process takes place, public bodies should review the risk assessment to ensure the risk level has not changed.

An internal strategy should also be put in place to trigger the suspension of any disposal activity as a result of:

- Pending legal cases;
- A disposal freeze is issued;
- Legislation, policy, or other requirements to retain the original source records are implemented;
- A significant administrative change occurs;
- A change to the disposal status of the records under an approved Records Retention and Disposition Schedule.

4.3 Metadata

Metadata is an essential element in the digitization process, as it is designed to help categorize and index records for faster and more efficient retrieval of information. It provides consistent identification of records, preserving their authenticity, and implementing retention and disposition requirements. It also plays a role in the long-term preservation of digital records as it helps identify key information for the conversion/migration of the records, and places them in context by providing data about the creator, public body, and subject.

Metadata can be defined as:

- *Data describing context, content, and structure of records and their management through time.*
- *Structured or semi-structured information which enables the creation, management, and use of records through time and within and across domains in which they are created. Recordkeeping metadata can be used to identify, authenticate, and contextualize records; and the people, processes and systems that create, manage, maintain and use them.*

In general metadata is 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. Elements one would find in a typical bibliographic record are examples of metadata.*
- 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, file type and other technical information, and who can access it. Very often, there are subsets in this category. The two most typically included are rights management and preservation metadata.*

The ***Government of New Brunswick Recordkeeping Metadata Standard*** (in progress) must be used and implemented during the digitization process. While technical imaging metadata, such as image resolution, colour depth, compression, file format and sub-format are captured by the imaging software, as part of the 'Format' element , other metadata will need to captured or generated at the point of digitization, including

- Date of digitization;
- Equipment used;
- Action Officer.

This metadata can be captured as part of the 'Event History' element of the ***Government of New Brunswick Recordkeeping Metadata Standard.***

The National Information Standards Organization (NISO) has developed a detailed data dictionary of technical metadata for Digital Still Images:

It is recommended that public bodies reference this standard to determine if additional technical metadata should be recorded.

4.4 Documenting Quality Assurance

Quality assurance testing is recommended for any digitization project. It ensures the records integrity, reliability, and accessibility. For recommendations on documenting quality assurance see Section 7:

Quality Assurance. Records approved for disposal after digitization, may not be destroyed until after they have been checked for quality assurance.

4.5 Documenting Destruction of Source Records

Once the public body is confident that it has undertaken the necessary requirements outlined in this standard, it may destroy the original source records. To document the destruction of original source records, public bodies must fill out a Records Destruction Form (See Appendix B).

5. TECHNICAL SPECIFICATIONS

The technical specifications adopted by the public body must depend on whether the digitized records will replace the source records as the record of business; and how long the digitized records need to be retained. Image quality and longevity are determined by factors such as

- Bit depth
- Resolution;
- Pixel array;
- Compression.

To ensure image quality and promote consistency across government, public bodies must meet the minimum technical requirements stated in this standard.

The image file formats recommended in this standard are based upon the following general principles:

- Image formats should be open source (non-proprietary) or widely used and accepted within the government of New Brunswick;
- Image formats should be supported across GNB systems;
- Master images should be created at the highest quality possible;
- Derivative images can be created at a lower quality to provide faster retrieval and access.

For a list of the acceptable Master Image and Derivative Image formats see sub-section 5.6: Recommended Scanning Parameters.

5.1 Mode of Capture
Mode of capture refers to the method by which your device will capture a digital image:
- Bitonal – refers to an image made up of pixels that are white or black;
- Greyscale – multiple bits per pixel representing shades of grey;
- RGB – multiple bits per pixel representing colour;
- CMYK (Cyan-Magenta-Yellow-Black) – a colour model in which all colour are described as a mixture of these four process colours. It is typically referred to in relation to the four basic colours used in ink. Because CMYK is a derivative colour space that is device-specific, it should not be used when digitizing records.

5.2 Bit Depth
Bit depth measures the number of colours (or levels of grey in greyscale images) that are available in an image's colour shade palette. "A bit is the basic digital building block with a value of either 1 or 0. Every pixel sampled is assigned a value that corresponds to the colour/shade it represents.

An image with a bit depth of 1 has pixels with two possible values: black and white [bitonal]. An image with a bit depth of 8 has 2^8, or 256, possible values. Greyscale mode images with a bit depth of 8 have 256 possible gray values.

RGB images are made of three colour channels [Red, Green, and Blue]. An 8 bit per pixel RGB image has 256 possible values for each channel, which means it has over 16 million possible colour values. RGB images with 8 bits per channel (bpc) are sometimes called 24 bit images (8

bits x 3 channels = 24 bits of data for each pixel). RGB images with 16 bits bpc are called 48 bit images (16 bits x3 channels = 48 bits of data for each pixel)."[2]

Below is a comparison chart, illustrating the different image types in terms of bit (bit depth), number of colours available, and common names:

Bits Per Pixel	Number of Colours Available
1 bit bitonal	2 (black and white)
8 bit greyscale	256
16 bit greyscale	65536
24 bit colour	16777216
48 bit colour	281 million

5.3 Resolution

Resolution is measured by pixel dimensions (height and width) for on-screen use or by physical size and PPI. When digitizing textual records, resolution is determined using the Digital Quality Index (QI). However, manuscripts and other non-textual material, such as photographs, maps, and plans, which have no obvious fixed metric for measuring detail, are better determined by the pixel array (See section 5.4). In general, images scanned at a higher resolution will produce a better image, but this also increases the file size. Depending on the record, the resolution will need to be adjusted based on the size, quality, condition, and use of the digital object. When digitizing records, the preferred term to use is pixels per inch (PPI), but most scanners define resolution in dots per inch (DPI).

5.4 Digital Quality Index (QI) Formula for Textual Records

The recommended scanning resolution for textual records is determined using a Digital Quality Index (QI) formula. This formula, developed by the C10 Standards Committee of the Association of Image and Information management (AIIM), was based on the Quality Index formula used in the micrographic industry. Both formulas are 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.[3] As in the preservation microfilming standard, the digital QI formula forecasts levels of image quality: barely legible (3.0), marginal (3.6), good (5.0), and excellent (8.0).

5.4.1 Digital Quality Index for Greyscale Images

A Digital Quality Index for greyscale images is calculated based on the formula:

$$dpi = 2QI/0.39h$$

---

[2] BCR's CDP Digital Imaging Best Practices Version 2.0 http://mwdl.org/docs/digital-imaging-bp_2.0.pdf (June 2008)

[3] Anne Kenney and Stephen Chapman, *Digital Resolution Requirements for Replacing Text-Based Material: Methods for Benchmarking Image Quality*, Commission on Preservation and Access (April 1995), pg. i. See also, Moving Theory into Practice Digital Imaging Tutorial, Cornell University Library Research Department, (2003)

dots per inch = (2 dots per line)(Quality Index)/ (1 inch is equal to 0.39 mm)(height of the smallest letter)[4]

Example 1: the following formula was used to calculate the dpi to reach a Quality Index of 3, with the smallest letter "e" (shown above) being 1 mm, scanning in greyscale;

$$Dpi = 2QI/0.39h$$
$$Dpi = (2 \times 3)/(0.39 \times 1mm)$$
$$Dpi = 6/0.39$$
$$Dpi = 15.3$$
$$Dpi = 153$$

Example 2: the following formula was used to calculate the dpi to reach a Quality Index of 3.6, with the smallest letter "e" (shown above) being 1 mm, scanning in greyscale;

$$Dpi = 2QI/0.39h$$
$$Dpi = (2 \times 3.6)/(0.39 \times 1mm)$$
$$Dpi = 7.2/0.39$$
$$Dpi = 18.4$$
$$Dpi = 184$$

Example 3: the following formula was used to calculate the dpi to reach a Quality Index of 5, with the smallest letter "e" (shown above) being 1 mm, scanning in greyscale;

$$Dpi = 2QI/0.39h$$
$$Dpi = (2 \times 5)/(0.39 \times 1mm)$$
$$Dpi = 10/0.39$$
$$Dpi = 25.6$$
$$Dpi = 256$$

Example 4: the following formula was used to calculate the dpi to reach a Quality Index of 8, with the smallest letter "e" (shown above) being 1 mm, scanning in greyscale;

$$Dpi = 2QI/0.39h$$
$$Dpi = (2 \times 8)/(0.39 \times 1mm)$$
$$Dpi = 16/0.39$$
$$Dpi = 41.0$$
$$Dpi = 410$$

5.4.2 Digital Quality Index for Bitonal Images

The formula for bitonal scanning provides a generous over sampling to compensate for misregistration and reduced quality due to thresholding information to black and white pixels:

---

[4] The smallest letter is typically a small "e" and is measured in millimeters.

$$Dpi = 3QI/0.39h$$

Example: the following formula was used to calculate the dpi to reach a Quality Index of 8, with the smallest letter "e" (shown above) being 1 mm, scanning as bitonal;

$$Dpi = 3QI/0.39h$$
$$Dpi = (3 \times 8)/(0.39 \times 1mm)$$
$$Dpi = 24/0.39$$
$$Dpi = 61.5$$
$$Dpi = 615$$

5.5 Pixel Array

Pixel Array refers to "the number of pixels across both dimension of an image in terms of rows and columns across the dimensions of the image."[5]

Text

When resolution cannot be accurately determined for textual records using the Digital Quality Index (QI), such as manuscripts that are handwritten, the goal is to have the widest edge as close to 6,000 pixels wide for 1 bit bitonal and 4000 pixels wide for 8-bit greyscale to 24-bit RGB files. The smaller the size of the original the higher the resolution you will need to select in order to achieve this width. For example, an 8.5 x 11" paper record needs to be scanned at 400 ppi (8-24 bit depth) in order to achieve a wide edge of 4400 pixels.[6]

Recommended scanning resolution and pixel dimensions		
Originals	Resolution in PPI	Dimensions in pixels
8.5 x 11"	600 (1 bit)	5100 x 6600
	400 (8-24 bits)	3400 x 4400
8.5 x 14"	400 (1 bit)	4400 x 5600
	300 (8-24 bits)	2550 X 4200
11 x 17"	400 (1 bit)	4400 x 6800
	300 (8-24 bits)	3300 x 5100

Photographs

When scanning photographs, the goal is to have the widest edge around 4000 pixels wide. The smaller the size of the original the higher the resolution you will need to select in order to achieve this width. For example a 35mm film which is about 1.5 inches wide needs to be scanned at 2700 PPI in order to get 4050

---

[5] Steven Pugila, "Technical Primer," *Handbook for Digital Projects: A Management Tool for Preservation and Access.* Northeast Document Conservation Center, Andover, MA (April 2004).
[6] U.S. National Archives and Records Administration (NARA) *Technical Guidelines for Digitizing Archival Materials for Electronic Access: Creation of Production Master Files – Raster Images* (June 2004).

pixels on the widest edge while an 8x10 negative or print only needs to be scanned at 400 PPI in order to achieve a wide edge of 4000 pixels.

Recommended scanning resolution and pixel dimensions		
Originals	Resolution in PPI	Dimensions in pixels
35mm film (1.5")	2700	4050 X 2700
2.75 x 4" film	1000	2750 X 4000
4 x 5" film	800	3200 X 4000
5 x 7"	600	3000 X 4200
8 x 10"	400	3200 X 4000

Maps and Architectural Plans

When scanning maps, plans, graphic illustrations/artwork the goal is to have the widest edge between 6000 to 8000 pixels wide for 16-bit greyscale or 48-bit RGB.[7]

5.6 Compression

Compression relates to algorithms designed to reduce the size of an image and is primarily expressed in terms of lossy or lossless.

- Lossy: Information is removed from the stored information during the compression process, resulting in information being lost. However, lossy images such as JPEG are smaller in file size and take less time to load.
- Lossless: No information is lost. Images appear the same as the original. However, lossless images such as TIFF are larger in file size and take longer to load.

Lossy compression is not recommended when the source record is being replaced by the digitized record as the official business record. Because information is lost during the process, the accuracy of the image could be called into question. However, when the source records are retained by the public body or transferred to the Archives, lossy compression can be used since the original source record can be retrieved if needed.

5.7 Master Image or Derivative Image

The master image file is the higher quality (archival) image. It must be capable of supporting a wide range of users' needs, including the creation of derivative images for printing, display, and image processing. Since user expectations are likely to be more demanding over time, the digital master file must be able to accommodate future applications. Master image files are therefore scanned at a 1:1 ratio. This means that the digital image can be reproduced as a legible facsimile when produced in the same size as the original source record.

---

[7] Library and Archives Canada *Digitization Strategy* (October 2008). See also *Yale University Digitization Shared Practices – Still Images Version 1.0*, *odai.yale.edu/node/252/attachment (August 2010);* BCR's CDP Digital Imaging Best Practices Version 2.0 http://mwdl.org/docs/digital-imaging-bp_2.0.pdf (June 2008)

If a record is scheduled for selective retention or archival retention, and the public body plans to destroy the original source records after digitization, and declare the digitized record as the official records of business activity, it must digitized as a master image file.

If a public body plans to keep the original source records after digitization, and retain the original source records as the official records of business activity, it is recommended that the original source records be digitized as a derivative image.

If a record is scheduled for destruction, and the public body plans to destroy the original source records after digitization, and declare the digitized record as the official records of business activity, it is recommended that the original source records be digitized as a derivative image. However, if a public body must maintain the digitized record for 10 years or more, it is recommended that the public body digitize the record as a master image file, due to the risks of technological obsolescence.

5.8 Scanning Parameters

The following scanning parameters specify the bit depth, resolution, and file format needed to produce an archival master image, and recommendations for a derivative image.

5.8.1   Textual records – Master Image

Characteristics of Original	Recommend Image Parameters for Master Files	Recommended File Format
Clean, high-contrast documents with printed type (e.g. laser printed or type set).	1-bit bitonal mode or 8-bit greyscale – adjust scan resolution to produce a QI of 8 for the smallest significant character; or 1-bit bitonal mode – 600 ppi for documents with smallest significant character of 1.0 mm or larger; or 8-bit greyscale mode – 400 ppi for documents with smallest significant character of 1.0 mm or larger. Note: Regardless of approach used, adjust scan resolution to produce a minimum pixel measurement across the long dimension of 6,000 pixels for 1-bit files and 4,000 pixels for 8-bit files.	PDF/A or TIFF
Documents with poor legibility or diffuse characters (e.g. carbon copies, Thermofax/Verifax, etc.), handwritten annotations or other markings, low inherent contrast, staining, fading, halftone illustrations, or photographs.	8-bit greyscale mode – adjust scan resolution to produce a QI of 8 for smallest significant character; or 8-bit greyscale mode – 400 ppi for documents with smallest significant character of 1.0 mm or larger. Note: Regardless of approach used, adjust scan resolution to produce a minimum pixel measurement across the long dimension of 4,000 pixels for 8-bit files.	PDF/A or TIFF
Documents as described for greyscale scanning and/or where colour is important to the interpretation of the information or content, or	24-bit RGB mode – adjust scan resolution to produce a QI of 8 for smallest significant character; or 24-bit RGB mode – 400 ppi for documents with smallest significant character of 1.0 mm or larger.	PDF/A or TIFF

desire to produce the most accurate representation.	Note: Regardless of approach used, adjust scan resolution to produce a minimum pixel measurement across the long dimension of 4,000 pixels for 24-bit files.	

5.8.2   Textual records – Derivative Image

Format – Original
1-bit bitonal mode – 300 ppi for documents with smallest significant character of 2.0 mm or larger;   or   8-bit greyscale or 24 bit colour mode – 300 ppi for documents with smallest significant character of 1.5 mm or larger.   Recommend File Formats: PDF or JPEG

5.8.3   Textual records – Microfilm

Format – Original
For duplicates, follow the scanning recommendations for the size that matches the actual dimensions of the duplicate. However, if original size is not known, adjust scan resolution to produce a minimum pixel measurement across the long dimension of 4000 pixels for 8-bit greyscale.   Scan resolution calculated on the actual dimension of the duplicate – Original 8.5 x 11" textual record on 35mm microfilm – 400ppi (8-bit greyscale, PDF/A or TIFF) for Master Image, 300ppi (8-bit greyscale, PDF or JPEG) for Derivative Image.

5.8.4 Photographs – Master Image

Format - Original	Master Image	Recommended File Format for Long-Term Preservation
35mm film (1.5") up to 8 x10"	Adjust scan resolution to produce a minimum pixel measurement across the long dimension of 4000 pixels for 16-bit greyscale, or 48-bit RGB. Scan resolution calculated on actual dimensions – approx. 2700 ppi for 35mm film (1.5") original and ranging down to approx. 400 ppi for 8 x 10" originals.	TIFF

5.8.5 Photographs – Derivative Image

Format – Original
Adjust scan resolution to produce a minimum pixel measurement across the long dimension of 4000 pixels for 8-bit greyscale or 24-bit RGB. Recommend File Formats: JPEG

5.8.6 Maps and Architectural Plans – Master Image

Format - Original	Master Image	Recommended File Format for Long-Term Preservation
Scanning maps may involve items that vary widely in size, condition, and amount of detail.	Adjust scan resolution to have the widest edge between 6000 to 8000 pixels for 16-bit greyscale, or 48-bit RGB.	PDF/A or TIFF

5.8.7 Maps and Architectural Plans – Derivative Image

Format - Original
Adjust scan resolution to have the widest edge between 6000 to 8000 pixels for 8-bit greyscale, or 24-bit RGB.
Recommend File Formats: PDF or JPEG

6. THE DIGITIZATION PROCESS

Source records must be assessed and prepared prior to imaging. The quality of the images is dependent on the quality and characteristics of the source records. Size, colour, and physical condition should be taken into consideration, as explained in sub-section 5.6 Recommended Scanning Parameters.

It is recommended that clearly defined business rules and policies be established to address the following:

- The management of fragile or damaged source records. Not all source records will withstand the physical handling required for automatic document scanners (feeders). If records are fragile or damaged consider the type of digitization equipment the public body is using, perhaps there is an alternative type of scanner that will do less damage to the source record (see sub-section 2.4). In some cases, such as maps, a clear protective covering or sleeve will be required. In cases where the public body is unsure of how to handle fragile or damaged source records, advice and assistance is available from the Conservation Unit of the Provincial Archives of New Brunswick;
- Removal of papers from bindings such as staples, file clips, or paper clips;
- Managing loose items to ensure that the original order of the records is maintained;
- Managing adhesive notes, white out, blank pages, faded, torn or illegible pages and reverse pages on the file;
- Other possible issues unique to the digitization process.

6.1 Processing the digitized image

Once the original source record is scanned, a digitized version of the source record will be produced. The initial capture of the digitized records may be in a proprietary format specific to the digitization equipment. In some cases, the images will be processed as RAW files, which will need to be converted to one of the preferred formats outlined in Sub-section 5.6 Recommended Scanning Parameters. In other cases, the software specific to the digitization equipment may allow the public body to scan directly as one of the preferred formats.

After the records have been digitized, public bodies may want to enhance the digitized record. It is recommended that the Master Image be left unaltered for purposes of authenticity, integrity, and reliability. Public bodies should make a derivative copy which all enhancements are applied to.

The digitized records will also need to be checked for quality assurance. For information on quality assurance see Section 7: Quality Assurance.

6.2 Processing Sensitive Information

Sensitive information requires a high degree of confidentiality. It must be protected against unauthorized access, disclosure, removal, modification, and interruption. During the digitization project, original source records must be protected to ensure the **confidentiality, integrity, and availability** of information. Both the original source record and digitized record must also be protected after the digitization process.

The Government of New Brunswick has four levels of security classification: Unclassified, Low, Medium and High. Each level of classification indicates an increasing degree of sensitivity.

- With **unclassified** information and assets, the disclosure or loss would have no effect on the enterprise. This information is often considered to be beneficial to the public and enterprise.

- With information or assets classified as **low**, the disclosure or loss could be expected to be of a limited nature. The information should remain within government and not be distributed to public.
- With information or assets classified as **medium**, the disclosure or loss would be expected to be an invasion of privacy, or cause hardship or monetary loss or could be detrimental to relations between governments. In general, information about individuals which is not widely available should be protected whether or not such information would be embarrassing to the person.
- With information or assets classified as h**igh**, the disclosure would be expected to be a major invasion of privacy, cause serious harm to citizens, corporations or employees; or could be detrimental to relations between governments. This information requires special handling and security procedures to adequately protect it.

When handling records classified as low, medium, or high, it is recommended that the public body review the *Government Information Technology Systems Security Policy* (GISSP) and its accompanying Standards and Directives, and the *Right to Information and Protection of Privacy Act* (RTIPPA).

Once the record has been digitized, the records classification can be added to the record metadata as part of the "Security Classification" sub-element of the "Access" element, of the *Government of New Brunswick Recordkeeping Metadata Standard.*

6.3 Integrating Digitized Records into Existing File Structure

Trusted systems and processes to capture and manage digitized images as records must be in place. These systems must only permit access to the digitized records and its metadata by authorized employees, and provide appropriate security controls and measures to preserve the records authenticity, integrity, reliability, and accessibility. Appropriate back-up and disaster recovery processes must also be in place. In the case where the records have a long life span in the office, appropriate migration strategies should be developed to reduce the risk of technological obsolescence.

7. QUALITY ASSURANCE

Quality assurance provides the means to measure, monitor, and analyze digitization projects, and improve the process if necessary. It ensures the integrity, reliability, and accessibility of the digitized records. Specific controls need to be established and maintained to assess the quality of the:

- Digitization equipment;
- Business process of creating the digitized image;
- Metadata.

All quality assurance procedures should be documented and approved by senior management. All quality control data (such as logs, reports, decisions) should also be retained by the public body. This supports the records authenticity and could assist with future long-term preservation decisions.

Quality assurance procedures should be established before beginning a digitization project. The internal procedures adopted by the public body should also be reviewed on an annual basis, or as needed to reflect changes in the process, or because of emerging technology.

7.1 Digitization Equipment

In any digitization project, the hardware purchased to digitize the records will need to be calibrated. Software will also need to be updated to ensure the best quality standards. To establish acceptable levels of quality to capture a digital image, the digitization equipment should be tested using test targets or charts.

Calibration testing involves using a test target, which is digitized, and then compared with the benchmark settings. To establish a benchmark, the test target should be scanned at a high resolution, and in full scale view. The International Standard ISO 12653 *Electronic Imaging: Test target for black-and-white scanning of office documents Parts 1 and 2* contains guidance on evaluating the output quality of a black-and-white scanning system against a test target. ISO 12641 – *Graphic technology – prepress digital data exchange – Colour targets for input scanner calibration* provides advice related to evaluating the output quality of colour scanning systems.

The International Council of Archives has outlined several quality control measures that public bodies should adopt relating to issues such as, spatial resolution, tonal reproduction, colour reproduction, noise, and artifact detection. In projects which are scanning oversized material, such as maps and plans, geometric accuracy is also an important factor. For detailed information see the International Council of Archives, *Guidelines for Digitization Projects for Collections and Holdings in the Public Domain, Particularly those held by Libraries and Archives*.

Spatial Resolution

When digitizing records, the preferred term to use is pixels per inch (PPI), but most scanners define resolution in dots per inch (DPI). It is important to note that a scanner with a maximum resolution of 600 dpi, in practice means that it scans at a maximum of 600 ppi. This optical sampling rate is only one component necessary for resolution. Other factors include the quality, focal range, and mechanical stability of the optical system (lens, mirrors, and filters); the input/output bit depth; the vibrations of the source document and charge-coupled device (CCD); and the level of image processing applied to the image.

It is important to determine the Optical (true) resolution and the Interpolated resolution. "Optical (true) resolution is the inherent resolution of the scanner based on the size of the imaging sensor and the magnification of the optical system. Interpolated resolution is synthetic or calculated resolution. Interpolation is a mathematical process that is used to increase or decrease the resolution of an image.

This can be done during or after scanning. Higher optical (true) resolution in a scanner will provide better image quality than interpolated resolution."[8]

The International Council of Archives recommends several methods of evaluating resolution:
- *Resolution targets,* which were originally made for use in micrographic and photographic industries. They are normally used to measure the reproduction of details, uniform capture of different parts of a source document, image sharpness etc. The results are not always trustworthy but resolution targets are still practical tools to use especially for bitonal conversion.
- *The Modulation Transfer Function (MTF),* in which the spread of light in the imaging process (line spread function) is measured. This is a more reliable and objective way to evaluate how well details are preserved and best suits greyscale and colour systems.
- *Spatial Frequency Response (SFR),* which means measuring the ability of the scanner to transmit high-frequency information by means of a specified transfer function (in practice equivalent to MTF).

Examples of targets in use for resolution are:
- IEE (Institute of Electrical and Electronic Engineers) standard Facsimile Test Chart
- AIIM Scanner Test Chart no 2
- PM-189 (A&P International) Resolution Target
- The Scanner SFR and OECF no 2 Target (Applied Image Inc.)

A standard for resolution evaluation is ISO 12233 *Photography, Electronic Still Picture Cameras, Resolution Measurements.*

Tone Reproduction

A visual inspection will show if an image has acceptable tone reproduction. In practice tonal reproduction determines how dark or light an image is as well as its contrast. Due to various (electronic) noises in the digitization equipment, there will always be losses in the bit depth. It is recommended to digitize an image at a higher bit depth than is needed for the digital output, for example at least 12 and 14 bits/channel to get an 8 bit output greyscale.

There are three common attributes that affect tone reproduction: the opto-electronic conversion function (OECF), dynamic range, and flare:
- The OECF shows the relationship between the optical densities of an original and the corresponding digital values of the file.
- Dynamic range refers to the capacity of the scanner to capture extreme density variations. The dynamic range of the scanner should meet or exceed the dynamic range of the original.
- Flare is generated by stray light in an optical system. Flare reduces the dynamic range of a scanner.

Tonal values are measured using a tone reproduction curve, which compares the paper record or microfilm with the digital image. Tonal values are also sometimes assessed using a histogram, which graphically illustration the tones in the image, between the brightest bright and the darkest dark.

Colour Reproduction

Colour reproduction is the process of ensuring that the coloured original record is reproduced in the same way on other screens or printouts. This is difficult to achieve, however, since monitors, operating systems and system applications, display colour in different ways. The colour humans see also differs between individuals.

---

[8] Steven Pugila, "Technical Primer," *Handbook for Digital Projects: A Management Tool for Preservation and Access.* Northeast Document Conservation Center, Andover, MA (April 2004).

Targets and Colour Bars are often used to help measure system resolution, tonal range, and colour. Colour targets are created using a set of colour patches with established numerical values and grey. Different targets exist for the records being scanned, such as paper, film, and photographs. Some digitization projects include a colour bar or greyscale bar beside the original record when being scanned, providing the users the opportunity to verify the records colour and greyscale accuracy. These bars can assist when adjusting the white balance to accurately represent the shading in an image.

Examples of targets' in use for colour and greyscale include:

- Kodak Colour Separation Guide and Grey Scale (Q13 and Q14)
- Kodak Q-60 Colour input Target (IT8)
- RIT Process Ink Garmut Chart.

Colour management software is also available to adjust the colour profile of various devices, including monitors, in order to obtain the same appearance across all devices used in the digitization workflow. Colour management software uses the International Color Consortium (ICC) standards. Recommended Colour management software includes:

- Adobe CMM;
- ColorSync (Apple);
- LittleCMS (open source software);
- Argyll CMS (open source software).

Artifacts

Artifacts are physical objects that may interfere with the quality of the scanned image, such as dust on the record, or scratches and streaks on the lens of the scanner or camera. In some cases, when digitizing photographs and negatives Newton Rings may be visible. Newton Rings are formed due to interference created by reflection of light from a flat surface onto a spherical surface. In the most extreme cases, a visual inspection will reveal these issues.

7.2 Visual Inspection

Colour bars and targets are designed to assist the operator performing the digitization process by providing a benchmark to determine the quality of the image. In the end, however, a visual inspection of the records will need to be conducted to determine the quality. Human judgment is often subjective, leading to different results from person to person. A policy should be put in place, and employees properly trained, to ensure the record was properly captured during the digitization process. The record should be visually inspected by at least two employees. In cases of discrepancy between the first two employees a third employee (supervisor) should be consulted.

Aspects against which the digitized record could be inspected and checked include:

- Has the smallest detail been legibly captured? (e.g. smallest type size for text; clarity of punctuation marks, including decimal points);
- Are all details complete? (e.g. acceptability of broken characters, missing segments of lines, missing information at the edges of the image area, images cropped or incomplete);
- Do the dimensions accurately compare with the original?
- Does the colour accurately compare with the original? (E.g. density of solid black areas – too light? Too dark?; colour fidelity);
- Is the sharpness of the image comparable to the original (e.g. lack of sharpness or too much sharpening; unnatural appearance and halos around dark edges)
- Where optical character recognition (OCR) is used, is the captured text accurate?

7.3 Validating the Digitized Image

Records are only eligible for disposal after digitization under an approved Records Retention and Disposition Schedule. Before disposal the digitized records must first be validated. Public bodies must validate the number of input paper documents against the number of digitized images created. For multi-page items, the number of pages within the record must accurately reflect the number of digitized images created, and are structured and arranged in the original order.

To validate the digitized image, a visual inspection of the records will need to be conducted. While testing all the digital images will ensure that the images meet the minimum requirements outlined in this standards, this can be time and resource intensive. The Provincial Archives of New Brunswick recommends that public bodies do a 10-25% random sampling of its digitized images. In some cases, such as following equipment repairs, or if using new staff, each image should be checked until there is confidence that the standard is being met. When sampling the records for quality care must be taken to ensure that the sample is representative of the range of records digitized.

In cases where the randomly selected sample was found to be defective the entire collection of digitized images since the last sample would need to be inspected. If less than 1% of the batch is found to be defective then only the specific defective images and metadata that are found in the re-inspection need to be re-digitized.

To document the quality assurance process public bodies should use an Image Quality Control Log (see example below). Public bodies may decide how they log these results internally to provide greater functionality and accessibility with its existing workflow. Image Quality Control Logs should be maintained for the life of the images.

Image Quality Control Log

Date	Records Retention and Disposition Schedule	QC Technician Name	Results (OK/Failed)	Identifier(s)	Issue(s)	Comments/Action Taken)
2012-01-30		Joseph Smith	OK			
2012-05-20		Jane Doe	Failed	00044	Poor Colour Poor Contrast	Rejected and sent for rescanning.

Possible Issues:

- Broken/illegible characters
- Poor contrast/tone
- Poor colour
- Repeating scan lines
- Portion of image missing
- Incorrect orientation
- Too much noise
- Images out of sequence
- Images missing
- Incorrect image size
- Other (please specify)

7.4 Validating Metadata

Metadata is an essential element in the digitization process. It is designed to help categorize and index records for faster and more efficient retrieval of information. As part of the quality assurance process checks should be determined, documented, and implemented to ensure that the metadata that was entered manually and or automatically generated was captured during the process. Metadata should be checked for:

- Accuracy of grammar, spelling, and punctuation, especially for manually keyed data;
- Consistency in the creation of metadata and in interpretation of the metadata;
- That all mandatory element and sub-elements are complete.

To document the quality assurance process public bodies should use a Metadata Quality Control Log (see example below). Public bodies may decide how they log these results internally to provide greater functionality and accessibility with its existing workflow. Image Quality Control Logs should be maintained for the life of the images.

Metadata Quality Control Log

Date	Records Retention and Disposition Schedule	QC Technician Name	Results (OK/Failed)	Identifier(s)	Element (sub-element)	Comments/Action Taken
2012-01-30		Joseph Smith	OK			
2012-05-20		Jane Doe	Failed	00045-00055	eventHistory (eventDate)	eventDate Not included in metadata

8. Terminology

Bitonal – refers to an image made up of pixels that are white or black. Bitonal scanning is often used for high-contrast documents, such as printed text.

Bit depth – a measure of the number of colours (or brightness in greyscale images) available to represent the colours(or shades of grey) in the original document.

CMYK – CMYK is short for Cyan-Magenta-Yellow-Black. It refers to an image made up of pixels that are made up of these four process colours. CMYK is often used in offset printing for full-colour documents.

Dots per inch (DPI) – often used interchangeably with ppi, but refers to measurement of the resolution for computer printers.

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.

JPEG – Joint Photographic Experts Group is a commonly used file format used for digital images. It is also commonly used for online access for image files because they require less storage space. JPEG files use lossy compression, and are therefore not recommended for files that require multiple edits.

Lossless Compression – has no loss of information. The decompressed object appears the same as the original.

Lossy Compression – information is removed from the stored information during the compression process, reducing the size of the image. Although the information may not be noticeable, a loss of quality occurs. The quality of lossy images also decrease each time the image is manipulated or edited – this loss of quality over time is called generational loss.

Optical density – measures the "brightest bright" and the "darkest dark" that a piece of equipment can capture. The total tonal measurement is on a scale of 0.0 (white) to 4.0 (black). Most consumer-grade scanners have an optical density of approximately 2.5. High end scanners have an optical density of upwards of 3.8 to 4.0, which represents about 95% to 100% of the original source's tonal rage.

PDF/A – PDF/A is a long-term preservation formation for digital 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, and much more." PDF/A – 1b is often applied to scanned images and preexisting PDF files that are converted to PDF/A, ensuring the "reliable reproduction of the visual appearance of a document."[9] PDF/A is an ISO standard, ISO 19005-1:2005.

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.

---

[9] PDF/A archiving standard, Adobe Systems.

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

RGB – A mode of capture which is 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.

TIFF – Tagged Image Format (TIFF) is a widely accepted, open standard image format, and is considered to be the preferred professional image standard. TIFF files are lossless, meaning they are in an uncompressed format. TIFF is an ISO standard, ISO 12639:2004.

9. Appendix A – Compliance Declaration – Disposal of Source Records

This form must be signed by senior management and retained as a permanent record.

Digitization Project Details

The following declaration relates to: (Insert name of public body)

_____

It covers digitization activities:

☐   Across all business areas/locations of the public body;

☐   In the following business area/locations of the public body (please specify).

_____

Original source records which have been deemed to be eligible for disposal after digitization are covered under: (Please insert relevant Records Retention and Disposition Schedule Number)

_____

Details of the source records to be destroyed are:

☐   Listed in an attachment to this Compliance Declaration;

☐   Listed in the internal digitization policy.

Recordkeeping Context:

☐   I confirm that the public body has met the following prerequisites:

1. A records management system or business system with recordkeeping functionality to ensure that electronic records, including digitized records, can be managed appropriately;
2. Internally approved recordkeeping policies and procedures to demonstrate that recordkeeping is managed in an orderly and systematic way across the full range of business activities undertaken by the public authority;
3. Appropriately trained and skilled staff to develop and implement policies and procedures and manage the system and tools in use.

Minimum Recordkeeping Requirements

☐   I confirm that the public body has met the following minimum recordkeeping requirements as outlined in the Provincial Archives of New Brunswick Digitization (Imaging) Standards:

1. That all original source records intended for destruction are covered under a current Records Retention and Disposition Schedule approved by the Provincial Archivist;
2. Any disposal direction issued through a disposal freeze will be followed;
3. A risk assessment to determine which source records are eligible for early disposal has been conducted based on documented criteria and legal advice sought on the application of the criteria if necessary;
4. An internal policy statement about the handling and disposal of original source records after digitization which is commensurate with the risk assessment has been developed and issued;
5. Processes are in place which monitor changes which may have an impact on the public bodies need to retain the source records and the steps that will be taken to maintain compliance with this standard;
6. Digitization plans, policies, and procedures have been developed and put into place covering the identification of records selected for digitization, how the digitization process will be managed, including staff responsibilities, and the capture and management of digitized records and its metadata;

7. A technical standard for reproduction, including approved image formats, compression algorithms, resolution and bit depth (colour depth), and the use of any image processing procedures has been set and put into operation;
8. Quality assurance procedures have been put into operation that include:
   - The timing of equipment tests and equipment calibration;
   - Procedures for checking output, such as what proportion of the digitized records will be subject to visual inspection and how long the source records need to be retained after digitization to ensure that quality assurance has been determined;
   - Procedures for re-imaging if quality standards are not met;
   - Roles and responsibilities for checking and approving output.
9. A well-managed system is in place with appropriate recordkeeping functionality to ensure that the digitized images can be managed as records for as long as they are required to be retained in accordance with a current Records Retention and Disposition Schedule approved by the Provincial Archivist. This includes:
   - Adequate physical and other security safeguards to ensure the digitized records remain inviolate and can only be changed in an authorized manner;
   - Audit logs of access to and use of digitized records are generated and retained as recordkeeping metadata;
   - Ability to capture and maintain mandatory recordkeeping metadata elements;
   - Capability to generate and capture appropriate technical imaging metadata at the point of digitization.
10. The system is covered by internally approved business continuity and disaster recovery plans and a migration strategy is in place to ensure that the digitized records are not placed at risk of loss through technological obsolescence.

I understand that the public body may be required to supply the Provincial Archives of New Brunswick, auditors, and others with documentation to support these claims.[10]

Signed _____   Date _____

Name (Please Print) _____

Position Title _____   Public Body _____

Contact Person _____

Position Title _____   Contact Information _____

---

[10] This form was adapted from Queensland State Archives Compliance Declaration.

This Page Left Blank

10. Appendix B – Records Destruction Form

DEPARTMENT / MINISTÈRE: _____

BRANCH/DIRECTION	RECORDS DESCRIPTION/DESCRIPTION DES DOCUMENTS	DATES OF RECORDS/ DATES DES DOCUMENTS	SCHEDULE NUMBER/ NUMÉRO DE CALENDRIER	DATE DESTROYED/ DATE DE DESTRUCTION	APPROVED BY/ APPROUVÉ PAR

11. RECOMMENDED RESOURCES

Anne Kenney and Stephen Chapman, *Digital Resolution Requirements for Replacing Text-Based Material: Methods for Benchmarking Image Quality*, Commission on Preservation and Access (April 1995), http://www.clir.org/pubs/reports/pub53.

Association for Information and Image Management, ANSI/AIIM MS44-1993, *Recommended Practice for Quality Control of Image Scanners.*

Association for Information and Image Management, ANSI/AIIM MS53-1991, *Recommended Practice for the Requirements and Characteristics of Documents Intended for Optical Scanning.*

Archives New Zealand, Government Recordkeeping Group, *Digitisation Standard*, 2010.

Cornell University, *Moving Theory into Practice Digital Imaging Tutorial*, Cornell University Library Research Department, (2003).

Government of Alberta, Government of Alberta Standards Oversight Committee*, Digitization Process*, Standard number A000015, 2010.

Government of Alberta, Government of Alberta Standards Oversight Committee, *Digitization Technical Requirements*, Standard number A000013, 2010.

Government of the Northwest Territories, Public Works Services, *Guidelines for Scanning Projects*, 2010.

International Council on Archives, *Guidelines for Digitization Projects for Collections and Holdings in the Public Domain, Particularly those hold by Libraries and Archives* (March 2002).

Library and Archives Canada, *Guidelines for the Digitization of Textual Information Resources DRAFT,* Version 0.65, www.collectionscanada.gc.ca.

National Archives of Australia, *Digitising accumulated physical records: A guide to initiating and planning digitisation projects* (2011).

National Archives and Records Administration, *Technical Guidelines for Digitizing Archival Materials for Electronic Access: Creation of Production Master Files - Raster Images,* 2004, http://www.archives.gov/preservation/technical/guidelines.pdf.

National Standards of Canada, *Microfilm and Electronic Images as Documentary Evidence* CAN/CGSB72.11-93.

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

Service New Brunswick, *Imaging Manual: SNB PLANET* (2010)

Queensland State Archives, *Digitsation Disposal Policy Toolkit* (May 2010), http://www.archives.qld.gov.au/Recordkeeping/Pages/Publications.aspx