Digitizing Government Information

Guide
# Table of Contents

## Introduction
- What is digitization? .......................................................... 4
- Intended audience .......................................................... 4
- About this version ......................................................... 5

## Step 1: Plan to Digitize
- Getting started .................................................................. 6
- Is this a new digitization process? ......................................... 6
- Digitization Quick Start Guide .............................................. 7
- Develop a rationale for your digitization process ................. 8
- Benefits ............................................................................ 9
- Challenges ........................................................................ 10
- Resources ........................................................................ 11
- Responsibilities ................................................................ 13

## Step 2: Assess Records
- Which records should be digitized? ..................................... 14
- Records management requirements ..................................... 16
- Frequently Asked Questions ............................................... 17

## Step 3: Develop a Defensible Process
- Select a digitization approach ............................................. 19
- Planning the workflow ...................................................... 20
- Bulk process or ongoing process? ......................................... 21
- Official file copy or not? ..................................................... 22
- How to use the Technical Requirements in the Standard ....... 23
- Documentation ................................................................ 24

## Step 4: Prepare & Digitize
- Physical preparation ........................................................ 25
- Digitizing options ........................................................... 26
- Digitizing in-house ......................................................... 27
- Outsourcing ..................................................................... 28
- Digitize source records .................................................. 29
- Digitize audio and video .................................................. 30
- Post-digitization processing ............................................. 33

## Step 5: Compile Metadata
- What is metadata? .......................................................... 34
- Metadata requirements for digitization ............................... 35
- How to apply metadata .................................................... 38
- Naming conventions ....................................................... 40
- Best practices for naming government information ............. 41
- Managing versions ........................................................ 42
- Organize digitized content ............................................... 43

## Step 6: Assess Quality
- Quality assurance .......................................................... 44
- What is an error? ............................................................ 45
- Implementing QA .......................................................... 48
- How to run QA ............................................................. 49
  - Sample scans .............................................................. 49
  - How much needs to be quality assured? ......................... 50
  - When to do QA .......................................................... 50
  - Who should do QA ...................................................... 50
  - Dealing with errors ..................................................... 51
  - Re-digitizing process .................................................. 51
  - Review the QA process ............................................... 51

## Step 7: Store & Manage Records
- Managing original records after digitization ..................... 52
- Transferring digitized records ........................................... 53
- How will the digitized records be managed, stored and accessed? ........... 55

## Tools & Resources
- Planning Checklists ........................................................ 57
- Risk Assessment Matrix ................................................... 58
- Roles and Responsibilities ............................................... 60
- Resources ....................................................................... 61
- Infographics .................................................................... 63
- Glossary ......................................................................... 64

## Digitizing Government Information Standard
- .......................................................... 74
Welcome to the world of digitization!

This guide will help you navigate your digitization process, whether your goal is to replace original records with authentic, defensible digital copies, to provide online access to useful documents, or to digitize physical documents destined for the digital archives.

A well-planned, properly-resourced digitization process will help ensure that your digitized records are trustworthy and that you will meet the requirements for disposing of the original records (usually referred to as source records in the context of digitization projects).

More and more, in recent years, government bodies have been choosing to digitize. There are many reasons to digitize. It is often done because managing digital information is more efficient and can bring additional benefits. Sometimes it is done to streamline business processes, and sometimes to save costly office space. Further, the Information Management Act (IMA) requires that archival information in physical form be digitized before transferring it to the government archives, subject to exemptions. The Managing Government Information Policy (MGIP) section 3.3 also requires digitization in accordance with the Digitizing Government Information Standard (DGIS).

The digitization process outlined in this guide will allow you to:

- protect the authenticity and integrity of the records
- ensure that the digitized records can be managed, retrieved and preserved over the long term
- enable the timely, defensible destruction of originals (in most cases)
Introduction

What is digitization?

Digitization is the process of copying physical records (usually paper) into digital form. The process of digitization includes planning, assessing, preparing, digitizing, compiling metadata, running quality assurance checks, and storing and managing the digitized records.

This guide covers the following types of digitization:

- **digitization as part of a business process**: routinely digitizing source records as they are created or received during a business process
- **digitization projects**: digitizing a specific set of existing records to preserve the information, enable digital access, and if appropriate, protect source records from damage (especially if destined for the government archives)
- **ad-hoc digitization of individual records**, such as a signed travel authorization form or another document that needs to be transmitted or filed digitally

Digitizing records is just one step on the road towards the digital transformation of the BC Government. While your organization may choose to conduct digitization projects in order to meet their business needs, the focus is to ensure a transition away from generating any future physical records. The majority of records are now born digital and should be kept in digital format, where possible. Some guidance for your organization on handling **born digital** records throughout their lifecycle is available on the Records Management website, and more comprehensive guidance for these records is under development.

Intended audience

- Employees of ministries and other government bodies subject to the *Information Management Act* (IMA)
- Service providers who provide digitization services
Introduction

About this version

If you have questions about how to use this Guide, please contact your Records Team.

Version control

<table>
<thead>
<tr>
<th>Version</th>
<th>Change(s)</th>
<th>Author(s)</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>First draft shared with project partners.</td>
<td>MCH</td>
<td>2018/01/25</td>
</tr>
<tr>
<td>0.2</td>
<td>New interactive format. Introduction updated. Added the Quick Start Guide. Steps 1-8 edited.</td>
<td>MCH</td>
<td>2018/05/09</td>
</tr>
<tr>
<td>0.3</td>
<td>Minor revisions to Quick Start Guide. Digitizing in-house vs. outsourcing moved to Step 5 (Digitize). File naming guidance added to Step 6 (Metadata). Glossary added to Tools &amp; Resources section.</td>
<td>MCH</td>
<td>2018/07/10</td>
</tr>
<tr>
<td>0.3</td>
<td>Version control table added. Minor wording changes.</td>
<td>MCH</td>
<td>2018/07/24</td>
</tr>
<tr>
<td>0.4</td>
<td>General edits and new content added to all sections. Table of Contents and page numbers added. New infographics section.</td>
<td>MCH</td>
<td>2019/02/08</td>
</tr>
<tr>
<td>0.5</td>
<td>General edits and update of DGI Standard citations.</td>
<td>BNS</td>
<td>2019/11/13</td>
</tr>
<tr>
<td>0.6</td>
<td>Incorporated stakeholder feedback, including GRS staff and Director of Archives &amp; Records Initiatives.</td>
<td>BNS</td>
<td>2020/07/09</td>
</tr>
<tr>
<td>1.0</td>
<td>Guide approved and aligned with approved Standard v.1.0</td>
<td>BNS</td>
<td>2020/07/13</td>
</tr>
</tbody>
</table>
Step One
Plan to Digitize

Getting started
You will need to have a plan and process in place before starting to digitize. Carefully thinking through the many components of a digitization process or project will help ensure that:

- your digitization processes are **defensible**
- you can protect the **authenticity** and **integrity** of the digitized records
- you meet the requirements for **destroying** source records, if appropriate
- you capture and maintain **metadata** about the digitized records
- you can store and manage the digitized records over the long term
- you can continue to meet business and information management requirements

*Is this a new digitization process?*

Use the **Digitization Quick Start Guide** when starting your planning process.

*Are you already digitizing?*

For existing digitization activities, review the **Digitizing Government Information Standard** and this guide to identify any potential gaps in your current digitization process. Gather the information you need by reviewing documentation and engaging with stakeholders. If it looks like you need to adjust existing practices, contact your **Records Team** for advice.
Digitization Quick Start Guide

01 Plan to Digitize
- What are the benefits, challenges and reasons for digitizing records?
- What resources will you need?
- Who will authorize the digitization activity?
- What is the level of risk for your organization?

02 Assess Records
- Are your records covered by an approved information schedule?
- What is the retention period and final disposition?
- What records need to be digitized, how many, and when?
- Can the records survive a digitization process?

03 Develop a Defensible Process
- What is your digitization approach?
- Who will scan the records?
- Where will the work take place?
- What are the technical requirements?

04 Prepare & Digitize
- How will the records be prepared for digitizing?
- Is special handling required?
- Will the work be outsourced?
- What are your digitizing and post-digitization procedures?

05 Compile Information About the Records
- What metadata needs to be captured?
- What are the document naming conventions?

06 Assess Quality
- What level of quality assurance is needed?
- What errors will you look for? How much error is allowed?
- Who will carry out the quality checks?
- How will you deal with errors?

07 Store and Manage Records
- How will the original records be managed after digitization?
- How will the digitized records be securely transferred?
- Where will the digitized records be stored and managed?
- Is there a secure office recordkeeping system?

Next: Develop a rationale for your digitization process
Step One
Plan to Digitize

Develop a rationale for your digitization process

Digitization should only proceed if it improves your organization’s ability to carry out its business functions, and meet its information management and compliance requirements. Document your rationale using the approach preferred in your organization (e.g. a business case, business process map, project initiation document [PID]) and obtain all required approvals. Whatever your approach, make sure you:

- clearly outline and quantify the benefits and anticipated business or cost efficiencies
- identify the necessary resource commitments
- include a realistic budget
- address the minimum requirements established in the Digitizing Government Information Standard
- comply with the Freedom of Information and Protection of Privacy Act (s. 69(5.1)) by completing a privacy impact assessment (PIA) whenever engaging in a new project, program or activity or working with a service provider

Consider factors such as how digitization fits into your organization’s service plan, technology plan and workflows. Also determine whether you will need to retain source records after digitizing, due to legal, archival, or other requirements. Digitization may be justifiable for accessibility and preservation of fragile source records. Cost benefits can be realized as quality assurance costs may be reduced, and originals may be transferred offsite.

To learn more about making a business case, talk to your organization’s business analyst or project planner (there is also good general advice on the Procurement Services Business Case page). For more information on the requirement to document your project or process, see Step 3 of this Guide.
Step One
Plan to Digitize

Benefits

In your digitization plan, identify the anticipated benefits of digitizing the source records. Digitization can help you:

- **Streamline Business Processes**
  - by enabling you to digitally manage all incoming and outgoing transactions

- **Save Space**
  - by reducing the need to store hardcopy records

- **Provide Ready Access**
  - to digitized records by making it easier to search and retrieve information during litigation and in response to other access requests

- **Enable Greater Sharing of Information**
  - (when appropriate)

- **Save Time**
  - retrieving digitized records when they are well-indexed

- **Store Records**
  - in your office's recordkeeping system (EDRMS or another system)

- **Preserve**
  - with intrinsic value by providing access copies to reduce handling

- **Create a back-up copy**
  - of source records that are vital for disaster recovery and business continuity purposes

- **Locate paper, digital and other formats together**

Next: Challenges
Step One
Plan to Digitize

Challenges
Consider how the challenges of digitization relate to your situation, before deciding to go ahead. You may even decide digitization is not the right path at this time. Think about these challenges:

- Preparing source records, compiling metadata and quality assurance are complex, time-consuming processes that require expertise.
- Potential technology, processing and digital storage costs will become obsolete over time and need replacing.
- Specialized digitization equipment may be required for bound volumes, large-format items, microfilm, slides and other formats.
- Legal, organizational, or archival requirements (see Step 2: Assess Records) may prevent you from destroying the source records after they have been digitized, especially if the quality of the digitized records does not meet expectations.
- You may need to store multiple copies of a record for distribution purposes (e.g., creating an access copy to post on a website).

An additional challenge is the fact that you and your program area may not have dealt with these issues before. Don’t be intimidated! With the help of this guide you can address them all.
Step One

Plan to Digitize

Resources

The digitization process may require a significant investment of resources, especially when preparing the source records, digitizing them, adding metadata and running quality checks. If you are new to digitization, you may find it helpful to start with a small pilot digitization project – either in-house or with a service provider – to identify what appropriate resources will be required to complete a larger-scale project.

Consider these factors in your digitization plan:

People

In addition to salary expenses, consider the cost of recruiting, training and retaining staff.

Technology

Determine what type of equipment is needed for digitization and if you can use existing equipment, in the context of the following factors:

- the volume and variety of source records
- equipment needs to be scalable to accommodate any new or expanded projects
- special arrangements may be appropriate to ensure full and accurate capture of material that cannot be digitized using standard office equipment – i.e. specialized imaging, audio or video equipment or may require outsourcing.
Step One  
Plan to Digitize

Resources (continued)

Space

Does your office or service provider have the right environment to securely store and handle the source records during the digitization process? You’ll need to figure out:

- how to safely transport and store the source records
- if there are formats (e.g. photographs and microfilm) that need to be stored in a cool, dry location
- physical security controls to ensure the source records are protected against unauthorized access

Time

How much time is required to carry out the digitization? Test your process on a representative sample of the source documents to estimate how much time is required for preparing and digitizing source records, compiling metadata, and conducting quality assurance checks – tasks that can be very time-consuming and costly. To reduce staff time and error rates, ensure that most or all of the metadata is automatically generated. If outsourcing, ask your service provider to provide estimates.

Digital storage

Consider the costs of network infrastructure, server storage space, security, and the eventual cost of migrating digitized records to new systems.

Next: Responsibilities
Step One
Plan to Digitize

Responsibilities

If you are carrying out a digitization project, it’s helpful to have a plan that clarifies roles, responsibilities, reporting lines, and communications for staff across all processes. It should be possible to identify the digitization project owner (i.e., the work unit responsible). Depending on your organizational structure, you will need to ensure the project or process is authorized appropriately. If you intend to destroy the source records once they have been digitized, you will also need authorization as required by the Redundant Source Records Schedule.

Digitization projects require a combination of skills from a variety of staff with different areas of expertise (see Roles and Responsibilities). Each digitization project should have a key contact who can act as a liaison between different groups of stakeholders.
Which records should be digitized?

Consider what records need to be digitized based on your business requirements and the source record format. Depending on the business need, you may decide only to digitize government information that is frequently accessed, or on a “going forward” basis. Another option is to digitize a long-term accumulation of existing paper records, in addition to newly created/received records.

Some records are more challenging to scan and may have content or formats not easily captured in digital form. It may not be possible to translate all the quality and characteristics of the source information into digital form, or the records may have intrinsic value. You may decide that these records are not physically suitable for digitization and should be retained in physical form.

Section 13 of the Information Management Act requires that records scheduled for transfer to the government archives (in other words, records covered by classifications marked “Selective Retention” or “Full Retention” in the relevant information schedule) must be transferred in digital format.

However, the Chief Records Officer has released a directive to government organizations exempting them from the section 13 requirement for eligible information to be digitized before it is archived. This means that until the CRO directs otherwise, your organization can continue to archive the information in a physical form.

Questions to Consider

Is digitizing justified? Have you considered questions like these?

Have you:
- optimized processes to keep documents in a digital format without printing?
- verified a schedule directs you to archive government information under a Selective or Full Retention rule?
- verified a schedule directs you to retain records for very long periods of time?
- confirmed your physical records will not be damaged by digitization?
- received direction from your supervisor or executive to digitize physical records for business reasons?

Contact GRS for more advice.
Which records should be digitized? (continued)

When evaluating source records, carefully assess whether they require any special physical preparation procedures (see Step 4) or specialized equipment. Here are some examples of records that may require particular attention:

- **Large format records**
  - including maps and plans (larger-sized documents may require specialized equipment)

- **Smaller-sized documents**
  - (may require special handling)

- **Files that contain a variety of documents**
  - such as documents with multiple page sizes, varying colour, paper quality and inks, graphic materials, photos, drawings, handwritten documents and annotations

- **Fragile records**
  - that may be put at risk in the digitization process (if they are important records, this could lead to a decision not to digitize)

- **Records that are creased, stapled, bound, rolled, brittle, etc.**
  - requiring extra preparation before digitizing

- **Records with unique finishes**
  - (transparent, semi-transparent or opaque), reflective surfaces, and embossing (usually in the form of seals), requiring extra preparation before digitizing or image optimization after digitizing

- **Audio-visual records**
  - on tape, disc, or hard drive and microfilm/microfiche, requiring different playback and recording equipment and handling procedures

- **Records with intrinsic value**
  - such as colonial-era papers (you may need to retain the source record as well as the digitized copy)
Step Two
Assess Records

Records management requirements

Ensure that records management requirements are being met for both source records and digitized copies throughout the planning and implementation phases of a digitization process so that:

- **source records** that need to be kept are not accidentally damaged or destroyed
- digitized records maintain their **authenticity** and **integrity**
- digitized records can be stored and retrieved over time

Consider these questions before starting your digitization process:

- What are the relevant legal, security, privacy, FOI, **information schedule**, and other requirements? For example, some legislation and information schedules contain requirements to keep records in their original formats.
- Must source records be retained due to their historical significance, **intrinsic value**, or legal requirements?
- Which version will be the official record - original or digital copy?
- What is the appropriate **disposal** process for source records once digital copies have been created and verified?
- How will the source records and digitized records be managed until they are destroyed or transferred to the government archives?

Next: Frequently Asked Questions
**Step Two**

**Assess Records**

---

**Frequently Asked Questions**

**Can I destroy source records after they have been digitized?**

The Redundant Source Records Schedule (RSRS) provides for the destruction of source records, where the source records have an approved information schedule. The RSRS requires GRS Records Officer approval for digitization projects and processes, but not for ad-hoc digitization activities. For more information, see the How to Use the Redundant Source Records Schedule guide.

Source records not covered by an approved information schedule cannot be destroyed without specific approval from the Chief Records Officer. Contact your Records Team if you need a new information schedule.

**What if records are scheduled for full or selective retention?**

When the information schedule for the records to be digitized lists them as “SR” or “FR,” that means that some or all of the records are of archival significance. Their information needs to be protected in perpetuity in the government archives, so extra measures are needed for the destruction of the originals. Contact your Records Team and ask for assistance and authorization under the RSRS to destroy the source records. In exceptional circumstances, you may need to keep source records that cannot be digitized without losing information (e.g. records with embossed seals). See Exemptions below. In most cases, if you use a defensible process, the digitized copies should be sufficient for archival purposes.

---

Please note: The Redundant Source Records (RSR) process is under review and will be updated shortly. If you need assistance evaluating your source records, Government Records Service is available to provide advice.
Step Two
Assess Records

Frequently Asked Questions (continued)

Will there be exemptions from the requirement to scan archival records?

As noted earlier, the Chief Records Officer has released a directive to government organizations to hold digital information scheduled for archiving until notified. Physical records ready to be archived are currently exempt from this requirement and can be transferred to the BC Archives at the Royal BC Museum. The exemption categories and the process for granting exemptions have yet to be established; this guide will be updated when that happens. Contact GRS for more information on exemptions.
### Step Three
Develop a Defensible Process

#### Select a digitization approach

A defensible digitization process shows that you have developed and documented a considered approach, and can demonstrate compliance with the [Digitizing Government Information Standard](#). Having a defensible process will help you avoid costly errors in the planning and implementation phases. This ensures records remain searchable and accessible. A defensible process is important to protect the government from unnecessary risk.

Select and document a digitization approach that aligns with your digitization rationale. Digitization approaches will vary across work units and over time, depending on the nature of the digitization activity or business process. Make sure you address the list of questions provided in the [Digitization Quick Start Guide](#).

Documentation is key to defensibility. This should include the digitization approach and procedures, as well as approval of the process by an appropriate authority for your organization. Embed this information into the digital files themselves. Or create a digitization case file for documentation. Link this case file to the digitized records and retain it for their lifecycle to prove their authenticity. If the records to be digitized are frequently the subject of litigation, legal advice may be desirable to ensure adequate defensibility of your digitization process.

To learn more see [How to Apply Metadata](#) and the [Risk Assessment Matrix](#) in the sections below.

---

**Standard #5.1.1**

All digitization processes must be defensible. This means it must be possible to demonstrate that:

- the digitized government information retains its integrity and can be depended on as a true and accurate version of the source record (i.e., the original physical record);
- digitization processes are documented and usable as evidence that all relevant requirements can be or have been met;
- documentation about the digitization process is retained for the life of the records digitized; and
- digitization does not expose government to any undue risk.
Step Three
Develop a Defensible Process

Planning the workflow

Implementing digitization will result in new processes that require integration with previously existing processes. You will need to decide on:

- the digitization location (you may decide to create a single location or multiple locations)
- when digitization will occur
- who selects and prioritizes records for digitization
- who is authorized to digitize the records or to send and receive source records that must be transported to be digitized
- how the source records will be retrieved and transported
- security and privacy requirements
- procedures for preparing source records, digitizing, capturing metadata, and assessing image and metadata quality
- where will the digital files be stored
- authorized destruction process of source records

Next: Bulk process or ongoing process?
Step Three
Develop a Defensible Process

Bulk process or ongoing process?

The type of process – bulk or ongoing – will impact your digitization workflow and available resources.

Bulk processing involves processing accumulations of source documents, either as a legacy project for older source records (legacy digitization), or as a work process for newly created or received records (e.g. setting aside source documents to digitize once per week or month).

Alternatively, you can digitize individual records as they are created or received, by making it part of ongoing business processes within the office or program area. With this approach, you may decide to make a practice of digitizing an entire file or other aggregate, rather than only the specific documents needed right away.
Step Three
Develop a Defensible Process

Official file copy or not?

Whether or not you are creating an official file copy determines whether you must follow the digitization standard.

Follow the Digitizing Government Information Standard if you are replacing originals with official file copies in digital form. An official file copy:

- is the version used for making business decisions
- is evidence of a business action or acts as supporting reference material
- will be sent to the government archives if appraised as archival
- can serve as a vital record for business continuity purposes
- should be maintained in a safe, secure environment
- for digitized records, the master copy (i.e. the primary, most complete digitized copy) becomes the official file copy; any additional copies are referred to as derivative copies or access copies.

There is no obligation to follow the Standard if you are not creating an official file copy, i.e. if the original physical record will continue to be the main record retained in your recordkeeping system. Copies of physical, born-digital, or digitized records that are made for personal convenience, easy distribution, or posting online do not need to comply with the Standard, as long as an official file copy exists and is managed elsewhere (e.g. a compressed copy may be posted online but the uncompressed master copy is filed in EDRMS Content Manager).

Learn more about derivative copies in Step 5: Compile Metadata: Managing versions.

Next: How to use the Technical Requirements in the Standard
Step Three
Develop a Defensible Process

How to use the Technical Requirements in the Standard

Decide in advance what file formats to use, whether to generate compressed access copies as well as master copies, their resolution, and the extent to which colour will be represented in the copies. Follow the Technical Requirements in the Standard (section 5) in combination with your own specifications as needed.

Establish your technical specifications based on the type and purpose of the record, its scheduled disposition, and your ability to store, preserve and manage the digitized record (don’t choose a file format that cannot be stored and managed within existing systems).

Elements you can adapt:

- digital file formats: **PDF/A** is preferred for textual records, **TIFF** for photographs
- **resolution**: 300ppi minimum; photographs may require higher resolution
- colour management: adapt **bit-depth** according to the type of record
- **compression** for **derivative copies**: lossless compression is required for master copies, but you may choose to make access copies for posting online that are “lossy” (e.g. **JPEGs**)

**Recommended practices**

- create **master copies** with the highest possible technical specifications and retain them in secure storage
- choose digital file formats that preserve the essential characteristics of the source records, are uncompressed or have lossless compression, and are supported by a variety of applications and outputs
- create access copies in formats most convenient to your business requirements (such as PDFs and JPEGs for online distribution)

Next: Documentation
Step Three
Develop a Defensible Process

Documentation

To ensure all staff understand the digitization process and that the digitized records are **authentic** and **defensible**, it is important to maintain documents explaining decisions, processes and procedures relating to digitization activities. For processes where part or all of the digitization is outsourced, some documents may be provided by the service provider (e.g. imaging procedures). Other documentation will need to be developed by your organization (e.g. planning documents). Ensure that all documentation is complete and accessible (preferably in one central location).

Digitization documentation may include the following:

- planning documents (e.g. a business case, project initiation document, project charter, project plan, authorization(s) for the digitization process)
- digitization contracts (if project is outsourced). Amendments, vendor evaluations, vendor procedure manuals, etc. must also be kept
- risk assessment documentation, including context analyses and risk registers
- **Privacy Impact Assessment** (PIA)
- metadata requirements
- chain of custody logs (e.g. records transmittals, delivery receipts, etc.) for both the source records and the digitized copies
- operating procedures (e.g. for preparing and digitizing source documents)
- project and/or vendor reporting (monthly volume and costs, invoices, resource hours, error and/or problem resolution reporting, annual reporting, etc.)
- loan agreements with service providers, or other forms of contract or service agreements, outlining terms and conditions for care and handling of records while in their physical custody
- data management practices relating to the stored digital images
- QA procedures, activities, and reports
- documents relevant to the project wrap-up
Physical preparation

Physical preparation procedures ensure that source records are ready for digitization. Typically, source record preparation only requires basic activities to allow for efficient digitization processes, such as removing staples or paperclips and aligning single pages. But, in some cases, more extensive work may be required.

You (or your service provider) may need to:

► include a standard box content file list (see Specifications for File Lists [RIM 422B]) for the digitization service provider
► remove any material that does not require digitizing such as exact duplicates, paper markers with digitization instructions, binders and folders (although their labels may be scanned)
► digitize fragile items individually rather than digitizing them in bulk (“bulk scanning”)
► disassemble bound volumes
► temporarily enclose fragile originals in plastic sleeves
► repair any paper that has tears or holes
► straighten folds or creases
► use a flatbed or overhead scanner for large-format, fragile, or other items (e.g. bound volumes, if they are not to be disassembled)

Note: When digitizing a file, multiple files, or large amounts of records, maintain their original order. If records have become disordered before digitization, correct this before proceeding (i.e. a misfiled document or folder should be put in its correct place before digitizing). In some circumstances (such as large format items requiring specialized equipment) it may be necessary to scan related records separately, and to ensure their original order is documented in metadata. Where source records are to be retained, return them to their original locations.

Standard #5.1.2

Government information to be digitized must be assessed to identify whether any special physical preparation procedures are required to protect its integrity, or if it is physically suitable for digitization.
Digitizing options

Digitization can be carried out in-house, by a service provider, or using a combination of the two. There are pros and cons to each of these options. Consider:

- size and format of the records being digitized
- privacy and security requirements of the information (details provided below)
- if the records need to be accessed during the digitization process
- geographic location of the records
- availability of service providers (use the Corporate Supply Arrangements (CSA) for BC Mail Plus Scanning and Data Capture services) available in the Goods and Services Catalogue, see Document Scanning and Data Capture
- other benefits and challenges of digitizing in-house vs. outsourcing

Outsourcing will require defining both the organization and service provider responsibilities throughout the digitization process as part of a service agreement.

Changes in processes, related or similar projects, and changes in technology are all factors that may lead to a reassessment of the original decision to digitize. Service agreements should include a process for monitoring compliance to ensure that all requirements are met.

It is good practice to review the digitization processes periodically, whether in-house or outsourced, to maintain efficiency and compliance with standards and best practices, which may change over time.

NOTE: Don’t forget the Privacy Impact Assessment (PIA) requirement. Depending on a number of factors, such as the location of the service provider, there may be privacy implications. A PIA can assess whether there is authority for the use of a service provider.
Step Four
Prepare & Digitize

Digitizing in-house

Benefits

- more control over the entire process, including handling, security, and storage of both the source records and the digitized copies
- grow in-house digitization knowledge and expertise
- hardware and software is available for future projects
- lower security risk - source records do not leave your organization
- flexibility to respond to changes in project requirements and parameters as the project develops

Challenges

- requires initial and ongoing financial investment in space, people and equipment
- may require longer, more flexible timelines to allow for training staff, establishing and implementing a process, and responding to conflicting priorities
- requires dedicating staff time to the project
- your organization must accept operational costs for network downtime, equipment failure, and upgrades to software and hardware over time

Next: Outsourcing
Step Four

Prepare & Digitize

**Outsourcing**

**Benefits**

- costs are more predictable: your organization pays for deliverables, (e.g. usually a set cost per unit), which helps with project planning and budgeting
- economies of scale: service providers can handle large volume and high output
- broad range of available options and services (e.g. special equipment for digitizing bound volumes and other non-standard formats)
- high production levels
- specialist expertise and experienced service providers absorb costs of technology obsolescence, system failure, downtime, staff changes, etc.

**Challenges**

- less control over how the digitization is carried out
- quality assurance cannot be fully observed at the time of digitization (the records owner must inspect files and metadata upon receipt, as it is essential to ensure that the work meets requirements)
- complex contractual process: digitization specifications must be clearly defined up front, solutions to problems must be negotiated, communication must be open and problems must be dealt with immediately as they arise
- knowledge gaps may cause delays and confusion, as service providers are unfamiliar with your specific business
- risk of the service provider going out of business or altering their practices
- risk of possible loss or damage to the records due to transportation and handling by the service provider (requirements for this need to be defined in the contract to help manage this risk)
- unforeseen circumstances or additional service requirements may cause budget overruns

Next: Digitize source records
Digitize source records

To digitize your records, you can use: the single and multi-function devices (MFDs) provided by Managed Print Services (MPS); as well as desktop scanners, digital cameras (not a smart phone camera), and other types of equipment depending on the physical item and its condition. Reproduce source records as faithfully as possible.

Recommended procedures:

- capture both sides of all documents (unless one side is blank throughout). This can be set up on most equipment prior to digitizing.
- validate the number of source records against the number of digital copies created to ensure that all source records have been digitized.
- for multi-page items, ensure the number of images in the digital file accurately reflects the number of pages in the input batch, and are structured and arranged in the correct order.
- create two scans for pages with appended sticky-note: one scan with the sticky note on the document, and another with the sticky note removed.
- ensure the entire document has been captured.
- ensure all required metadata is accurate and complete.
- confirm document is digitized directly to the correct digital file type – e.g. directly to TIFF, not JPEG copied to TIFF.
- verify TIFF compression settings are lossless (e.g. LZW or G4).
Step Four
Prepare & Digitize

Digitize audio and video

If you are planning to digitize analogue audio recordings, video recordings or motion picture files, you will need to choose file formats that can be sustained over time.

The best practice is to create lossless preservation files, compressed or uncompressed. However, limited financial or technical resources may make this choice unsustainable. It is important to ensure a balance between quality and sustainability.

Preferred and acceptable file formats are listed on the next page. For additional information on Master and Derivative copies see Step 5 “Managing versions” (pg. 42).

Please contact GRS with any questions about audio and video digitization.

Technical Requirements for Digitizing Analogue Audio and Video

Use a digital resolution that equals or exceeds the original, bit depth as appropriate, lossless compression, and file formats such as BWF for audio and MXF or MOV for video.

For current details on the variety and complexity of changing formats, see the rest of this guide and resources such as the National Heritage Digitization Strategy - Digital Preservation File Format Recommendations.
## Digitize audio and video (continued)

<table>
<thead>
<tr>
<th>Content Types</th>
<th>Preference</th>
<th>Format</th>
<th>Master</th>
<th>Derivative</th>
<th>Notes</th>
</tr>
</thead>
</table>
| **Video**     | Preferred  | Codec: JPEG 2000  
Audio: PCM | X      |          | Lossless compression 10-bit, variable bit rate Minimum average 50 mbps |
|               |            | Codec: H.264 or H.265  
Wrapper: MPEG-4 pt 14 (MP4)  
Audio: PCM |        | X        | Lossy compression 8-10 bit, variable bit rate |
|               | Accepted   | Codec: FFV1  
Wrapper: Matroska (MKV)  
Audio: PCM | X      |          | Lossless compression 8-10 bit, variable bit rate |
|               |            | Codec: JPEG 2000  
Wrapper: Audio Video Interleaved Format (AVI)  
Audio: PCM |        | X        | Lossy compression 10-bit variable bit rate |
|               |            | Codec: Uncompressed  
Wrapper: QuickTime (MOV)  
Audio: PCM | X      | X        | Lossess compression 8-10 bit, variable bit rate |
|               |            | Codec: DV-NTSC (DV AVI)  
Wrapper: Audio Video Interleaved Format (AVI)  
Audio: PCM |        | X        | 8-bit, 29 mbps |
|               |            | Codec: MPEG-2  
Wrapper: QuickTime (MOV)  
Audio: PCM | X      | X        | Lossy compression 8-10-bit, variable bit rate |
|               |            | Codec: Apple ProRes  
Wrapper: QuickTime (MOV)  
Audio: PCM | X      | X        | Lossy compression 8-10-bit, variable bit rate |
| **Film**      | Preferred  | Codec: Uncompressed  
Wrapper: Digital Moving Picture Exchange  
Bitmap (DPX)  
Audio: PCM | X      |          | Uncompressed  
Scan resolution: 35mm: 4K; 8-16mm: 2 K; 10 bit RGB |
## Digitize audio and video (continued)

<table>
<thead>
<tr>
<th>Content Types</th>
<th>Preference</th>
<th>Format</th>
<th>Master</th>
<th>Derivative</th>
<th>Notes</th>
</tr>
</thead>
</table>
| **Audio**     | Preferred  | Codec: Linear Pulse Code Modulated Audio (LPCM)  
Wrapper: Broadcast Wave (BWF) | X      |         | 96 kHz / 24 bits recommended  
48 kHz/24 bits minimum |
|               | Accepted   | Codec: Linear Pulse Code Modulated Audio (LPCM)  
Wrapper: Waveform Audio (WAV)  
Codec: Linear Pulse Code Modulated Audio (LPCM)  
Wrapper: Multichannel Broadcast Wave | X | X | 96 kHz / 24 bits recommended  
48 kHz/24 bits minimum |
|               |            |        |        | X | Lossy compression |

Next: Post-digitization processing
Step Four
Prepare & Digitize

Post-digitization processing

Recommended procedures for post-digitization processing and enhancements (where not done during digitization):

- optimize images to improve legibility and quality by performing and documenting tasks such as de-skewing, sharpening, or despeckling (removing marks from the digitizing

- manage annotations to the source record (such as highlighting, stamps, redaction or addition of sticky notes) as overlays that do not change the actual image. You should be able to view the image with or without the annotations

- assign links between associated documents to be regarded as a single item: e.g. combining a document and appendices into one pdf, or describing the relationship between associated documents in the metadata properties field

- run text recognition on digitized records to support retrieval. Using Optical Character Recognition (OCR) in software such as Adobe Acrobat helps you to search within the document using keywords, copy text, and extract metadata. A scanned document is just an image with no text and cannot be searched. OCR creates text, which is then searchable in your office's recordkeeping system.

Standard #5.1.3

To ensure digitized copies are accurate, any enhancements (e.g. de-skewing, sharpening, or despeckling) must not remove any of the original content and must be documented.

Text recognition options include:

- using a scanner that comes with OCR software
- using software that can perform OCR on a single file or batch set of files

NOTE: alterations that cause a loss of information (such as cropping) are not acceptable in a defensible process. (Such alterations are acceptable for derivative copies, as the master copy remains unaltered.)
Step Five
Compile Metadata

What is metadata?

Metadata is structured information that describes and/or allows users to find, manage, control, understand, or preserve other information over time. It is data about data. It provides digitized records with searchability, meaningful context, and supports their management and retrieval in your organization’s recordkeeping system. Metadata also maintains the authenticity and integrity, as well as the context of creation, of the digitized record.

You must assign metadata to all digitized images. Capture the following types of metadata:

- metadata associated with the particular image/aggregate of images (e.g. a file/folder)
- metadata about the record/file and the business activity it documents
- metadata about the digitization process

Digitization equipment and information systems should automatically capture this metadata where possible. While this may require greater upfront development costs, it will reduce staff time and error rates.

Metadata needs to be maintained along with the digitized records to which it relates.

Standard #5.1.4

Metadata must be added during the digitization process. Digitized records must include:

- core metadata as defined in the Core Metadata Standard [a LINK to the Core Metadata Standard will be provided here when that is available]; and
- metadata that documents digitization processes and context (e.g. the date of digitization).

Next: Metadata requirements for digitization
Step Five
Compile Metadata

Please note that a government-wide Metadata Core Standard is in development. This guide will be updated where required, following its approval.

Metadata requirements for digitization

The metadata required for digitized records is essentially the same as that which is required for all digital information stored in government information systems (i.e. recordkeeping systems and any digital systems used to manage government information). Most of this metadata should be automatically captured during the digitization process, but some may need to be added or updated manually (e.g. the title).

Where applicable, the following recordkeeping metadata must be created for each file, folder or other information aggregate for digitized records. Note that more than one instance of some elements may be required; for example, the date element will need to be repeated each time there is a new digitization action.

<table>
<thead>
<tr>
<th>Metadata Element</th>
<th>Specifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Title</td>
<td>A meaningful title that aligns with authorized naming conventions. (See naming conventions below)</td>
</tr>
<tr>
<td>B. Creator</td>
<td>The name of the ministry and branch/division responsible for creating, receiving and managing the record(s). (Note this is the original creator, not the digitizer)</td>
</tr>
<tr>
<td>C. Dates</td>
<td>Date associated with an event in the lifecycle of the digitized record(s):</td>
</tr>
<tr>
<td></td>
<td>▶ Creation date (both start and end dates may be required)</td>
</tr>
<tr>
<td></td>
<td>▶ Digitization date</td>
</tr>
<tr>
<td></td>
<td>▶ Quality assurance check date</td>
</tr>
<tr>
<td></td>
<td>▶ Dates relating to any modifications of the digital image</td>
</tr>
<tr>
<td></td>
<td>▶ Project authorization date</td>
</tr>
</tbody>
</table>

Next: Metadata requirements for digitization (continued)
Step Five
Compile Metadata

Metadata requirements for digitization (continued)

<table>
<thead>
<tr>
<th>Metadata Element</th>
<th>Specifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>C. Dates (cont.)</td>
<td>This element must be provided every time there is a change in one of the other elements. This should be expressed as YYYY-MM-DD unless there is a need for more or less specificity, e.g. an exact time. In addition, when any of the other metadata elements of the digitized record are changed (e.g. if you rename the digital document), then the date of that change should be recorded.</td>
</tr>
<tr>
<td>D. Identifier</td>
<td>A unique combination of characters, numbers, and/or letters used to identify each record or aggregate of records. This information may be automatically generated when files are added to an office record keeping system.</td>
</tr>
<tr>
<td>E. Information schedule information</td>
<td>Relevant information schedule number and (if applicable) ARCS/ORCS primary and secondary number, or indication that the record(s) are unscheduled.</td>
</tr>
<tr>
<td>F. Security classification</td>
<td>The relevant information security status of the record(s), if known, in accordance with the Information Security Policy.</td>
</tr>
<tr>
<td>G. Format (if needed)</td>
<td>The original format of the record(s) and the new format, if either format is not typical (e.g. rolled map, Polaroid photograph, VHS tape).</td>
</tr>
<tr>
<td>H. Rights (if non-standard)</td>
<td>Intellectual property rights or any other relevant information about specific rights to the information held by individuals or organizations.</td>
</tr>
<tr>
<td>I. Language (if not English)</td>
<td>The specified language of a record, if it is other than English.</td>
</tr>
</tbody>
</table>

Next: Metadata requirements for digitization (continued)
### Metadata requirements for digitization (continued)

<table>
<thead>
<tr>
<th>Metadata Element</th>
<th>Specifications</th>
</tr>
</thead>
</table>
| **J. Digitization activities** | Provide information about the digitization activities. Digitization activities metadata may include:  
  - Number of items digitized  
  - Digitization enhancements  
  - Device name  
  - Calibration settings (resolutions, colour, dimensions, etc.)  
  - Batch reference  
  - The entity responsible for the digitization process or project  
  - Redundant Source Records (RSR) schedule application number  
  - Integrity check log files |

| **K. Additional information (if needed)** | In addition to the minimum metadata, you should assess what additional metadata is needed for business purposes. Additional metadata may include:  
  - Alternative title  
  - Description  
  - Geographical location  
  - Subject  
  - Usage considerations  
  - Extent  
  - Original recordkeeping context |

Next: [How to apply metadata](#)
Step Five

Compile

Metadata

How to apply metadata

The main approaches to capturing and storing metadata about digitized records are embedded metadata and linked metadata.

Embedded metadata (as the name suggests) is embedded within the file itself, becoming an integral part of the record (can normally be found under a menu item such as “properties”, though for email and some other documents, metadata is automatically added to the document header). Linked metadata, in contrast, is information that is stored separately from the newly created digital files. It could be linked to the digitized records with an entry in a register or database, with a hyperlink, or by being contained in the same case file. For digitized records, we recommend embedding the metadata within the digitized file if possible.

Some of the benefits of embedded metadata include:

- metadata stays with the file throughout its lifecycle
- no need to rely on database software to link the metadata and file
- metadata is human and machine readable
- open standard (will be readable for a long time)
- searchable
- legal benefits (i.e. copyright notices)

When you create a digital file, a large amount of metadata is automatically captured and embedded within the file. An example of this is the technical metadata added by your digital camera when you take a photograph (e.g. date and time taken, camera model, file type, file size, etc.). Other metadata elements (e.g. the subject line or an email) need to be manually added. Best practice is to develop standard practices for populating manually added elements to ensure searchability, such as naming conventions.
Step Five
Compile Metadata

How to apply metadata (continued)

From the list of required metadata elements, only two elements are typically applied by the scanning software or digital camera: dates and identifier.

The following descriptive elements will usually need to be manually added: title, creator, source file format, records management code, security classification, digitization actions.

A recommended way to do this is to use dedicated digital assets management software such as Adobe Bridge, which is part of the Adobe Creative Suite and available for download via the iStore (IDIR required).

Bridge is particularly useful for large digitization projects because it can apply metadata to multiple files at the same time. To do this, create a metadata template for your project that applies basic identification metadata to a batch set of files. You can include the program area, project or the individual documents. There is also the option of adding detailed descriptions of content, keywords, locations, etc. to help with search and retrieval. Rights metadata can be added to describe copyright information for legal purposes.

You can also use Bridge scripts to copy embedded metadata to a CSV file for use in a database or as an external backup.

For more information about using Bridge see:
https://helpx.adobe.com/ca/bridge/using/metadata-adobe-bridge.html
Step Five
Compile Metadata

Naming conventions

Naming conventions are standard rules applied to documents, files and other information aggregates. Applying consistent and predictable naming conventions to documents helps distinguish documents and files from one another and facilitates their storage and retrieval.

Follow the naming conventions used by your organization wherever possible. The names should be meaningful and borrow from naming conventions already in place. If your office does not have an established naming convention, consider adopting and/or adapting the suggested naming conventions outlined below.

Key elements

▶ **Title:** choose a name that accurately reflects the content of the document

▶ **Versions:** where applicable, differentiate between versions, by adding terms such as CURRENT, DRAFT, REVIEW COPY, SUPERSEDED, CLOSED, FINAL, or “APPROVED” (e.g. 2018_UFO_Minutes_draft, 2018_UFO_Minutes_final)

*Tip:* Ensure titles are updated so that you don’t end up with three “CURRENT” versions of the same document or file.

▶ **Date:** when a date is necessary (creation date or date of update) use a hyphen to separate the year, month and day, e.g. YYYY-MM-DD

▶ **Creator:** write the name of the ministry/branch/division organization, and use standardized acronyms if available

Next: Best practices for naming government information
Best practices for naming government information

Names are key metadata: consistency in naming can greatly improve the findability of scanned documents.

- Keep document names short, but meaningful
- Use consistent wording, punctuation and spelling and arrange data elements in a logical order
- Use standard terms, codes, abbreviations, acronyms, and structure. Where government standards or system-generated codes already exist, use them. Where departures from these or new standards are needed, document them and update them as necessary.
- Where an acronym is necessary, spell out the first usage, or provide the full name in metadata or another accessible source

Recommended practices:

- Hyphens or underscores to separate words
- Providing leading zeros when using a number in a document title to keep the documents in numerical order, e.g. 01, 02, 03

Avoid:

- Spaces in file names for records NOT managed in the Enterprise Document Records Management System (EDRMS)
- File names that have more than 255 characters
- Repetition & redundant words
- Words that are excluded from searches, e.g. the, if, but, so, for, etc.
- Symbols that are incompatible with common software, e.g. ~ ” # % & * : < > / \ { | }
Managing versions

For various reasons, you may choose to keep one or more derivative copies along with your production master file. Derivative copies may include:

- the same image prepared for different output intents (e.g. a compressed version suitable for posting online)
- versions with additional edits (e.g. contrast may be sharpened in a photo, specks may be removed from a textual document)
- layers (e.g. a digitized map may have multiple layers to display data themes such as roads or boundaries)
- versions scanned on different scanners

To distinguish between different versions, add qualifiers to the document/file name that reflect the nature and intent of each version (e.g. draft, superseded, current, reviewed by X).

Follow these guidelines:

- distinguish versions from one another using the same name with a qualifier (but if you are using numbers, keep them distinctive - e.g. if 001 and 002 are page numbers, they shouldn’t also be used for version numbers)
- add a qualifier to distinguish derivative files from other files (for example, “p” for published version, “t” for thumbnail)
- for derivatives intended for posting online, provide enough descriptive or numerical information to provide easy reference to other digital versions
- provide master file identification information in the derivative file names or other metadata, so that users can easily find the master file if needed
Step Five
Compile Metadata

Organize digitized content

Arrange the digitized records to mirror the source records, unless there is a compelling reason not to. In most cases, this will result in a set of files organized according to the relevant *Operational Records Classification System* (ORCS) or *Administrative Records Classification System* (ARCS), with the documents in each file arranged in order by date created or received, with the latest documents appearing first and the earliest at the end.

The ARCS/ORCS classification codes provide key metadata for your records. There are a growing number of tools available to help you manage unstructured data by tagging digitized records with meaningful metadata.
Step Six
Assess Quality

Quality assurance

Quality assurance (QA) is the process of verifying whether the digitized record meets or exceeds expectations. It involves checking the operation and output of digitization processes against agreed benchmarks to ensure that these benchmarks have been met. These measures will form part of your defensible process and need to be documented and incorporated into ongoing digitization processes.

The digitized record can only meet your office needs and be defensible if it has integrity and is a true and accurate copy of the source record. Effective quality assurance is especially critical if the source record will be destroyed and the digitized version will become the only record of the business activity. Without strong quality controls the digitized version is at risk of being inaccurate, incomplete, and illegible. Quality assurance ensures the digital copy is a true and accurate copy of the source record.

Quality assurance measures also need to be documented for in-house digitization. This can be as simple as recording the date of the QA check, but can also involve signatures or other modes of identification (e.g. stamps, marks, electronic identifiers). Digitization service providers should be required in contractual agreements to complete and document quality assurance activities.

Standard #5.1.5

Quality assurance measures must be established and documented for all digitization activities. Quality assurance measures must include:

- providing appropriate training for all staff who create, manage or work with digitized government information;
- ensuring that scanning equipment is fully functional;
- verifying completeness, quality and accuracy of images and metadata;
- handling and resolving errors, including any necessary re-digitization, within a set time period; and
- regular review of quality assurance procedures to ensure they continue to meet business requirements.
What is a digitization error?

You need to consider and decide what constitutes an error, and what level of error is permissible. Decide what you will accept as “good enough” and what will result in a “fail”. Use the table below to determine what errors to look for in your quality assurance process:

<table>
<thead>
<tr>
<th>Error type</th>
<th>What to look for</th>
<th>Causes</th>
</tr>
</thead>
</table>
| Completeness | Missing records: Have all the records been digitized? Verify that the number of digitized records matches the quantity of source records | ▶ Incomplete or inaccurate specifications or process documentation  
▶ Faulty capture hardware (incorrectly calibrated devices)  
▶ Faulty software (inaccurate image processing or faulty image links within database) |
| Completeness/ Image Accuracy | Missing content: Are there any broken characters or missing segments of lines? Don’t forget to check if there is any missing information at the edges of the image area | ▶ Incomplete or inaccurate specifications or process documentation  
▶ Faulty software (inaccurate image processing or faulty image links within database)  
▶ Basic capture errors  
▶ Cropping that has cut into the image, is too loose, or is uneven |
| Image Accuracy | Dimensions: Do the dimensions accurately compare with the original? | ▶ Cropping that has cut into the image, is too loose, or is uneven |
| Image Accuracy | Legibility: Has the smallest detail been legibly captured? (e.g. smallest type size for text; clarity of punctuation marks, including decimal points; handwritten annotations, visibility of watermarks and embossing) | ▶ Orientation of the image is the wrong way around, or upside down  
▶ Exposure of the image is too light or too dark  
▶ Focus, where the image is out of focus  
▶ Basic image processing errors  
▶ Capture device has been incorrectly calibrated |

NOTE: It’s a good idea to identify poor quality source records before digitizing them. When the source record is of very poor quality (such as faded type), then the digitized images will also be of very poor quality. Consider taking special measures (including digitizing material at a higher resolution) to increase the legibility of poor quality images.
### What is an error? (continued)

<table>
<thead>
<tr>
<th>Error type</th>
<th>What to look for</th>
<th>Causes</th>
</tr>
</thead>
</table>
| **Image Accuracy** | **Colour fidelity**: Do the colours accurately compare with the original? Are black areas too light or too dark? | ▶ Incorrectly calibrated colour management systems  
                      ▶ Exposure of the image is too light or too dark  
                      ▶ File optimization errors, where incorrect adjustments are made to the colour, contrast and brightness of the image during processing |
| **Image Accuracy** | **Sharpness**: Is the sharpness of the image comparable to the original? Is there a lack of sharpness or too much sharpening? Are there any halos around dark edges? | ▶ Basic capture errors  
                      ▶ Focus, where the image is out of focus  
                      ▶ Low quality source record |
| **Image Accuracy** | **Artifacts from scanning process**: Has scanner-generated speckle been removed? | ▶ Dirty originals  
                      ▶ Capture device has been incorrectly calibrated |
| **Image Accuracy** | **Image enhancements**: Assess image enhancements to ensure information is not lost (for example, if the tolerances are set too high, the dot above the letter ‘i’ may be removed). Image optimization processes should also be documented to help ensure the **authenticity** and completeness of the digitized copies are not at risk of being challenged (a particular concern for documents that will eventually be transferred to the government archives) | ▶ File optimization errors  
                      ▶ Images modified by watermarks or fingerprints (watermarking and fingerprinting techniques are not acceptable; use metadata instead to convey any needed information) |
## What is an error? (continued)

<table>
<thead>
<tr>
<th>Error type</th>
<th>What to look for</th>
<th>Causes</th>
</tr>
</thead>
</table>
| **Metadata quality and accuracy** | **Adherence to standards** set by institutional policy or by the standards of the digitization project | ▶ Inaccurate source metadata  
▶ Incorrect use of controlled vocabulary  
▶ Incorrect data entry  
▶ Incorrect metadata attribution faults  
▶ Incorrect file-naming, where image files are incorrectly named or use names that are not unique |
|                               | **Completeness**: All mandatory metadata fields should be complete               |                                                                        |
|                               | **Relevancy and accuracy of the metadata** (check OCR accuracy, file name, correct grammar, spelling and punctuation – especially for manually-keyed text) |                                                                        |
|                               | **Usefulness of the metadata** being collected                                   |                                                                        |
|                               | **Synchronization of the metadata** stored in more than one location: make sure metadata is updated in a synchronized manner across more than one location (for instance, information related to the image might be stored in the TIFF header, the digital asset management system, and other databases) |                                                                        |
| **Data loss/corruption**      | **Data integrity**: Assess if any data was lost when transferred between systems (e.g. during transfer from the service provider to the office that owns the records). One method for checking the integrity of files after they are transferred is using checksum software to ensure your copy of the record is genuine and error free. Another option is to use SSL certificates to safeguard sensitive data and prevent transferred information from being modified. | ▶ Incomplete file transfer  
▶ Data corruption |
Step Six
Assess Quality

Implementing QA

If you are digitizing one or a few documents, QA can be as simple as glancing at the copy to verify that it matches the original. For digitization that is part of an ongoing business process or a large project, you will need to incorporate and document quality assurance measures throughout the digitization process.

The complexity and detail of quality checks will depend on your needs, the nature of the records’ content or document format, and the risks associated with the records. Apply strict and detailed quality control if you intend to destroy the source records. High value or high risk records – such as vital records, records with long retention periods, records with enduring historic value, records scheduled for full retention, or fragile records – need special consideration.

If you are unsure what level of quality assurance is sufficient, consider the potential risk of not being able to provide trustworthy records in situations requiring documentary proof. Management will need to look at the documentation and decide what level of accuracy is required, and whether checking more or all of the digitized files is needed. (See the Risk Assessment Matrix for more guidance.)

Once you have a good understanding of the records’ nature and associated risks, you can define how much error to allow. Recommended practice is to accept an error rate of 1% or 0.1%, depending on the documents being digitized. For high risk records, fewer errors should be allowed (e.g. 0.1% or 1 error in a sample batch of 1000).

Recommended practices

- 1% or 0.1% error rate, depending on the level of business need for accuracy, given the nature and use of the source records being digitized
- test a sample of all document/media types being scanned to create a quality benchmark
- at a minimum, check 5% of the total volume of digitized records (see section “How much needs to be quality assured?”)
- perform quality assurance checks throughout the digitization process
- correct errors by allowing sufficient time for QA to be carried out and for any re-digitization to be completed before source records are returned to storage or destroyed.

Apply your QA ratio (the number of errors that are acceptable in a subset of records) to your digitization project. Review the sample batch of records you want to digitize and create an estimate based on the volume and type of documents (note: count double-sided pages as 2 units). For example, if you apply a QA ratio of 0.1% to 2,000 sheets of paper (with content on only one side of the sheet), 2 errors are acceptable.

Next: How to run QA
Step Six
Assess Quality

How to run QA

Sample scans

Before you start a digitization process, it is a good idea to test your digitizing, metadata, and QA measures with a sample. Depending on the type of records being digitized, you may choose to inspect every image or a percentage of the images. Best practice is to sample at least 5% of the images. To do this, pull a representative sample of documents from the batch and scan them at a high resolution, and in full scale view to set a quality benchmark. After digitizing the sample, review the results, and document the technical settings that allow an appropriate level of legibility, clarity, range of tones and colours of the digitized image. If there are errors, you need to review their potential causes (see What is an Error?) and associated processes and procedures. Adjust as needed and test again.

Make sure the representative sample includes:

- all types of records in the aggregate (e.g. single page documents, multipage documents, photographs, etc.)
- all sizes of records (especially if there are non-standard document sizes)
- all conditions (include poor quality, creased, stapled, and fragile documents to determine if they require special treatment)
- all formats (textual records, microfiche, maps, bound volumes, etc.)

Test digitization equipment on a regular basis to monitor system performance and check that digitized records are within agreed benchmarks established in existing procedures (you can also use the results of previous tests).
Step Six
Assess Quality

How to run QA

How much needs to be quality assured?
At a minimum, you or your service provider should check 5% of the total volume of digitized records. However, if the records are high risk or have high value, you may need to do more QA. When any work factors change, such as equipment repairs or replacement, new staff, or a new service provider, best practice is to check each image until you are confident the QA standard is being upheld. You may need to check all digitized records if the source records are particularly detailed, an unusual size, difficult to scan, or contain information near edges of the page.

When to do QA
QA should occur throughout the digitization process as an integral part of the regular workflow. This way, issues can be resolved as they occur. For some processes, it may be appropriate to perform quality assurance on a daily or weekly basis (example: 5% of all images created each day) or to use random sampling throughout the project. Ensure that the total volume of digitized images meets the required benchmark.

If it is not possible to conduct QA soon after digitizing, you will need to establish a dedicated QA process.

Who should do QA
The person who carries out the quality checks needs to be different from the person doing the digitizing. This may be another employee within the digitization team or somebody independent of the team. If digitization is outsourced, it is recommended that some QA checking be carried out in-house, in addition to QA checks carried out by the service provider.
Step Six
Assess Quality

How to run QA

Dealing with errors
Create a process for handling and resolving errors. All errors should be documented and reported immediately. If the errors raise concern, review your process for preparing source records and incorporate more stringent requirements if needed. Re-inspect other images from the same batch for errors to determine if the error was a one-time mistake or a bigger issue that needs to be resolved.

Re-digitizing process
Correct any specific errors found during the digitizing process or in the random sampling by re-digitizing. Choose an acceptable error rate in advance, and specify in your documentation that error levels over this rate will require re-digitizing the whole batch.

Review the QA process
Periodic review of the quality assurance process is important to ensure that benchmarks and quality assurance measures continue to meet business needs, current technology, and reflect changes to legislation.
Step Seven
Store &
Manage
Records

Managing original records after digitization

Consider how the source records will be managed after digitizing or if they will be destroyed. For projects, you will need to decide how long to keep source records before they are returned to offsite storage, transferred to the government archives or are destroyed, to allow you time to run quality assurance checks and re-digitize any that are not acceptable. Your organization may choose to allow two years or more before destruction occurs, to allow more time to find errors. In other cases, a few weeks or months may be adequate.

If source records are to be disposed of after digitization, contact your Records Team and ask for authorization to destroy the source records under the Redundant Source Records (RSR) schedule.

Standard #5.1.6

If there are no legal requirements to maintain the original source records, applications to destroy source records can be authorized in accordance with:

• this standard; and
• relevant information schedules, including the Redundant Source Records Schedule (206175).

If source records are retained after being digitized, the original order and context must be maintained, and digital versions must be linked to source records.

Please note: Digitization projects already approved under the Redundant Source Records Schedule do not require new approvals; they are encouraged but not required to meet this standard.
Step Seven
Store & Manage Records

Transferring the digital images

If you are working with a service provider, you will need to ensure that the digitized files of the source records are securely transferred from the service provider to your organization. Options include using a Secure File Transfer Protocol (SFTP) or setting up a shared drive with the service provider.

BC Government ministries are encouraged to use the the Office of the Chief Information Officer (OCIO)’s Secure File Transfer (SFT) Service. This web-based service enables ministries to exchange files over the internet with service providers and other organizations. For more information on this service, visit the Secure File Transfer website.

Whatever method you use for transferring digitized records, you will need to ensure that the digitized files are not altered during transfer. Collecting fixity information for digitized records will help you:

- validate that you have received the files you expected (e.g. when a Service Provider provides fixity information with the digitized files)
- determine if the data is corrupted or altered from what you expected (e.g. comparing the baseline fixity information with future fixity check information)
- prove that the digitized files you have are authentic and trustworthy

Next: Transferring digitized records (continued)
Step Seven

Store & Manage Records

Transferring digitized records (continued)

There is a wide and evolving range of approaches available for checking the fixity of digital files, including those that are built into storage systems, those that can be automated through scripts and applications, and those that might involve manual workflows. One recommended method for creating fixity information is using checksum software (e.g. “Bagit”) that packages the files, and creates a manifest listing their names and checksums. Fixity information can be stored in metadata, in databases and logs, and alongside the digitized files. For more information on fixity and checksums see the Digital Preservation Coalition Handbook.

As an additional measure, you can use “write blocker” tools; this is any hardware or software that permits viewing file contents to identify changes, without permitting any further changes to the contents.

Do not transfer files using:

► mobile storage devices (such as memory sticks)
► unsecured drives
► services that have not been approved by the Office of the Chief Information Officer (OCIO)
How will the digitized records be managed, stored and accessed?

Digitized copies that replace source records need to be retained in an appropriate recordkeeping system, in accordance with the appropriate information schedule.

See the Office Recordkeeping System Guide for an explanation of the relationship between an “appropriate system” for managing government information and a “recordkeeping system”.

Best practice is to use the government standard Enterprise Document and Records Management System (EDRMS) or a system tailored to the business needs of the program area. It may be necessary to adapt or plan for improved information management controls in line of business systems. As an interim approach pending use of a fully functional EDRMS, your organization may store limited volumes and types of digital records on a LAN Drive structured according to Administrative Records Classification System (ARCS) and Operational Records Classification System (ORCS), with security measures put in place to protect the integrity of the information. To set up secure folders, complete an iStore (IDIR required) application.

Next: How will the digitized records be managed, stored and accessed? (continued)
How will the digitized records be managed, stored and accessed? (continued)

Your organization should ensure the government information in your custody and control is managed in accordance with legislation and policy, and in a manner that meets operational and information system requirements. Remember to manage the new master copies according to processes outlined in the Information Management Act (IMA) and the Managing Government Information Policy (MGIP). For more information see the MGIP internet page.

If your current recordkeeping system appears to be inadequate, develop a strategic plan for improving it. Contact your Records Team for assistance.

To learn more about records management training opportunities, check out GRS Learning on the GRS Intranet.
### Planning Checklists

The following checklists will assist you with assessing the viability of **business process digitization** for original source records, and with planning digitization projects.

**Business Process Digitization Planning Checklist**
(print-friendly pdf)

<table>
<thead>
<tr>
<th>What are your reasons for digitizing?</th>
</tr>
</thead>
<tbody>
<tr>
<td>☑ I need to digitize to manage all incoming and outgoing transactions electronically</td>
</tr>
<tr>
<td>☑ I need to transmit the records across internal or external networks</td>
</tr>
<tr>
<td>☑ I need to make the records available to staff at centralized and remote locations</td>
</tr>
<tr>
<td>☑ Other reasons: For guidance, see Step One: Plan to Digitize in the Digitizing Government Information Guide</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>What will be digitized?</th>
</tr>
</thead>
<tbody>
<tr>
<td>☑ The records are covered by an approved/pending information schedule</td>
</tr>
<tr>
<td>☑ All incoming source records will be digitized</td>
</tr>
<tr>
<td>☑ Only a selection of incoming source records will be digitized</td>
</tr>
<tr>
<td>☑ For guidance, see Step Two: Assess Records in the Digitizing Government Information Guide</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>When will you digitize?</th>
</tr>
</thead>
<tbody>
<tr>
<td>☑ The records will be digitized before being used for work processes</td>
</tr>
<tr>
<td>☑ The records will be digitized after being used for work processes</td>
</tr>
<tr>
<td>☑ Sales and responsibilities have been clearly documented, identified, and communicated for staff across all processes</td>
</tr>
<tr>
<td>☑ For guidance, see Step Three: Develop a Defensible Process in the Digitizing Government Information Guide</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>How will the records be digitized?</th>
</tr>
</thead>
<tbody>
<tr>
<td>☑ A documented procedure for preparing and digitizing the source records and post-digitization processing</td>
</tr>
<tr>
<td>☑ Have determined whether any special physical preparation procedures are needed and developed these if so</td>
</tr>
</tbody>
</table>

**Digitization Project Planning Checklist**
(print-friendly pdf)

<table>
<thead>
<tr>
<th>What are your reasons for digitizing?</th>
</tr>
</thead>
<tbody>
<tr>
<td>☑ The records document a process or function that continues to be done</td>
</tr>
<tr>
<td>☑ The equivalent information in the records is now recorded in digital form (i.e. “born digital” information)</td>
</tr>
<tr>
<td>☑ The records contain a rich information source by current business processes</td>
</tr>
<tr>
<td>☑ The information in the records is enhanced by being available in digital form (e.g. additional indexing, sorting capabilities)</td>
</tr>
<tr>
<td>☑ Other reasons: For guidance, see Step One: Plan to Digitize in the Digitizing Government Information Guide</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>What will be digitized?</th>
</tr>
</thead>
<tbody>
<tr>
<td>☑ The records are covered by an approved/pending information schedule</td>
</tr>
<tr>
<td>☑ Are a coherent and complete set</td>
</tr>
<tr>
<td>☑ Need to be retained for a long time or will be transferred to the government archives (i.e. due to legislation or an information schedule)</td>
</tr>
<tr>
<td>☑ For guidance, see Step Two: Assess Records in the Digitizing Government Information Guide</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>When will you digitize?</th>
</tr>
</thead>
<tbody>
<tr>
<td>☑ Roles and responsibilities have been clearly documented, identified, and communicated for staff across all processes</td>
</tr>
<tr>
<td>☑ For guidance, see Step Three: Develop a Defensible Process in the Digitizing Government Information Guide</td>
</tr>
</tbody>
</table>

Government Records Service, Province of British Columbia

Last revised: June 29, 2020 2:07 PM
## Risk Assessment Matrix

<table>
<thead>
<tr>
<th>Risk criteria</th>
<th>Questions to consider when assessing the level of risk</th>
<th>Comments</th>
</tr>
</thead>
</table>
| Digitized image of the record may not be a full and accurate copy. | Have you implemented robust [technical](#) and metadata specifications, and [digitization](#) and [quality assurance processes](#) to ensure full and accurate digitized images of the originals are being captured? | Risks to consider:  
- Records being:  
  a) incomplete,  
  b) not digitized at all,  
  c) of poor quality, or  
  d) missing metadata |
| Digitized records may not be accessible in the long term. | Do you have a strategy in place for the ongoing preservation of digital records? | The lifespan of most technology is shorter than the period of time you may need to keep your digitized records. Systems and storage media quickly become obsolete. |
| Digitization may result in privacy/security issues. | Do the records include any personal information? Are there any security requirements? | Completing a [PIA](#) would mitigate this risk. |
| Digitized records will not survive if not maintained in a secure, well-managed [recordkeeping system](#). | Do you have a reliable digital recordkeeping system for managing the digitized records, such that all recordkeeping and information security controls are managed, and through which the authenticity of the record can be demonstrated? | Recordkeeping systems must be reliable and secure. If your digital storage system is inadequate, you need a strategy to improve/replace it. |
| The authenticity of the digitized record could be challenged after the source record is destroyed. | Have you [documented](#) your digitization process adequately? | Without trusted systems and processes, the authenticity of the records may be compromised. |

Next: [Risk Assessment Matrix (continued)](#)
### 8.2 Risk Assessment Matrix (continued)

<table>
<thead>
<tr>
<th>Risk criteria</th>
<th>Questions to consider when assessing the level of risk</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>The source records contain physical elements that cannot be replicated in the digitized copies (e.g. corporate seals, watermarks, original signatures).</td>
<td>Do the records contain physical elements which attest to their <strong>authenticity</strong> or evidential value?</td>
<td>Physical elements are frequently used to demonstrate authorship and authenticity and may have <strong>intrinsic value</strong>. Ensure source records are only destroyed in compliance with information schedules and government policy. If necessary, also seek legal advice.</td>
</tr>
<tr>
<td>Citizens may expect that certain source records they deem significant will be retained in their original form due to perceived intrinsic value (e.g. original designs and works of art, maps with overlays, handwritten letters that may hold historical or personal significance).</td>
<td>Do the source records have value as physical objects?</td>
<td>Ensure that considerations of intrinsic value are accounted for and documented in the digitization process documentation, if not already adequately documented in an <strong>information schedule</strong> or elsewhere.</td>
</tr>
<tr>
<td>While the digitized record is sufficient for the organization’s immediate business purposes, the source record is still required for legal, archival, or other reasons.</td>
<td>Have you documented the justification for retaining both the source record and the digitized record?</td>
<td>Do not <strong>destroy</strong> the original source record if it is still required to comply with legislation, an information schedule, or for another reason. Document the reasons for retaining both.</td>
</tr>
</tbody>
</table>
## Roles and Responsibilities

<table>
<thead>
<tr>
<th>Role</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Management</strong></td>
<td>▶ Assessing the business case for digitization</td>
</tr>
<tr>
<td></td>
<td>▶ Negotiating and monitoring purchase of services, equipment, and supplies</td>
</tr>
<tr>
<td></td>
<td>▶ Managing staff resources</td>
</tr>
<tr>
<td></td>
<td>▶ Authorizing the digitization plan and procedures (see Responsibilities)</td>
</tr>
<tr>
<td><strong>Business Analysis</strong></td>
<td>▶ Defining the workflow for the digitization process</td>
</tr>
<tr>
<td></td>
<td>▶ Defining the integration of digitized records into existing business systems/workflows</td>
</tr>
<tr>
<td></td>
<td>▶ Selecting image format</td>
</tr>
<tr>
<td></td>
<td>▶ Determining image enhancement requirements</td>
</tr>
<tr>
<td></td>
<td>▶ Identifying information architecture for business-process support</td>
</tr>
<tr>
<td><strong>Systems Analysis</strong> <em>(unless project uses standard MFDs or is entirely outsourced)</em></td>
<td>▶ Selecting and integrating imaging equipment and software</td>
</tr>
<tr>
<td></td>
<td>▶ Providing ongoing support of digitization equipment and software</td>
</tr>
<tr>
<td></td>
<td>▶ Defining storage requirements</td>
</tr>
<tr>
<td></td>
<td>▶ Integrating digitization requirements into existing organizational IT standards</td>
</tr>
<tr>
<td></td>
<td>▶ Testing configurations</td>
</tr>
<tr>
<td></td>
<td>▶ Following government information management policy and guidance</td>
</tr>
<tr>
<td><strong>Recordkeeping</strong></td>
<td>▶ Ensuring compliance with relevant policy, standards, and information schedules</td>
</tr>
<tr>
<td></td>
<td>▶ Integrating digitized records into organizational records and business systems</td>
</tr>
<tr>
<td></td>
<td>▶ Defining the naming conventions</td>
</tr>
<tr>
<td></td>
<td>▶ Defining and implementing disposal process</td>
</tr>
<tr>
<td></td>
<td>▶ Defining metadata and monitoring its quality</td>
</tr>
<tr>
<td></td>
<td>▶ Managing the source records and the digitized copies after digitization</td>
</tr>
</tbody>
</table>
Resources

B.C. Government resources

- Core Policy and Procedures Manual, Chapter 12
- Digitizing Government Information Standard (DGIS)
- *Electronic Transactions Act (SBC 2001, c. 10)*
- *Information Management Act (SBC 2015, c. 27)*
- Managing Government Information Policy (MGIP)
- Privacy Impact Assessment Templates
- Recorded Information Management Manual (RIMM)
- Redundant Source Records Special Schedule (206175) and Guide

External Standards *(GRS holds reference copies)*

- AIMM TR 34: Sampling procedures for inspection by attributes of images in electronic image management and micrographic systems
- ISO 12653 – Electronic Imaging: Test target for the black-and-white scanning of office documents – Part 2: Method of use
- ISO 15836-1 - Information and documentation - the Dublin Core metadata set - Part 1: Core elements
- ISO/TR 13028:2010 Information and documentation. Implementation guidelines for digitization of records
Tools & Resources

Best practices in other jurisdictions

- Government of Alberta, Digitization Quality Assurance, 2018
- Government of Northwest Territories, Directive - Digitizing, 2018
- Government of the Northwest Territories, Guideline - Digitization, 2018
- Government of Northwest Territories, Standard - Digitizing, 2018
- National Archives of Australia, Preservation Digitisation Standards, 2018
- Queensland Government, Digitise Records, 2018
**Infographics**

**How to Digitize Your Records: Ongoing Business Process Digitization**

1. Develop a rationale for your digitization process.
2. Address the minimum requirements in the Digitizing Government Information Standard.
3. Ensure that records management requirements are being met.
4. Select a digitization approach.
5. Plan the workflow.
6. Determine the digitization location (bulk process or ongoing process?).
7. How to transport the records?
8. Security & privacy requirements.
10. Where the digitized files will be stored.
11. Authorized destruction of source records.
12. Who does what (official file copy or not?).
13. When the digitization will occur.
15. Prepare source records.
16. Determine whether the digitization will be carried out in-house or by a service provider.
17. Document procedures for digitizing and post-digitization processes (including OCR).
18. Determine metadata requirements.
19. Apply consistent & predictable naming conventions.
21. How errors will be resolved (e.g., redigitize).
22. Review the QA process periodically.
23. Consider how the source records will be managed after digitization or if they will be destroyed.
24. Ensure that digitized records are securely transferred.
25. Retain digitized copies in an appropriate recordkeeping system in accordance with the appropriate information schedule.

**SUCCESS!**

**How to Plan a Digitization Project**

1. Develop a rationale for your digitization process.
2. Identify the necessary resources.
3. Include a realistic budget.
4. Address the minimum requirements in the Digitizing Government Information Standard.
5. Consider the challenges.
6. Clarify Roles, responsibilities, reporting lines, and communications for staff across all processes.
7. Ensure that records management requirements are being met.
8. Select a digitization approach.
9. Plan the workflow.
10. Determine the digitization location (bulk process or ongoing process?).
11. How to transport the records?
14. Where the digitized files will be stored.
15. Authorized destruction of source records.
16. Who does what (official file copy or not?).
17. When the digitization will occur.
18. Document your decisions, processes and procedures.
19. Prepare source records.
20. Determine whether the digitization will be carried out in-house or by a service provider.
22. Determine metadata requirements.
23. Apply consistent & predictable naming conventions.
25. How errors will be resolved (e.g., redigitize).
26. Review the QA process periodically.
27. Consider how the source records will be managed after digitization or if they will be destroyed.
28. Ensure that digitized records are securely transferred.
29. Retain digitized copies in an appropriate recordkeeping system in accordance with the appropriate information schedule.

**SUCCESS!**

Next: **Glossary**
**Glossary**

**A**

**ACCURACY**

The qualities of a record that render it precise, correct, truthful, free of error or distortion.

**ADMINISTRATIVE RECORDS CLASSIFICATION SYSTEM (ARCS):**

The government-wide standard for classifying, filing, retrieving and disposition scheduling of administrative records. ARCS also includes freedom of information and protection of privacy designations. ARCS is a block numeric system, reflecting function and subject.

**AUDIO FORMATS**

An audio file format is a file format for storing digital audio data on a computer system. The bit layout of the audio data (excluding metadata) is called the audio coding format and can be uncompressed, or compressed to reduce the file size, often using lossy compression. Examples of audio formats include:

- **Broadcast Wave (BWF)**
  BWF has been universally adopted throughout the audio, computer and broadcast industries as a digital audio format. It is compatible with any hardware or software that supports the WAVE format. Note: This format is limited to a 4 gigabyte file size.

- **Linear Pulse Code Modulated Audio (LPCM)**
  Linear PCM captures and encodes lossless audio. Fidelity for LPCM is enhanced by higher sample rates and sample sizes.

- **MPEG-1/2 Audio Layer III (MP3)**
  MP3 is a compressed audio file format developed by the Moving Picture Experts Group (MPEG). MP3 files are often about one tenth the size of an uncompressed file format.

- **Multichannel Broadcast Wave (MBWF / R64)**
  MBWF is a BWF-compatible file format which is designed to meet the requirements for multichannel sound in broadcasting and audio archiving. MBWF is very widely used for encoding bitstreams.
Glossary

- **Pulse Code Modulated Audio (PCM)**
  Pulse code modulation (PCM) with linear quantization. PCM is a digital representation of an analog signal where the magnitude of the signal is sampled regularly at uniform intervals, then quantized to a series of symbols in a digital (usually binary) code.

- **Waveform Audio File Format (WAV)**
  Waveform Audio (WAVE) is a chunk-based audio format developed by Microsoft, based on the generic Resource Interchange File Format (RIFF) specification. The format supports a wide variety of compressed and uncompressed audio encodings.

**AUTHENTICITY**

The quality of being genuine, not a counterfeit, and free from tampering or corruption. Authenticity alone does not automatically imply that the content of a record is reliable or accurate; it merely establishes that a record is what it purports to be and has verifiably been created by the person who claims to be the creator.

**B**

- **BIT DEPTH (digital image)**
  Refers to the number of bits used to represent each pixel in a digital image. The more bits that are used by each pixel, the more closely a digital image can approximate a photograph (and the more memory the image occupies).

- **BIT DEPTH (video/audio)**
  Bit depth in video is the number of bits in each sample. Increased bit depth equals increased precision of the sample and therefore increased quality of the preservation files. When digitizing audio, a greater bit depth will enable a greater ratio between the quietest and loudest possible values of a signal. Digitizing at 24 bits bit depth ensures a dynamic range greater than any audio equipment and format.

- **BORN DIGITAL**
  Digital information which is created/received and managed only in digital form. This term is used to differentiate from digital information which has been created as a result of converting physical source records (i.e. digitized records).
Glossary

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

BUSINESS PROCESS DIGITIZATION

Digitization of records, often on receipt, and incorporation of the digitized record into a digital system or work process, where most of the other records are born digital.

CHECKSUMS

A checksum is an algorithm used to confirm that a digital record has not changed when transferred from one location to another, or over time; in other words, it is a digital fingerprint that confirms the integrity of the record.

CODEC

Hardware or software capable of encoding and/or decoding the essence of an audio or video file. The coded format of the essences in a file is often identified by the type of codec that is required to decode the essence for playback. The codecs that support lossless compression (uncompressed, JPEG 2000 or FFv1) preserve all the information of the source signal, but they also require high bitrates, large file sizes, and extensive computing power to process.

Note: the majority of video compression codecs are lossy, meaning some level of information from the original source signal has been lost. There is always a trade-off between visual quality, file size and processing power needed to encode and decode a given video compression codec. Examples of video codecs include:

  This codec is widely supported in broadcast post-production and professional editing software. Playback of the codec is included in the base QuickTime codec package and open-source playback compatibility exists. Note: ProRes is a lossy video compression codec developed primarily for use in video post-production. It has not been standardized by any international body.

- **DV-NTSC**
  This codec was first standardized through the International Electro-technical Commission (IEC) family of standards in 1995. This standard outlined how DV video was to be recorded on videotape. DV-NTSC is a good alternative option for recordings that are on digital tape formats. It will also create smaller target files than lossless or uncompressed codec options.
Glossary

A   B   C   D   E   F   G   H   I   J   K   L   M   N   O   P   Q   R   S   T   U   V   W   X   Y   Z

▶ **FFv1 Version 3**
The FFv1 codec is open-source. Version 3 of the FFv1 supports frame level CRC fixity information and more self-descriptive properties at the codec level such as field dominance, aspect ratio, and colour space information.

▶ **H.264 and H.265**
H.264 and H.265 are commonly used for video compression formats that support both lossless and lossy image compression. They have a wide range of applications, covering all compressed video ranging from low bit-rate streaming applications (YouTube, Vimeo, Facebook) to Blu-ray Discs, DVD storage and HDTV broadcasts.

▶ **JPEG 2000**
An image compression method that provides much better image quality at smaller file sizes than the original JPEG format. The JPEG 2000 file format supports both lossless and lossy image compression. A full range of metadata may also be bundled within the file.

▶ **Digital Moving Picture Exchange Bitmap (DPX)**
The DPX file format is an open standard, defined by the Society of Motion Picture and Television Engineers (SMPTE). A wide range of motion picture film scanners support output to DPX.

**D**

**DERIVATIVE COPY**
A user or access copy created from the master copy of a digitized record through an editing process. The process usually involves compressing the file size or altering an image using enhancement techniques.

**DESTRUCTION OF RECORDS**
Inactive government information can only be destroyed in accordance with information schedules approved under the *Information Management Act*. Destruction may be carried out using various methods, such as deleting, shredding, incineration, or pulping, but must always comply with relevant policy. Destruction is only complete when the information is obliterated and cannot be reconstituted. See also *Dispose*. 
**DIGITIZATION PROJECT**
Retrospective or project-based digitization of existing sets of source records to enhance accessibility, maximize re-use, or for preservation purposes.

**DISPOSE**
The Information Management Act defines the term “dispose” to mean to destroy, or render impracticable to decipher, recorded information. See also [Destruction of Records](#).

**DISPOSITION**
The process which enables government to dispose of records which no longer have operational value, either by permitting their destruction, by requiring their transfer to the government archives, or by agreeing to their alienation from control of government.

**FIXITY INFORMATION**
Fixity is the property of a digital file or object being fixed or unchanged. Fixity information offers evidence that one set of bits is identical to another. Types of information include checksums, cryptographic hashes and other methods such as expected file size and file count.
Glossary

INTEGRITY
The quality of being whole and unaltered through loss, tampering, or corruption. In the context of records, integrity relates to the potential loss of physical or intellectual elements after a record has been created. As one of the components used to determine a record's authenticity, integrity is a relative concept that assesses whether the essential nature of a record has changed.

J

JPEG 2000
An image compression method that provides much better image quality at smaller file sizes than the original JPEG format. The JPEG 2000 file format supports both lossless and lossy image compression. A full range of metadata may also be bundled within the file.

K

L

M

MASTER COPY
A faithful digital reproduction of a document, optimized for longevity and for production of a range of delivery versions (derivatives). Masters are captured at the highest practicable quality or resolution and stored for long-term use. They may serve as official file copies, if authorized.

MEGABITS PER SECOND (Mbps)
Refers to file size or the amount of data transferred per second between two points.

METADATA
Metadata is a set of data that describes and gives information about other data. Metadata is structured information about the characteristics of an information resource that helps identify and manage that information resource.
QUALITY ASSURANCE (QA)
Quality assurance (QA) is the process of verifying or determining whether the digitized record meets or exceeds expectations. It involves checking the operation and output of digitization processes against agreed benchmarks to ensure that these benchmarks have been met.

RECORDKEEPING SYSTEM
A recordkeeping system is a shared filing system in which records are captured, protected, retained and destroyed in accordance with approved information schedules. A recordkeeping system, when used in conjunction with recorded policies and procedures, defined roles and responsibilities, and on-going training, constitutes an appropriate system for managing government information.
Glossary

RESOLUTION (DPI and PPI)
The density of graphic information in an image, expressed in dots per inch (dpi) or pixels per inch (ppi). Pixels per inch is a unit of measurement for the quality of a digital capture relating to the resolution of a screen image or scan. Often incorrectly used interchangeably with DPI (dots per inch), a term which more accurately refers to the resolution and image quality of an output such as a print. When referring to on-screen representations and image capture, ppi is the preferred term.

SECURE FILE TRANSFER PROTOCOL (SFTP)
A method of securely transferring files between two remote systems. The advantage of using SFTP is the ability to leverage a secure connection to transfer files.

SOURCE RECORD
A record that has been copied, converted or migrated or will be the input for such a process. A source record may be the original record or it may be a reproduction that was generated by an earlier copying, conversion or migration process.

SSL CERTIFICATE
Secure Sockets Layer (SSL) is a computer protocol that helps secure communications over computer networks, and is most often used with email. An SSL certificate is a digital document that ensures the content provided is from the correct (verified) sender.

SAMPLE RATE
When digitizing audio, sample rate is the number of samples per second taken from a continuous, analog signal to make a discreet, digital signal. Analog audio that has been digitized at a sample rate of 96 kilohertz (kHz) has been sampled 96,000 times per second. A greater sampling rate enables the digitization of higher audio frequencies, resulting in a preservation file that captures more of the qualities of the original recording.
A file format for audio or video, also referred to as a container, serves to package the coded essence and associated metadata together in a specified file format. Types of wrappers include:

- **Audio Video Interleaved Format (AVI)**
  A file format for moving image content that wraps a video bitstream with other data chunks and supports synchronous picture-sound playback. This format is well documented and supported natively in Windows.

- **Material Exchange Format (MXF) OP1A**
  An object-based file format that wraps video, audio and other bitstreams ("essences") and is intended for implementation in devices, ranging from cameras and video recorders to computer systems.
Tools &
Resources

Glossary

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

- **Matroska File Format (MKV)**
  An open source file format that can wrap an unlimited number of video, audio, still picture, and/or subtitle tracks into one single file. There has been a modest adoption in heritage repositories.

- **Quicktime (MOV)**
  A file format that wraps video, audio, and other bitstreams. This file format has been widely adopted in large institutions, video post-production and including many consumer level devices. 8-bit QuickTime uncompressed 4.2.2 is supported by any system that can support the base QuickTime codec package and system/storage requirements to handle large media files.

X Y Z
How to access the Standard

Access the DGI Standard on the IM/IT Standards internet page.

Contact your Records Team for assistance.