Section A: Digitization Overview

Digitization can be performed by scanning records in paper form, or by capturing analogue material with a digital camera and by digitizing media, such as audio or video, using an analogue to digital converter.

By implementing and following sound practice procedures, we are able to:

- Ensure that the digital files we create are of high quality and meet national standards;
- Maintain the integrity and longevity of the digital files;
- Ensure the digital objects in the collection are discoverable and shareable;
- Save time and costs by avoiding duplication of work.

A1: Definition of Key Terms

Access Master Copy – is a working copy of the preservation master and is the source file for all other derivatives.

Checksum - is a file generated from digital data for the purpose of detecting accidental errors that may have been introduced during its transmission or storage. The integrity of the data can be checked at any later time by recomputing the checksum and comparing it with the stored one. If the checksums match, this shows that the file has not been altered (either intentionally or unintentionally).

Collection – is a general term to describe a body of records, and may include documents, photographs, audio/visual material, maps, etc., in both physical and electronic forms.

Content – is information contained in or on a resource that can be copied by traditional copying processes or digitization. For audio/visual material, the content is the signal encoded in the sound recording. For a photograph, it is the image itself and not the medium the image is held on, i.e. paper, glass, or plastic. For a digital photograph, content is the image and embedded metadata.

Data integrity - refers to the trustworthiness of system resources over their entire life cycle.

Derivative/Surrogate Copy/Version – is a copy that has been derived from the access master for purposes of access such as screen viewing, web delivery, printing, thumbnail galleries, etc.

Digital Asset/Resource – is a digital file that is considered to have value. It can be either "born digital" or the result of the digitization of information content held on an analogue medium, e.g. audio tape, film, etc.

Digital Object – Digital data stored in binary format, consisting of a bit-stream and relevant metadata. Digital objects can include text, photographs, audio files, and videos.

Digital Preservation – is the managed activities necessary for ensuring both the long-term maintenance of digital information and the continued accessibility of its contents. To support digital preservation, it is critical to capture administrative, structural and technical data associated with each object (see metadata).

Digitization – is the process of copying analogue material in any form (text, photographs, voice, etc.) to a digital file form using a device such as a scanner, a camera, or any other electronic device.

Electronic document management system (EDMS) - is software that manages the creation, storage, and control of semi-structured documents.

Lossless formats - are file formats that are stable and therefore compatible with long-term preservation efforts. In general, these formats have the following characteristics: openly documented; supported by a range of software platforms; widely adopted; lossless data compression or no compression; non-proprietary; and does not contain embedded files or embedded programs.

Master File - represents the highest quality version/copy possible, and has permanent value and should be managed in an appropriate environment. Once preservation masters are produced and an access copy created, they are not handled. Access masters are the working copies of the preservation master from which all other derivative files are created.

Metadata – Metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource. Metadata is often called data about data or information about information. It includes all cataloguing or indexing information created to locate, describe and manage the preservation of a resource. For example, metadata recorded for a digital image or photograph would include data about the content of the image, the photographer, the date of creation, date(s) of modification, technical information such as resolution, file type, file format, and its relationship with other related files and their locations.

Metadata can be grouped into general categories, including, but not limited to:

> **Administrative metadata** provides information to help manage a resource, such as when and how it was created, file type and other technical information, and who can access it. There are several subsets of administrative data; two that sometimes are listed as separate metadata types are:
>
> i. *Rights management metadata,* which deals with intellectual property rights,
>
> ii. *Preservation metadata,* which contains information needed to archive and preserve a resource.
>
> **Descriptive metadata** describes a resource for purposes such as discovery and identification. It can include elements such as title, abstract, author, and keywords. For digital resources, descriptive metadata is the information used for the indexing, discovery and identification of a resource. Examples include Dublin Core (DC), and Canadian Rules for Archival Description (RAD).
>
> **Structural metadata** indicates how compound objects are put together, for example, how pages are ordered to form chapters. For digital resources, structural metadata is the information used to display and navigate digital resources; information on the internal organization of the digital resource; information on viewer or reader plug-in needed to open the digital resource.

Non-proprietary format - a file format where the mode of presentation of its data is transparent and/or its specification is publicly available. Open formats are ordinarily standards fixed by public authorities or international institutions whose aim is to establish norms for software interoperability. There are cases of open formats promoted by software companies which choose to make the specification of the formats used by their products publicly available.

Preservation – refers to activities undertaken to repair or treat damaged materials, activities undertaken to prevent future damage or degradation of materials, and activities associated with maintaining the content of materials for use.

Preservation Master Copy –is the "archival quality" digital copy of material that is stored securely on a physical format or carrier, e.g. compact disc, DVD, magnetic tape, or digital file format, which is likely to be accessible in the future. It may be duplicated in an emerging physical or digital format, to protect its content and structure over time and can serve as a template for producing derivatives or performing other preservation actions.

Preservation Methods/Strategies – include 3 main methods for preserving digital media: migration, encapsulation, emulation, and software and hardware archiving.

> **Migration** – involves ensuring that the digital information is re-encoded in new formats before old formats become obsolete.
>
> **Emulation** – involves programming computers to emulate older, obsolete computer platforms and operating systems.
>
> **Software and Hardware Archiving** – involves preserving the original software and hardware that was used to create the information so that it can be accessed in the future.

Proprietary format - a file format where the mode of presentation of its data is opaque and its specification is not publicly available. Proprietary formats are developed by software companies in order to encode data produced by their applications: only the software produced by a company who owns the specification of a file format will be able to read correctly and completely the data contained in this file. Proprietary formats can be further protected through the use of *patents* and the owner of the patent can ask royalties for the use or implementation of the formats in third-party's software.

A2: Purpose of Digitization

It is necessary to determine the purpose of digitization and the intended use of the digital items (i.e. safeguard access through digital management, preserve the original materials through access of digital copies, improve usability of material through digital manipulation, etc.).

Unless the material at hand is unique and/or a master copy is required (i.e. hand written margin notes in the book or manuscript that are of value, etc.), the material may not need digitization if an acceptable digital copy can be found elsewhere.

A3: Digitization Policies

A clear policy should be a key component of a digitization project. Some elements to consider when creating your policy include:

- Description of digitization activities and the management of digitization projects;
- Description of digital objects, file types and file formats for both preservation and access;
- Description of metadata schema(s);
- Definition of essential characteristics of the original (i.e. curatorial, archival, technical);
- Description of approach and quality levels for digitization.

A4: DIGITIZATION POLICY TEMPLATE

Place Logo Here

_____ Nation

Name of Policy: Digitization (Version ##)

INTRODUCTION

Countless First Nations materials are at risk of being lost if not properly stored. Storage space is often at a premium. Digitization can facilitate access to and preservation of these invaluable materials. Digitization can be complex and costly but the long term benefits can outweigh initial investments.

PURPOSE

The purpose of this policy is to provide an overall framework for the _____ Nation and its member communities for initiating and carrying out digitization projects.

This policy aims to ensure ongoing access to the content of _____ Nation records, regardless of the physical media or digital file format on which they were originally created or acquired. The policy provides a foundation for achieving best practice outcomes when undertaking digital preservation and digitization activities. These outcomes include:

- Prioritizing digitization and digital preservation according to _____ Nation and community requirements;
- Minimizing the risk of information loss;
- Meeting international standards for digitization and preservation;
- Capturing metadata required for ongoing preservation, access and rights management of digital assets;
- Meeting legal obligations, such as copyright, privacy, and intellectual property;
- Ensuring digitization work is compatible with ongoing preservation requirements.

SCOPE

This policy encompasses all media types generated by _____ Nation which may include but is not limited to:

- Manuscripts and printed text
- Photographs
- Film, negatives, and slides
- Graphic art
- Maps
- Audio recordings
- Video recordings

POLICY STATEMENT

The policy is intended to ensure the preservation of and access to all _____ Nation materials according to _____ Nation and community requirements.

RESPONSIBILITY

The overall responsibility of information management lies with the _____ Nation Administrator. Digitization projects require a combination of skills from a variety of staff with different areas of expertise. Individual digitization projects (either Grant-Based or Ongoing) require a team approach, may be initiated and managed by any _____ Nation employee. Each project should have a project manager, who is responsible for coordinating and planning the activities associated with the digitization project.

ASSOCIATED DOCUMENTS

Associated with this policy is a Digitization Toolkit that includes information on

- Project Planning
- Best Practices
- Metadata and Description
- Additional Resources

IMPLEMENTATION

State here how the Policy is to be implemented, or rolled out, throughout the _____ Nation staff. This implementation could take the form of:

- Information Sessions, and/or;
- Training Sessions, and/or;
- Documentation distribution (e.g. posters, brochures) and/or;
- An Announcement Notice to alert the staff of the approved Policy, and/or;
- Other means.

Authorized by: [_____] Original Issue: [DD/MM/YYYY]

Staff position or body responsible for authorizing the policy *Date of the original authorization and issue of the policy*

Document Owner: [_____] Current Version: [DD/MM/YYYY]

Staff position responsible for keeping the policy up to date and error free *Date of the most recent amendment to the policy*

Review Date: [DD/MM/YYYY]

State here next date of review [3 - 5 years]

NOTE

In developing policy, it is prudent to put a DRAFT watermark on any policy that is not approved. The *Issue Date* at the bottom of this document is the date of the APPROVAL of the Policy and the *Current Version Date* is the date of the latest amendment; these fields should be left blank when in draft stage. The document title should contain the version of the draft that you are working on.

A5: Digitization Principles

A number of general principles have been identified to help provide a solid foundation for all digitization projects:

- Collaborate across the organization and with external partners, where possible;
- Digitize at the highest resolution appropriate to the nature of the source material;
- Digitize at an appropriate level of quality to avoid re-digitizing and re-handling of the originals in the future;
- Digitize an original or first generation (i.e. negative rather than print) of the source material to achieve the best quality reproduction possible;
- Create and store a master file that can be used to produce surrogate files in accordance with the requirements of current and future users;
- Ensure originals are maintained, as digital copies are not substitutes for originals, although some original analogue media may not remain viable in which case the digital master will become the best available copy and should be treated as the "original";
- For master files, use lossless file formats whenever practical;
- Use components that are not specifically dependent on proprietary products and formats;
- Use file formats and compression techniques that conform to existing cultural heritage standards;
- Create backup copies of all files on servers and have an off-site backup strategy;
- Create meaningful metadata for files and collections;
- Store digital files in an appropriate electronic document management environment (e.g. external hard drives, servers, etc.)
- Monitor data as necessary (e.g. data integrity, checksums, etc.)
- Document a migration strategy for transferring data across generations of technology;
- Plan for future technological developments.