

# NATIONAL FILM BOARD OF CANADA DIGITIZATION PLAN – A CASE STUDY

**Julie Dutrisac, Eng.**

**Luisa Frate, CA**

**Christian Ruel**

National Film Board of Canada

3155 Côte-de-Liesse, Montreal, Quebec, Canada

The digital revolution has completely changed how audiences use and interact with audiovisual media. The National Film Board of Canada (NFB) has been preparing for this inevitable revolution for several years now, developing partnerships and carrying out research on image and sound processing, innovative transfer techniques, accessibility and distribution to facilitate the transition to digital technology. In recent years, the technical and operational infrastructure has undergone significant upheavals. And the NFB has taken up the gauntlet. The new digital reality is much more sophisticated than past technologies, but also much more open and promising. Our aim is to exploit the full potential of these new and constantly evolving technologies.

## 1. INTRODUCTION

In this context, the accessibility of the works the NFB produces and distributes is a major priority as well as part of our mandate.

The institution serves Canadians by making its rich collection and productions available to them when and where they want and on the platform of their choice. The NFB's digitization plan is an important step in achieving those goals.

One of the NFB's main objectives is worldwide on-line accessibility of its extensive collection, either through excerpts or full-length streams. The NFB must be able to offer a range of formats and platforms, from D- or E-Cinema in movie theatres to HD television and Internet broadcasts, and from downloads on different platforms – including mobile devices like the iPhone and the iPad – to traditional DVD or Blu-ray discs for home viewing.

The objectives of the NFB's digitization plan are to improve current and future accessibility of NFB works in digital formats, store and preserve the NFB's works on new media, and restore works that have deteriorated.

The NFB digitization plan applies primarily to finished audiovisual works. This collection, dating back to

1939, is made up of 12,963 titles on film and video in a wide range of formats, from 16mm, 16mm and 35mm to 70mm, and in a variety of magnetic, optical and digital sound formats.

Most of the NFB collection has been produced in multi-language versions (usually in French and English), and since early 1990, all films have been closed-captioned.

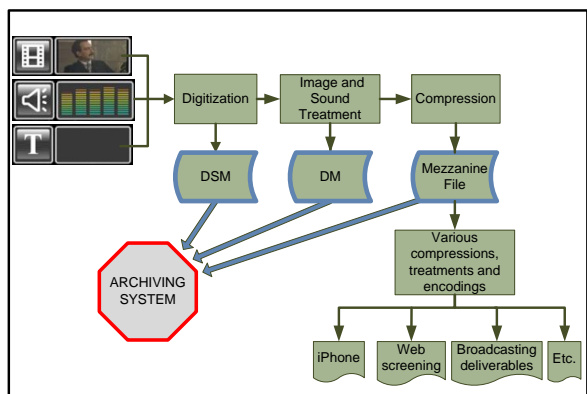
Faced with the challenge of digitizing its huge collection, the NFB has created workflows based on the type, physical condition and variety of source material to be processed for a given work. In addition, the technology current when the work was produced must be taken into account so that as much relevant information as possible is captured, all the while respecting the creative choices made at the time.

The NFB's plan to digitize its collection requires innovative workflows. The purpose of the workflows is to process the largest number of works possible. Specific workflows will be established to handle exceptions and other difficult cases. This approach should ensure efficiency, but also allow restoration in a way that respects the originals.

## 2. DIGITIZATION PLAN VISION AND TECHNICAL CONCEPTS

Our vision of the digitization and distribution of the NFB's collection may be summarized as follows: for each work in the collection, a Digital Source Master (DSM) will be created to preserve the work. Each DSM will be made up of its individual component parts in an uncompressed format: the image, sound, metadata and effects. Every segment of an NFB work (all image and sound segments, titles in all existing languages, subtitles, credits in all existing languages, closed-captioning files, etc.), found in every one of its versions, will be digitized and processed only once at a sufficiently high resolution to allow delivery in all of our distribution and access formats as well as to create a digital master for digital preservation. Damaged or deteriorated works can be digitized at a higher resolution (4K or 6K) to ensure that they can be

restored and preserved. The assembly rules followed in producing each version will be saved along with all other data that will help us understand the processing performed on each component. This content becomes the work's Digital Source Master (DSM), and all of these components will be archived. It should be noted that this practice of digitizing and processing each work's individual parts only once will make the whole process more efficient and flexible.



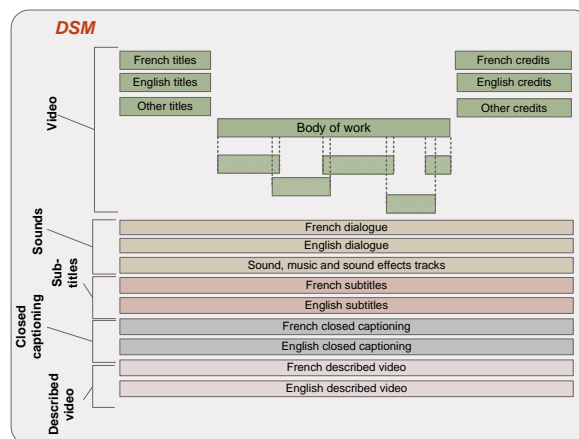
**Figure 1.** Major elements of the digitization plan.

Another advantage to this approach is that it reduces the volume of data archived for each work, particularly since image data files are much larger than their sound data counterparts. Furthermore, for preservation purposes, this approach allows us to select the best original source for each component. Thus, for works on film, an assessment of all available sources in each version and every format that was produced (negative, interpositive, internegative) will help determine the best source material for each segment digitization.

Producing appropriate descriptive and technical metadata for all audiovisual material and all processing involved in the material's life cycle is both attractive and essential in order to pursue fully file-based production and distribution, as well as for automation of workflows.

Metadata then needs to be closely attached to the media essence and embedded in the headers of the various files. Metadata must also be ingested into the archival system database so it can be easily searched, read and manipulated.

Metadata is the key to producing multiple versions of works from their many digital components, because it provides us with a better understanding of a wide range of parameters and characteristics.



**Figure 2.** Digital Source Master (DSM) approach.

Once the digitization and metadata capture have been performed and checked, the completed DSM and its metadata for the work are archived. Integrity of the content is checked when accessed or ingested into the archival system database using MD5 checksums. To ensure the security of our collection, data replication is achieved using rules defining the number of duplicates to make on different media and their storage locations. These rules, based on our collection preservation strategy, rely on the state and availability of the original source elements, the availability of the playback equipment and the heritage value of a title.

At this stage, the uncompressed DSM components are processed in order to return the work to a state as close as possible to the original and to recreate the various versions of the work. Processing includes restoration, colour calibration and component synchronization. This processing will result in an uncompressed Digital Master (DM) that will be archived with all of the metadata collected at each processing step. The uncompressed and unaltered DSM will be kept for future restoring and further processing improvements as digital-image and sound-processing technology evolves.

The archiving process of the DSM generates over 400,000 metadata entries associated with 1.5 TB of content and 100,000 files for an hour-length work, with a very simple version in 2K resolution. The same figures are used for archiving the DM.

### 3. CREATION OF MEZZANINE FILE AND ACCESSIBILITY DELIVERABLES

At this stage, the finished, uncompressed DM contains all the unassembled segments and the various assembly lists allowing the creation of all the versions of the

work. Before being archived, it will be run through a two-phase quality control check. The first phase is a fully automated process of planned and systematic verification and validation operations to ensure that all elements are up to established standards.

In the second phase of the quality control workflow, all files (DPX, Broadcast Wave, subtitle text, etc.) of the various segments are ingested with the various versions' assembly lists, and a complete viewing of the work and its various versions is done to confirm and validate all choices made. Upon approval of the DM, the unassembled segments of media essence and metadata are automatically wrapped as an MXF-AS02 mezzanine bundle. Automatic processes are also initiated to archive the DM and to archive the mezzanine file.

The use of the MXF standard with the AS02 specification for the mezzanine files offers the possibility and flexibility to manage multiple versions of a work within a single mezzanine file bundle without creating multiple copies of the media essence. The MXF file format is widely used within the industry, and supports multiple video compression standards and the management of multiple segments and their reassembly.

This compressed MXF-AS02 digital bundle, generated from the DM, is used to create the NFB's main production and distribution deliverables. Its degree of compression is determined by the types of deliverables to be produced. The mezzanine will meet our primary deliverable needs. On demand, the required deliverable will be created from the mezzanine file MXF-AS02-wrapped content by rendering the specific simple version in an MXF-AS03 delivery format.

We plan on using the mezzanine file for Internet, DVD (traditional and Blu-ray), download, mobile platform and television deliverables. For some types of deliverables, it will be necessary to return to the DM, particularly for D-Cinema or transfers back to film. Rules will be put in place to manage files based on their usage history. In addition to being adopted by the industry's major players, this approach will also meet our current and future accessibility requirements.

#### **4. METADATA AND ARCHIVING**

Long-term retention and management of digital assets is another key requirement for this digitization plan, which must contend with exponential growth in data volumes, long-term retention requirements and on-

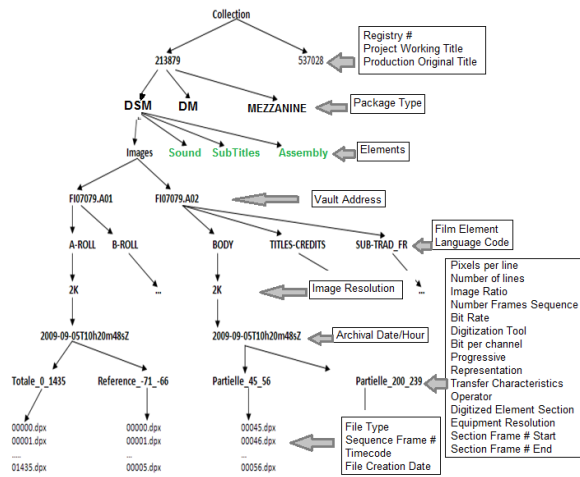
demand-deliverables mandates on a per-customer basis. When content is archived, its specific parameters must be recorded to ensure proper categorization and allow effective and efficient search and analysis.

A robust, reliable archiving solution is a key component of our media operations for the collection, preservation and distribution of our digital assets. The archiving solution needs to ensure long-term retention of the digital collection and efficient management of the exponential growth in data volumes from all new productions. It also needs to provide efficient tools for the complete asset life and allow efficient search, retrieval and cataloging functionalities along with scalability to handle the growing data volumes required to support the digitization plan and each year's new productions.

One challenge is to change the organizational structure of information to adapt it to the new opportunities and flexibilities possible with the digital shift. Notably, we have refashioned the structure to link all information to a work in all of its possible versions instead of the usual way of addressing single titles of a work. This approach helped to achieve the digitization plan's vision of creating a single DSM per work and a seamless integration of the archiving solution with our existing asset management, which stores all descriptive metadata of each work and its specific versions.

For each digitization and processing activity, the maximum amount of metadata, within reason, is collected. All descriptive metadata resides in the asset management database and all administrative metadata (i.e., of technical and digital provenance) is kept within the archival database system, with some parts being embedded in the various files.

In order to enable the organization of our digital assets in a hierarchy tailored to our workflows, it was important that the archival solution not dictate a specific archive structure. The metadata associated with the digitization of a film was implemented from the Digital Picture Exchange (DPX) file format ANSI/SMPTE standard (268M-2003). The data model implemented is presented in figure 3.



**Figure 3.** Film digitization data model.

For the audio workflow, metadata related to the DSM is collected for the source audio object at the recording level describing the technical attributes of all physical audio sources (physical characteristics of the original material, format of the source and specific information on transfer and creation). In each activity related to the processing of the sound components of a work, from recording to finalizing the DM, the process history metadata describing all choices and processes of the activity is generated. The technical metadata and digital provenance (process history) metadata implementation use the Audio Engineering Society's AES-X098B: (D098B) Audio object structures for preservation and restoration, and AES-X098C: (D098C) Administrative metadata for audio objects – Process history schema, encapsulated in a METS (Metadata Encoding & Transmission Standard) schema. A limitative amount of sound metadata is also included in the Broadcast Wave file (in the Bext-Chunk). Since this metadata is insufficient for sound preservation, it serves for sound data exchange for automation and information exchange between applications.

Given the huge volume of files and metadata generated, an XML plug-in to enable the full automation of the ingest capability according to the established structure was developed. It allows data and metadata to be validated before ingest and stored in a specific level in the hierarchy. Search results are thus more efficient and can be returned at a very precise level. This helps to optimize work processes and increases flexibility, productivity and efficiency.

Digitization generates a huge amount of data that must not only be saved and archived, but must also be available for reuse in order to improve the accessibility

of our collection. This solution also manages archiving media based on their use and life cycle, and will allow orderly and possibly automated data migration.

## 5. STRATEGIES

To be successful, the NFB's mass digitization project requires innovation, a review of current procedures, new work methods and possibilities, and appropriate processing choices based on the content and source medium of each work.

Process automation is one of the keys to success. Although some processes and choices require, and will continue to require, a technician's involvement and manual operation, a large portion of this work will be automated. Automation helps to optimize work processes through formalization and standardization. Automation is implemented progressively with the digitization plan and improves efficiency, particularly in information sharing, data integrity and processing throughput. New ways and methods are explored and implemented in the various workflows, especially for automating colour grading, sound assembly working from sound-signature recognition, image and sound restoration, and certain aspects of quality control.

For example, the implementation of quality assurance processes in various phases of workflows ensures planned and systematic controls for verifying and maintaining levels of quality according to objectives, and for validating operations to ensure that all elements are up to our standards and correspond to the metadata. The reports generated by these processes serve as performance indicator measurements.

A quality control workflow involving a complete viewing of the work to confirm and validate all choices is necessary. The integration of digital technology and the implementation of automation will transform our traditional approach in production and distribution and change the way the quality controls interact with the rest of the production chain.

The goal of the various ongoing automation developments is to organize a series of methods and technics to achieve efficiencies, to have employees work on creative initiatives, and to bring the finalized DM work directly to the quality control level. Only at this point will a human resource validate the automation and, if necessary, bring back the work to manual correction using a more traditional approach. This approach will not be entirely feasible for all

works, but it will provide a more efficient way of processing the huge amount of data.

To bring these strategies to fruition, the NFB will implement a number of innovative workflows. The purpose of the workflows is to process the largest number of works possible. Specific workflows will be established to handle exceptions and other difficult cases. This approach looks promising and will enable us to restore works in a way that respects the originals.

## **6. DIGITIZATION WORKFLOWS**

The digitization plan workflow model allows us to represent future optimal workflows with current procedures. These workflow models are organic and will serve as a baseline for work, training, assessment, refinement and continual improvement. Over time, processes will be adjusted, optimized and evaluated for their potential to be automated.

We also detail several underlying procedures to these workflows that support collection digitization, processing and accessibility. When developing workflows, the methodology used is:

- Investigate industry standards, analyze their use in similar contexts and select which to use;
- Analyze all information on the selected standards in order to determine relevant metadata and information structure;
- Analyze and understand existing processes and the limits and potential of current technology in order to handle mass digitization of the collection;
- Check digitization and processing tools to ensure they efficiently capture all metadata connected with the choices made;
- Implement work processes that will allow us to preserve the work's component parts in their current state, but also allow us to reprocess them in the future if necessary;
- Define and implement a data structure for each process with the goal of optimizing our knowledge of the digitized content and search capabilities;
- Design applications that automate processing, perform quality checks and ensure operational efficiency while limiting manual errors;
- Analyze and define the links between work processes and their interoperability.

To facilitate interoperability between systems, technologies and partners, and to ensure efficient access to data in the future, we have selected open, non-proprietary standards that are well established and accepted in the industry. We will install ready-to-use, commercially available infrastructure that will be customized to our specific needs.

The digitization plan opens new opportunities in terms of innovative workflows, advanced technologies, infrastructure modernization, resource training, organization and culture change, and process review. All these developments will progress into all areas of production and distribution.

## **7. IMPLEMENTATION OF DIGITIZATION PLAN**

The integration of digital technology that has been underway for several years now is transforming our production and distribution chains and giving us an opportunity to implement a dynamic digitization and accessibility solution. End-to-end integration of a fully digital production chain also gives us more flexibility. With this flexibility, we can automate workflows when appropriate and modify them as technological capabilities and business needs change.

Over the past five years, the NFB has made more than 6,000 titles accessible in a variety of digital and encoding formats. The NFB has always responded to the accessibility needs of its various clients, and its digitization plan is no different in this regard.

## **8. TECHNOLOGICAL CHANGE IN YEARS TO COME**

Digital screening and audiovisual technologies are revolutionizing the film, entertainment and education industries. We are seeing an explosion of new devices that enable consumers to access the content of their choice. For content distributors, the array of formats that must be prepared for each distribution channel is increasing, while the time available to control each version is diminishing. We must therefore seek more efficient methods to meet this demand.

The choice of using a mezzanine file with metadata and files containing assembly information will enable us to efficiently meet all current and future demand. No file specific to a given distribution channel will be kept, since it can be recreated on request from the mezzanine file. We are confident that this strategy will reduce the volume of data archived for each work and

provide an efficient method for automatically generating any new deliverable.

## 9. REFERENCES

1. Advanced Media Workflow Association, [aafassociation.org](http://aafassociation.org).
2. AES-X098B: (D098B), "Audio object structures for preservation and restoration", [aes.org](http://aes.org).
3. AES-X098C: (D098C), "Administrative metadata for audio objects - Process history schema", [aes.org](http://aes.org).
4. ANSI/SMPTE 268M-2003, Digital Picture Exchange (DPX), [smpte.org](http://smpte.org).
5. Casey M., Gordon B. *Sound Directions - Best Practices for Audio Preservation*, Indiana University and Harvard University, 2007, <http://www.dlib.indiana.edu/projects/sounddirections/bestpractices2007/>.
6. Devlin, B., Wilkinson, J., *The MXF Book*, Focal Press, Burlington, MA, 2006.
7. EBU Recommendation R111-2007, "Multichannel use of the BWF audio file format (MBWF)", [ebu.ch](http://ebu.ch).
8. EBU – TECH 3306, « MBWF / RF64: An extended File Format for Audio », 2009, [ebu.ch](http://ebu.ch).
9. EBU Tech 3285, « BWF - a format for audio data files in broadcasting », 2001, [ebu.ch](http://ebu.ch).
10. EBU Tech 3285-s1, "EBU Tech 3285 Supplement 1: BWF - MPEG Audio", 1997, [ebu.ch](http://ebu.ch).
11. EBU Tech 3285-s2, "EBU Tech 3285 Supplement 2: BWF - Capturing Report", 2001, [ebu.ch](http://ebu.ch).
12. EBU Tech 3285-s3, "EBU Tech 3285 Supplement 3: BWF - Peak Envelope chunk", 2001, [ebu.ch](http://ebu.ch).
13. EBU Tech 3285-s4, "EBU Tech 3285 Supplement 4: BWF - link chunk", 2003, [ebu.ch](http://ebu.ch).
14. EBU Tech 3285-s5, "EBU Tech 3285 Supplement 5: BWF - chunk", 2003, [ebu.ch](http://ebu.ch).
15. EBU Tech 3285-s6, "EBU Tech 3285 Supplement 6: BWF - Dolby Metadata", 2009, [ebu.ch](http://ebu.ch).
16. METS - Metadata Encoding & Transmission Standard, [www.loc.gov/standards/mets](http://www.loc.gov/standards/mets).
17. SMPTE 277-1-2009, "Material Exchange Format (MXF) – File Format Specification", [smpte.org](http://smpte.org).