https://doi.org/10.2352/issn.2168-3204.2018.1.0.29

Development for Audio-Visual Archiving System of The National Archives of Korea: A Case Study

Ji-Yong Lee, Sang-Kook Kim; National Archives of Korea; Seongnam, Gyeonggi-do, Republic of Korea

Abstract

The National Archives of Korea (NAK) has developed and is currently operating "Audio-visual Archiving System" to ensure easy accessibility of users to the records through digital conversion of analogue type audiovisual records and more systematic management of records using a system.

This paper covers the whole process of the system development and the actual operation of the system by describing the actual cases of records management such as building DB through digital conversion of audio-visual records, and user´s search and utilization.

Specifically, this paper contains the background of the development of the audiovisual records management system, the process of development, the whole process of the system, introduction of major management functions, introduction of automation functions for efficient work processing, linkage with existing systems and migration.

It is hoped that it will be used as a case study in the course of introduction and operation of audiovisual record management system in the future by presenting effect analysis obtained by operating the audiovisual records management system and suggesting constructive directions for management and utilization of audiovisual records.

Ⅰ. Introduction

Audiovisual documents are works comprising reproducible images and sounds embodied in a carrier whose. Recording, transmission and perception require a technological device. Visual and/or audio content has linear duration. By nature, moving images and sound recordings cannot be perceived instantly, no matter how short or in what context.[1]

Digital audiovisual records have different features from text-type electronic records. One of the features is variety of format. Digital audiovisual records are produced in various formats according to filming methods, equipment, storage devices and editing programs and so on. Therefore, a function to handle these various formats is required to ingest and manage digital audiovisual records. Another feature is that they are produced in large volume. It is said that making a single digital film generates approximately 60-2000 TB (terabyte) or more of born digital files. Accordingly, audiovisual records require an economic management method considering the exponential growth in required capacity and the annual cost of maintaining the digital data.

As the archive environment changed with shift from analog to digital records and increase in volume of digital audiovisual records, the need to systematically manage the massive audiovisual records arose. And the NAK worked to improve preservation and restoration system of massive volume of audiovisual records as an effort to systematically manage important audiovisual records. It started introducing required facilities in 2010 and it expanded facilities and added functions step by step from 2011 to 2016.

This paper will introduce process of system development and main functions of the archive management system established by the NAK. It will also contain the considerations in system development, additional functions developed to improve work environment in the NAK, linkage to the existing system, automated functions developed to improve the system more efficiently. It is expected that this paper will serve as a case study in introduction and operation of an archive system for audiovisual records by presenting a better direction to ingest, preserve and use audiovisual records.

Ⅱ. Main Body

Ⅱ-1. System Development Background

Since audiovisual records started to be digitized in 1997, all digital audiovisual records have been ingested and managed in CAMS (Central Archives Managements System). The CAMS, developed in 2007. is still in use by the NAK. The archive management system at the central records management agency performs long-term safeguarding of various types of records that are automatically transferred from repositories.

But, the CAMS has difficulty to handle special features of digital audiovisual records because it was designed to mainly mange not only audiovisual records but also text-type records. For example, it was able to ingest limited file formats of video, image and audio records and its viewer was able to support only limited formats. (Formats like mpeg for video, jpeg for image, mp3 for audio were only possible)

As article 19 (creation of audiovisual records) of Enforcement Decree of Public Records Management Act was added in May 2010, it laid a legal ground to mandatorily produce a huge amount of audiovisual records of broadcasting and film recordings, and national events and projects worth preserving nationally and permanently. It meant the collection of a massive number of audiovisual recordings. The MAM (Media Assets Management System) was introduced in 2010 to prepare the massive collection of audiovisual records and to properly preserve, store and manage them.

1 Ray Edmondson Director, Archive Associates Pty Ltd, "AUDIOVISUAL ARCHIVING : A VIEW FROM AUSTRALIA"

The MAM was a commercial system that was used to manage media files in broadcasting companies. It resolved some of problems that the existing CAMS had. Details are as follows.

. Ingesting and storing large sized files – size of records is not a problem in the MAM.

. Multi-ingest of large quantities of records

. Handling various formats and generating formats for preservation and access files simultaneously

. Cataloging function to extract key images of video records

. Multi-view function to see ingested image and video records at a glance

. Ingest and management function of audio, image and mixed records

These features were the important points that were considered in the process of development of audiovisual record archiving system.

The digital audiovisual records stored in the CAMS from the beginning of digital conversion to the introduction of the MAM system have the following features from the technical point of view.

. Compression method applied to digital conversion process

. Encoding video records to a low-resolution SD level

. Using a system mainly targeted to manage non-electronic paper records

In the time before the CAMS was introduced and after the CAMS was used actively to ingest and manage records in 2007, SD was mainstream of broadcasting technologies. HD technology that appeared later has a feature of creating a large volume file in digital conversion. Therefore, the storage method of SD resolution applied to CAMS was the best option at that time. But there was a problem that rework was required to convert important records with historical value, which were stored in the CAMS after digital conversion to SD resolution, into a high-resolution files equivalent to original quality. It is because physical media such as films and tapes not only degrade over time, but also make it difficult to handle born files as devices which they depend become obsolete.

Therefore, it was needed to solve the problem of integrating the dual programs, the CAMS and the MAM system, into one.

Ⅱ-2. System Development Process

The first goal that the NAK had when it established the audio-visual archiving system was to systematically and safely preserve, use and manage large volume files of audiovisual records. And it was to preserve and use important national audiovisual records in a safe way by preserving and managing digitized files of analog type records and building DB of audiovisual records.

The NAK implemented a project to enhance a preservation and restoration system for audiovisual archives to systematically preserve and manage nationally important audiovisual records. It started introducing basic facilities in 2010 and expanded facilities and added functions gradually until 2016. The followings are the step-by-step establishment and extension progress of the system

<Table 1> Step-by-step establishment and extension poregress of system

Division		Details
System Establishm	2010	Basic project : Construction of the MAM and securing storage (50TB)
ent (Infrastructure)	2011	Extension (1st) : Extension of servers and storage increase (200TB)
	2014	Introduction of archive : Tape library (500TB), research on automated quality check function
	2016	Extension (2nd) : Tape library (3PB), dual server, development of link function between CAMS and MAM
System Functions (Requirements)	Commercialized S/W default functions	° Automated generation of preservation file and access file at the same time when a born digital file is ingested ° Download upon authorization ° Cataloging of video files
	Additional requirements	° Batch file ingest, modification and delete function ° Collective ingest of video data and metadata simultaneously ° ° Quality check of collectively ingested files ° Application of quality standard for preservation and access ° Backup for long term preservation format ° Dual system management → Integration

Ⅱ-2-1. System Configuration Principles

The considerations to build a digital audio-visual archiving system are that it should be able to effectively process high-resolution video data, and to make users more easily access and use large-sized video data, and to preserve and manage a large amount of large volume data in a safe way. These considerations are reflected in the following system configuration.

<Table 2> System Configuration Environment

Division		Details
HW	Ingest PC	CPU : 3.40GHz Memory : 4GB OS : Win 7 64bit
	CMS Server	CPU : E5-2630 v3 2.40GHz Memory : 16 GB OS : Windows Server 2012 R2 64bit
	Transcoder Server	
	Cataloger Server	
	Streaming Server	CPU : E5540 2.53GHz Memory : 6 GB OS : Windows Server 2012 R2 64bit
	Web Server	CPU : E5640 2.67GHz

	Archive Server	Memory : 6 GB OS : Windows Server 2012 R2 64bit
	Search Server	CPU : E5640 2.67GHz Memory : 6 GB OS : Windows Server 2012 R2 64bit
	INF Server	CPU : E5506 2.13GHz Memory : 4 GB OS : Windows Server 2012 R2 64bit
	Storage	8TB × 64ea(512TB)
SW	DBMS	Oracle 10G Std Windows 2012 64bit C ++
	O/S	
	Programing language	

Clear principles for system construction were required as follows, to build a system that enable systematic management of digital video data and user's easy access and use, based on the system environment shown above.

First, it should be able to store and mange transferred born digital data in a safe way. It means that it should be able to handle systematically various formats of audiovisual records, store transferred born files in a way of reflecting their own features, and maintain the quality of the files.

Second, it should enable effective process and management of a large amount of large volume audiovisual data. It means that it should be able to handle a large-sized data like video files rather than text or image files by system itself.

Third, it needs to enhance convenience so that users can easily access information service. It means that the system needs to be built in a way that users can easily and effectively search and use information.

These three principles were considered in building the system to manage audiovisual records. They were partly reflected by introducing the commercial MAM system.

Ⅱ-2-2. System Requirement Analysis

We added functions necessary for the work environment of the NAK with introduction of a commercial system. First we identified the requirements for effective archiving work such as a function to collectively ingest a large amount of audiovisual data and their metadata transferred in a file format, a function to automatically generate preservation files and access files according to quality standard when ingesting born files, and a function to check file quality and damages upon ingest. We made efforts to reflect these requirement to the system over years.

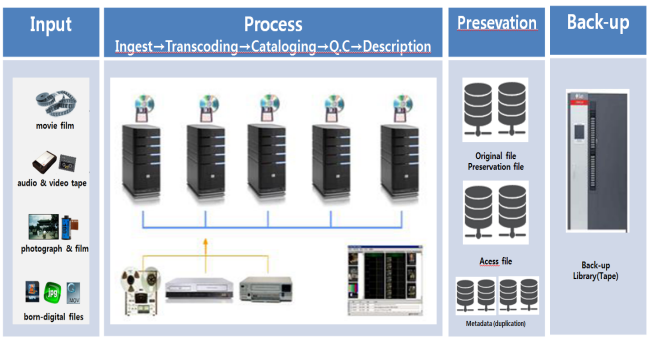
<Table 3> System Development Progress

Division		Details
Functions	2012	Physical check function
	2014	Backup function
	2016	CAMS-MAM linking function
	2017	SSO function, CAMS-MAM migration tool, inquiry function in CAMS

Ⅱ-3. Process of NAK's audiovisual archiving System

Figure 1 shows the process of audiovisual archiving system of the NAK. The system has a series of process from collection, preservation treatment, file conversion, ingest, format conversion to preservation and access files, content description, quality check and to backup. Main functions consist of ingest, format conversion, scene analysis, extraction, quality check and backup.

<Figure 1> Concept of Audiovisual Archiving System of the NAK.



Ⅱ-3-1. Storage Construction

This system has a storage capacity of 605TB to store and manage audio-visual records data. The storage allocates 100TB for a temporary storage to ingest collected files, and 50TB for born file storage, and another 50TB for digitization of analog type records. High-resolution digital files of movie films, especially, are stored here temporarily. It also allocates 200TB for digitally restored files of damaged data, and 5TB for DB table storage.

The system ingests born files (video, image, and audio files) and digitized files. Preservation and access files that are generated automatically together with ingest files are stored in 300TB storage. This storage(300TB) is logically divided into 150TB each. Ingest files and preservation files are stored in a 150TB storage for 15 days. After 15 days, the files are backed up into tape. Access files are stored in the other 150TB storage permanently.

Ⅱ-3-2. Development of Key Management Functions

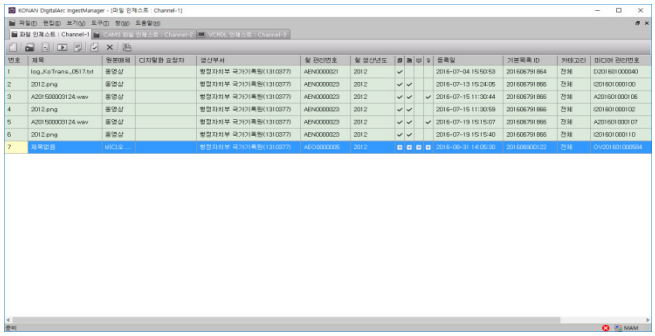
<A. Ingest Function>

The lexical meaning of Ingest is 'input', and here it means inputting audio-visual records data into a system. In the audio-visual archiving system, ingest means a process of receiving SDI signals of digital or analog type audiovisual sources and storing them in the central storage through an ingest server, and registering their related information to CMS (Contents Management System). In short, Ingest plays a role of registering collected images and videos to the archiving system.

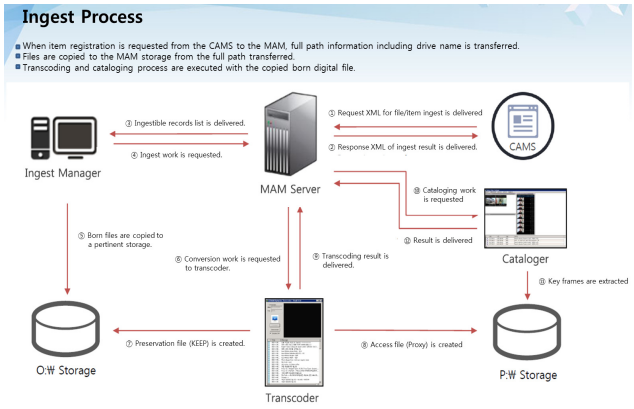
Ingest configurations are in operation using 4 ingest servers. The person in charge ingests video, image, audio and photograph files using Ingest Manager. Ingest Manager provides a function to ingest collected audiovisual data into storages. Video, audio and photograph files are ingested through File Ingest Channel. Analog tapes are ingested through VCRDL Ingest Channel. Batch ingest

(collective registration of a large number of files and metadata) is processed through CAMS Ingest Channel. There are 6 Ingest channels in operation. Each ingest server operates File Ingest, VCR Ingest and CAMS Ingest. Figure 2 and 3 show the snapshot and detail process of CAMS Ingest.

<Figure 2> Ingest Channel View



<Figure 3> Ingest Process

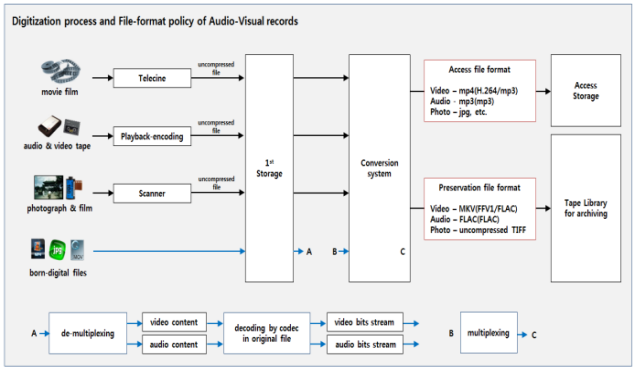


< *B. Transcoding (Format conversion)*>

Transcoding means a format conversion process to generate preservation format (mkv) and access format (MPEG-4) from high-quality and high-resolution digital born files. The system automatically generates preservation files of the same resolution and access files at the same time as the born file is ingested. Through this function, users can easily search and use a small-sized format of a high-volume born file using an ordinary desktop computer.

The NAK is using guidelines for audiovisual records digitization as its organizational standard after setting out the guidelines as a standard task in 2014 to resolve the issue of long-term preservation and use of audiovisual records both in the physical media and in the born digital format. The purpose of digitization is to prepare the loss of the original audiovisual records or obsoleteness of applied technologies and to make it possible reproducing the original records at the same level of quality in the future and use the records for public service. In case original records are lost or no longer possible to access, digital version of the records need to replace them. To prepare the situation, the contents and properties of original records should turn digital without missing as much as possible. In this regard, the NAK reviewed digitization standards from the perspective of preservation and utilization. Figure 4 shows the digitization process and file format policy of the NAK.

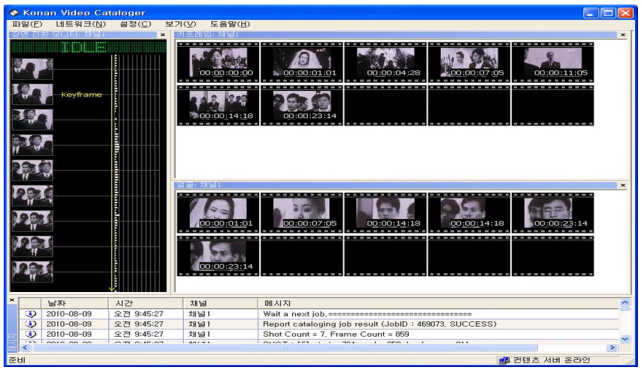
<Figure 4>. Digitization Process and File-format policy of Audio-Visual Records



< *C. Cataloging(Scene Extraction)*>

Cataloging means a process of making storyboard by extracting key information such as scene change, key frames and characters by analyzing access files. Storyboard means showing sketch or photos of key contents of an original sequence. For scene change, it comprehensively considers how similar the images between neighboring frames, color information between two images, shape information and texture information. The storyboard made in this way provides a better way for users to search and access to what they want as it helps to identify the whole contents of a video data.

<Figure 5> Operation of Transcoder



<Figure 6> Storyboard Screenshot

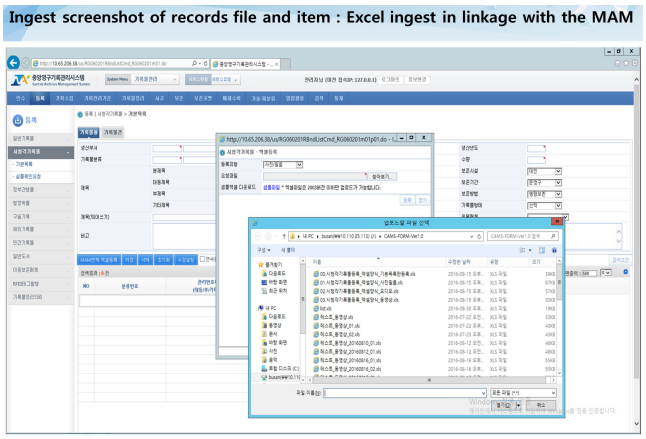
Ⅱ-4. Automation for Efficient Process – Batch Ingest, Quality Check, Backup

This chapter introduces automated functions for better efficiency in the digital audio-visual archiving system.

A. Batch Ingest

As massive digital audiovisual files were collected in earnest, the NAK had to think about ways to easily ingest large number of audiovisual files collectively from the collection stage. And we came up with a work process to ingest multiple files and metadata at the same time by production organizations or collection units. Such batch ingest function was additionally developed when the CAMS and MAM were linked to each other. The following is the screenshot of batch ingest. If the location information (file path) of files is registered in the list information, the files are also ingested at the same time as the corresponding items are created.

<Figure 7>. Batch Ingest Screenshot, Batch Ingest Form



B. Quality Check

The NAK conducted a research and development project (R&D title: development of a quality inspection tool for audiovisual records) for two years from 2014 to 2015. The quality checking system developed in 2016 is currently in operation with the MAM system. This is to analyze the cause of errors in the audiovisual files and create a DB of error status of the files which will be used to select the target files for future digital restoration and maintain file quality. This helps to overcome visual inspection limitations by applying an automated quality verification process and to improve the service quality through the quality control of critical video records held by the NAK.

C. Backup

In 2014, a tape-library device equivalent to 500TB of large-capacity storage (tape-library) was introduced. In 2016, 3PB was added and now in total 3.5PB is in operation. Ingested audiovisual files are backed up on tape as a set of born digital file, preservation file and metadata through an automated schedule targeted to records not less than 15 days after ingest via automatic backup jobs. First, when a backup request is executed within the system by selecting a backup target record, the backup solution backs up the backup target data to the 3.5PB tape library and creates a copy. The tape of copies will be preserved separately.

Ⅱ-5. Work In Progress: Migration

The NAK is currently migrating from its CAMS system, which has been managed for the past 10 years, to the MAM system for digitized audiovisual file. The migration targets are about 890,000 files of approximately 32TB. To achieve this, procedures and tools were developed and tested in 2017. The migration procedures were designed to have pre-implementation, implementation, verification, and final stages according to the relevant standards (ISO 13008 Information and documentation - Digital records conversion and migration process). The implementation planning phase was the preparation for the implementation of migration, mainly including process design, test environment establishment, preparation for migration target confirmation, and migration tool. During the test phase before implementation, test migration, result analysis, verification and supplementation process were carried out. Now in 2018, the real migration is in progress. The following table describes the migration procedures in 2017. As a method for verifying the migration result of a large number of files, not only a comparison of basic elements like file sizes but also a method of extracting hash values and comparing hash values of the before and after files were reviewed and they are now in progress .

<Table 4> Migration Process and Details

Stage	Procedure	Details
1. Preparation (Pre-implementation)	1.Process design	Migration process review
	2.Development of test environment	Development of test environment of MAM * Checking linkage status of CAMS-MAM
	3.Analysis of migration target list	Providing migration list (Raw data)
		Review and analysis of list of files and items
	4.Confirmation of migration target list File download	Confirmation of analyzed list
		Providing audiovisual files
	5.Preparation and verification of migration files	Dividing video files and items Checking file errors
	6.Development of migration procedure and method Development of migration tool	Development of migration tool Migration tool: procedure and method for migration

2.Test (Pre-implementation)	1.Confirmation of Test migration target	Creating and confirmation of test migration list
	2.Test migration	Test implementation
3. Verification and Supplementation (Pre-implementation)	1.Result Analysis of test	Analyzing success and failure causes of test Analyzing errors and results and developing countermeasures
	2.Supplementation of migration tool	Handling error causes and supplementing migration tool Adding monitoring function for migration
	3.Developing responses by error types	Analyzing error rates by error types and developing countermeasures
4. Migration (Implementation)	Implementation of migration	Implementing migration for 890,000 files (32TB)

Ⅲ. Conclusion

Ⅲ-1 System Effect Analysis

The NAK had digitized and preserved analog audio-visual records even before it constructed the audio-visual archiving system. In the past, offline and manual digitization processes were less efficient in time and manpower input, and it took more time to check the contents of collected audiovisual records. There was also limit in sharing and using the digitized audio-visual data.

The establishment of the audio-visual archiving system provided an environment that internal users can use large-sized audio-visual data in an easy and convenient way using their PC in the workplace, and anyone can share and use audio-visual data that they need on their own. In other words, the audio-visual archiving system for a large amount of high-volume data could reduce the time needed for utilization and improve usability by creating an environment where users can directly access the audiovisual records.

In addition, it has developed additional functions that differ from the previous environment. Search leads to an exact location and it also provides storyboard to overview the contents at a glance in a review screen to reduce time necessary for search and use. Search can be available as soon as ingest is done. In addition, the functions of collective ingest and file-based quality check in an environment that electronic files are generated are greatly helpful from a business perspective.

Ⅲ-2 Future Directions

There are still problems to be solved, such as preservation of back-up tapes in separate places, ‘is it necessary to migrate metadata after file migration?’, confirmation of metadata for long-term preservation, ‘how would modified metadata be applied to the long-term preservation format?’, ‘What are the management points for maintaining authenticity?’. Technology cannot determine a reliable and accurate generation of digital records or a solution for long-term, authentic preservation. Therefore, we would have to be able to deduce the authenticity of the records based on their creation, maintenance, preservation, and chain of custody. It would be very important to come up with a way to keep the realistic authenticity of records through practical analysis while operating the system. These tasks will be updated step-by-step, which will ensure the high-level stability of audiovisual records and ensure objective and fact-based key audiovisual information is efficiently served.

References

[1] Ray Edmondson: “ Audiovisual Archiving : A view from Australia”, Preservation and restoration Journal, v. 93-107,

[2] Sang-Kook Kim, “A Case Study of the audio-visual archives system development and management”, Preservation and restoration Journal, v.6, pg. 41-63, 2013

[3] Dong-Hyron Shin; Se-Young Jung, “A Case Study of th audio-visual archives system development and management”, Korean Society of Archival Studies Journal, v.9, no.1, pg. 33-50, 2009

Author Biography

Ji-young Lee received her MS degree in Forestry Engineering(Paper Preservation Chemistry) from Kookmin University, Rep. of Korea, in 2005. From 2005, She is working for Archival Preservation and Restoration Center of National Archives of Korea as a researcher with audio-visual archival works. Her work has focused on the development of audiovisual archiving preservation system and digital restoration issues.

Sang-Kook Kim is working for Archival Preservation and Restoration Center of National Archives of Korea as a director of electronic and audio-visual archives team.