

IPEA: The digital archive use case

Robbie De Sutter¹, Stijn Notebaert¹, Laurence Hauttekeete², and Rik Van de Walle¹

¹ Ghent University – IBBT,

Dept. of Electronics and Information Systems – Multimedia Lab
Sint-Pietersnieuwstraat 41, B-9000 Ghent, Belgium

² Ghent University – IBBT,

Dept. of Communication Sciences – MICT

Korte Meer 7, B-9000 Ghent, Belgium

e-mail: {robbie.desutter,stijn.notebaert,laurence.hauttekeete,rik.vandewalle}@ugent.be

Abstract

Now is the time to migrate tape-based media archives to digital file-based archives for television broadcasters. These archives not only address the issue of tape-deterioration, they also create new possibilities for opening up the archive. However, the switch from tape-based to file-based is something only the very big television broadcasters can manage individually. Other broadcasters should work together to accomplish this task. In the Flemish part of Belgium, the two largest broadcasters in Flanders, namely the commercial broadcaster VMMA and the public broadcaster VRT, the television facilities supporting company Videohouse, and different university research groups associated with the Interdisciplinary Institute for Broadband Technology joined forces and started the “Innovative Platform on Electronic Archiving” project. The goal of this project is to develop common standards for the exchange and archiving of audio-visual data. In this paper, we give a detailed overview of this project and its different research topics.

Introduction

As the preservation of media archives is seen as a difficult and expensive task, its importance as cultural heritage for future generations is well-recognized. Digitization of these media archives is a tremendous but necessary task for all businesses working with media, such as television broadcasters. Currently, most archives store their content on (digital) tapes, however, these mediums are doomed to deteriorate in such a way that the content eventually will no longer be readable and usable. The challenging task to create a file-based digital archive is something only the very big television broadcasters can manage individually; others will have to team up to perform this task.

In the Flemish part of Belgium, the two largest television broadcasters (the public broadcaster “Vlaamse Radio- en Televisieomroep” (VRT)¹ and the commercial broadcaster “Vlaamse Media Maatschappij” (VMMA)²), the television facilities supporting company Videohouse³, and different university research groups associated with the Interdisciplinary Institute for Broadband Technology (IBBT)⁴ have joined forces in order to tackle the file-based digital archive issue. This resulted in the “Innova-

tive Platform on Electronic Archiving” (IPEA)⁵ project.

The IPEA project not only focuses on the creation of a digital archive, but also aims to create a system that allows easy querying and exchanging of archived media content. In a first phase, this project is focused on all technical and usability aspects of a digital archive for television broadcasters. In a later phase, the project will focus on digital archiving of the audio-visual data of other players, such as advertisement agencies, government, educational institutions, and so on.

In order to realize the goals of the IPEA project, different research topics are investigated by the different consortium partners. This paper gives an overview of these research topics and an outline of the goals of the project.

The remainder of the paper is organized as follows. In the next section, we will further elaborate on the IPEA project and give an overview of its goals. Hereafter, we disseminate the project by its work packages and discuss each of these in detail. Finally, we conclude this paper.

Innovative Platform on Electronic Archiving What is IPEA?

The Innovative Platform on Electronic Archiving is developing common standards for the exchange and archiving of audio-visual data in Flanders. In addition to the practical difficulties of storing non-digital formats, the increasing demand for the access of archived information is also a significant impetus for this project. Exchanging means establishing agreements, and the collaboration between the market leaders in the world of the audio-visual broadcasting is therefore an absolute condition for the realization of these objectives.

Within Flanders there is interest from various directions for a community initiative that encompasses media, culture and government services, not only from the government sector, but also within the private sector. It is clear that “digital archiving” is important in various sectors. IPEA focuses first of all on all technological and user aspects of digital archiving for production houses and TV broadcasters. In a later phase the project can also focus on the digital archiving of audio-visual data with other players such as press agencies, advertising agencies, publishers, the government, educational institutions, etc.

¹Information about VRT at <http://www.vrt.be>.

²Information about VMMA at <http://www.vmma.be>.

³Information about Videohouse at <http://www.videohouse.be>.

⁴Information about IBBT at <http://www.ibbt.be>.

⁵Information about IPEA at <http://ipea.ibbt.be>.

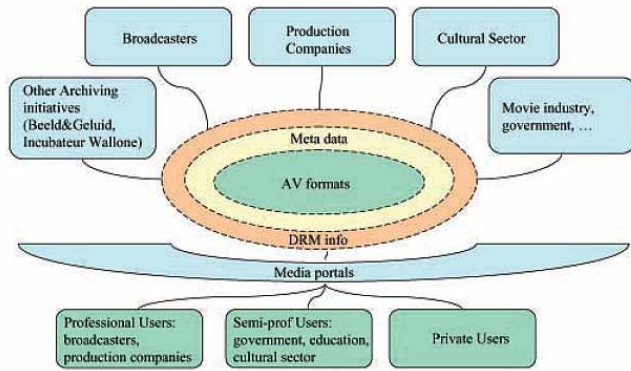


Figure 1. Overview of the IPEA architecture.

Why the need for IPEA?

Broadcasters and media production houses are faced with a great challenge. Through convergence between traditional broadcasting and information technology, both the production of programs and the access to the media is undergoing far-reaching changes. New distribution possibilities are available for media houses. The current challenges for these organizations include the following: How can we make production models more efficient? How should we deal with a massive quantity of visual content, with constantly changing digital standards and more and more fragmented productions?

A structured and complete digital archiving system permits broadcasters to make their existing archives accessible in a more efficient manner to B2B users (production houses, facilitating companies, other domestic and foreign broadcasters, etc.) and B2C users (art centers, schools, governments, citizens, etc.). In addition a digital archiving system can play an increasingly important role in the context of “ownership rights” by making the protection of data not only easier but also more secure and easier to monitor.

What is the goal of IPEA?

In order to develop the envisaged digital archiving technology, competency and additional research is a necessity in the following domains:

- Research with regard to the identification of “user needs,” in the context of which professional, semi-professional and individual end users are eligible.
- Development of a standardized metadata model for the description of audio-visual data.
- Development of a standardized metadata model for the exchange of audio-visual data.
- Establishment of guidelines for the use of specific media formats in the context of a digital archive.
- Securing multimedia data and networks, bearing in mind both B2C and B2B user scenarios.
- Specification of a general technical infrastructure that permits digital archives to be accessed in an efficient manner.
- Study of possible business models.

The idea is to convert the acquired knowledge into a joint digital archive prototype. This prototype will be housed at the

IBBT test center. After the conclusion of the project, the participating companies (and/or other companies) can of course build further on this prototype, albeit within the context of their own business operations.

VRT, VMMA and Videohouse are all participating in this project and are in this way contributing their own expertise and knowledge of the problem area. The participation of the market leaders in the broadcasting world will result in a generally applicable standard for the exchange of audio-visual data. In addition, all the partners in the project are also aware of the fact that Flanders is not an island, and that agreements with the countries and regions surrounding us are also desirable, if not indeed necessary. This is why within this project we also strive toward active collaboration with similar initiatives at home and abroad.

Work Packages

In this section, we disseminate the project by its different work packages. For each work package, we discuss its tasks and expected results.

The IPEA project started in January 2004 and runs for two years, with a total manpower of forty full-time equivalents for each year. The IPEA approach is based on the following common principles:

- If possible, use existing technology, otherwise perform research about these missing or inadequate technological areas.
- Through international collaboration, achieve efficiency in research and implementation.
- Learn from related (international) archiving projects and focus on the uncompleted or missing features.
- Attempt to create a standardized solution supported by all participants of the consortium in order to achieve economical practicability.
- Set up common hardware infrastructure for back-up and disaster recovery.

WP1: Project Management

WP1 is the “classical” management work package with as main task to coordinate the actions in and between the other work packages; to give feedback at the IBBT supervisor and other partners in the consortium; and to organize steering committee meetings.

WP2: State-of-the-Art Study of Electronic Archiving

The main goal of the second work package is to identify and evaluate existing technologies that are useful for the digital archiving project. This is done nationally as well as internationally. The Dutch use case *Beeld & Geluid* is the starting point for this work package. Nevertheless, the scope of this work package is not limited to the technology and vision that were created for *Beeld & Geluid*. Indeed, a global state-of-the-art view is worked out with regards to the perceived benefits of electronic archiving in the (inter)national audio-visual industry.

The topics that are addressed in this work package are:

- What is currently being done with regards to electronic archiving?

- What are the initiatives of the larger public and commercial broadcasters?
 - What are the business cases developed for the digital libraries? Are these based on distributed or centralized archives? Is the dissemination of these archives intended for internal, professional, or public usage?
 - Which metadata are used when archiving the audio-visual material? What is the architecture of the archive?
 - What is the government policy with regards to audio-visual archives?
- What is the economic, social, and cultural potential usage of the archives?
 - Is there a need for professionally or publicly available archives?
 - What are the issues concerning copyright and digital rights management (DRM)?

Next to the study of existing use cases, the work package will study standards, guidelines, and research that are focused on archiving procedures, metadata, and archiving file formats.

Halfway the project, this work package has finalized its report. The report discusses eight Belgian television broadcasters (namely VRT, VMMA, VT4, TV-Brussels, ATV, Vitaya, Kanaal Z, and Focus-WTV), two production houses (D&D Productions and Woestijnvis), and three international archives (BBC Motion Gallery, BCE, and Beeld & Geluid). For each archive an overview is given on the following aspects:

- The history of the archive.
- The (hardware) infrastructure.
- The principal users of the archive.
- The current method of archiving and dissemination.
- The used metadata model.
- If applicable, the plan of migrating to a digital archive.
- The management of the juridical and legal aspects.

Next, an overview and comparison of different metadata standards for audio-visual material is given. The discussed standards are Dublin Core [1], the IFLA model [2], ISO/IEC MPEG-7 [3, 4, 5, 6], SMPTE Metadata Dictionary [7], the FIAT/IFTA model [8], EBU P/Meta [9, 10], BBC SMEF [11], ETSI TV-Anytime [12].

Finally, an overview of related national and international digital archiving projects for audio-visual material is discussed, namely the projects EUROMEDIA⁶, Cedars⁷, DAVID⁸, cDAVID⁹, ECHO¹⁰, PRESTO¹¹, AMICITIA¹², Primavera¹³,

⁶Information about EUROMEDIA at <http://www.foyer.de/euromedia>.

⁷Information about Cedars at <http://www.leeds.ac.uk/cedars>.

⁸Information about DAVID at <http://www.antwerpen.be/david>.

⁹Information about cDAVID at <http://www.cdavid.be>.

¹⁰Information about ECHO at <http://pc-erato2.iei.pi.cnr.it/echo>.

¹¹Information about PRESTO at <http://presto.joanneum.ac.at>.

¹²Information about AMICITIA at <http://www.amicitia-project.de>.

¹³Information about Primavera at <http://www.primavera-ist.net>.

FIRST¹⁴, BIRTH¹⁵, PrestoSpace¹⁶, and PACKED¹⁷.

WP3: Research with regards to Users

The third work package deals with the identification of the user expectations and user requirements. It has a lot of common ground with work package 2 and 4. During the analysis of the expectations and needs of professional users, it is important to bear similar (foreign) research in mind. Furthermore, it is useful to evaluate the results to foreign developed application models.

Two kinds of users are taken into account and examined independently, namely the professional user (e.g., the television broadcasters) and the semi-professional user (e.g., the government and educational institutions). Through questionnaires, the needs and expectations of the storage and consultation of digital media, internal or external use, and centralized versus decentralized systems are determined.

Other tasks of the work package are to perform usability tests on the IPEA test platform at the IBBT laboratory and to advise the local governments on the optimization of an archiving policy.

The results of the work package will be used as an input to create different business cases done by work package 8. WP3 will also give a more detailed view on the user expectations from a legal point of view (e.g., privacy concerns and intellectual property rights).

Currently, this work package mainly focuses on the analysis of the user expectations and user requirements. A thorough analysis of foreign projects is already finalized and the in-depth interviews of possible users of the archive are being executed.

WP4: Creation of a standardized metadata model

A crucial aspect within the IPEA project is the creation of a common metadata model that all partners can relate to. The following tasks are assigned to this work package:

- Make an inventory and a study of the currently used metadata technologies within the participating broadcasters.
- Create a common metadata model which includes:
 - A standardized semantic description.
 - A standardized syntax.
 - The definition of useful ontologies.

Special attention goes to the incorporation of juridical metadata in order to ensure that the legal rights and obligations are completely fulfilled for the digital archive.

After the first year, WP4 has investigated and compared different metadata standards. The result hereof was incorporated in the state-of-the-art report of WP2. Next, WP4 has created two metadata standards: an exchange model and an internal model. The former is utilized by an user external to the digital archive and this in order to retrieve media from and ingest media into the archive as depicted in figure 2. This exchange standard is a subset of the EBU P/Meta standard 1.1. As such, the syntax, being XML and checked by an XML Schema, and the semantics of all elements are strictly defined. While creating this subset, it was felt that this is usable for larger broadcasters, but that the smaller

¹⁴Information about FIRST at <http://www.film-first.org>.

¹⁵Information about BIRTH at <http://www.birth-of-tv.org>.

¹⁶Information about PrestoSpace at <http://www.prestospace.org>.

¹⁷Information about PACKED at <http://www.packed.be>.

(local) broadcasters needed an even more tuned-down version. Hence, WP4 created a *regular* IPEA profile that is a subset of the P/Meta standard which can be used for larger corporations and a *minimal* IPEA profile that is a subset of the regular IPEA profile which can be used by smaller broadcasters or semi-professional users.

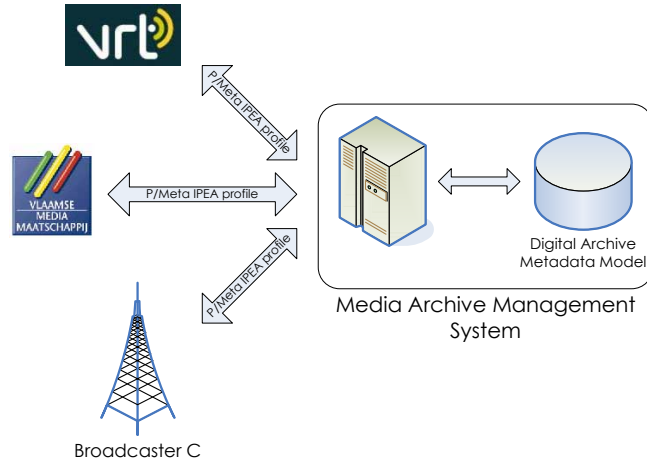


Figure 2. Overview of the IPEA exchange model.

The second metadata standard — the internal model — is an Entity-Relationship Diagram (ERD) that defines the layout and the relations of the digital archive database. This model was based on the existing models of the two participating broadcasters and was extended with the necessary juridical information entities. A high-level structural overview of the model can be found in figure 3.

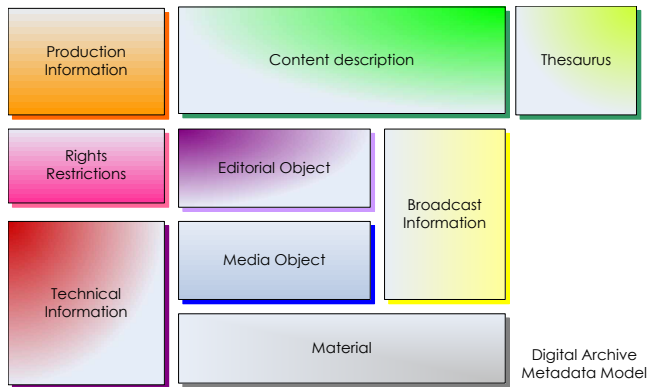


Figure 3. Structural overview of the IPEA internal archive model.

WP5: Selection of the audio-visual format(s)

In work package 5, best-practice guidelines are defined for good usage of audio-visual (encoding) formats. This project topic is required to ensure (future) access to the stored content. This is achieved by the following tasks:

- Enlist the audio-visual data formats that are currently used within the participating broadcasters.
- Study and compare these identified audio-visual data formats.

- Select the audio-visual data formats that will be supported by the project.

It is not primordial that this work package selects only one unique data format, contrary to the common metadata model. Indeed, different broadcasters will use different data formats depending on their (current and future) hardware infrastructure. Nevertheless, selecting one or more data formats for the browse resolution of the data will be necessary.

Currently, WP5 has finalized the first tasks and discussed audio-visual data formats in detail. The visual data formats are MPEG-2, H.264/MPEG-4 AVC, JPEG2000, Windows Media 9 / VC1, DV25/DVCAM, D-10 / IMX, VHS, Betacam SP, Digital Betacam, Betacam SX, HDCam, and Avid DNxHD. The audio data formats are various MPEG formats, WAV/PCM, Windows Media Audio, Dolby Surround (AC-3-broadcast, E-production), and NICAM.

WP6: Security Facets of Digital Archiving

For each of the archiving activities (ingest, management, and usage), adequate security procedures are required. As a result, an integrated security model that is not only technically but also legally sound is a major research topic. WP6 researches the following topics:

- Create a state-of-the-art overview of security systems and an overview of juridical aspects with regard to digital archives.
- Techniques to secure the content of digitized objects by using, for example, watermarking and fingerprinting.
- Accountability of the archive: how to ensure the correctness and integrity of the archive.
- Study of access-control mechanisms for the different users of the archive.
- Techniques to protect sensitive data with regard to the privacy of the users of the archive and the personal details of the data stored in the archive.
- Securing data streams with fast encryption algorithms.

Currently, WP6 created the state-of-the-art overview.

WP7: Study of the IPEA Architecture

Using the results of the previous work packages, the consortium defines in WP7 a technically sound architecture which allows the creation of a digital archive for the Flemish part of Belgium. A demonstration platform will be set up to verify its feasibility and to illustrate the possibilities. WP7 will investigate following research topics:

- Choose between a centralized and a decentralized architecture.
- Use one or more portals for different users.
- Develop disaster recovery strategies.
- Use or create search engines.
- Set up the demonstration platform.

WP7 has finalized the IPEA architecture and is currently creating the demonstration platform.

WP8: Study of Feasible Business Models

In the last work package, different business cases will be created and evaluated. WP8 will use the information from the user expectation and user requirements study (i.e., the market analysis) in order to create different cases. These business cases will be evaluated using the following techniques:

- Market research.
- Commercial possibility.
- SWOT analysis.
- Cost-benefit analysis.
- Financial feasibility.
- ROI analysis.

Initial market research studies are currently being performed by WP8. It is expected that there is a demand for an archive as proposed by the IPEA consortium.

Conclusions

In this paper, we motivated the need for a digital file-based archive for audio-visual material from television broadcasters. The main reason to switch from tape-based to file-based is the preservation of the cultural heritage that is at stake because of the deterioration of tapes. In order to create a digital file-based archive, it is necessary for television broadcasters to join forces.

This paper discusses in detail the “Innovative Platform on Electronic Archiving” project, which has as goal the development of common standards for the exchange and archiving of audio-visual data. After an overview of the IPEA consortium and its work methods, we gave a detailed discussion of the different research topics, organized in eight work packages.

Acknowledgments

The research activities that have been described in this paper were funded by Ghent University, the Interdisciplinary Institute for Broadband Technology (IBBT), the Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT), the Fund for Scientific Research-Flanders (FWO-Flanders), the Belgian Federal Science Policy Office (BFSP), and the European Union.

References

- [1] Dublin Core Metadata Initiative. Dublin Core Metadata Element Set, version 1.1: Reference description. Technical report, 12 2004. Available at <http://www.dublincore.org/documents/dces/>.
- [2] International Federation of Library Associations and Institutions. Functional requirements for bibliographic records. Technical report, 9 1998. Available at <http://www.ifla.org/VII/s13/frbr/frbr.pdf>.
- [3] B. S. Manjunath, P. Salembier, and T. Sikora. *Introduction to MPEG-7: Multimedia Content Description Language*. John Wiley & Sons, 6 2002.
- [4] J. M. Martínez, R. Koenen, and F. Pereira. MPEG-7: The Generic Multimedia Content Description Standard, Part 1. *IEEE MultiMedia*, 9(2):78–87, 2002.
- [5] J. M. Martínez. MPEG-7: Overview of MPEG-7 Description Tools, Part 2. *IEEE MultiMedia*, 9(3):83–93, 2002.
- [6] P. Salembier and J. R. Smith. MPEG-7 Multimedia Description Schemes. *IEEE Transactions on Circuits, Systems and Video Technology*, 11(6):748–759, 2001.

- [7] Society of Motion Picture Television Engineers. Metadata dictionary structure – for television. Technical Report RP 335M-2001, 2001. Available at <http://www.smpete.org>.
- [8] Fédération Internationale des Archives de Télévision - International Federation of Television Archives. FIAT/IFTA minimum data list. Technical report. Available at <http://www.fiatifta.org>.
- [9] R. Hopper. Metadata exchange standards. Technical Report No. 284, European Broadcasting Union, 9 2000.
- [10] P/Meta Metadata Exchange Scheme v1.1. Technical Report 3295, 6 2005. Available at http://www.ebu.ch/en/technical/metadata/specifications/notes_on_tech3295.php.
- [11] British Broadcasting Corporation (BBC). SMEF data model version 1.1. Technical report. Available at <http://www.bbc.co.uk/guidelines/smf/>.
- [12] The TV-Anytime Forum. Metadata specification version 1.3. Technical Report COR3 SP003V13, 8 2003. Available at <http://www.tv-anytime.org>.

Author Biography

R. De Sutter received the master degree in computer science from Ghent University, Belgium, in 1999. He joined the Multimedia Lab in 2001 where he is currently working toward the Ph.D. degree. His research interests include video coding technologies, usage context modeling and negotiation, and metadata.

S. Notebaert received the Master degree in Industrial Sciences from KaHo Sint-Lieven, Belgium, in 2001 and received his M. Sc. degree in Engineering from Ghent University, Belgium, in 2004. He joined the Multimedia Lab in 2004 where he is currently working toward the Ph.D. degree. His research interests include digital video compression and content adaptation.

L. Hautekeete received her Ph.D in Communication Sciences from Ghent University, Belgium, in 2004. In her Ph.D she looks at tabloidisation tendencies in the Flemish press, the development of a measurement model. She joined MICT (media and ICT) in 2005. Her research interests include the printed and audio-visual media, media economics, and in particular new media.

R. Van de Walle received his M.Sc. and Ph.D. degrees in Engineering from Ghent University, Belgium in 1994 and 1998 respectively. After a visiting scholarship at the University of Arizona (Tucson, USA), he returned to Ghent University, where he became full professor of multimedia systems and applications, and head of the Multimedia Lab. His current research interests include multimedia content delivery, presentation and archiving, coding and description of multimedia data, content adaptation, and interactive (mobile) multimedia applications.