A STRUCTURED METADATA MODEL FOR SUPPORTING NEW AUDIO-VISUAL CATALOGUING

Giuseppe Amato, Donatella Castelli, Serena Pisani, Paola Venerosi

Istituto di Elaborazione dell'Informazione – CNR
Via Alfieri 1, 56010, Ghezzano (PI), Italy
Email:{amato,castelli,pisani,venerosi}@iei.pi.cnr.it

1. INTRODUCTION

This paper presents the AV-Metadata Model (AV-MM) for historical documentary audio-visual resources. Its definition originated from the experience that we acquired carrying out the European Chronicles On Line (ECHO) project[1] (http://pc-erato2.iei.pi.cnr.it/echo/) [4], [8]. The main objectives of this project were:

- provide a Web-based *cross-access* to collections of historical documentary audio-visual resources of great international value;
- develop a long term reusable software infrastructure to support digital films archives;
- develop an architecture able to support service extensibility and interoperability.

The model was designed in such a way that it may support both *traditional archive access services* and *advanced access services*. The first category of services allows the retrieval of audio-visual documents by using descriptive fields about the content or the physical copy of the resource (e.g. name of the producer, title of the series, the tape collocation). The second category of services allows the retrieval of audio-visual documents by using information automatically generated processing the documents (e.g. keyframes, video abstracs, words in the transcript, visual features, etc.).

[1] ECHO - Project number 11994, Information Societies Technology (IST) Programme, funded by the European Commission. It aims to build a digital library of audio/visual historical documentary films. The first release contains selected materials from Institut National del'Audiovisuel (INA), Netherlands Audiovisual Archive (NAA), Istituto Luce (IL) and Memoriav.

The AV-MM was designed, step by step, taking into account the requirements and the specific usages of the traditional and potential users of historical documentary archives, and all information needed to support the realization of more advanced access services [3].

The development of the model passed through three important phases. (i) An important part of the work, was to understand the effects of the massive transformation of audio/visual archives due to the digitalisation process. (ii) We subsumed the efforts sustained by authoritative groups such MPEG7 [7], DC [2], ABC [1] in delivering reference metadata models for audio-visual resources. (iii) We gathered all information necessary to implement well-tried technologies [5] improving access, search, media transfer, and interoperability.

As a result of this analysis we have noticed that, given the new possibilities offered by the digital support, AV content management systems, commercial strategies, searching methods, access capabilities, and cataloguing rules are all being reconsidered, revised, and transformed.

From the content management system perspective, i.e. our perspective, the need of reorganization is being due essentially to the increasing demand of more content information, even more detailed; to the range of new technologies digital form make now applicable; and to the market competition. AV content management reorganization had to take into account very specific users requests going from the simple AV document representation to the identification of programme version, revision and extension; recompilation of extracts; renewal of their presentation considering different points of view or similar themes; original footages cuts; "refreshing" execution by means of colour enhancement, soundtrack re-dubbing, presenter substitution, etc.

All the needs above require new advanced search methods. Traditional search based on textual description and browsing seems to be insufficient. It is now required to express requests able to identify, in the visual component, portions and structures (shot, sequences, frames, shapes, colours), and in the audio component, words and sounds from the soundtrack. The description of documents, enriched by visual-feature based information, permits to exploit advanced searching methodologies, such as *similarity searc*, and tools as the *automatic extraction of key-words*, the *visual abstract*, the *story-board* etc.

The diversification of access capabilities become even more important as the audience grows. A variety of formats is essential to satisfy the various usages, the commercial exigencies, the copies conservation. Their management needs to assure the visibility of all formats available and the immediate information of the limitations and the cost of replaying them.

Also, the possibility of cross-language search become relevant. As the technical possibility of accessing archives maintained in other countries become a reality, users should be provided with tools that allow them to effectively use these archives.

2. MOTIVATIONS

With the above premises, cataloguing methods need to be rethought. The traditional description level is insufficient to describe the information needed to find programme, extracts, etc. for different usages and to assure metadata interoperability. The entire production cycle of the audio/visual document should be described from the creation, physical embodiment to its final collocation. Furthermore, the document should be described on its components, and/or their parts; the textual description should be harmonised with physical features representation. Given this scenario, modelling audio/visual documents needs to cross their process history, the context of their production and dissemination, their feasible uses, and the barrier between archives and languages.

Even if emerging classification schemes may support efficiently the AV cataloguing task, the most authoritative AV metadata model solutions are currently under development. Each of them tends to cover specific aspects and does not supply a provision for applicability to other frameworks. The attempts to harmonize various kinds of description and requirements are in course, but a standardized solution is not at hand.

This open situation stimulated the launch of our proposal, a metadata model for AV historical documentary archives. The main aims were: to assure metadata interoperability and extension, to represent AV historical document at a good level of detail, to enable the application of advanced finding aids and searching methods. Following the logic proposed by MPEG7, the main criterion in defining the model was to use two description typology (textual and feature based), in parallel or in a exclusive way, according to circumstances; and to leave model open and subject to change in response to new improved solutions. In fact our model can be described by using the MPEG-7 DDL and can become a specific MPEG-7 description schema.

3. THE METADATA MODEL

The model we propose extends the IFLA-FRBR model [6] to be able to describe audio visual documents and to cope with the requirements discussed above.

IFLA-FRBR suggests to describe resources using four different entities: (1) *Work* entity to describe the abstract idea of a resource, (2) *Expression* entity to describe different versions of the same work, (3) M*anifestation* entity to describe different physical supports where an expression can be embodied, (4) I*tem* entity to describe different copies of the same manifestation.

We have added new sub entities for each of the previous IFLA-FRBR entity to offer more specialised mechanism to describe audio visual documents. The overall schema of the metadata model is sketched in Figure 1.

The Work entity was extended by defining the sub entity *AVDocument* that contain attributes specialised for describing abstract ideas of audio visual documents as, for example, Director, Event, Date, Person, Location, Description. Some of these attributes can be defined in such a way that cross-language search can be performed effectively. For instance, a controlled vocabulary can be used where terms are chosen in such a way that a translation exists in all supported languages.

A specific version of a work can have several forms. For instance, it can be a silent movie, an audio stream, a black and white movie, the Italian translated version, etc. To this aim, the *Expression* entity was extended by defining the *Version* entity, which contains specialised attributes like VersionTitle, Duration, etc. The *Version* entity was also extended by other entities specialised to describe specific versions: the *Video* entity, corresponding to a silent video expression, *Audio*/*Video* entity, corresponding to an audio/video expression, *Audio* entity, corresponding to an audio only expression, and *Transcript* entity, corresponding to the word spoken in an audio stream.

It is also important to describe separately relevant portions of the entire expression as, for instance,

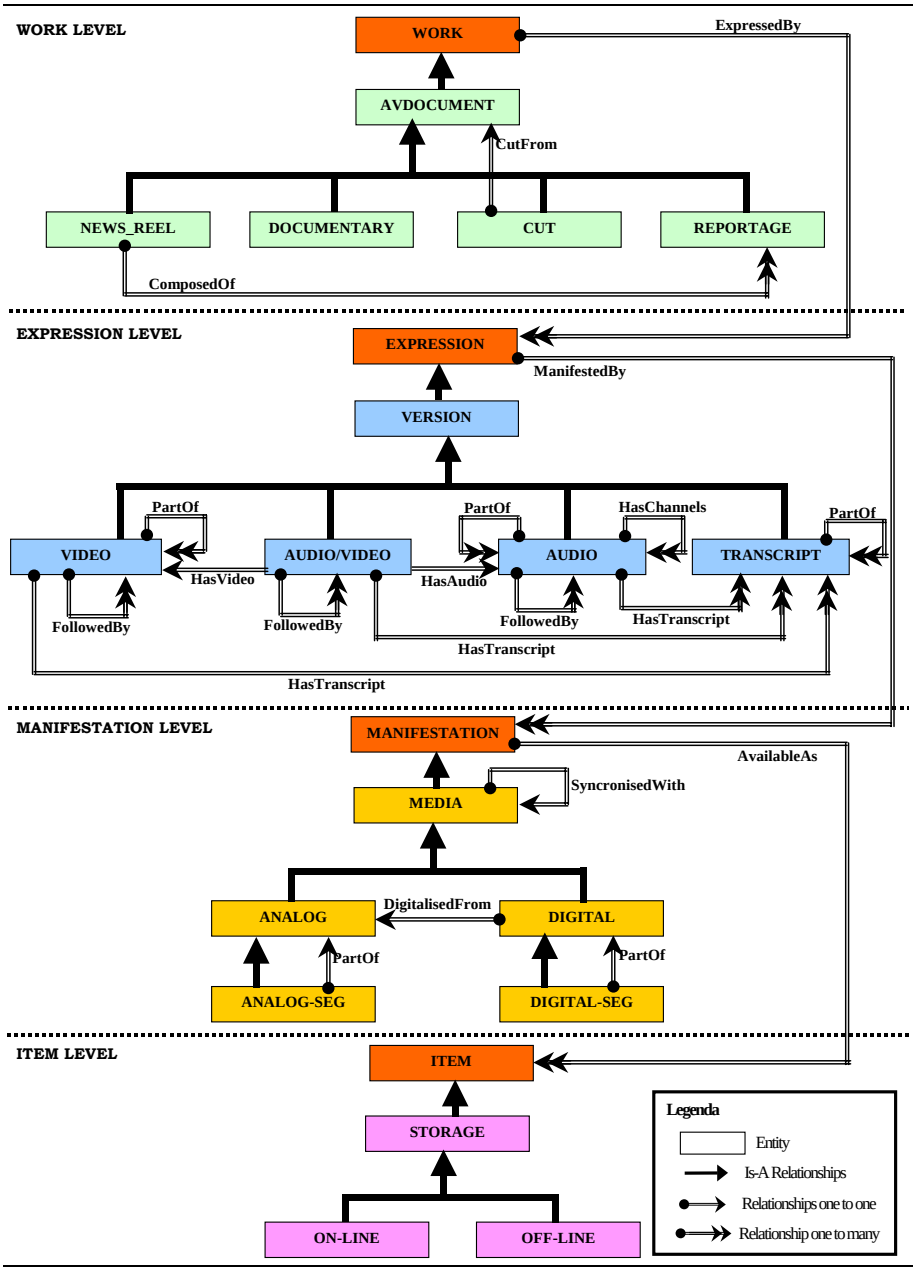

Figure 1**:** The metadata model schema

sequences and shots. Scenes and shots of a video are represented by the *Video* entity itself (since portions are just videos). *PartOf* relationships relate composite videos and video components. The same technique is used to represents portions of Audio/Video, Audio and Transcript portions. Other appropriate relationships model the possible links among these expressions such as for example *FollowedBy*, *HasAudio*, *HasVideo*, *HasChannel* and *HasTranscript*. Using these entities and relationships, it is possible to separately describe audio streams and corresponding video streams. It is possible, for instance, to separately keep track of audio streams produced in different languages, associated with the same video stream, and their corresponding transcripts. It is also possible to describe the fact that the same sequence (for instance a video sequence) was used in different documents.

Expressions are associated with features that are extracted from them in order to support similarity retrieval. Typically, these features are extracted automatically by means of image, audio, and text processing. Examples of features for visual

documents are colour histograms, shapes, motion vectors, etc.

Audio video documents can be stored on analog and digital supports. Examples of analog supports are VHS and Betacam, while examples of digital supports are DVD and MPEG files. This was modelled by extending the *Manifestation* entity with the *Media* entity that was further refined by means of the *Digital* and *Analog* sub entities.

In order to relate expressions, corresponding to relevant portions of entire documents to their physical position in the corresponding manifestations, *Digital* and *Analog* sub entities were further refined by defining the sub entities *Analog-Sequence* and *Digital-Sequence*. A relation *SinchronisedWith*, included in the *Media* entity, permits to relate two manifestations that may be played synchronously as, for instance, a movie and its soundtrack, in the case they are stored separately.

A library may have several copies of the same manifestation. For instance the same MPEG file may be stored on several video servers, or the library may have available several copies of the same VHS tape. To represent this situation, the *Item* entity was refined by the *Storage* entity, containing attributes for the right access management and ownership, which was further refined by *On-Line* and *Off-Line* entities. The first represents and describes resources that can be accessed directly by a computer, the second, resources that are stored, for instance, in some shelf of the library. In the first case, it is possible to describe that on-line resources may have different access speed, and the systems where they are stored have different workload. In the second case, it is possible to describe the quality of preservation of different copies.

In designing this metadata model great care was taken to the provision for interoperability. We are currently experimenting the mapping of this model into other metadata models to evaluate its interoperability level. Another aim in designing the model was to render it extensible. This is a key features of any metadata model since it permits a successive expansion to cover new emerging needs.

The current version of the model does not cover yet the typical requirements for an efficient “right approach”. Future work will inquire at what degree of detail of rights description a content information system should supply and how: by providing a right issue specialization, or by activating a systematic flux with a right management system.

4. REFERENCES

[1] ABC A Logical Model for Metadata Interoperability. http://www.ilrt.bris.ac.uk/discovery/harmony/docs/abc/abc_draft.html

[2] Dublin Core Metadata Element Set, Version 1.1: Reference Description, 1999-07-02. http://purl.org/DC/documents/rec-dces-19990702.htm

[3] ECHO User Requirement Report. Annemieke de Jong, Johan Oomen (NAA); Pasquale Savino, Paola Venerosi (IEI-CNR), Deliverable Number D1.2.1 of the ECHO Project, June 2000

[4] ECHO Metadata Modelling Report. Giuseppe Amato, Donatella Castelli, Serena Pisani, Paola Venerosi (IEI-CNR); Philippe Poncin, Laurent Vinet (INA), Deliverable Number D3.1.1 of the ECHO Project, September 2000.

[5] [A Comparison of Schemas for Video Metadata Representation. Jane Hunter, Liz Armstrong. WWW8, Toronto, May 10-14, 1999. http://archive.dstc.edu.au/RDU/staff/jane-hunter/www8/paper.html

[6] Functional Requirements for Bibliographic Records. Final Report. K. G. Saur München, 1998. http://www.ifla.org/VII/s13/frbr/frbr.pdf

[7] MPEG-7 Overview (version 3.0), Jose’M Martinez (UPM-GTI, ES) http://www.darmstadt.gmd.de/mobile/MPEG7/Documents/W3445.htm

[8] Building an Audio-visual Digital Library of Historical Documentaries: the ECHO Project, Pasquale Savino, D-Lib, Vol 6, N. 11, November 2000