A metadata model for historical documentary films

Giuseppe Amato, Donatella Castelli, Serena Pisani

Istituto di Elaborazione dell'Informazione – CNR – Pisa (Italy)
{amato,castelli,serena@iei.pi.cnr.it}

Abstract. This paper presents a metadata model for historical audio-video material able to describe information for supporting both the traditional archival functions and the advanced applications, like video summary, speech recognition, and automatic semantic content extraction.

1 The Metadata Model

The metadata model presented in this paper originated from our experience within the European Chronicles On-line (ECHO) project[1]. The aim of this project is to build a digital library to permit cross-archives searching and retrieval of historical documentary films by exploiting very advanced audio-visual technologies.
In the ECHO digital library, historical documents regarding the war and post war period are considered; in particular, *newsreels*, *documentaries, reportages* and *cuts* are collected. Newsreels are the parents of today's news television, that were broadcast at cinema before the beginning of a film and the typical language used is propaganda language. Documentaries are more detailed and longer films mainly focused on one particular aspect. Reportages are subparts of several newsreels. Cuts corresponds to the video material that was collected, but never broadcast.
The metadata model is depicted in Figure 1. It resulted from the analysis of the answers to three different questionnaires tailored for the historical documentary film archivists, the technology providers and the speech recognition and digital video abstracting researchers. This analysis led us to conclude that a multi-layered and hierarchical description of audio-video documents was needed. The former characteristic allows describing different aspects of the same document. The latter permits to support both (i) the interoperability among different archives at the corresponding hierarchical levels and (ii) to satisfy the needs of special interest user communities. This last point is simply obtained by refining the model by adding specialized descriptions.
This multi-layered and hierarchical structuring is in accordance with the IFLA model [3]. IFLA defines the four entities WORK, EXPRESSION, MANIFESTATION and ITEM describing different aspects of the resource.
In this paper we present the structure of the proposed metadata model and omit the metadata fields associated with each entity. They can be found in [1]. In our model

[1] ECHO - Project number 11994, Information Societies Technology (IST) Programme, funded by the European Commission.

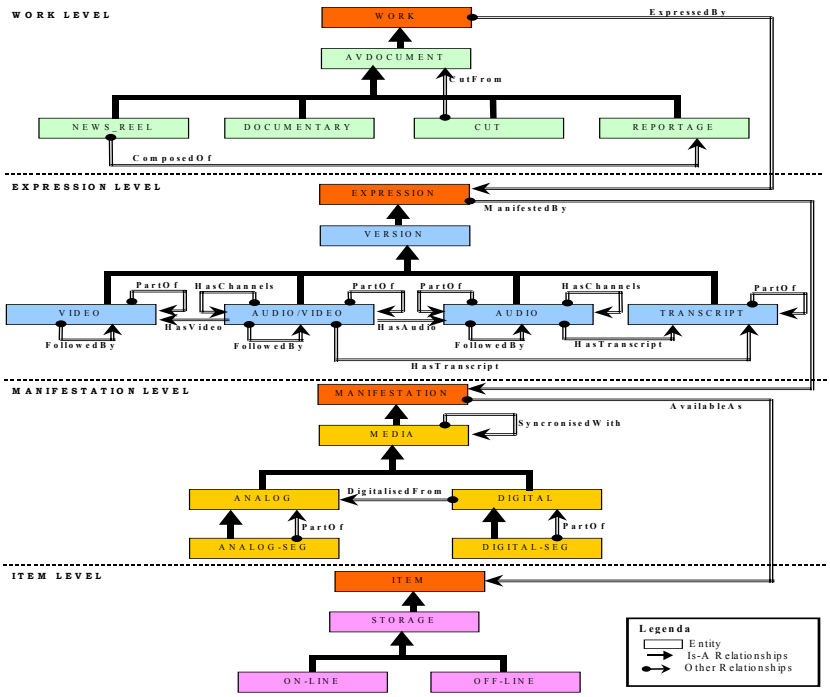


Fig. 1. The metadata model structure

each of the four IFLA entities has been considered as the most general entity of the corresponding level. Each entity has then been specialised to perform the structure of the metadata fields. At the Work Level the entity AVDOCUMENT is a generic entity representing an intellectual or artistic film work, while its subentities NEWSREEL, REPORTAGE, CUT, and DOCUMENTARY represent the types of historical audio-video documentary films. A newsreel may be composed by a sequence of reportages that correspond to various parts of the whole audio-video document itself. This situation is modelled by the relationship *ComposedOf*. This structure emphasizes the fact that some reportages are used as part of different documentary films. The elements in AVDOCUMENT may be related to the cuts from which are extrapolated through the relationships *CutFrom*.

Each document, belonging to the entities of the Work Level, may have one or more versions. For instance, an element of VERSION can be the audio realization of a newsreel, another the audio-video realization of the same newsreel. A decomposition of VERSION in VIDEO, AUDIO, AUDIO/VIDEO and TRANSCRIPT was introduced to endow these particular kind of versions of their own set of metadata fields.

At the Manifestation Level, the entity MEDIA represents different supports that can be used to maintain documents. The sub-entities corresponding to segments are ANALOG-SEG and DIGITAL-SEG. The sub-entities ANALOG and DIGITAL represent, instead, entire documents.

The entity STORAGE at the Item level describes general information about individual copies of the same document. Specific information about on-line and off-line

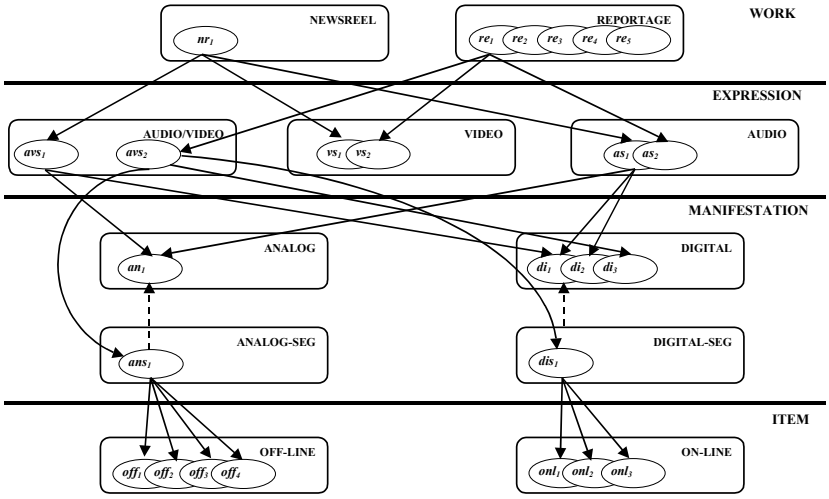


Fig. 2. A running example

document copies is, respectively, modeled by the sub-entities ON-LINE and OFF-LINE. This partition permits, for instance, to represent a situation in which the same MPEG1 file may be stored on servers with different load and access times, or in which several copies of the same CDROM or of the same VHS tape have different quality of preservation.

2 A Running Example

Figure 2 outlines this example. In the figure, boxes correspond to entities, ellipses to objects, arrows highlight relationships between objects at different levels, dashed arrows represent part-of relationships.

Let's suppose that we have a newsreel about the Landing of the allied forces in Normandy. This newsreel is represented by an object nr_1 of the NEWSREEL entity of the work level. The "Landing in Normandy" newsreel is composed of several reportages. These components are objects of the REPORTAGE entity. We suppose that there is the object re_1 representing the reportage where the US president Roosevelt announces the attack to the press; the object re_2 that presents the reaction of Churchill; the objects re_3, re_4, re_5, that are about various actions taken directly during the attack.

Let's suppose that our newsreel is available in different versions. Different objects, at the expression level, represent these versions. These versions may belong to different entities, or may be different objects of the same entity. For instance, the object avs_1, of the AUDIO/VIDEO entity represents the Italian spoken version of the newsreel (this is obtained by inserting the value "Italian" in the field language of these objects); the object vs_1, of the VIDEO entity, is a video only expression; the object as_1, of the AUDIO entity, is an audio expression, corresponding to a radio program announcing the attack. Of course, reportage objects may also have several expressions. For instance we suppose that re_1, is represented by three expressions.

Let's suppose that the audio/video expression avs_1 is available as VHS videotape and also as a MPEG file. This is modeled by two objects at the manifestation level: the object an_1 of the ANALOG entity and by the object di_1 of the DIGITAL entity. The audio expression as_1, is embodied both in the objects an_1, and di_1, so it can also be represented by these objects. In addition, we suppose that there is a separate MP3 file that only contains the audio of the newsreel. The di_2 object of the DIGITAL entity may represent it. Let's now consider the audio/video sequence avs_2 corresponding to the reportage with President Roosevelt. We suppose that there is a QuickTime file that contains it. The file is represented by the object di_3 of the DIGITAL entity. In addition, since it is also contained in the whole newsreel, there are two objects ans_1 and dis_1, respectively of the ANALOG-SEG and DIGITAL-SEG, that represent the portion of the objects an_1 and di_1, corresponding to VHS and MPEG supports of the entire newsreel, where the reportage appears.
Let's now suppose that the library has four copies of the VHS videotape an_1 that contains our newsreel. These copies are associated, at the item level, with the objects $offl_1$, $offl_2$, $offl_3$, and $offl_4$ of the OFF-LINE entity. Continuing with the example, we can suppose that the MPEG file represented by di_1 is mirrored on three different servers. These files are then associated respectively with the objects onl_1, onl_2, and onl_3 of the ON-LINE entity.

3 Concluding Remarks

This paper introduced a metadata model for historical documentary films. This model is the result of an interaction with the data, technology and services providers involved in the ECHO project. This interaction outlined two realities: (i) the reality of the traditional archivists which were accustomed to maintain information about the content of the video, its location and its cataloguing process; (ii) the reality of the technology providers that describe the video mainly in terms of its physical characteristics. Most of the existing metadata models, taking into account audio-visual resources, tend to focus on one of these two realities. The model presented is an attempt to cover both of them. It differs from other similar attempts [2] for its approach towards interoperability. The metadata fields, representing different aspects of the audio-visual resource, have been described at different levels of abstraction. This structure provides a great potential for interoperability. We are currently specifying the model as a simple RDF schema [4].
As concluding remarks, let us outline that even if this work focused on the historical documentary films, we are confident that our approach may be generalised to define models for other kind of audio-visual resources.

References

[1] The ECHO Metadata Model. G. Amato, D. Castelli, S. Pisani, P. Venerosi. IEI-CNR Technical Report, 2000.

[2] A Comparison of Schemas for Video Metadata Representation. J. Hunter, L. Armstrong. WWW8, Toronto, May 10-14, 1999.

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

[4] Resource Description Framework (RDF) Home Page: http://www.w3.org/RDF/