# VandA: a metadata model for supporting new usages of historical audio-visual archives material

*Giuseppe Amato* [*], *Donatella Castelli* [*], *Paola Venerosi* [*]

(*) IEI-CNR –Italy

E-mail: {G.Amato, D.Castelli, P.Venerosi}@iei.pi.cnr.it

## ABSTRACT

The paper proposes a new metadata model (VandA) directed to support innovative exploitation of Historical AV Archives content. After an analysis of the needs of traditional and potential users, the model is illustrated through running examples to show at what extent the new usages can be supported.

**KEYWORDS:** metadata, digital libraries, video content description

## INTRODUCTION

There is a growing demand for different and new ways of using the content of historical audio-visual (AV) archives. These range from different forms of TV programs, new on-line outlets in the Internet, CD-ROMs, etc., that illustrate and interpret events from the past, to research source of information exploited not only by historiographers, but also by teachers, that use it as interactive learning materials, and by program makers and film directors that are interested in observing how a particular actor/film director performed.

The aim of this paper is to highlight which are the cataloguing needs raised by the new supporting technology for historical AV archives and to illustrate how the use of an appropriate cataloguing metadata model can contribute to the satisfaction of these needs.

The paper begins with the analysis results of the new needs brought by the different classes of historical AV archives users (Section 2). Then, it proposes a new metadata model, called VandA, (Video and Audio) that combines descriptive features of traditional AV historical archives with features employed by advanced AV digital applications (Section 3). This model is illustrated through running examples that show to what extent it can support the new usages of the historical AV archives (Section 4). The paper ends by discussing how the model could be extended to overcome it limitations (Section 5).

## USAGE AND REQUIREMENTS

The digitalization of historical audio visual archives allows the application of Internet advanced technologies, coming from interdisciplinary research, like the integration of historical archives enhancing new ways of content usage and distribution.

Three major classes of historical AV archives users, both usual and potential, were considered [6]. Here below we list these classes and for each of them we outline the Specific Functionality Required (SFR).

**Educational institutions, teachers, specialists, students, researchers**: they are characterized by an high level of domain expertise, but often a low level of video management systems knowledge.

- In the secondary education the audio visual material is employed both in lectures and in experimental activities.

SFR: simple query interface (graphical appearance), reuse of content (browse-copy availability), multilingual capability.

- Historical researchers interests focus on exploiting historical archives of different countries with common access codes.

SFR: finding aids, cross-search capability, personalization of presentation results, cross-language search.

- Cinema historians deal with film content and its expressive forms.

SFR: information on framing typology, framing occurrence frequency, accelerated / decelerated camera movements, cutting effects, fades, etc.; feature based search; tools supporting automated analysis on film structure.

- Students experimenting film editing, the mixing of sounds, or the combination of both, learn how to handle the complex relations among media.

SFR: highly qualified descriptors enabling "linking mechanism" for media synchronization.

**Broadcasting companies, documentary makers, news editors, journalists, program researchers**: they are the most involved in content management due to the need to produce new programs and reuse content in a competitive situation; they are also the most concerned with the "right issue".

- Journalist and documentarist activity leads often to the reconstruction of current or retrospective history about some subject, by composing video fragments.

SFR: advanced finding aids, cross-document search for places, person, event, concept, etc.; annotation functionality; editor functionality.

- Programme researchers (broadcasting context) are involved to plan new manners of displaying archives, to organize the content in different views to satisfy the requests [16].

SFR: efficient repackage and scheduling methods of archive content; fast document/fragment identification; fast copyright issue investigation procedure [10].

**Large Audio Visual Archives, archivists, librarians, marketing and sales managers**: they are experts in video archive management systems and content, heavy users of existing systems and connoisseurs of articulated user needs.

- Archivists, librarians are engaged to modify their role from that of intermediary par excellence between documents and end-users to that of finding aids providers (inventories, registers, indexes, guides, collections), allowing people to search without the costly mediation of reference staff [2].

SFR: robust classification schemes providing to describe: all the aspects of an AV document in the whole, in its decomposition (audio, video, transcript), in its segmentation (sequence, shot, frame), in its physical structures (features); documents movement (migration) and preservation (refreshing); editor functionality.

- Large historical archives researchers are engaged to improve the archive

holding and to provide quality products. SFR: granular description of content; finding aids (summaries, key-words, visual abstracts, story-board, thematic repackages of content, collections).

## THE VandA MODEL

The new form of usages of historical AV archives reviewed above impose a re-thinking on the way of cataloguing documents. The VandA model presented in this section was designed to satisfy this need for a more complete cataloguing description.

VandA is an enlarged and revised version of the model designed within the European Chronicles On Line (ECHO) project[1] [19], [7], (*http://pc-erato2.iei.pi.cnr.it/echo/*).

The model is intended as a support for information discovery and as a descriptive surrogate of the AV resources disseminated to the users. It provides a necessary ground for supporting both the *traditional archive access services* - based on the descriptive fields about the content or the physical copy of the resource (e.g. access by name of the producer, by title of the series, by the tape collocation) - and a set of *advanced access services* - based on information automatically acquired through the manipulation of digital video (e.g. access by keyframe,

---

[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.

by video abstracting, by words in the transcript, etc.).

The more authoritative reference metadata models for audio-visual resources [1], [5], [17] focus on specific aspects of the audio-visual resource description, and they do not provide a standardized harmonisation. VandA contributes to support the innovative AV historical archive organisation and usages.

VandA is an extension of a general metatada model proposed by the International Federation of Library Associations (IFLA) [14]. The extension allow to describe audio visual documents and to cope with the requirements discussed above. IFLA 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) *Item* entity to describe different copies of the same manifestation.

VandA, depicted graphically in Figure 1, extends IFLA with a set of sub entities and fields that are specifically tailored for describing historical audio visual documents. The entities are implicitly distributed on four description levels: each level corresponding to an IFLA base entity. The entities Work, Expression, Manifestation and Item are considered as the most general entity of the corresponding level. At the Work Level the entity *AVDocument* and its subentities *Newsreel*, *Documentary*, *Cut*, and *Reportage* are introduced for modelling an abstract AV historical work. AVDocument is a generic entity representing an intellectual or artistic

work, while its subentities represent the different genres of historical audio-video documentary films.

The subentities Newsreel, Documentary, Cut, and Reportage model different kinds of works. These subentities turn out to be sufficient for our experimentation. However, other entities may be added at this level to represent other types of audio-visual resources (as fiction, entertainment, etc.) that require a specialised management. The important point here is that any extension of this type does not invalidate the descriptions made according to the previous model.

A newsreel may be composed of 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 relation permits to describe independently the whole document and its individual parts. The elements in AVDocument may be related to the cuts from which are extrapolated through the relationship *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 realisation of a newsreel, another the audio-video realisation of the same newsreel. This is expressed by the relationship *ExpressedBy* between AVDocument and Expression. The entity Version is specialised by the subentities *Video, Audio, Audio/Video* and *Transcript*. These subentities represent, respectively, the following possible realisations of a work: audio only realisation, video only realisation, audio/video realisation, and transcription of spoken words. As the whole document and its parts share the same set of descriptive fields, the same entity is used for modelling both concepts. A relationship *PartOf* permits to associate a component of a document to the corresponding whole document; whereas the relationship *FollowedBy* express the temporal succession of the document parts and the type of transition between a part and the following one (cut, fading, etc.). The relationship *HasChannels,* defined on audio and audio-visual documents, models the different audio channels that an audio document may have. The relationships *HasAudio* and *HasVideo* create a link between an audio-visual documents and its audio and video components when these exist as separate resources. The relationship *HasTranscript* links the video document to the transcript of its audio (in the case that the audio document of that video document is missing) or links the audio document to its transcript.

At the Manifestation Level, the entity *Media*, with its specialisation *Analog* and *Digital*, represents different kinds of analog and digital supports that can be used to maintain audio visual document versions. Each version of the Expression Level can be stored on one or more Manifestation as described by the relationship *ManifestedBy*. The synchronisation among different elements of Media are modelled by the relationship *SynchronisedWith*. The entities Analog and Digital, that represent the entire document, are specialised in the entities *Analog-Seg* and *Digital-Seg*, that represent segments. The relationship *PartOf* permits to associate the part to its whole document.

At last level, the entity *Storage* describes general information about the individual copies of the documents.

Specific information about on-line and off-line document copies are modelled by the *On-Line* and *Off-Line* sub-entities. This partition permits, for instance, to represent a situation in which the same MPEG1 file may be stored on different servers, or in which several copies of the same CD-ROM or of the same VHS tape have different quality of preservation. The relationship *AvailableAs* links a manifestation to its items.

A set of specific descriptive fields are associated with each entity. They model the characteristic properties of each entity, i.e. they represent a description of the AV documents as perceived when looking at the AV document at a certain abstraction level. For example, a Version, which models a conceptual realisation of a work, is characterised by fields like: Title of the Version, Edition, Themes, Working Notes, Copyrights owners, etc. A Video, which is a particular Expression, is characterised by fields like: Subtitle Language, Video Abstract, Keyframe, Camera Movements, etc; whereas a Digital entity, which is physical embodiment of an expression, is described by fields like: Support Type, Format, BitRate, etc. The complete list of the descriptive fields can be found in [7].
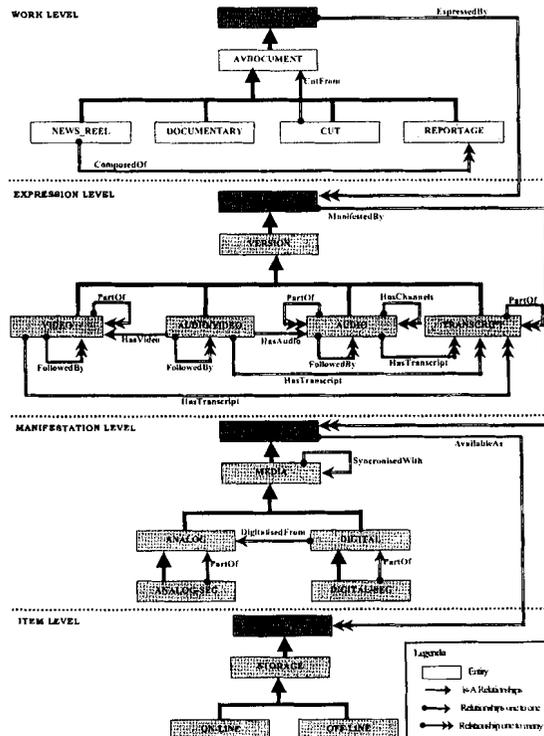


**Figure 1: The VandA model**

## RUNNING EXAMPLE

The example discussed in the following is shown in Figure 2.

### Work level

Let us suppose that we have a newsreel about the Landing of the allied forces in Normandy. This is represented by an instance of the Newsreel entity. The "Landing in Normandy" newsreel is composed of several reportages. For instance one where the US president announces the attack; one where the UK prime minister read an official declaration; one showing the soldier disembarking during the attack; one where aerial bombardments are shown. These are represented by instances of the Reportage entity.

We also suppose that we have another newsreel regarding the Second World War. This is again represented by an instance of the Newsreel entity. This newsreel is also composed of several reportages. One of these reportages, the one about bombardments, is the same used in the Landing in Normandy newsreel. In this case, the works corresponding to the landing in Normandy and the Second World War share the work corresponding to the bombardment. It is important to point out that this only means that the description (metadata) is shared. Indeed since the reportage is the same there is not reason to replicate this information. However, it could happen, and in most case it will really happen, that the files containing the two entire newsreels have the same reportage replicated. The result is that the data corresponding to the audio/video stream is replicated, while we have only one description of the reportage.

The existence of an explicit relationship between reportages and newsreels allows users to retrieve and have access to all contexts in which A/V documents or portions of them have been used. This feature, as previously stated, is one emerging type of usage of video material.
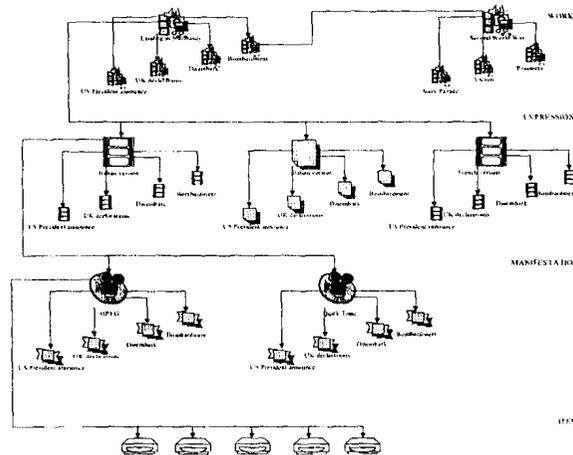


Figure 2: Running example

## Expression level

Let us concentrate on the Landing in Normandy newsreel. Distinct version of a work can be modelled. These versions are implicitly related since they are conceptual realisations of the same work. Let us suppose that, in our example, it was produced in two different versions. One version was produced to be broadcasted in Italy, while the other to be broadcasted in French. The language is Italian in one version and French in the other. However, it is also possible that the Italian version and French version do not differ only for the language, but, for some reasons, some parts may have been cut or modified in both versions. This way, it is possible to inspect, not only how different phrases were translated in different language, but also the differences in which the same episode was reported in different countries. This type of usage is promising for video material accessed by teachers and students, when observing different versions of documents from a conceptual/analytic point of view.

In correspondence of the two versions, we also have two audio/video expressions and of course the corresponding audio only, video only and transcript expressions. In the figure, for simplicity, only the transcript expression corresponding to the Italian version was reported.

The description, contained at the expression level, characterises a particular version of a work. Among the others, here we can find attributes that allows us to perform similarity retrieval as for instance visual features (colour histograms, motion vectors, textures, etc.) of the video documents. In addition, attributes that give information on the style used for a particular version are included. For instance, in case of a video expression, information on the camera movement in a shot are specified along with information on the type of transitions between scenes (e.g. fade in, fade out, etc.). Using this information is possible to enhance the way in which documents can be retrieved, using the concept of similarity retrieval. As previously stated, specialists and researcher may need to analyse A/V material by using this kind of information.

## Manifestation level

At the manifestation level, information on the physical support, on which expressions are available, is included. For instance, in our example, the Italian version of the Landing in Normandy newsreel is available in MPEG and QuickTime formats. This means that the "same" expression can be played using MPEG players and QuickTime players. The content is exactly the same, but is different the way in which it was encoded in the file. Different formats means also to favour different usages. If a document is requested for broadcasting maybe BETACAM is the right format.

Here also portions of the entire documents can be identified. For instance, the reportage corresponding to the bombardment can be precisely identified in the MPEG and QuikTime file.

The possibility to identify portions of A/V material also allows people (documentary makers, news editors journalists, etc.) to reuse these components by creating new documents that are created by extracting segments from other existing documents.

## Item level

At the item level different copies of the same manifestation are described. In our example we supposed that the MPEG file, corresponding to the Italian version of the Landing in Normandy news reel, is mirrored on five different sites. That is, the file or its physical support (CD, tape, etc.) can be found in five different locations. Notice that these different copies may have, for instance, different access speed, preservation quality, therefore different access prices, and different authorisation information associated with them.

## FINAL REMARKS

In designing this metadata model great care was taken to render it extensible. This is a key feature of any metadata model since it permits a successive expansion to cover new emerging needs.

Expansion may regard an addition of a new genre of *Work* form (in our case a new genre of film), or to be an expansion by specification when a certain type of usage claims for more articulated description level of a metadata (metadata qualification), or to simply represent the incorporation into the model of some further concept (metadata).

Now, we are involved to inspect the coverage level allowed by the model in respect to the stated usages and to provide the extensions when needed. In our model, for example, the metadata and attributes with which cinema historians are heavily concerned are only the essential ones.

Another aim in designing the model was the provision for interoperability. We are currently experimenting the mapping of this model into other metadata models to evaluate its interoperability level, starting with those belonging to the IFLA family.

About the reuse of AV material we will also remark that the current version of the model was designed taking into account the semantic approach rather the right approach of reuse. In fact the organization of the model permits to represent quite well the history of the resource content (programme, extracts) along the reuse phases and their context; meanwhile the right requirements for the reuse weren't kept in consideration in our inspection intents.

The reason is that right information often encompasses intellectual property rights, copyright, and other rights. Thus, if rights element is absent, no assumptions can be made about the status of these and other rights with respect to the resource.

To understand if a metadata model, directed to describe AV resources, should supply a rights specialization or an integration among models, applying each on a different aspect of the same domain (description of the resource and rights management), or furthermore assume the form of a e-commercial version [15], is a matter to inquire.

## REFERENCES

1. ABC A Logical Model for Metadata Interoperability.
   *http://www.ilrt.bris.ac.uk/discovery/har mony/docs/abc/abc_draft.html*

2. American Heritage Virtual Archive Project.
   *http://sunsite.berkeley.edu/amher/pr oj.html*

3. A Metadata Model for Historical Documentary Films. Giuseppe Amato, Donatella Castelli, and Serena Pisani. ECDL 2000, LNCS 1923, 320-331.

4. ConText: Towards the Evolving Documentary. Glorianna Davenport, David Murtaugh. ACM Multimedia '95, 580-593.

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

6. 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

7. 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.

8. The FIAF cataloguing rules of Film Archives. *http://www.cinema.ucla.edu/fiaf/english/book.html*

9. Code of Conduct for Access to audiovisual Archives. Steve Bryant. FIAT/IFTA NewsLetter. *http://www.nb.no/fiat/2000-08/bryant.html*

10. How can Content and Asset Management systems solve the problematic issue of providing clarity on rights and copyright clearence for television archives? Catherine Lacken. FIAT/IFTA Vienna World Conference, 2000.

11. The Application of Metadata Standards to Video Indexing. Jane Hunter, Renato Iannella. Second European Conference on Research and Advanced Technology for Digital Libraries, Crete, Greece, September, 1998. *http://archive.dstc.edu.au/RDU/staff/jane-hunter/ECDL2/final.html*

12. 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*

13. The IASA cataloguing rules. *http://www.llgc.org.uk/iasa/icat/icat001.htm*

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

15. Latest version of the model. *http://www.indecs.org/washington/offer.pdf*

16. The needs for and the means to exploit archives of television material. Sue Malden. DELOS Workshop on Audio Visual Digital Libraries, Santorini, Greece, 1998.

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

18 The TV Programmes no one believed existed: Cooperation between Scholars and the Broadcasting Archives. Lasse Nilsson. 19th International Congress of Historical Sciences, Oslo, 2000

19 Building an Audio-visual Digital Library of Historical Documentaries: the ECHO Project, Pasquale Savino, D-Lib, November 2000

20 Trial Publication of RP 210A proposed SMPTE Recommended Practice-Metadata Dictionary. *http://www.smpte.org/stds/*

## ABOUT THE AUTHORS

**Giuseppe Amato** graduated in Computer Science at the University of Pisa, Italy, in 1992. He is a member of the CNR-IEI in Pisa, working in the area of multimedia information systems. His main interests are model for content based retrieval of multimedia documents, access methods for similarity retrieval and approximate similarity search. He has published papers in many international conferences and journals in the areas of database information systems.
E-mail: *G.Amato@iei.pi.cnr.it*

**Donatella Castelli** graduated in Computer Science at the University of Pisa, Italy, in 1983. He is a member of the CNR-IEI in Pisa, working in the area of Digital libraries. Her current research interests include digital libraries systems and related interoperability problems, metadata models and formal methods. She has published papers in many international conferences and journals in the areas of information systems.
E-mail: *D.Castelli@iei.pi.cnr.it*

**Paola Venerosi** gratuated in Political Sciences in 1974 at the University of Pisa, Italy. She was member and now collaborator of the CNR-IEI in Pisa, working in the area of indexing, knowledge organization and library automation. She is interested in the following topics: document representation, e-learning. She published several papers and technical reports.
E-mail: *P.Venerosi@iei.pi.cnr.it*