

# Enriching Multimedia Content Description for Broadcast Environments: From A Unified Metadata Model to A New Generation of Authoring Tool

Wilfried Jouve, Boris Rousseau, Laure Berti-Équille

*INRIA Rennes, Campus de Beaulieu, 35042 Rennes cedex, France*

*{Boris.Rousseau, Wilfried.Jouve, [Laure.Berti-Equille](mailto:Laure.Berti-Equille@irisa.fr)}@irisa.fr*

*Tel : +33 299 847 100 Fax : +33 299 847 171*

**Abstract:** In this paper we propose a novel way for authoring a diversity of multimedia resources (audio, video, text, images, etc). We introduce a prototype authoring tool (called M-Tool) relying on a metadata model that unifies MPEG-21 and TV-Anytime descriptions to edit and enrich audiovisual contents with metadata. Additional innovative functionalities extending the basic functionalities of the M-Tool are also presented. This new generation of metadata authoring tools is designed and currently used for scenarios of TV and News broadcasting, and video on demand broadcasting in the framework of the IST integrated European project ENTHRONE.

**Keywords:** Metadata, Multimedia Authoring, MPEG-21, TV-Anytime, XML, Digital and Audiovisual Resources

# 1 Introduction

The recent explosion of the number of digital sources has profound implication in the domain of multimedia applications. Such a quantity of documents enforces the requirements for efficient ways for describing and exploiting them by various enhanced services with the aim of ensuring end-to-end high quality of service. Furthermore, with the emergence of new services for Video on Demand and Live or Pay TV Broadcast scenarios, current descriptors such as proprietary Electronic Program Guides (EPG) are no longer sufficient. An electronic program guide is a digital representation of a paper-based television program guide used with set-top boxes. It provides the list of current and scheduled TV or video programs associated with a short summary per channel, optional annotations and related functionalities (e.g., for setting up parental controls, ordering pay-per-view programming, searching for programs based on theme or category, or setting up a VCR to record programs).

Most of multimedia applications rely on such content-descriptive metadata. Metadata is essential for understanding content as it describes the resource with regards to its accessibility, organization, configuration, rights, etc. In other words, metadata is essential for understanding and exploring information and more specifically for enabling the representation of multimedia document properties in a structured way that simplifies document management and retrieval.

This paper presents a subset of the work achieved as part of the IST Integrated European Project ENTHRONE (*End-to-End QoS through Integrated Management of Content, Networks and Terminals*) (IST-507637) where content-descriptive metadata management and orchestration are one of the major transversal tasks. In this context, we focused on modeling, designing and developing an innovative metadata authoring tool (called M-Tool) to enrich content-descriptive metadata in a multimedia broadcast environment.

Our contribution in this paper is twofold: firstly, we propose an integrated approach to model and manage multi-media content-descriptive metadata for a wide range of usage broadcast scenarios. This approach is original in the sense that, compared to *ad hoc* or applications

specific, it conciliates the usage of several relevant standards in the same framework. Secondly, we develop a metadata authoring tool (M-Tool) that can be considered as a proof-of-concept of our unified metadata model and that fully implements our vision of metadata management and orchestration with innovative functionalities.

The rest of the paper is organized as follows: Section 2 will present the background of this work namely the Enthroned project and the metadata objectives. Section 3 will cover metadata standards and work related to multimedia document annotation. Section 4 will depict our approach on the one hand for modeling in a unified way content-descriptive metadata and on the other hand for processing metadata. Section 5 will detail the design and implementation of such a multimedia application. Finally, future work and conclusions will be presented in Section 6.

## 2 Background

Among numerous initiatives that have recently emerged in the area of multimedia delivery and applications, ENTHRONE (IST-507637)<sup>1</sup> is an IST project which considers the provision of an integrated management based on the end-to-end *quality of service* (QoS) over heterogeneous networks and terminals. The ENTHRONE project proposes an integrated management solution which covers an entire audio-visual service distribution chain, including content generation and protection, distribution across networks and reception at user terminals. The main objective of the ENTHRONE project is to investigate and develop an integrated solution that is able to manage the functionality of various entities in the digital information distribution chain, from content/service generation to user terminals, using heterogeneous networks and based on the end-to-end QoS approach. Such a digital information distribution chain requires mechanisms to efficiently manage, exchange and adapt distributed audiovisual resources. This is best achieved through the use of various metadata (e.g. information about the identification, the content description, the adaptation of QoS, the characteristics of networks and terminals, complex information concerning structure

---

<sup>1</sup> IST European Integrated Enthroned Project: <http://www.enthrone.org>

description, semantics and contents of data items, and the distribution of digital resources, etc). Metadata must describe the data sufficiently well as to be used as a surrogate for the data when making decisions regarding description, adaptation and use of resources. In the MPEG-21 framework [ISOM21] adopted by the ENTHRONE project, metadata is considered as a transversal keystone of the ENTHRONE project. Metadata is a global resource for achieving interoperable and transparent access to multimedia resources from any type of client terminal or network and for providing an implementation to the *Universal Multimedia Access* (UMA) concept. Metadata describing *digital items* (DI), network, users, services, terminals, and sub-systems configuration is then advantageously used in a distributed way throughout the whole architecture of the project [ENTD03][ENTD15]. Furthermore, metadata is concerned with a wide range of issues such as video on demand and live broadcast scenarios and should gather different requirements from user's to service provider's and various content types. In the next section, we'll present a state of the art of metadata standards for multimedia resources and different authoring approaches found in the literature and related standards.

### **3 Related Work**

#### **3.1 Describing Multimedia Contents with Metadata Standards**

MPEG-21, MPEG-7 and TV-Anytime Phase 1 are established standards for defining multimedia contents. They answer to various needs and thus, they are complementary on several levels. MPEG-21 [ENTSoA] enables a hierarchical representation of heterogeneous contents and provides tools to protect and adapt these contents to the changing characteristics of networks and of client terminals. MPEG-7 [MSS02] supplies a set of tools to describe a wide range of multimedia contents. Although MPEG-7 [ISOM7] covers a wide range of abstraction level, the MPEG-7 strength is inherent in the description of low level information such as texture, color, movement and position information and also in the specifications of description schemes. Its primary goal is to facilitate the management of huge amount of multimedia sources by proposing powerful tools for indexing and searching. In contrast, the TV-Anytime specifications [TVA1][TVA2] provides very little low level information but

provides a large panel of semantically higher level information (e.g., title, synopsis, genre, awards, credits, release information etc.). Besides, since TV-Anytime is initially dedicated to TV services, it handles the segmentation and the repeatable broadcast of multimedia resources or programs. It is also interesting to note that TV-Anytime uses a sub-set of the MPEG-7 standard for low-level information.

As every metadata specification answers to specific requirements, they cannot perfectly meet them all. Therefore, hybrid approaches have been designed to adapt to particular situations. Hybrid approaches (such as the ones covered by the Enthrone project shown in Annex, Table 1) associate complementary features of multiple standards to gain flexibility. Indeed, specific needs imply refining these standards. Durand et al. [DKL05] encapsulate MPEG-21 inside TV-Anytime to realize a service description for scalable interactive TV. The Madeus [JLR98] model associates the MPEG-7 with a SMIL-like mechanism to create complex multimedia documents. Other approaches are especially designed to create interactive multimedia presentation such as the MHEG standard [ME95]. In this context and to cope with the various Enthrone scenarios, we chose a hybrid approach for the metadata model used in the M-Tool.

### **3.2 Other Multimedia Authoring Tools**

Multimedia authoring tools are designed to achieve multiple purposes which can be described as browsing and visualizing, analyzing, segmenting, annotating and integrating. The common purpose of most authoring tools is the capacity of browsing and visualizing the metadata together with the multimedia content. Some authoring tools enable analyzing multimedia contents to provide users with information facilitating temporal segmentation. From these video analyses, semi-automatic or automatic scene boundary detection can also be performed as in the *Video Retrieval and Sequencing System (VRSS)* [CR95] which semi-automatically segments the video for later retrieval. Temporal segmentation can also be done manually by specifying the beginning and the end of each segment similarly to media production applications [ADO05]. In addition to segmentation information, authoring tools enable adding semantic information to the content description. Finally, the integration authoring tools

consist in creating structured multimedia contents by establishing temporal synchronization, spatial layout and hyperlinks to bind multiple contents in advanced multimedia presentations.

Many authoring tools combine several of these purposes, often fostering one aspect. Most of these tools focus on the temporal segmentation aspect disregarding the semantic annotation one. Thus, the IBM's MPEG-7 authoring tool called *VideoAnnex* [IBM02] supplies automatic temporal segmentation of audiovisual contents but only enables the semantic annotation of four elements: events, static scenes, key objects and keywords. In the same way, the family video archive [AGL03] proposes advanced browsing and temporal segmentation features however only three kinds of semantic annotation are available: date, freeform text, and a metadata tag. On the contrary, the semantic annotation aspect constitutes an important M-Tool's feature for videos, still images and multimedia documents.

Advanced browsing features have been developed to perform searching. Some researches focus on video feature-based search; the *ViMeta-Vu* [TCJ01] authoring tool put forward a query system based on automatic segmentation and analysis of low-level information. In contrast, the authoring tool *Silver* [MCS01] uses semantic information from the *Informedia Digital Video Library* [HS95] to perform searching. *Informedia* provides titles and textual transcripts of audio tracks for each segment of audiovisual contents. However, *Silver* only allows users to search exact words in a single free form box in a Google-like fashion and the non-structural textual transcript is not adapted for advanced search.

By providing a full and comprehensive description solution for multimedia contents, the MPEG-7 standard has emerged as the primary choice for mono-media approach in numerous authoring tools [RSK02][IBM02][TR03]. Conversely, very few authoring tool relying on TV-Anytime could be found at the design stage. In contrast to the mono-media approach, interactive hypermedia authoring tools endeavor to integrate several multimedia contents inside a same structure. *Enikos* [ENI05] designed a MPEG-21 authoring tool, called *DIEditor*, allowing users to link multiple resources inside a MPEG-21 structure. However, this *DIEditor* does not enable temporal and spatial segmentation. Moreover resources descriptions are

restricted to a freeform text. The *Mdefi* authoring tool [TR03] combines two models: the MPEG-7 standard and a SMIL-like multimedia model. MPEG-7 is used to describe audiovisual resources and the SMIL-like model to integrate these resources inside a complex multimedia presentation. Owing to the specificity of our requirements in the framework of the ENTHRONE European project, no current solution could answer to our needs (see the various scenarios given in Annex, Table 1). As a consequence, our approach was to combine and unify the TV-Anytime and the MPEG-21 standards. MPEG-21 provides content protection, network adaptation, client terminal adaptation and a structure to link several multimedia contents while TV-Anytime provides content-descriptive metadata and temporal segmentation.

## **4 From the Unified Metadata Model to the Design of the M-Tool**

### ***4.1 A Unified Model for Content-Descriptive Metadata***

TV-Anytime and MPEG-21 are complementary and cover the requirements specified by our partners (RBB, City of Metz, France Telecom, Thales Broadcast and Multimedia, Optibase) for Video On Demand, TV broadcast, and news broadcast scenarios. Although initially dedicated to TV services, the TV-Anytime specification enables the description of all these services and provides powerful tools to perform advanced search on a huge amount of heterogeneous contents. Our approach is that TV-Anytime alone is not sufficient as it neither defines transport mechanism nor the hierarchical structure to represent a mix of multimedia contents. The use of MPEG-21 adds the flexibility mandatory to adapt to these various services and situations. MPEG-21 enables the hierarchical representation of multimedia contents which is useful to create advanced and interactive multimedia contents. Every resource is described using the TV-Anytime standard and then, resources are gathered and synchronized inside a MPEG-21 structure that includes: 1) network quality of service and adaptation metadata and 2) right protection metadata.

- 1) MPEG-21 *Digital Item Adaptation* (DIA) metadata [ENTSoA] describes how multimedia contents should be adapted to network characteristics and client terminals. DIA metadata consists of *adaptation QoS* (aQoS) and *Usage Environment Description* (DIA UED) metadata. The aQoS provides the required information to select optimal adaptation parameters. The description of terminal capabilities given by DIA UED metadata is primarily required to satisfy consumption and processing constraints of a particular terminal. DIA UED also specifies network characteristics in terms of network capabilities and conditions, including available bandwidth, delay and error characteristics.
- 2) MPEG-21 *Right Expression Language* (REL) [ENTSoA] specifies whether a given group of people can perform a given right upon a given resource under a given condition. MPEG-21 *Intellectual Property Management and Protection* (IPMP) manages rights and intellectual property of a specific resource.

The metadata description schema is illustrated in Figure 1; TV-Anytime metadata (1) is encapsulated inside the MPEG-21 structure and each MPEG-21 descriptor contains REL, DIA or TV-Anytime information. Metadata is seen as a global resource that will be used in a distributed way throughout the architecture of the project and will feed or be generated by different component of the Enthroned system. Content-descriptive metadata describes a resource (e.g. title, synopsis, etc.) in addition to dedicated fields for content providers' specific needs and with potentially all the fields proposed by TV-Anytime. Right protection metadata (2) manages rights and intellectual property of the specific resource (REL and IPMP). Network QoS metadata (3) is information necessary to adapt content delivery with network and QoS parameters (DIA). User metadata (4) is information on the end user such as preferences, interest, etc. Terminal metadata (5) summarizes terminal capabilities such as display resolution. Content providers (CPs) and subsystems configuration allow content providers and Enthroned subsystems metadata (6) to be configured with parameters and commands.

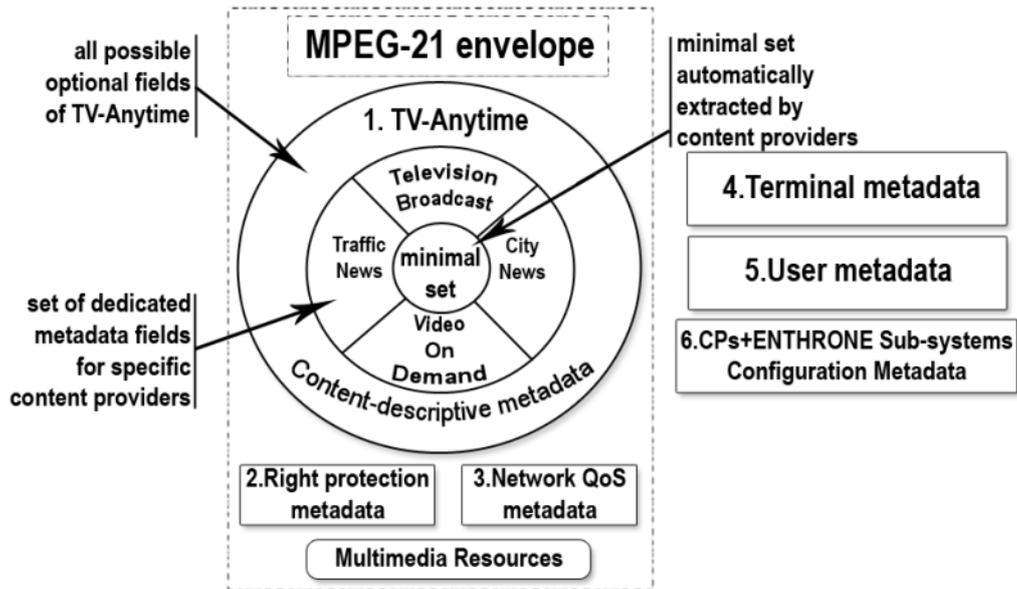


Figure 1: Metadata Model in the Enthroned Project

This model is well adapted to enrich content-descriptive metadata and to integrate annotated resources inside the MPEG-21 structure for various broadcast scenarios.

## 4.2 Innovative Processing of Metadata

The conciliation of several relevant standards in a unified model for content-descriptive metadata emphasizes our vision of metadata management, composition and orchestration. Our prototype authoring tool takes advantage of such model to provide an easy to use and flexible user-facing application that facilitates composition of metadata with possible associated adaptation (MPEG-21 DIA) and rights (MPEG-21 REL) information. Furthermore, annotation is transparent to users. In other words, the user neither needs any prior knowledge of MPEG-21 nor TV-Anytime specification.

An essential point is that the M-Tool handles and enriches metadata for a mix of resources (video, text, audio, images, etc) which often implies different content providers' needs. Also, the fact that our prototype offers a number of editable fields from both TV-Anytime and MPEG-21 standards will facilitate advanced searches.

Several innovative aspects under current development and validation extend the M-Tool with the following enhanced functionalities:

- ***Enhanced mixed search coupling exact keywords search and more complex XQueries on available XML metadata and content-based information retrieval.*** Beside exact keyword search and queries on XML metadata files, searching images and videos is based on a library of similarity searches on low-level features on videos, images and multimedia documents (e.g., MPEG-7 image global descriptors such as *ColorLayout*, *ScalableColor* for colour, *HomogeneousTexture*, *EdgeHistogram* for texture and *RegionShape* for shape; local image descriptors such as interest points; segmentation and face recognition tools for video, etc.).
- ***Multi-view navigation and visualization of large collections of digital items:*** browsing and synchronized visualization of media and metadata is a challenging task for large collections of data, and in this context, we propose self-organizing graph-based views of digital item nearest-neighbors. Navigation and visualization in the digital item collections will be declined in three kinds of views : *i)* the structural DI organization view based on the MPEG-21DID hierarchical structure of items, sub-items and components; *ii)* the locality-based DI organization view that describes the collection of data on devices that exist around a specific context (e.g., location or user or community of users) useful for location-aware queries; and *iii)* the semantic proximity view based on LRU (*Least Recently Used*) and history policies that captures recent history (such as caches for example) and orders the semantic neighbours such that the most recent DI annotation is placed at the top of the list and removes the least recent one if needed.
- ***User profile-based personalization for metadata annotation and multimedia information retrieval:*** based on our previous work in the e-learning community [RBM04], we propose a user profile-based personalization model of the annotation and search processes and their associated client interfaces by tracing, retrieving and inferring information from the users and their preferences in a non-intrusive way.
- ***Fusion of annotations and XML metadata synchronization:*** Merging several XML trees of metadata files and several versions of annotations is a very useful enhanced

functionality for the annotators; this may allow integration, re-utilization and propagation of some descriptive parts of existing available annotations.

- **Metadata monitoring and control:** Because metadata and annotations will be transmitted along with the content over a variety of physical media using appropriate protocols, these metadata should be valid, up-to-date and consistent with their related media content. Controlling, refreshing and synchronising metadata files in conformance with their content include detection and elimination of XML metadata (sub-)item (or component) duplicates, and checking freshness and consistency of both metadata associated to distributed digital resources.
- **Metadata orchestration:** one of the long-term goals of the M-Tool is to allow the description and formalization of metadata orchestration mechanisms being the editor and syntax interpreter of an OCL-like programming language to express various constraints for DI adaptation based on metadata describing digital contents, network, adaptation QoS, terminal characteristics, right protection and user preferences.

In the next section, we describe the main functionalities of the metadata authoring tool (M-Tool) divided in two main applications that can be either independently used or combined to create a MPEG-21 file containing TV-Anytime information: the first application consists in authoring TVA metadata and the second consists of authoring MPEG-21 digital items. These applications are designed to be the core of the metadata management module.

## **5 M-Tool Design and Implementation**

Our metadata authoring Tool (M-Tool) is a prototype that offers the functionalities required to create or manage metadata information relevant to a particular resource. As described in Section 4, its aim is to provide the end-user and service providers with modules that enable annotating, browsing, maintaining, monitoring and querying content-descriptive metadata. The motivation for developing the M-Tool is to address a wide range of issues regarding metadata database-backed management and digital resources manipulation under the MPEG-21 framework. It is designed to help users in annotating and authoring multimedia resources.

In other words, this application allows users to create objects' relationships to each other in order to enrich multimedia resources with metadata both formatted and stored in XML or in a relational database. This type of service is useful for applications that present a mixture of textual, graphical, and audio data. The M-Tool adaptation is performed by two ways: The M-Tool is a modular authoring tool (the user chooses displayed areas and their sizes) and configurable (the user chooses editable information on account of a profile edition module). The next section will provide a thorough description of both applications and how they work together. More details can be found on the "Metadata Authoring Tool TVM Processor" document [ENTD15].

## **5.1 M-Tool for MPEG-21**

The M-Tool for MPEG-21 serves as the keystone to the prototype multimedia authoring tool as it centralizes all download of manually or automatically extracted metadata from various content providers (CPs). Once downloaded, these metadata files are locally stored and its list is displayed to the user. The user is then allowed to select one or more metadata document which is analyzed and visualized using the relevant (MPEG-21 or TV-Anytime) graphical user interface (GUI). Users have the possibility to (1) create a new MPEG-21 document, (2) edit, delete, convert or (3) send this metadata document to a specific (local or external) metadata database.

### **5.1.1 Creating a New MPEG-21 Document**

When a user requests to create a new MPEG-21 DID document, the M-Tool is invoked to create a new MPEG-21 structure. In this case, no download request to get metadata is required as the user is the actual service provider. The DID generation is achieved through a graphical representation of the MPEG-21 structure as shown in Area 1 in Figure 2. New elements are simply added by performing a drag and drop from the toolbox (Area 2) to the MPEG-21 structure. These elements consist of MPEG-21 structure elements (Container, Item and Component) and Descriptor elements (TVA schedule event, TVA program information, REL and DIA metadata). The M-Tool for MPEG-21 is a DID editor and generator. A DID is

edited or generated by inserting/deleting or modifying TV-Anytime metadata on account of the M-Tool for TV-Anytime or DIA and REL metadata on account of DIA and REL generators.

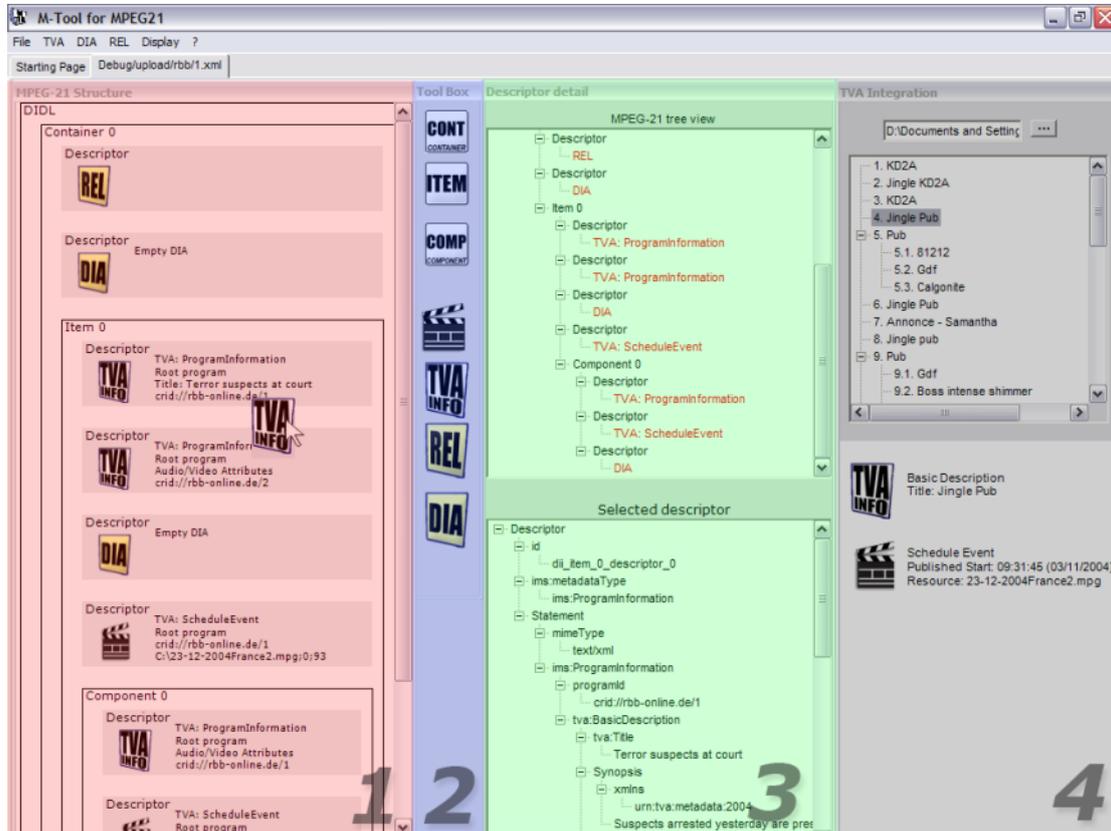


Figure 2: M-Tool MPEG-21 Editor

### 5.1.2 Edition and Annotation

The MPEG-21 DID editor is composed of four areas as represented in Figure 2: the DID structure visualization (1), a toolbox (2), the tree visualization (3) and the TV-Anytime integration area (4). Similarly to the creation of a new MPEG-21 document, updates to the current document are achieved by dragging and dropping different icons from the Area 2 to Area 1. Basic information about descriptors is provided by the graphical representation: the type of descriptor (DIA, REL, Program Information, and Schedule Event), the type of program information (general information or only Audio/Video attributes), the type of program (root or sub program) and the CRID.

Such events change the DID structure and generate relevant dialog boxes accordingly. Selecting a particular descriptor triggers the appropriate tool for edition. For example, TV-Anytime information can be edited straight away by a restricted part of the M-Tool for TVA. Area 3 (tree view) is complementary with the DID representation; it gives a detailed XML representation of all descriptors. Selecting fields in the top tree display its full content in the bottom tree. Area 4 provides the functionality to open a TV-Anytime file separately and drop selected fields into the DID structure. In other words, the TV-Anytime document structure in sub-programs is shown to allow users, upon selection of a program or sub-program to drop its content into the DID structure.

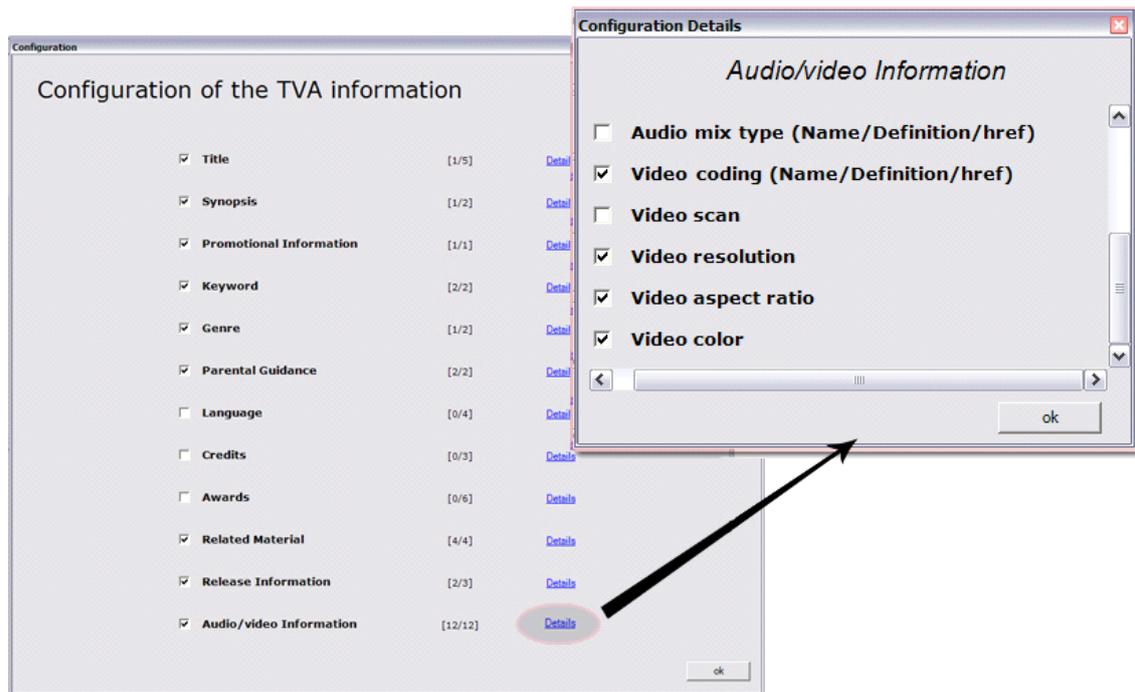
## **5.2 *M-Tool for TV-Anytime***

The M-Tool for TV-Anytime has functionalities to load, annotate, enrich or create a new TV-Anytime document while keeping any edition compliant with the schema and in synchronization with the audiovisual resource. It could be executed as a standalone application and is composed of two main modules: the configuration module and the TV-Anytime editor itself.

### **5.2.1 Configuration and Adaptation to User**

The configuration module (as shown in Figure 3) allows user to select relevant TVA fields depending on their requirements for annotation. Selection of the relevant TV-Anytime elements permits personalizing its optional sub-elements.

The current selection is recorded as a profile, loaded automatically for future use of the M-Tool. Modifications to the current profile are possible through an option in the menu. Additionally, this functionality is flexible enough to allow easy addition of new elements to this set of TVA fields.



**Figure 3: M-Tool for TV-Anytime Configuration Module**

## 5.2.2 TV-Anytime Editor

The M-Tool for TV-Anytime allows edition of multiple files at the same time and text mode edition. It also adapts its interface according to the fact the audiovisual content is available or not; besides, the interface is easily modifiable (hiding, moving, maximizing modules, etc).

The TV-Anytime editor presents the video through several views (see Figure 4): the program description view (Areas 1 and 2), the video player view (Area 3), the sub-programs view (Area 4), the TV-Guide view (Area 5), the timeline view (Area 6), the thumbnails view (Area 7), the history (Area 8) and the text mode edition (Area 9).

The Annotator is a form consisting of several editable fields as requested by the user with the configuration module or taken from the profile. The annotator is divided into two sections: program information (Area 1) and program location (Area 2) of the current program schedule as shown in Figure 4. The remaining areas enable browsing and synchronized visualization of the resource. The media player (Area 3) displays relevant program based on the selected program in synchronization with Area 1, the sub-programs view (Area 4), the TV-guide representation (Area 5), the time line (Area 6) and thumbnails view (Area 7) of the metadata document. The time line visualization gives an overview of such document through a color

caption (annotated region in blue, non-annotated region in red, overlapping programs in violet and blue patchwork and current program or sub-program in white and blue patchwork). As shown in Figure 4, when the mouse is over the mosaic, basic information is shown to the user. Finally, the history (Area 8) provides ways to undo/redo previous actions such as addition, deletion and modification of programs while the text mode visualization allows changing the XML document directly for advanced users.

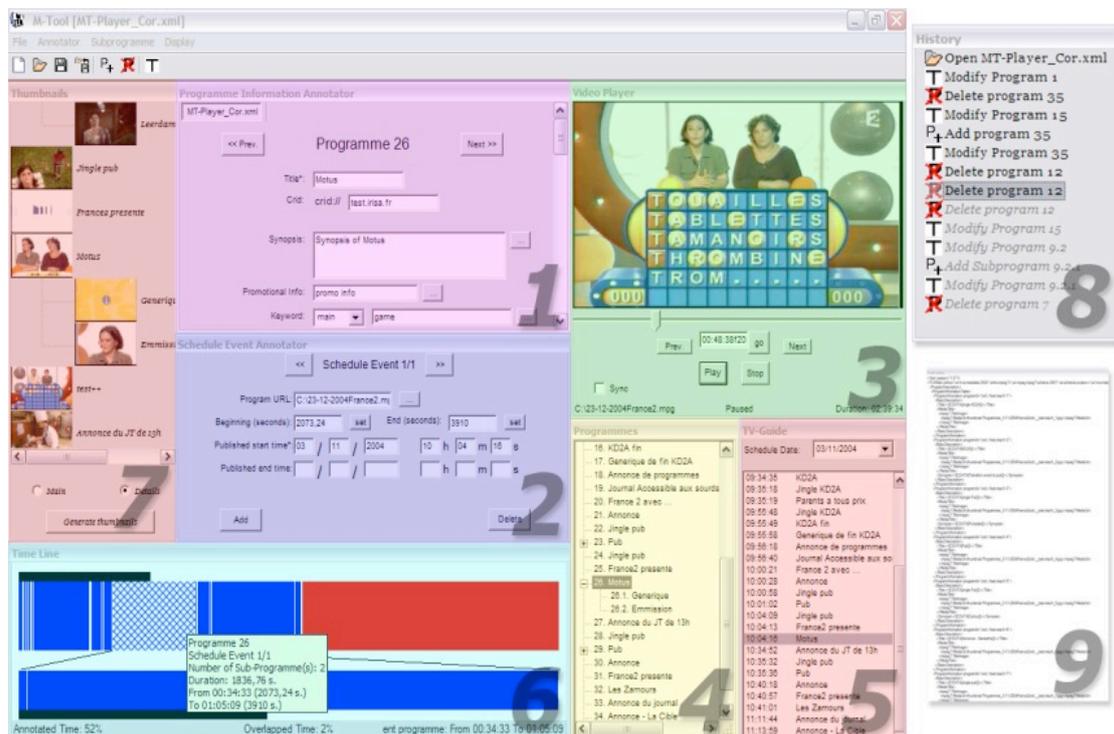


Figure 4: M-Tool TV-Anytime Editor

## **6 Conclusion and Future Work**

In this paper we have proposed an innovative approach that unifies MPEG-21 and TV-Anytime descriptions for modeling in a unified way content-descriptive metadata associated to audio-visual resources. We also described the basic and enhanced functionalities of the authoring tool (called M-Tool) that serves as a proof of concept implementing our vision of metadata management and orchestration for TV, News and video on demand broadcasting scenarios in the framework of the European ENTHRONE Project. Future versions of the M-Tool will fully implement advanced features and will take into account new possibilities offered by TV-Anytime Phase 2 [TVA2]

## **7 Acknowledgments**

The authors would like to thank the IST ENTHRONE project consortium for their collaborations, fruitful exchanges and discussions.

## 8 References

- [ADO05] Adobe Systems Inc., Adobe Premiere Pro 1.5. <http://www.adobe.com/>. Last accessed on the 13<sup>th</sup> of July 2005.
- [AGL03] G. D. Abowd, M. Gauger and A. Lachenmann, “The Family Video Archive: An Annotation and Browsing Environment for home movies”, Proceedings of the ACM conference on Multimedia Information Retrieval, November 2003.
- [CR95] T.S. Chua and L.Q. Ruan. “A video retrieval and sequencing system”. Proceedings of the ACM Transactions on Information Systems, pages 373-407, October 1995.
- [DKL05] G. Durand, G. Kazai, M. Lalmas, U. Rauschenbach and P. Wolf, “A metadata model supporting scalable interactive TV services”, Proceedings of the IEEE conference on Multi-Media Modeling, January 2005.
- [ENI05] Enikos DI Creator and Browser, <http://www.enikos.com/home.shtml>. Last accessed on the 13<sup>th</sup> of July 2005.
- [ENTD03] Enthroned deliverable D03 (WP2) – “Metadata definition and specification”, May 2004, INRIA.
- [ENTD15] Enthroned deliverable D15 (WP4) – “Metadata Authoring Tool TVM Processors”, June 2005, INRIA.
- [ENTSoA] Enthroned WP3 tutorial, “MPEG-21: A State-of-the-art Survey”, March 2005, INESC
- [HS95] A. G. Hauptmann, and M. Smith, “Text speech and vision for video segmentation: The informedia project”, Proceedings of the AAAI Symposium on Computational Models for Integrating Language and Vision, 1995.
- [IBM02] IBM research. “VideoAnnEx - IBM MPEG-7 Annotation Tool”, 2002.
- [ISOM21] ISO/IEC PDTR 21000-1, “Information technology — Multimedia framework (MPEG-21) - Part 1: Vision, Technologies and Strategy”, December 2003
- [ISOM7] ISO MPEG-7, part 5 - Multimedia Description Schemes. ISO/IEC JTC1/SC29/WG11/N4242, 2001.
- [JLR98] M. Jourdan, N. Layada, C. Roisin, L. Sabry-Ismaïl, and L. Tardif. “Madeus: an Authoring Environment for Interactive Multimedia Documents”, Proceedings of the ACM conference on Multimedia, 1998.
- [MCS01] B. Myers, J. P. Casares, S. Stevens, L. Dabbish, D. Yocum and Albert Corbett, “A Multi-View Intelligent Editor for Digital Video Libraries”, Proceedings of the ACM/IEEE-CS joint conference on Digital libraries, June 2001.
- [ME95] T. Meyer-Boudnik and W. Effelsberg: “MHEG - An Interchange Format for Interactive Multimedia Presentations”, IEEE Multimedia Magazine, Spring 1995.
- [MSS02] B. S. Manjunath, P. Salembier and T. Sikora, “Introduction to MPEG-7: Multimedia Content Description Interface”, ISBN: 0-471-48678-7, Wiley, April 2002.
- [RBM04] B. Rousseau, P. Browne, P. Malone and M. ÓFoghlú, “User Profiling for Content Personalisation in Information Retrieval”, Proceedings of the ACM Symposium on Applied Computing, March 2004.

- [RSK02] J. Ryu, Y. Sohn, and M. Kim, "MPEG-7 Metadata Authoring Tool", Proceedings of the 10<sup>th</sup> ACM International Conference on Multimedia, pages 267-270, December 2002.
- [TCJ01] J. E. Tandianus, A. Chandra and J. S. Jin, "Video Cataloguing and Browsing", Proceedings of the workshop on Visual Information Processing, 2001.
- [TR03] T. Tran-Thuong, C. Roisin, "Multimedia Modeling Using MPEG-7 for Authoring Multimedia Integration", Proceedings of the ACM conference on Multimedia Information Retrieval, November 2003.
- [TVA1] European Telecommunications Standards Institute, "ETSI TS 102 822-3-1 v1.2.1; Broadcast and On-line Services: Search, select and rightful use of content on personal storage systems ("TV-Anytime Phase 1"); Part 3: Metadata. Sub-part 1: Metadata Schemas", September 2004.
- [TVA2] European Telecommunications Standards Institute, "ETSI TS 102 822-3-3; Search, select, and rightful use of content on personal storage systems ("TV-Anytime Phase 2"); Part 3: Metadata; Sub-part 3: Extended Metadata Schema", May 2005

## 9 Annex

<i><b>Enthroned Applications</b></i>	<i><b>Metadata format</b></i>	<i><b>Content decoding</b></i>	<i><b>Terminal Platforms</b></i>
Free TV Broadcast	MPEG-21 DID TV-Anytime	MPEG-2 over DVB-T	Television set, set top box, mobile DVB-T device.
iTV-MHP Service	MPEG-21 DID TV-Anytime	MPEG-2 over DVB-T	Television set, set top box with PSTN connection over ADSL. Mobile devices with PSTN connection over GPRS/UMTS. Support DVB MHP-compliant services.
On demand video services to mobile devices with QoS over cooperating DVB-T / GPRS	QoS metadata MPEG-21 DID	MPEG-2 over DVB-T	Mobile devices with PSTN connection over GPRS/UMTS.
Broadcast TV service to mobile devices over DVB-T	QoS metadata MPEG-21 DID MPEG-21 DI	MPEG-2 over DVB-T	Mobile devices with PSTN connection over GPRS/UMTS.
information services on public transportation	MPEG-21 DID and DI Metadata for enhanced information (GPS, time)	MPEG-2 over DVB-T	Public PC terminal (a.k.a. Collective Info-Mobile Terminals).
Individual on demand information	MPEG-21 DID and DI Metadata for enhanced information (GPS, time)	MPEG-2 over DVB-T	PDA/smart phone with WLAN (802.11x) connectivity.
Video on Demand	MPEG-21 DID and DI HTML	MPEG-2/4 over DVB-T/S, UMTS/GPRS or IP.	Private PC/STB terminal with broadband PSTN connectivity (ADSL, cable-modem, optical fiber). PDA or mobile phones with UMTS/GPRS connectivity.
Multimedia content download	MPEG-21 DID and DI. HTML browser IP protection, licenses needs	MPEG-2/4 A/V decoders, MPEG-21 DI and DID parsers and encoders. MPEG-7 parser and encoder.	Fixed (PC/STB) or mobile (PDA or mobile phones) terminals either with or without broadband PSTN connectivity (ADSL, cable-modem, optical fiber).
Pay-TV	MPEG-21 DID and DI HTML TV-Anytime	MPEG-2/4 over DVB-T/S or UMTS/GPRS or IP.	Fixed (PC/STB) or mobile (PDA or mobile phones) terminals either with or without broadband PSTN connectivity (ADSL, cable-modem, optical fiber).
Distance learning	MPEG descriptors and Digital Item or other format that are supported by the infrastructure (FTP, chat, whiteboard).	MPEG-2/4 over DVB-T, UMTS/GPRS or IP.	Fixed (PC/STB) or mobile (PDA or mobile phones) terminals either with or without broadband PSTN connectivity (ADSL, cable-modem, optical fiber).

**Table 1. M-Tool's Usage Scenarios**