INTERNATIONAL ORGANIZATION FOR STANDARDIZATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC1/SC29/WG11
CODING OF MOVING PICTURES AND ASSOCIATED AUDIO

ISO/IEC JTC1/SC29/WG11
MPEG99/M5267
October 1999, Melbourne, Australia

Title: **Limitations of the MPEG-7 Generic DS: Reorganizing the Syntactic/Semantic DS's**
Status: **Proposal**
Source: **University of Brescia**
Authors: Riccardo Leonardi, Giovanni Paltenghi and Lorenzo Rossi

1 Introduction

In this document, we propose some modifications to the MPEG-7 Description Scheme (DS) [1] in order to enrich the structure of the Syntactic DS and Semantic DS by addressing some functionalities for semantically characterizing segments and for highlighting and ordering key-items in a multimedia document.

In our opinion, the Generic DS and in particular the syntactic DS can demonstrate some weakness in describing hierarchically organized documents. In other words, even if it is enunciated [1] that the Syntactic DS should act as the Table of Contents for the multimedia document being described, the description of the document temporal structure seems complicated. Therefore we start our discussion by implementing the ToC DS part of our old MPEG-7 proposal (the ToCAI DS [3]) using the MPEG-7 Generic DS [1]. In our opinion, due to the simpler structure of the ToC DS, this implementation allows to show the complexity of the MPEG-7 DS. For overcoming such a problem, we propose a simble extension of the *Syntactic DS* of the MPEG-7 Generic DS in order to handle semantic aspects of each segment directly at the *Segment DS* level.

Another issue that we analyze in this document is a possible extension of the MPEG-7 Generic DS for the inclusion of some important functionalities: the capability to (1) highlight description items (e.g. images, sounds, events, objects etc.) most relevant to the purpose for which a certain content description of a multimedia (MM) document has been created and (2) the capability of description information ordering[1]. In other words, due to a possible large amount of description items, an entity who will create descriptions of multimedia (MM) documents, according to MPEG-7 specification (i.e. a **description provider**), shall highlight certain items most representative for the kind of document being described in order to facilitate user queries. Besides we consider the need of providing users with ordering mechanisms a very relevant issue for MPEG-7. Such ordering mechanisms are derivable

[1] The requirements for highlighting and ordering key-items were introduced, among the functional requirements (# 11 and 12), at the Vancouver meeting.[2].

from descriptors (e.g. a set of key – frames ordered on the basis a color descriptor or a set of sounds ordered by means of a loudness D). However a possible large variety in the types of descriptors composing a description could lead to a consequent high number of ordering criteria to arrange description items. Therefore we propose that the description provider should also select which set of descriptors should be combined to order a subset of description elements (e.g. key frames, events etc.) most pertinent to the MM document being described.

The document is organized as follows: in Section 2 we explain the motivations for a representation of the ToC DS based on the MPEG-7 Generic DS. In Section 3 after a brief overview of the ToCAI DS, we present the implementations of the *ToC DS* according to the MPEG-7 Generic DS specifications; we also suggest in this sections some changes to the current specifications to better handle the ToC DS functionalities. Section 4 provides an example of implementation of such a DS. In Section 5, we explain, after a quick overview of the current Generic DS, the motivations behind the proposal for adding highlighting and ordering functionalities. In Section 6, we show the structure of the DS that enable these functionalities. In Section 7, we give an example in order to clarify the concepts of key-items and ordering keys. Finally in Section 8, we provide a brief summary of the contribution.

2 Motivation for a representation of the ToC DS based on the MPEG-7 Generic DS

The MPEG-7 Generic DS was generated as a visual DS during the MPEG Seoul Meeting (March 1999) [4]. After the Vancouver meeting, it has being extended to describe the content of audio-visual documents [1]. The Generic DS is composed of three main components: the *Syntactic structure DS*, the *Semantic structure DS*, and the *Syntactic-semantic links DS*. The *Syntactic structure DS* consists of region trees, segment trees, and segment/region relation graphs. The *Semantic structure DS* is composed of object trees, event trees, and object/event relation graphs. The *Syntactic-semantic links DS* provide a mechanism to link the syntactic elements (regions, segments, and segment/region relations) with the semantic elements (objects, events, and object/event relations), and vice versa.

The ToCAI idea comes out from the structure used for technical books. One may easily understand a book sequential organization by looking at its table of contents while quickly retrieve elements of interest by means of the analytical index. This DS provides therefore a hierarchical description of the temporal structure of a multimedia document from a semantic point of view at multiple level of abstraction (thanks to the ToC), so as to have a series of consecutive segments which are coherent with the semantic of information at that level. This type of indexing procedure allows a rapid navigation through the multimedia document. The "Analytical Index" (AI) of key-items of the document is suitable instead for effective retrieval. It allows an easy way to effectively retrieve relevant information, such as relevant images, sounds or events. To ease retrieval tasks, it is important that these items be arranged in the AI according to various ordering criteria, so as to ease the retrieval task.

In this document, we show how the ToC part of the ToCAI DS can be represented by the MPEG-7 Generic DS. Since the ToC part includes both syntactic and semantic concepts, it needs for implementation the *Syntactic DS*, the *Semantic DS*, and the *Syntactic-semantic Link DS*. As it will be shown, this representation turns out to be very complicated. We propose thus a possible extension of the *Segment DS* introducing a simple mechanism to handle semantic aspects directly at the segment level. As a result a hierarchical representation of the temporal structure of an AV programme would become much simpler, thus facilitating

navigation tasks[2].

3 Implementations of the *ToC DS* by means of the MPEG-7 Generic DS

3.1 ToCAI overview

In Figure 1, is shown the ToCAI DS, it is created by the aggregation of four main description schemes: the Table of Contents (ToC), the Analytical Index (AI), the Context and the Meta-descriptors description schemes. The roles of ToC and AI part have been introduced in the previous paragraph. The *Context DS* describes the category of the audio-visual material. This *Context DS* includes descriptors such as title of programme, actors, director, language, country of origin, etc. Indeed these informations are necessary for retrieving purposes to restrict the search domain. In the Generic DS, these informations are carried out by the *Meta information DS,* which contains descriptors carrying out author-generated information about an AV program that cannot usually be extracted from the content itself. The *Meta-descriptors DS* has the role to incorporate in the ToCAI DS a set of descriptors carrying information about how accurate is the description and by which means it has been obtained.

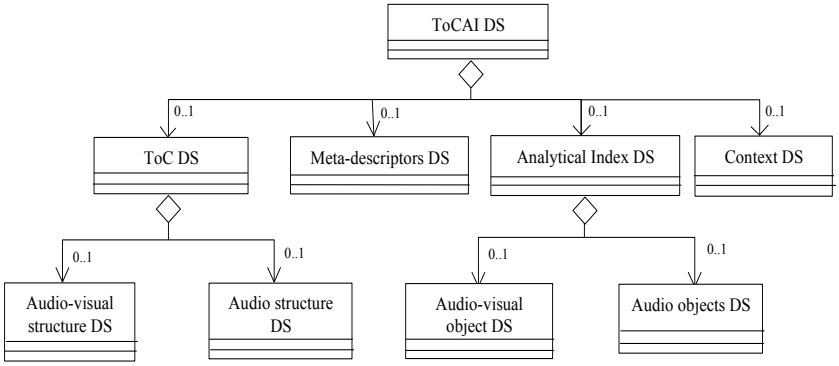

Figure 1: Structure of the ToCAI DS.

The *ToC DS* describes the temporal structure of the AV document at multiple level of abstraction. It contains two DSs, explained below, namely *Audio-visual Structure* and *Audio Structure*. In Figure 2, is shown the *Audio-visual Structure DS*. The two *Time-code Ds* specify the start and the end position of the AV document. The core of this DS is the *Scene DS*. A scene is a temporal segment having a coherent **semantic** at a certain hierarchical level. It is composed by a various number of sub-scenes, a time reference (two time-code Ds) and a *Type of Scene D* (a string and, if useful, a characteristic icon for visualization purpose). The elementary component of a scene is the shot (**syntactic** element). The *Shot DS* contains the type of editing effects at its boundaries.

[2] This proposed modification can suggest to rename the *Syntactic DS* as "structural" DS, as semantic concepts are now included in the Syntactic DS as well.

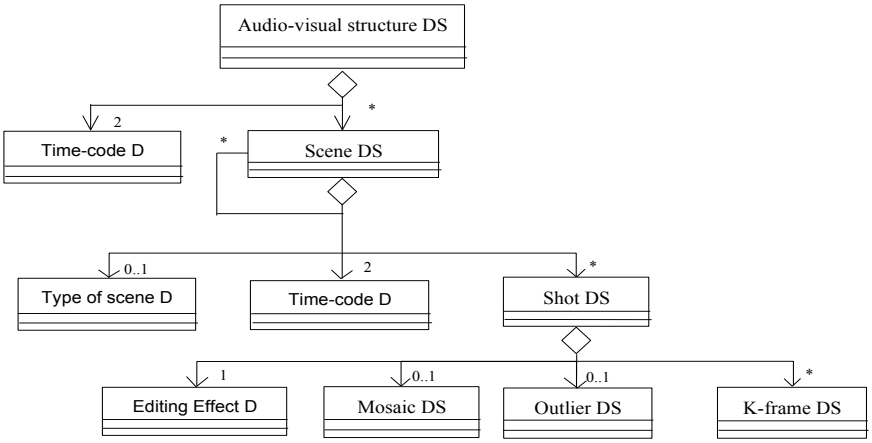

Figure 2:**The Audio-visual structure DS.**

3.2 Representation of the ToC DS using the Generic DS.

For implementing the *ToC DS,* we can use the *Semantic DS* for the representation of the semantic part of the *ToC DS*, the *Syntactic DS* for the representation of the syntactic part of the *ToC DS*, and linking the two parts by the *Syntactic-semantic link DS*. In Figure 3, this implementation is suggested.

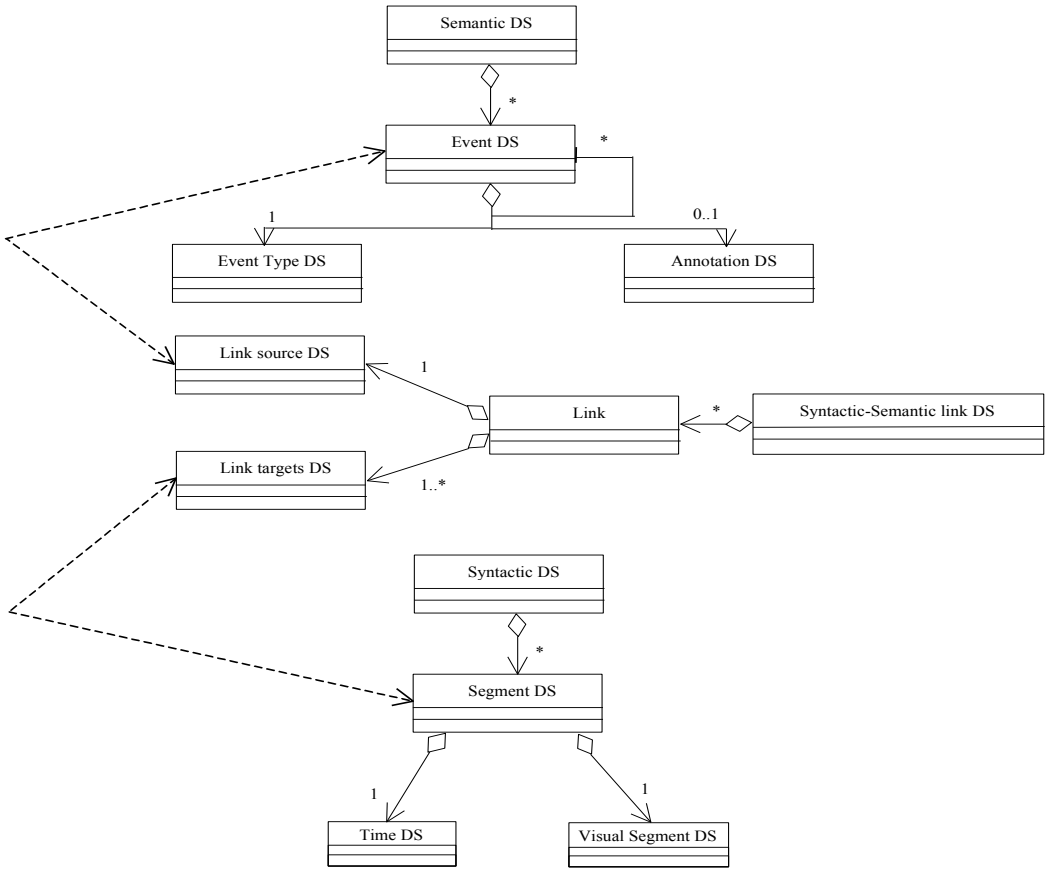

Figure 3: Syntactic-Semantic DS.

As we can see from Figure 3, we link every terminal scene (event) of the ToC tree with the respective shots (segment) that compose the scene. The proposed structure is very intricate for practical use. In particular it does not allow to easily search for information as a back and forth traversing strategy through the syntactic-semantic link between data stored in the syntactic and semantic DS's. For this reason, we propose a modification of the *Segment DS* that allows a simple implementation of the *ToC DS*.

3.3 The proposed modification

With the aim of a simpler implementation, of the *ToC DS*, we introduce the ***Type of Segment DS*** that allows to describe the semantic content of the segment, as shown in Figure 4. The *Type of Segment DS* specifies whether the segment is a "Scene" (and what type of scene), a "Shot", a "Program item", …

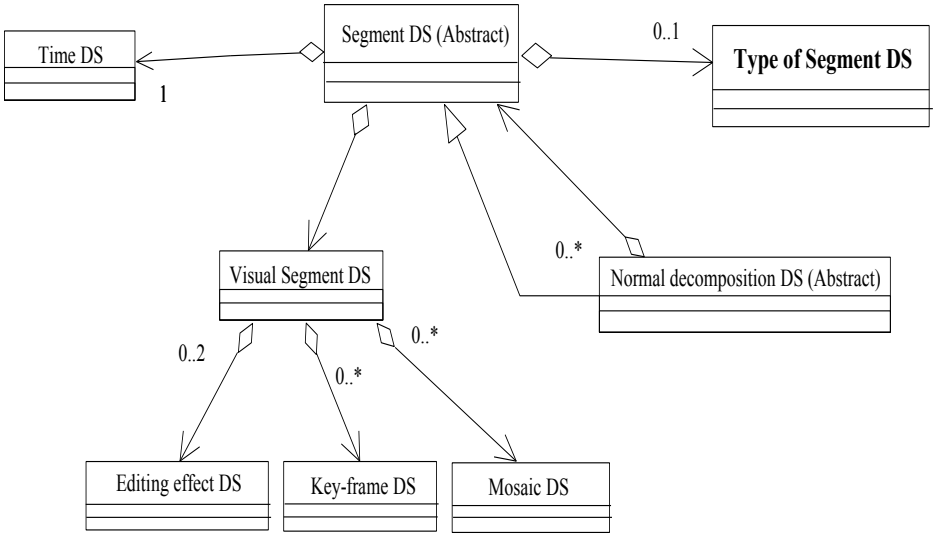

Figure 4: Modified Segment DS.

We can see the *Type of segment DS* as an evolution of the *Segment level DS*, which indicates the position of the associated segment in the tree hierarchy. The *Visual segment DS* describes the editing effects of the shots (it plays the same role of the *Shot DS* in the *ToC DS*).

4 Example

In this section, we provide an example of an audio-visual document described by the *ToC DS* and we show the representation of such a structure by using initially the Generic DS and then the proposed modified *Segment DS*._The considered audio-visual document is a broadcast news programme which presents the following structure:

- Summary (0:2'30'')
- Internal affairs (2'31':7')
  - Speaker presentation (…)
    - Shot 1 (…)
    - Shot 2 (…)

- …
- First reportage (…)
  - Shot 1 (…)
  - Shot 2 (…)
  - …
- Speaker presentation (…)
- ….

- International affairs (…)
  - ….

where the data within parenthesis specify the temporal location of the segment, while the label indicates its semantics. Every ToC item may be summarized by key-frames and audio segments.

In Figure 5, a UML-like structure which represents the ToC description of the proposed example is shown. The leaves of the structure are the basic shots forming the document.

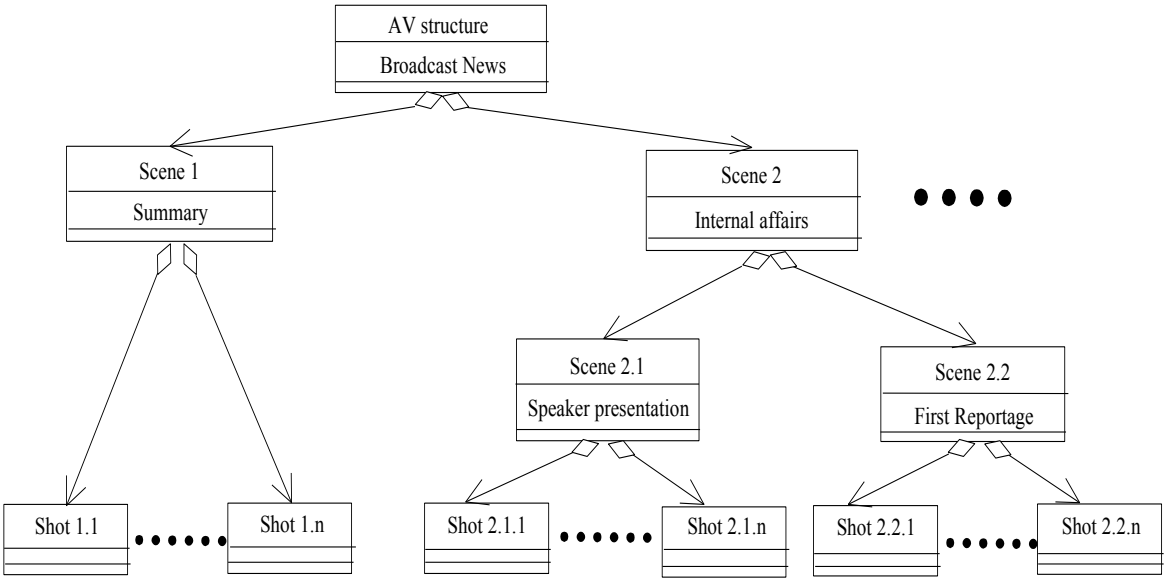

Figure 5: Representation of the AV document, used as example, based on the Toc DS.

In Figure 6, the representation of the AV document 'Broadcast News', by means of the Generic AV DS is shown. As we can see, the *Semantic DS* (*Event DS*) is adopted for the representation of the semantic concepts (Summary, Internal affairs, Speaker presentation, First reportage, …), and the *Syntactic DS* (*Segment DS*) for the representation of the syntactic concepts (Shots). Then we link the semantic and the syntactic concepts by the *Syntactic-semantic link DS* (dotted line in figure).

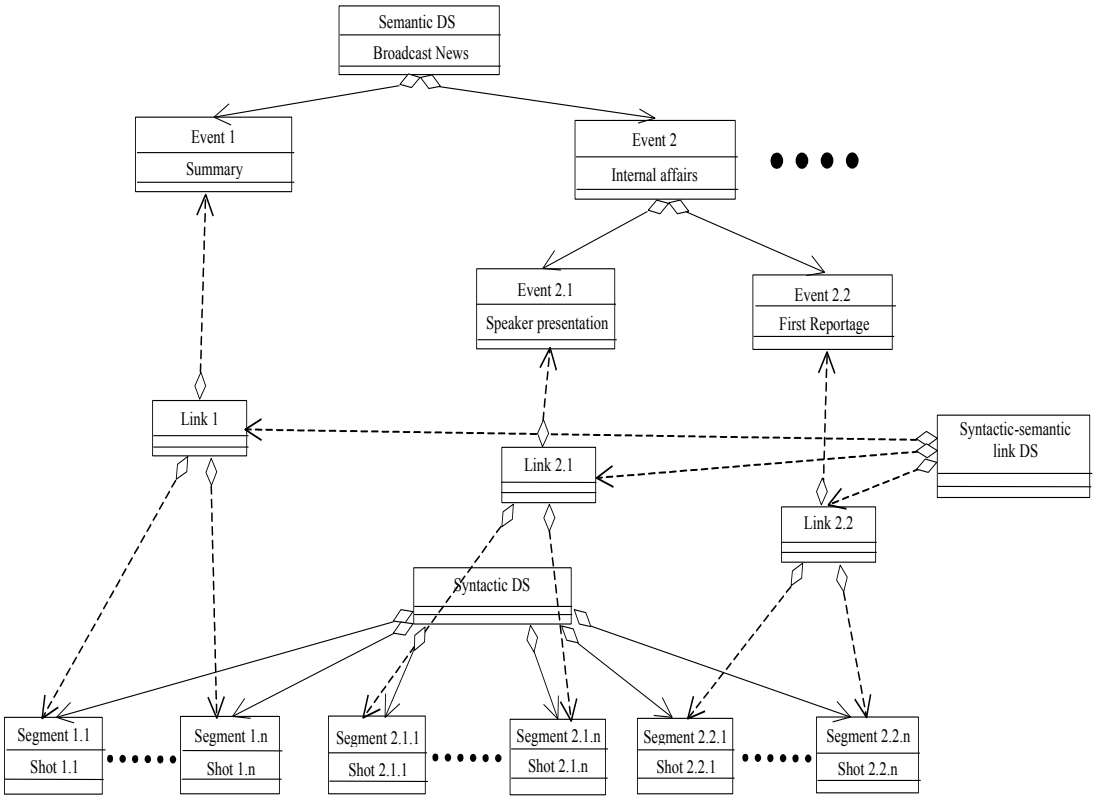

Figure 6 Implementation of the AV document using the Generic AV DS.

Figure 7 shows the representation of the AV document 'Broadcast News' by means of the modified *Segment DS*. As we can see, the structure becomes much simpler, since we can implement the *ToC DS* by using only the modified *Segment DS*. The *Type of segment DS* (T.o.S. in figure) carries out the semantic aspect of the associated segment.

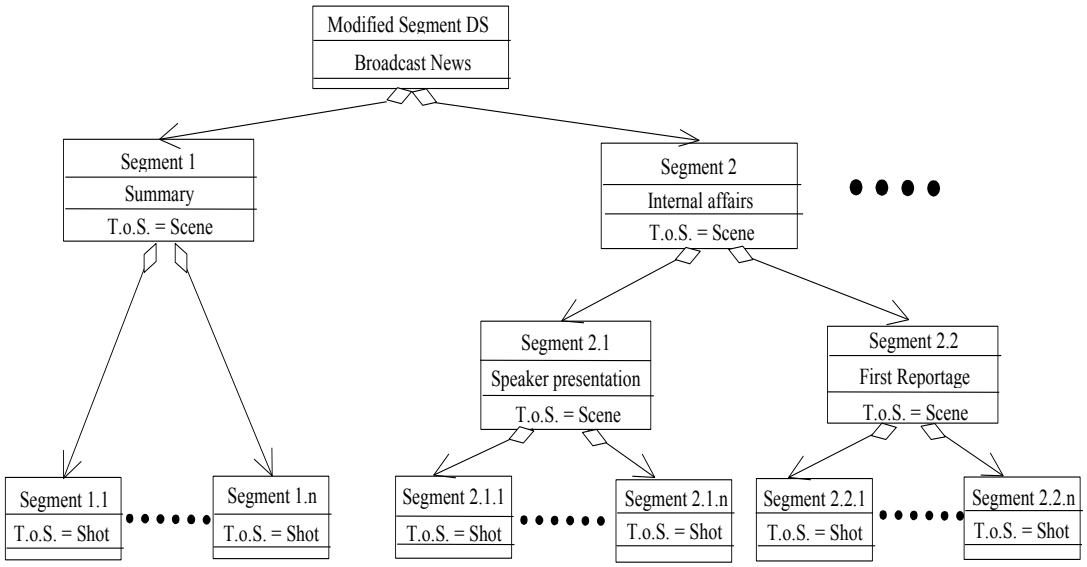

Figure 7 Implementation of the AV document using the modified Segment DS (T.o.S. stands for Type of Segment).

5 Motivation for the introduction of the key-items and ordering keys

The MPEG-7 Generic DS substantially reflects the structure of a book where the Syntactic DS can for example represent the table of contents of an audio-visual document (i.e. its temporal structure). The Semantic DS aims to represent a set of indices of the MM document with the locations of the items being provided by the syntactic-semantic links DS. The elements forming the semantic DS are the events and the objects. However there can be descriptions with an arbitrarily high number of events, objects, key-frames, mosaics, etc. leading to some difficulties in order to be exploited by users. It does not seem sensible to propose a limitation in the size of the description since a high level of details could be useful in several cases. We believe that such a problem could be overcome instead by introducing a DS to highlight the description items relevant to a particular document, given the purpose for which the description has been created or used. In this way, users can be facilitated in their queries by knowing, e.g., the most relevant topics of the documents being described while preserving at the same time all the necessary levels of detail. The kinds of highlighted items (i.e. **key-items**) could vary according to the category of document being described. For instance, in a documentary about flowers, the visual key items, defined as key-images, could be represented by images of the flower being contained in the program.

The concept of key-items for MM documents can be seen as a generalization of the concept of key-words for text documents. As in a book or a scientific paper, the key-words are highlighted by the editor and/or the author (rarely by the readers themselves), for MM documents as well key-items should be highlighted by the description provider according to the category of the described document and the purpose for which the description is used. A criterion for selecting the key-items could be based on a clustering of the corresponding description items, according to the semantic they carry. In other words, a subset of description items having the same semantics, can be represented by one key-item (linked to all the items it represents).

Another issue deals with the ordering mechanisms for the key-items. It is easy to understand that a suitable ordering of items can help user queries and that the ordering can be based on the values of several descriptors (e.g. color histogram, audio loudness etc.). However in MPEG-7 descriptions, there can be so many kinds of Ds making quite difficult the selection of ordering criteria for the items to be arranged, also depending on the kind of key-items itself. Hence we believe that in this case also a DS presenting a set of suitable **ordering keys** should be incorporated in the Generic DS so as to facilitate selection processes.

6 Structure of the modified DS for the inclusion of key-items and ordering keys

We have grouped the two DSs addressing the aforementioned functionalities in a DS called **Analytical Index DS**. Figure 8 shows the UML diagram of the Generic DS. In Figure 9, the proposed extension to the Generic DS is presented.

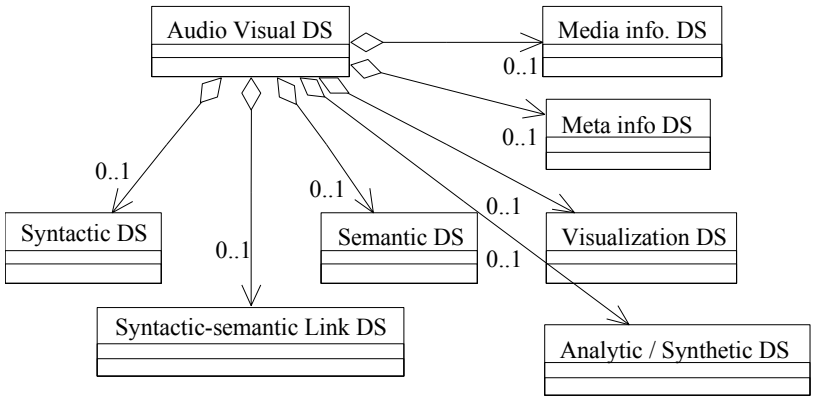

Figure 8: The Generic Audio-Visual DS.

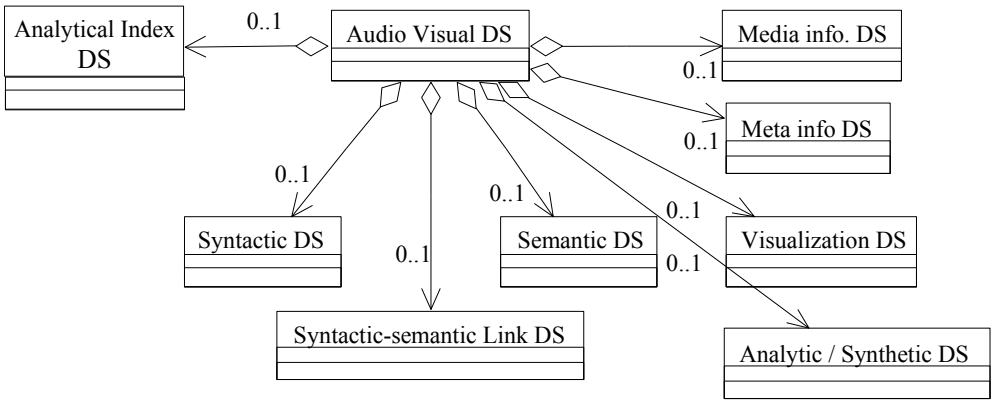

Figure 9: The Generic DS with the proposed extension.

The *Analytical Index DS* consists of the ***Key-Items DS*** (see Figure 10). The *Key Items DS* can be seen like a generalization of the concept of key-words in the context of MM documents. It is composed of several DSs (*Key-events DS*, *Key-objects DS*, *Key-images DS*, *Key-sounds DS*), each of them consisting in a set of key-items representative for certain type of description element (e.g., events, objects etc.).

In Figure 11, the structure of the *Key-images DS* is shown. The other key-items DS (*Key events DS*, etc.) reflects the represented structure. As we can see, each *Key-images DS* can contain an arbitrary number of sub-key-images, and therefore forms a hierarchy (tree) of key-items. For example, an AV document of a soccer match can contain, at a higher level of this hierarchy (then at a higher level of abstraction), a key-image representing *goals*, one representing *penalties*, another one representing *corners*, etc. At a lower level of the hierarchy a goal key-image can contains a sets of key-frames, each of them representing the shots describing that goal.

The *Key-images DS* consists of the *Ordering Key DS,* the *Attribute DS*, the *Links to segment DS*, and the *Representation DS* (see Figure 11). The *Ordering Key DS* presents a list of applicable ordering mechanisms for the listed key-items (e.g., having defined a set of key-images in a violent movie; the associated ordering key may be the level of underlying audio loudness descriptor). According to the tree structure of the *Key-images DS*, we can assign different sets of ordering-keys at different key-items pertaining at different level of the

hierarchy (see example of Section 7). Obviously these ordering keys can be applied to order other description items (for example, all key-frames and mosaics). The *Attribute DS* characterizes the key-item itself. Note that it could be (at least partially) accessed thanks to the *Links to segment DS*. The *Links to segment DS* identifies the parts (e.g., temporal segments, K-frames, …) in the sub-DS the key-item refers to. *Links to segment DS* allows clearly to refer to descriptors associated with such parts of the sub-DS. The purpose of the *Representation DS* is the visualization and the presentation of the key-items.

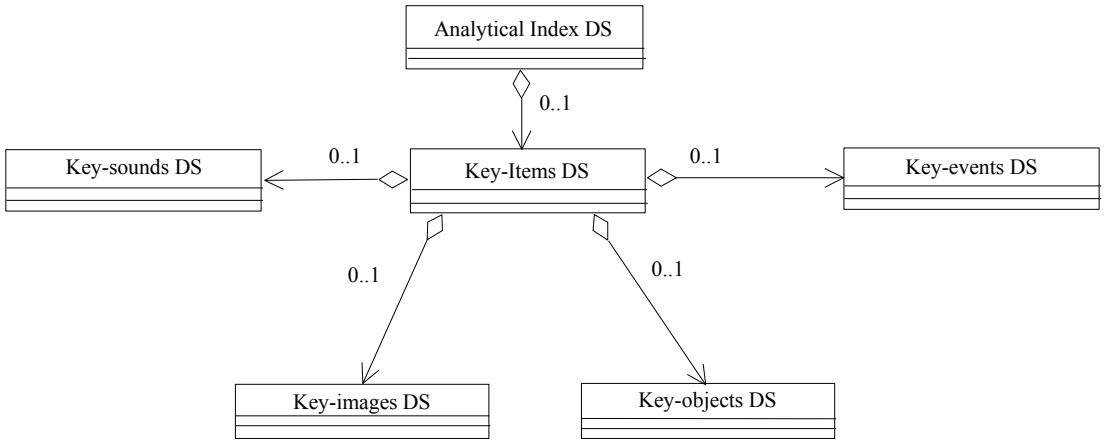

Figure 10: UML representation of the proposed DS.

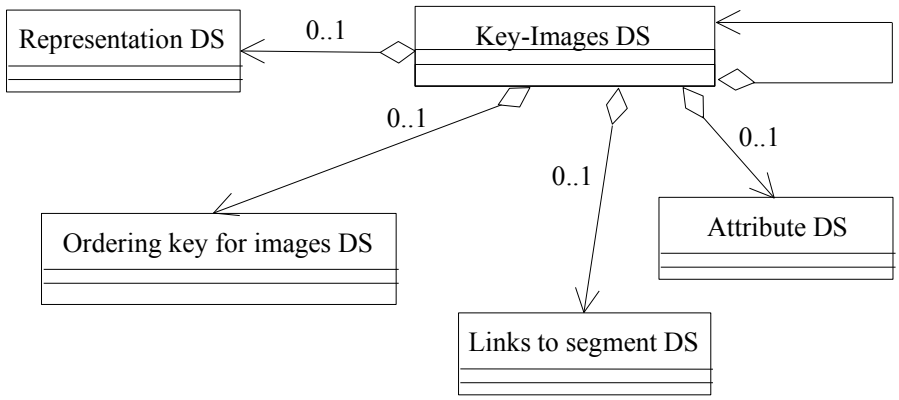

Figure 11: Key-images DS.

7 Example

Let us consider again an AV document of a soccer match. At a higher level of abstraction there can be a key-image linked to all key-frames associated to shots dealing with *goals*, another one linked to all key-frames associated to shots dealing with *penalties*, another one dealing with *corners*, etc. We can provide ordering keys for such high level key-images (for example an alphabetical ordering criteria). At the same time we can provide another sets of ordering keys for the sets of key-frames linked to the key-images. For example we can provide a ordering criteria based on the underlying audio-loudness level for the sets of key-frames representing the goals of the match linked to the key-images representing the item goal.

8 Summary

In this document, we have dealt with the implementations of the ToC DS using the MPEG-7 Generic DS. Since the solution found using the Generic DS appears complicated we have proposed a modified version of the *Segment DS* that permits a more simple representation.

Then we have presented some open issues linked with the MPEG-7 Generic DS: arbitrarily high number of description items and ordering criteria. Therefore we have proposed two new DSs. The former allows to highlight the more relevant description items (descriptor values) of a MM document. The latter allows to order the description items according to a list of criteria selected according to the application context. The aim of the aforementioned DSs is to facilitate user queries.

9 References

[1] AHG on MPEG-7 DS, "MPEG-7 Generic Description Scheme", Proposal to ISO/IEC JTC1/SC29/WG11 MPEG99/N2844, Vancouver, Canada, July 1999.

[2] AHG on MPEG-7 Requirements, "MPEG-7 Requirements Document V.9", Proposal to ISO/IEC JTC1/SC29/WG11 MPEG99/N2859, Vancouver, Canada, July 1999.

[3] N. Adami, A. Bugatti, P. Migliorati, R. Leonardi and L. Rossi. JTC1/SC29/WG11 MPEG99/M4586, Seoul, Korea, March 1999.

[4] Video Group, "Generic Visual Description Scheme for MPEG-7", ISO/IEC JTC1/SC29/WG11 MPEG99/N2694, Seoul, Korea, March 1999.