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/M9233
Awaji, Japan
December, 2002

Title: **Report of CE on Metadata Adaptation - Integration**

Source: **Nicola Adami, Marzia Corvaglia, Riccardo Leonardi**
(University of Brescia)
Hirofumi Nishikawa (Mitsubishi Electric Corp.)

Status: **Report**

1 INTRODUCTION 1
1.1 CE context and Previous works 2
2 DESCRIPTION OF THE CE 3
2.1 Software design and implementation 4
2.1.1 DIMAE features and supported functionalities 6
2.1.2 DIMAE validation 6
2.2 Digital Item design 6
3 RESULT EVALUATION AND DISCUSSION 8
4 CONCLUSION AND FUTURE WORKS 8
5 REFERENCES 9

1 Introduction

This paper reports the results of the CE on MPEG-21 Digital Item Adaptation – Integration defined in document N5182. In the original work plan definition, two main objectives were defined. The first one is to better evaluate if the current DIA Metadata Adaptation tools can fully support the integration of multiple MPEG-7 descriptions of a given content. The second objective of this CE is to evaluate if the current DIA tools allow to disable the metadata adaptation process in order to permit the delivery of "non adapted" content descriptions (e.g. for storage propose). Due to the few time available, only the first goal has been reached. The main contribution of this CE is a preliminary implementation of a metadata integration engine able to adapt Content DI. Based on this implementation, several integration experiments have been conducted to point out which integration tool can be useful.

1.1 CE context and Previous works

To introduce the concept of metadata integration, a use case scenario will be considered. In Figure 1, a simple, but enough complex, multimedia home network system is depicted. As it can be seen, the multimedia network is composed by a handheld and a Personal Computer, communicating by a wireless connection.

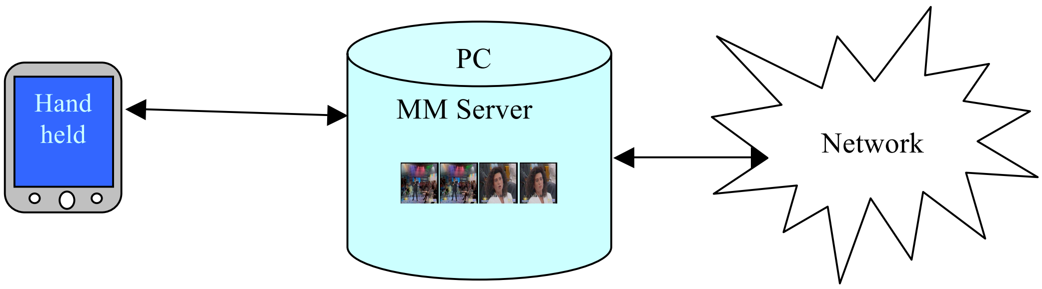

Figure 1: A simple MultiMedia sub-network composed by a handheld and a content-descriptions server.

Assuming the PC is a multimedia content-descriptions repository, we want to access this information from the handheld. More specifically, the objective is understanding the content of a specific document, like a video sequence, using the associated MPEG-7 descriptions.

The MPEG-21 aim is defining a multimedia framework in order to allow the easy and transparent access of any multimedia resource from any terminal in the network. In the use case here described, we focus the attention on the word "easy". This means that, because of the limited visualization capabilities of the handheld, some adaptation of the content-descriptions has to be performed. For example, if the user would like to explore the content of a video using the k-frames associated to the relative shots, at least the k-frame size and the color representation have to be adapted in order to fit the display characteristics of the handheld.

However, in general, more than one description associated to the same document can be available; for example, we can have many descriptions of the same content obtained by different extraction methods. In this case, the user would like to choose a particular description reflecting his/her characteristics or to select the integration of all available descriptions. In this second case, the descriptions must be adapted in order to obtain a unique and, generally, richer description with respect to the initial ones; also the resulting description should be lighter than considering all the initial descriptions. For example, if several k-frame series, relative to a specific video, are present in the network, a unique description can be generated by using the technique described hereafter.

In a previous CE on DIA Metadata Adaptation (m9024), University of Brescia has shown how integration of multiple descriptions of a given content can be useful for the reduction of the number and the size of metadata, consequently for the reduction of DIs. In that CE, the AdaptationHint DS has been tested to support the integration task; the conclusion was that the AdaptationHint DS can be useful to speed up the process and to reduce the memory occupancy. As a result of the experimental evaluation, the concept of "elementary unit" was also introduced; this new concept has to be considered in the integration and is the object of this report.

2 Description of the CE

As introduced in a previous CE (m9024), one of the main problem for the integration of metadata concerns the "elementary unit" that have to be considered during the process. To better understand this concept, let us consider the example depicted in Figure 2, where two video segment associated to the same content are presented. The two segments exhibit a different number of decomposition levels, two for VS1 and one for VS2.

1. Considering a general integration process, two main operation needed to be performed:

1. find the intersection between the to descriptions (ex: $I = VS1 \cap VS2$),
2. add to the first description the part of the second one that does not overlap (ex: $R = VS1 \cup (VS2 - I)$), or vice versa.

This means that, in order to find the intersection of VS1 and VS2, the engine has to find the best matching between pairs of sub segments, operation that present a complexity of M*N, where M and N are respectively the number of sub segment in VS1 and VS2.

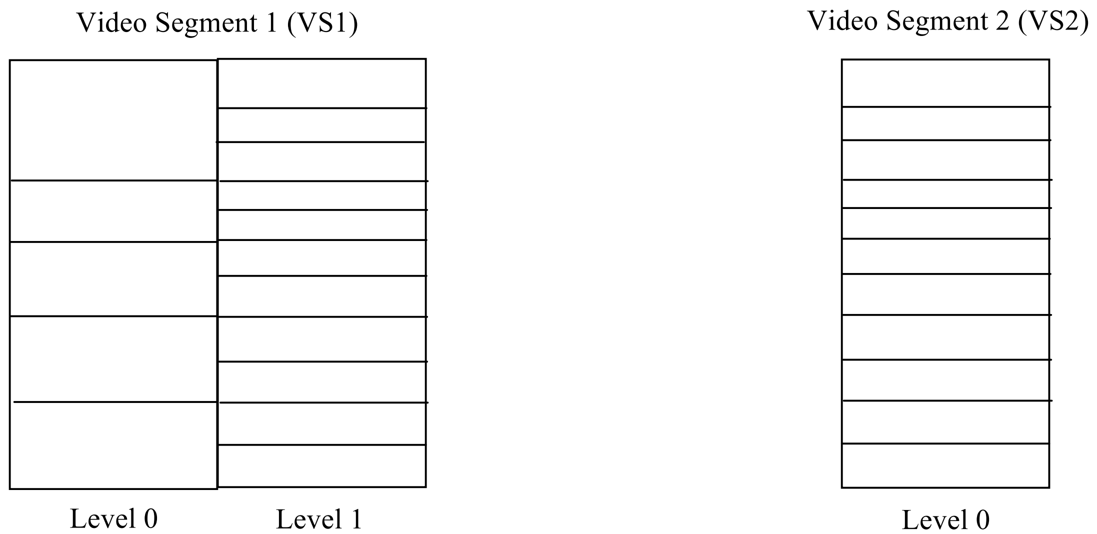

Figure 2: Segment comparison; segments of VS2 have to be compared with segment at level 1 of VS1.

A way to speed up the integration process is to consider some a priori knowledge of the descriptions that have to be integrated. In our example, such information can be the average number of segment over the time associated to a given level. With this information, the engine is able to automatically conclude that segments associated to the first level of VS1 are unlikely to have the same granularity of segments at the same level in VS1, and hence they can not overlap.

This consideration lead us to define the "elementary unit" as an indicator of the average information associated to a set of descriptions or descriptors in a descriptions instance.

The main objective of this CE is to demonstrate how the use of the "elementary unit" information can speed up the integration process of compatible description of a given content. For this purpose, a first prototype of DI metadata integration engine has been developed according to the guideline reported in the CE work plan document.

The integration process described is only a part of the complete integration because it just considers the structural level, that is the Description schemes (DSs) and not the Descriptors (Ds). A complete integration methodology is proposed in [9] and is summarized in the block diagram of Figure 3; the phase described in this CE is the phase 1.

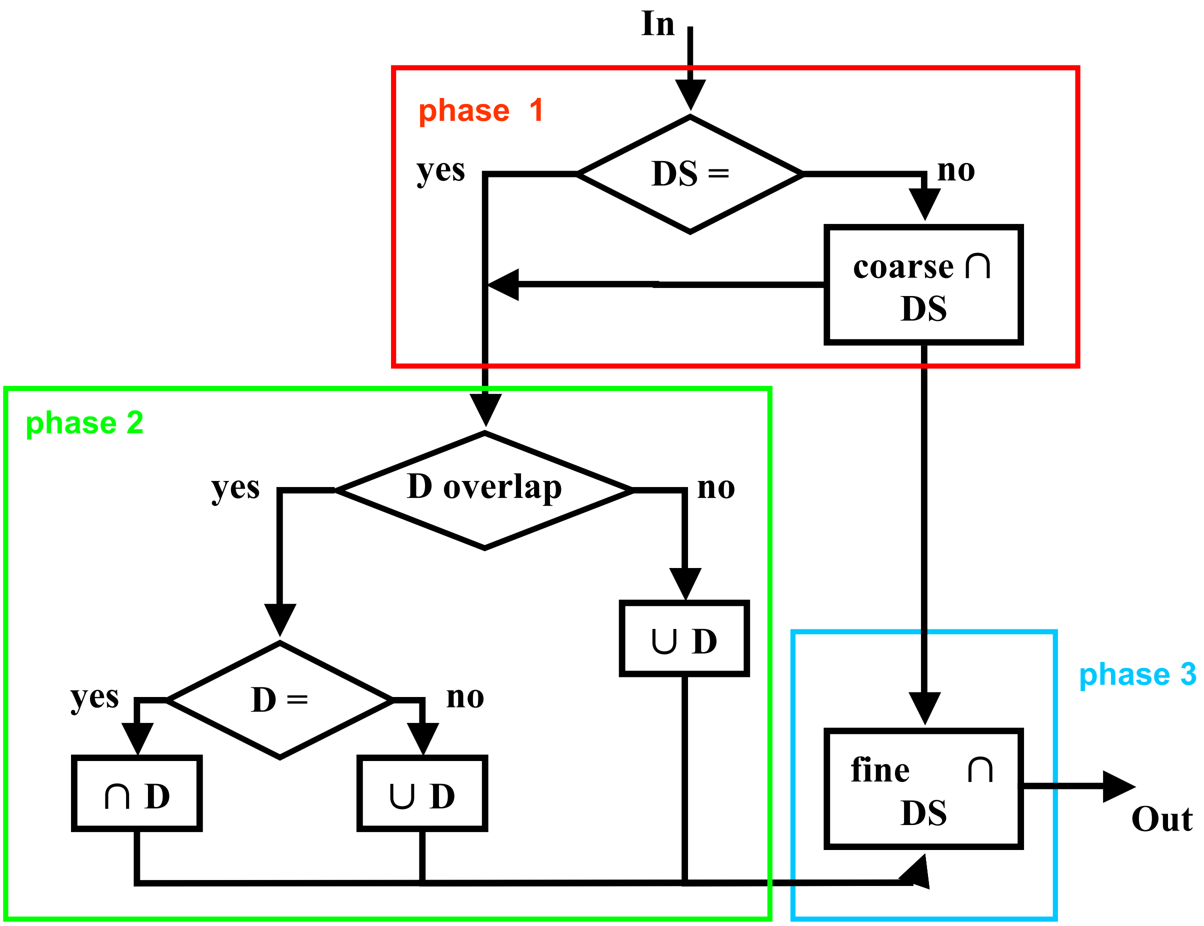

Figure 3 Complete integration process:

- *phase 1 (coarse DSs integration): integration at the structure level with redundant information*
- *phase 2 (Ds integration): integration of the attribute information*
- *phase 3 (fine DSs integration): the result given by phase 1 is re-defined on the basis of the result of phase 2.*

2.1 Software design and implementation

The architecture of the adaptation – integration engine has been designed following the guidelines defined in document N5180. As it can be seen in Figure 4, the adaptation module receives a number K of Context Digital Item **B**=(B_1, …, B_K), useful to support and define the required adaptation and the DI that has to be adapted (A). The produced output consist of the adapted Digital Item (A').

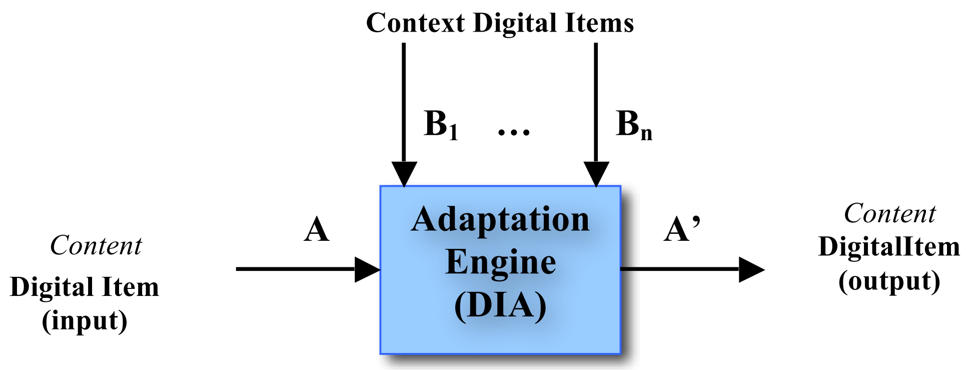

Figure 4: MPEG-21 DIA processing

In Figure 5, the main components of the DI Metadata Adaptation Engine (DIMAE) are shown. The "DIA Server" is the core of the engine; it controls the adaptation operations flow. The process starts parsing the input DIs A and **B**. This is realized using the functionalities provided by the "DOM" module. After the DIs have been parsed, the descriptions, that have to be integrated, are first extracted from A and; then, the appropriate functionality is requested to the "Metadata Integration" module. In the last step, the description generated by the integration process is packaged in the output DI A'.

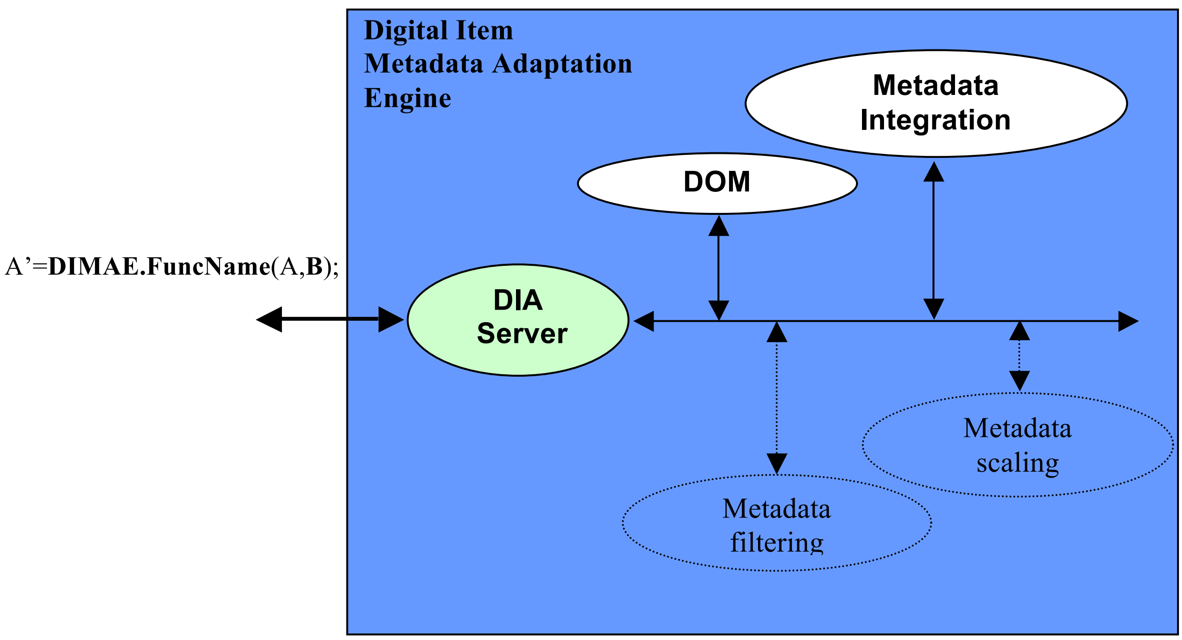

Figure 5: Digital Item Metadata Adaptation Engine elements.

2.1.1 DIMAE features and supported functionalities

The DOM module of the Engine has been implemented by using the DOM APIs provided by the Xerces software. In the current engine version only few basic integration functionalities are supported. More specifically, it supports the integration of multiple VideoSegment DS at the structure level; the descriptors integration phase has not yet implemented.

2.1.2 DIMAE validation

The engine has been validated by conducting several tests.

First, the correctness of the integration process has been checked using a set of input descriptions, each of them formed by a subparts of a same VideoSegment instance. In this case, it is easy to check if the integration of all the sub parts is equal to the original description instance or not.

A second test has been conducted in order to validate the integrated output descriptions against MPEG-7 schema.

Finally, the output DIs have been validated against MPEG-21 DID schema v.2.0.

2.2 Digital Item design

The general structure of the DI, that can be used as input for the adaptation engine, is presented in Example 1. In order to easily generate and manage the DIs used in the experiments, metadata have been stored into separate files and imported as resources. A descriptor, identified with a name “dia:ToIntegrationItemID”, has been used to let the engine know which items/descriptions have to be integrated. This is not a DIA tool and it can be removed in future implementations of the engine. All DIs used in the experiment have been validated against the DID v.2.0.

```
<?xml version="1.0"?>
<!-- edited with XMLSPY v5 rel. 2 (http://www.xmlspy.com) by Nicola Adami
(University of Brescia - Italy) -->
<DIDL xmlns="urn:mpeg:mpeg21:2002:01-DIDL-NS"
      xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
      xmlns:xml="http://www.w3.org/XML/1998/namespace"
      xsi:schemaLocation="urn:mpeg:mpeg21:2002:01-DIDL-NS
      D:\home\Adami\EUProject\SPATION\63rdMPEG\doc\XML\did_schema.xsd">

 <Item id="MULTIPLE_DS_DI">
   <Descriptor>
     <Statement mimeType="text/text">
       Example of DI with one video and three
       VideoSegment decomposition associated to it
     </Statement>
   </Descriptor>
```

```
   <!-- Video and Desciptions resources -->
   <Item>
   <!-- Video description and location -->
   <Descriptor>
     <Statement mimeType="text/text">Video Portugese</Statement>
   </Descriptor>
   <Descriptor>
     <Component>
       <Resource ref="repository/videos/tgporto.mpg" mimeType="video/mpeg"/>
     </Component>
    </Descriptor>
    <Descriptor>
      <Statement mimeType="text/xml">
        <dia:ToIntegrationItemID>
          <dia:tiiid>
             VS1 VS2 VS3
          </dia:tiiid>
        </dia:ToIntegrationItemID>
      </Statement>
    </Descriptor>
    <Item id="VS1">
      <Descriptor>
        <Statement mimeType="text/text">First Temporal decomposition</Statement>
      </Descriptor>
      <Descriptor>
        <Component>
          <Resource ref="repository/descriptions/VS1.xml" mimeType="text/xml"/>
        </Component>
      </Descriptor>
    </Item>
    <Item id="VS2">
      <Descriptor>
        <Statement mimeType="text/text">Second Temporal decomposition</Statement>
      </Descriptor>
      <Descriptor>
        <Component>
          <Resource ref="repository/descriptions/VS2.xml" mimeType="text/xml"/>
        </Component>
      </Descriptor>
    </Item>
    <Item id="VS3">
      <Descriptor>
         <Statement mimeType="text/text">Third Temporal decomposition</Statement>
      </Descriptor>
      <Descriptor>
        <Component>
          <Resource ref="repository/descriptions/VS3.xml" mimeType="text/xml"/>
        </Component>
      </Descriptor>
    </Item>
   </Item>
  </Item>
</DIDL>
```

Example 1: DI including multiple VideoSegment DSs of a same video.

3 Result evaluation and discussion

The integration of several different VideoSegment instances have been performed in order to evaluate the benefit in terms of computational time, introduced by using the elementary unit information during the integration process. In this experiment, we considered the integration of only two VideoSegment DS instances; in general, the integration of a generic number of DSs can be realized by recursively applying the integration function.

As shown in Table 1, the use of the Average Number of Segments over Time (ANST) significantly affects the computational time required to find the overlapped part of two descriptions. Beside, this auxiliary information is not really relevant in the second phase of the algorithm where the output description is created by combining the first description with the non overlapping part of the second one.

	With ANST information ***[number of machine cycles]***	***Without ANST information*** ***[number of machine cycles]***
Matching time	60000	150000

Table 1: Computational time required for the matching phase

This indicators represent the average duration required to integrate of ten different VideoSegments. It is important to note that when the ANST is used, the matching time can reach zero if the two descriptions do not partially overlap.

The computational time required to calculate the ANST value it is not really high (average value of 40 machine cycle) but still different from zero and it strongly depends on the description structure and size. Considering that a given description can be involved in an integration process more than once, we can conclude that it is useful to have such DIA Metadata Integration tool in order to provide this auxiliary information to the integration engine.

4 Conclusion and future works

In this document only some of the objectives defined in the work plan have been reached. In particular, a first implementation of a DIA engine, able to adapt DI carrying multiple descriptions of the same content , has been achieved. Using this engine, it has been shown how the elementary unit information is relevant to speed up the integration process. However more experiments have to be performed in order to extend this concept that has been applied here only to temporal segments (ex: average number of Dominant Color over the area). It is our intention to continue this experiment and add the following features to the current engine implementation:

1. Fully support the VideoSegment DSs integration
2. Support the Metadata Adaptation tools included in the DIA WD 2.0 (Hint DS)

3. Support the Descriptor integration of (phase 2 of the integration algorithm).

5 References

[1] N. Adami, M. Corvaglia, R. Leonardi, “CE on Metadata Adaptation – Integration of multiple descriptions”; ISO/IEC JTC1/SC29/WG11 M9024, Shanghai, October 2002.

[2] H.Nishikawa, “Report of CE on Metadata Adaptation”, ISO/IEC JTC1/SC29/WG11 M8589, Klagenfurt, Austria, July 2002

[3] K. Kazui: “Result of CE on Metadata Adaptation”, ISO/IEC JTC1/SC29/WG11 M8576, Klagenfurt, July 2002

[4] M. Sasaki: “Report and Study of CE on Metadata Adaptation”, ISO/IEC JTC1/SC29/WG11 M8560, Klagenfurt, July 2002

[5] N. Adami, R. Leonardi, “Content and Description Scalability Issues in a Home Network Environment”; ISO/IEC JTC1/SC29/WG11 M8693, Klagenfurt, July 2002.

[6] Description for Metadata Adaptation Hint, Mitsubishi, ISO/IEC JTC1/SC29/WG11 m8324, Fairfax, May 2002

[7] MPEG-21 Requirements on Digital Item Adaptation, ISO/IEC/JTC1/SC29/WG11 N4684, Jeju, March 2002

[8] Final Call for Proposals for Digital Item Adaptation, ISO/IEC/JTC1/SC29/WG11/N4683, Jeju, March 2002

[9] N. Adami, M. Corvaglia and R. Leonardi, "Comparing the quality of multiple descriptions of multimedia documents", *MMSP 2002 Workshop*, St. Thomas (US Virgin Islands), Dec. 2002

[10] N. Adami, “Work plan for CE on Metadata Adaptation – Integration”; ISO/IEC JTC1/SC29/WG11 N5182, Shanghai, October 2002.