# Designing intelligent content delivery frameworks using MPEG-21

Samir Amir, Ioan Marius Bilasco, Thierry Urruty, Jean Martinet, Chabane Djeraba

## 1. Introduction

Digital home systems are a reality. The possibility of switching between access devices while accessing online and multimedia contents is becoming common usage [13][24]. Deploying the content regardless of the heterogeneity of devices capable of accessing multimedia contents is a key factor for producers. Once created, the contents should be deployable on various configurations without subsequent transformation performed manually by content creators, or completely redesigned by authors to make them deployable on new access devices. With regard to the current state of the art, the variety of devices as well as deployed technologies, it is *naive* to consider that a content can be transmitted anywhere as it is.

Automatic transformations are necessary to modify the encoding, the structure or the content itself to make it accessible in constraint contexts. Among these transformations, we can mention: *transrating* (i.e. changing the bitrate of the content), *transmoding* (i.e. changing the modality to convey the content), *transcoding* (i.e. changing the encoding format). These three operations are closely linked to the encoding of the content. Basically, they are characterized by a set of parameters describing, respectively, the expected output rate, the output format (video, audio etc.), and the output codec. They do not need any explicit knowledge about the content itself. The obtained output is a degraded version of the initial content.

More intelligent transformations like summarization, filtering or reorganization of the content, require specific resources and information. Such techniques need a deep access to

the knowledge embedded within the content and to the characteristics of the broadcasting context. The broadcasting context is composed of access devices, networks, environmental configurations (indoor/outdoor, noisiness, visibility, etc.), user and community interests and preferences, as well as available adaptation services deployed on intermediate proxies. The description of all this content-related and context-related knowledge must be made following a standard format accepted by all the devices and services involved in the delivery and adaptation process.

Due to multimedia popularity and the wideness of its application domain, the amount of information describing the multimedia content and context has become essential. Information serves to describe all entities involved in multimedia systems in order to facilitate multimedia delivery and consumption. It starts with the description of the content itself (size, type, etc.), its semantics (objects appearing in a picture, place where a picture was taken, etc.), the characteristics of the devices transmitting or consuming the content (TV, networks, etc.) and finally the consumer profile (preferences, interests, etc.).

Currently, several standards just cover some part of the information. MPEG-7 [15][14] & MPEG-21 DID [16], for instance, deal with the description of the structure and semantics of a media object. MPEG-21 DIA [19], CC/PP [20] or Device Independent Activity[1] offer tools for modeling the utilization context. WSDL [8] and OWL-S [21] standards deal with the characterization of Web Services that might be involved in some adaptation process on the network. However, up to now, no integrated solution has clearly emerged. The MPEG-21 [7] set of tools (namely DID [17] and DIA [19]) seems the most prolific candidate as it proposes tools to describe the content, the context and the transformation to be applied to a

1 http://www.w3.org/2001/di/

given media by using BSDL [1][2]. But still, the difficulty of acquiring a deep knowledge on these tools might not encourage the designers, who try to find friendly solutions for specific applications. We underline here that, as far as we know today, there is no off-the-shelf solutions hiding to the designers the complexity of standards (in particular in terms of encoding and retrieving information) like the one proposed above.

In this chapter, we report on the creation of a new metadata framework that embeds information related to the content, context and adaptation services. This is part of the CAM4Home ITEA2 project[2]. A group of twenty multimedia academic and industrial practitioners from TV, 3G and Internet application fields defined a restricted set of metadata requirements in order to support the convergence of multimedia content in Digital Home environments. A unified model, called CAM Metadata model [5][26], has been designed to cover the categories of information introduced above, making no reference to the effective type of metadata encoding used to represent information. In the following, we explore first an alternative encoding by using multimedia native standards, like MPEG-7 and MPEG-21, in order to illustrate how the mapping of application-specific requirements can be projected to existing descriptors of MPEG standards. We show also some other specific requirements that cannot be described natively by MPEG standards. They have been encoded using specific constructs developed by CAM4Home project partners. These requirements enrich the semantic and the context of information in order to allow the development of a rich multimedia experience. However, the original MPEG21/7 constructs lay the basis of the semantic, structure and the context descriptions.

2 http://www.cam4home-itea.org

The chapter is organized as follows. The next section is dedicated to the presentation of metadata requirements. The third section presents the abstraction of the requirements by means of a metamodel organizing metadata information by nature and usage. An overview about XML multimedia metadata standards for encoding content description is given in the fourth section; we show here the lacks of existing techniques with regard to the identified requirements. The fifth section illustrates how this metamodel can be mapped to the descriptors existing in MPEG-7 and MPEG-21. We have chosen MPEG as it covers most of our requirements. We provide also a series of new description needed to meet the requirements related to Web 2.0 specificities like social tags, services, etc. We conclude by discussing our encoding choices.

## 2. CAM Metadata Framework Requirements

In this section, we present the delivery architecture that we have selected in the frame of the CAM4Home ITEA2 project, and we expose the requirements in terms of metadata for the proposed framework.

### 2.1 CAM4Home Framework Overview

The purpose of the CAM4Home project is to create a metadata-enabled content delivery framework that allows end users and commercial content providers to create and deliver rich multimedia experiences. They are based on a novel concept of collaborative aggregated multimedia (CAM) that is the main contribution of the CAM4Home project. The Collaborative Aggregated Multimedia (CAM) refers to the aggregation and composition of individual multimedia contents (called *objects*) into a content bundle that may include references to content-based services and can be delivered as a semantically coherent set of contents and related services over various communication channels.

The project develops one common metadata framework for CAM content that can be applied to both personal and commercial applications, and is interoperable with relevant standard metadata and content representation technologies. In order to better illustrate the role of the metadata framework within the project, the general architecture is presented in Figure 1.

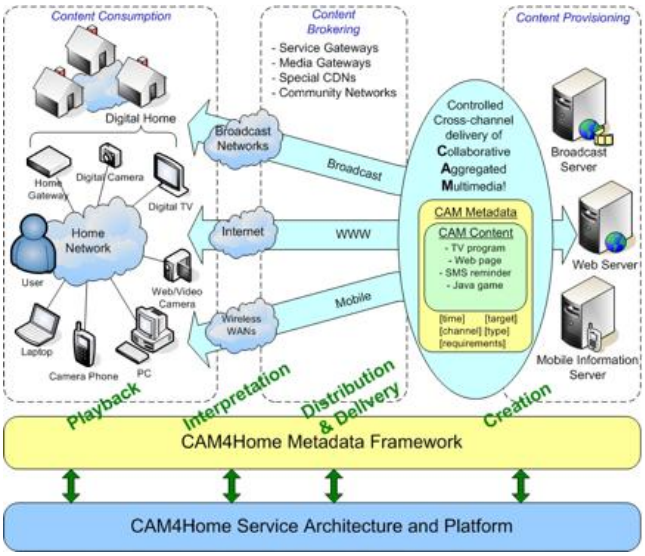


**Figure 1 CAM4Home architecture**

This CAM metadata framework enables a novel way of content provisioning by bundling different types of multimedia objects and services into bundles on the level of metadata. The CAM metadata framework is able to encapsulate existing metadata technologies for multiple types of contents and to incorporate references to content-related services. These content bundles are delivered into the digital home environment over multiple communication channels and mediums in a controlled (e.g. time synchronized) and adaptive (e.g. adaptation to environment capabilities and user preferences) manner.

The functional aspects of the metadata framework are performed by the CAM4Home service architecture. The service architecture defines the necessary service components to support the content lifecycle of collaborative aggregated multimedia from content bundle creation to distribution, and finally interpretation and playback of the content bundles.

As explained previously, the metadata framework serves two purposes: providing metadata representation format for CAM content and enabling the processing of such metadata. However, the focus in this paper is to represent the metadata model of the metadata framework which constitutes the specification of the data format providing a medium for the applications and services to distribute and deliver aggregated content. The description of the service platform is out of the scope of this chapter.

### 2.2 Metadata Requirements

In the following, we discuss some requirements that we observed while constructing the metadata framework. The requirements are related to content adaptation, content aggregation and metadata extensibility topics that have been gathered in the project.

### 2.3 Content Adaptation requirements

In the process of content adaptation, the framework selects the most suitable content according to the consumer profile and respective device characteristics; this operation is done through a content delivery step. In order to achieve such operation, all elements included in this process must be well described. For instance, if a given content is not suitable for playback by the consumer's device, it needs to be adapted to the discovered content adaptation services which analyze the consumer's device characteristics via the interpretation of its metadata. The conceived metadata framework ensures a proper description of the user profile, preference and usage context. This description covers:

- *Device*s and theirs audio capabilities, embedded decoding system, display capabilities, installed software and hardware description, etc;
- *Network* and its QoS characteristics (e. g., maximum and minimum of data throughput, reported packet lost, etc.);
- *Services* deployed within the framework and theirs QoS characteristics;

- *Users* and theirs user's personal information (e. g., name, date of birth, etc.), environment and preferences descriptions;
- *Communities* and their common denominators with regard to user interests.

### 2.4 Content Aggregation requirements

The CAM metadata framework enables a novel way of content-provisioning by bundling different types of content and content services into content bundles on the level of metadata. The framework describes the relationship between different contents; several relationships have been defined in order to help the consumers to select the related contents. For instance, we have defined a relationship named *thematic relationship* which links all the contents sharing the same theme. A user watching a video on a given subject has the ability to display all the contents on this subject (e. g., all videos on football). Many metadata relationships have been defined in the project, and some of them are mentioned here:

- Part-of relationship;
- Aggregation-type relationship, showing why different contents are aggregated in the same bundle (e. g., same event, same date, etc.);
- Derivative-of relationship, when the content is a derivative work of another one (e. g., content#1 is a re-mix of content#2);
- Alternative relationship, when the content is an alternative representation of another one;
- Rendering relationship, when a multimedia content embeds the representation of another one (e. g., a web page with a video rendered with a plug-in).

## 2.5 Extensibility

In order to make the framework extensible and allow CAM4home users to enrich the contents with other metadata, the designed metadata framework contains specific properties defined so as to link the metadata framework to other standards (e. g., users can upload a video and add metadata on the date of creation and the subject, for instance; or they may also add metadata which is encoded in an already existing format like MPEG-7).

# 3. CAM Metadata Model

The metadata requirements collected and detailed in the previous section have been organized in three main categories: metadata on the content – addressed as *core metadata*; metadata on the context – addressed as *supplementary metadata*; and metadata defined in existing standards – addressed as *external metadata* (external with regard to the current metadata framework). The design of the CAM Metadata model [5] is illustrated in Figure 2.

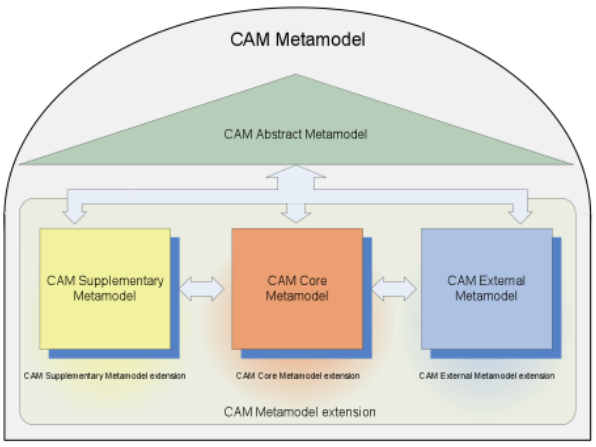


**Figure 2 CAM metadata metamodel**

The CAM Metadata model provides the core concepts and required metadata level information for collaborative distribution of multimedia and software contents as a structured model which can be partially or fully instantiated as metadata, and used in the system. In addition, the CAM Metadata model is designed to allow an easy encapsulation of existing metadata formats into the structures of the instantiated metadata. Furthermore,

the abstract from the CAM Metadata model enables to define new structures and associations that a system might need in its operations. The CAM Abstract Metamodel defines a generic categorization of concrete metadata entities and associations between them on an abstract level. In the following, we ignore the presentation of the metamodel's abstract part. We focus our attention on the concrete part as we will study how the concrete set of metadata selected to cover the requirements can be mapped to MPEG-7 and MPEG-21 structures.

We briefly describe the role of the core, supplementary and external metadata. Then, we present into more detail the simple and structured metadata belonging to each category. The *core* part of the metamodel is structured around two notions: the CAM Object and the CAM Bundle. A CAM Object corresponds to a set of metadata describing an atomic media object regardless of its precise type. A CAM Bundle is a metadata container representing information about the aggregation of several atomic media objects. These two concepts are the main deployment units within the framework.

The *supplementary* part of the metamodel introduces five entities: the access device, the network, the services available on the network, the user and its communities. For each of these entities, several profiles can be associated in order to support time-dependent (in the morning, in the afternoon) and usage-dependent (at home, at work) characteristics.

The *external* part defines the structures into the CAM Metadata model which acts as an interface towards external metadata formats (e.g. TV-Anytime [25], WSDL [9], CC/PP [20], etc.) and encapsulates them into the CAM Metadata model

### 3.1 CAM Core Metamodel

Constructed around CAM Object and CAM Bundle notions, the Core Metamodel supports the representation of a wide variety of multimedia contents: downloadable applications,

software services, images, videos, etc. The metadata describes both the content file (the *essence* of the content) and the actual content that is provided.

A CAM Object or CAM Bundle contains simple or structured metadata organized in six categories:

- Content feature metadata: describe the content itself. They are split into simple metadata (author, creator, copyright, description, legal notice, target domains, thumbnail, title etc.) and complex metadata (access restrictions, appearing concepts, content genre, creation context metadata, cue tones etc.);
- Community-created metadata [26]: social tags, user comments, user ratings;
- Essence of the content (applying only to CAM Objects): content location, nature of the deployment (streaming or downloadable content), essence file metadata, etc.;
- Aggregation metadata (applying only to CAM Bundles): semantic, spatial and temporal relationships between composing objects within a bundle;
- System metadata – used to manage the instances of a CAM Object and CAM Bundle : UID, version number, creation date time;
- Supplementary reference metadata – used to specify the deployment contexts for which the content was initially designed: target community reference, target device reference, target domains.

A partial view of the CAM Object and CAM Bundle classes containing complex metadata only is illustrated in Figure 3. The community-created metadata is attached to both object and bundles. The other complex metadata are directly associated to CAM Objects only,

but there can be inference for bundles by considering the association between bundles and composing objects.

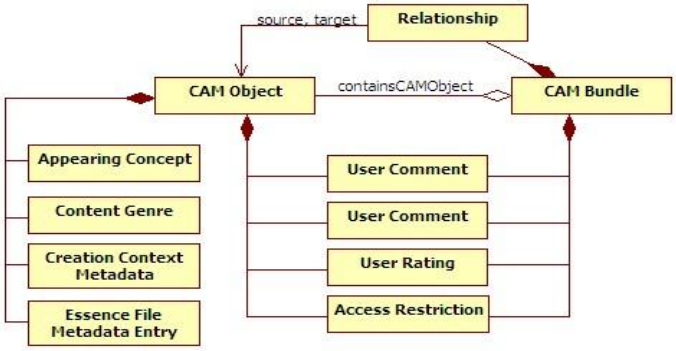


**Figure 3 Partial view of CAM Object and CAM Bundle structured metadata.**

## 3.2 CAM Supplementary Metamodel

The supplementary part of the metamodel is built on three main concepts: the entity, the profiles and the supplementary metadata. An entity can be described by several profiles. Each profile gathers a predefined set of supplementary metadata. For instance, devices physical capacities, browser capabilities, display capabilities, audio capabilities, available software, decoding systems, embedded decoders are some examples of structured metadata used to describe an access device. Preference description, environment description, personal descriptions are structured metadata to describe the users of the systems. Figure 4 reflects the organization of the supplementary metamodel.

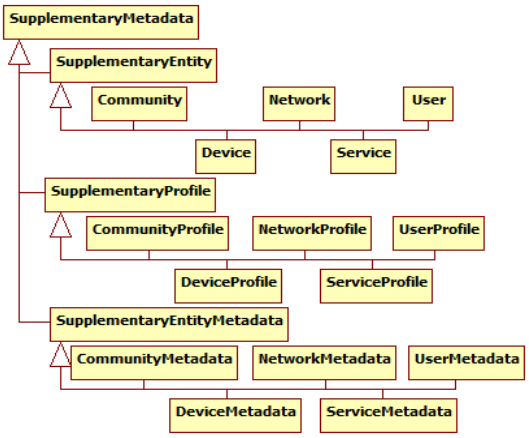


**Figure 4 CAM abstract supplementary metamodel**

The entities are also linked to each other. A user entity can belong to a community entity. A user entity uses a given device entity on a given network entity. A device entity uses, at a given time, a network entity. Those relations are represented in Figure 5. The dash lines indicate that the links are implemented by means of references.

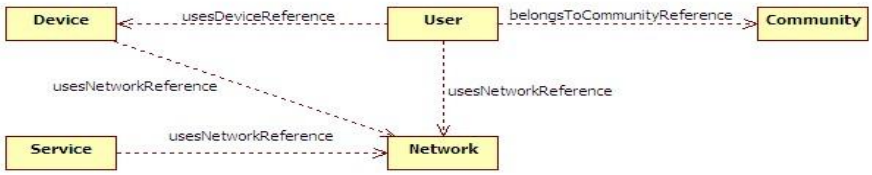


**Figure 5 Relationships between supplementary entities**

### 3.3 CAM External Metamodel

This part of the abstract model provides the basic structures that allow the integration of existing standards within the CAM metadata framework. The possibility to include external metadata description in CAM metadata framework is a key feature as it permits CAM4Home practitioners to benefit from existing metadata standards.

Two categories of external metadata descriptions are considered: external core metadata and external supplementary metadata. The external core metadata descriptions are related to the core aspects of a CAM Object or a CAM Bundle. The external supplementary metadata descriptions are related to the characterization of supplementary entities. In order to underline the type of external metadata classes that can be considered for integration within the CAM metadata framework, we have defined several subclasses for each type of external metadata. Hence, we have core-related external metadata classes that support the integration of content feature metadata or content aggregative metadata. With regard to the supplementary-related external metadata classes, we have considered external metadata classes that address the community, the device, the user, etc.

In this section we have presented an abstraction of CAM4Home delivery framework requirements. Effective instances of the abstraction have to be represented and encoded.

We naturally look for XML-based multimedia standards. In the next section, we discuss some existing standards that seem suitable for encoding CAM4Home metadata requirements.

## 4. Study of Existing Multimedia Standards

The ubiquitous presence of multimedia data requires an extensive use of metadata for multimedia content retrieval, filtering and adaptation. The main metadata specifications efforts are focused on the description of the multimedia content and context [3]. The MPEG-7 standard represents the most successful result in this field. It standardizes the description format (syntax and semantics) and decoding of a broad set of features of the multimedia assets at many different levels of abstraction. MPEG-21 DIA and TV-Anytime take MPEG-7 controlled terms to extend the metadata description to some more specific purposes. MPEG-21 DIA standardizes the description of multimedia adaptation procedures in a generic way, whereas TV-Anytime targets applications of digital TV, such as electronic program guides (EPG). Content structure is also another information that metadata takes into account. A multimedia presentation is a structured collection of elements, such as video and audio clips, images, and documents. The bundling of these elements is also described by multimedia metadata. Among existing metadata standards used for the structural description we can note METS [12], IMS Content Packaging [10] and SCORM [9]. But the most generic approach for this purpose is proposed by the MPEG-21 DID standard.

Another field of application of multimedia metadata is the description of the multimedia lifecycle. This information about content creation, modification, search, delivery and consumption is also described by metadata. MPEG-7, for instance; is one of the standards that include tools for the description of the user interaction with contents.

Multimedia content adaptation is also an interesting field of application of metadata. The adaptation is made according to the context: where and by whom these resources will be used. The context includes the information about devices consuming or transmitting these contents (e.g., networks, TV, mobile, etc.) and user characteristics (e.g., user profile, user preference, etc.). MPEG-21 DIA allows the description of device and network profiles. W3C for device descriptions for web content adaptation uses CC/PP.

Several content related standards exist but none allow the homogenous description of multimedia content, services and use context (as illustrated in Table 1).

MPEG-7 standardizes the description of content features and aggregation but does not cover other information type such as user created metadata (comment) or networks characteristics.

Tv-anyTime standard encloses specifications for the controlled delivery of multimedia content to a user's digital video recorder. It seeks to exploit the evolution in convenient, high capacity storage of digital information to provide consumers with a highly personalized TV experience. Users will have access to content from a wide variety of sources, tailored to their needs and personal preferences. Tv-anyTime does not cover all required information related to content and to context description. For instance, information describing device, service, network characteristics are not covered by Tv-anyTime.

METS is a metadata standard designed to encode metadata for electronic texts, still images, digitized video, sound files and other digital materials within electronic library collections. In doing so, it attempts to address the lack of standardization in digital library metadata practices which is currently inhibiting the growth of coherent digital collections. METS offers a coherent overall structure for encoding all relevant types of metadata (descriptive,

administrative, and structural). However, this standard does not offer a description of the use context. Besides this lack, the semantic description of this standard is very limited.

The MPEG-21 standard deals with most of the issues that we have enumerated above. Still, it only concerns multimedia content and context description but does not allow the description of several other information. For instance, it does not provide native support for user created metadata (such as comments), for the aggregation of content and services, etc. Still, it can be extended using its DIDL part.

| | Standards Information | MPEG-7 | MPEG-21 | Tv-anyTime | METS |
|---|---|---|---|---|---|
| **CONTENT** | Essence | Yes | Yes | Yes | Yes |
| | Feature | Yes | Yes | Yes | Yes |
| | User created metadata | | Yes | | |
| | Content aggregation | Yes | Yes | Yes | Yes |
| | Service aggregation | | | | |
| **CONTEXT** | User | Yes | Yes | Yes | |
| | Devices | | Yes | | |
| | Services | | | | |
| | Networks | | Yes | | |
| | Communities | Yes | Yes | | |

**Table 1 Metadata standards and the information they cover**

In the next section, we consider in detail the classes of the metadata metamodel, and we propose an MPEG-7/21 encoding. We think that these two standards are pertinent for encoding, as we can find intuitively the resemblances between the MPEG-21 Digital Item and CAM Objects and CAM Bundles. Besides, the supplementary part of the metamodel can be reflected into the environment description tools included into the MPEG-21 Digital Item Adaptation standard.

## 5. CAM Metadata Encoding using MPEG-21/7

As we have mentioned previously, the Core Metamodel regroups all properties describing metadata structures and semantics. In this section we show how the structural and semantic information can be encoded using MPEG-21/7 standards. Sine MPEG-21/7 do

not cover some requirement (e.g., complex relationships between CAM Object, physiological state of a user, community created metadata, services seen as content, etc), we show how to enrich MPEG-21/7 by adding some other information embedded from CAM4Home standard.

Before describing the encoding for the CAM Metadata model, we recall briefly some features of the MPEG-21 tools we use. MPEG-21 is based on two essential concepts: the definition of a fundamental distribution and transaction unit (Digital Item) and the concept of users interacting with Digital Items. The composition, structure and organization of a Digital Item are specified by a Digital Item Declaration (DID).

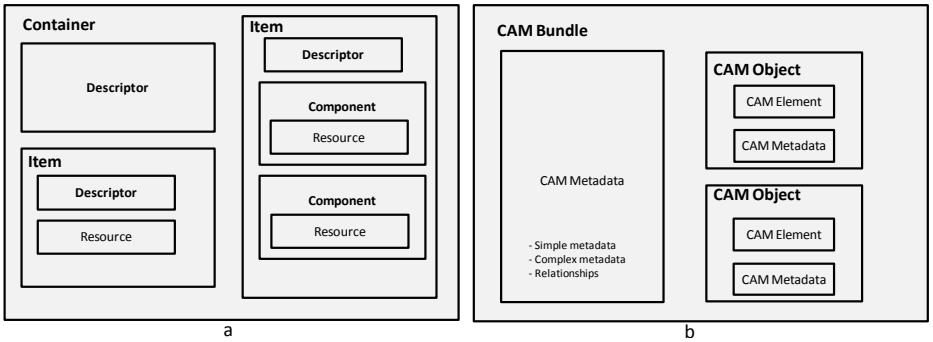


**Figure 6 DID structures and C4H structures**

Figure 6a shows a Digital Item Declaration (DID) which encloses a set of abstract terms and concepts to form a useful model for defining Digital Items. A Digital Item is described by a container which aggregates descriptors, and items. Items are defined as a collection of descriptors, components or other items. The expressive power of MPEG-21 has already been proven [6]. The structure of the CAM Metadata model fits the MPEG-21 DID specification properly (see Figure 6b). In CAM4Home, CAM Bundles are described by metadata (Descriptor) and they aggregated several CAM Objects (Item). Furthermore CAM Objects enclose CAM elements (Resources) and metadata (Descriptor).

We detail the encoding of the main classes from the CAM metadata core model: CAM Object and CAM Bundle. We also present part of the encoding of CAM metadata supplementary classes modeling the context by using MPEG-21 DIA constructs. The code

fragments that we presented in the following sections includes – besides standard MPEG namespaces (dii [17], didl[16], dia[19], mpeg7[15]) – the following C4H specific namespaces: *core*, *suppl* and *inst* – introducing respectively core and supplementary schema elements and default instances namespace.

In general, metadata are introduced by appropriate MPEG-21/7 constructs or specific `didl:Statement` attached to `didl:Descriptor` or DIDL extensions (for metadata not existing natively in MPEG-21/7) describing either CAM Objects or CAM Bundles. For instance, so that to define the creator of a content, we can reuse the `mpeg7:Creator` element. To encode the display characteristics of a device, we can reuse the `dia:Display` element.

In the following section, we first describe how the two main metamodel classes CAM Object and CAM Bundles are encoded. Then, we present some examples illustrating how object or bundle properties (simple or complex) can be encoded directly or by proposing specific extensions.

## 5.1 CAM Object Encoding

Figure 7 presents a MPEG-21 fragment describing a CAM Object. The CAM Object is encoded by the `didl:Item` element (lines 2-15).

```
01: <didl:DIDL xmlns:didl="urn:mpeg:mpeg21:2002:01-DIDL-NS">
02:  <didl:Item> ...
03:   <didl:Descriptor>
04:    <didl:Statement mimeType="text/xml; charset=UTF-8">
05:     <dii:Identifier>urn:c4h:inst:0-39-36-1</dii:Identifier>
06:                                  <!-- encodes core:camObjectUID -->
07:    </didl:Statement>
08:   </didl:Descriptor>
09:   <didl:Component>
10:    <didl:Resource mimeType="image/jpg"
11:           ref="http://www.cam4home-itea.org/graphics/logos/cam4home.png"/>
12:                    <!-- encodes core:essenceFileIdentifier -->
13:   </didl:Component> ...
14:  </didl:Item>
15: </didl:DIDL>
```

**Figure 7 MPEG-21 Fragment describing a CAM Object instance**

It contains a first descriptor to define the identifier of the CAM Object (line 5). This descriptor corresponds to the *camObjectUID* metadata of a CAM Object. Since the CAM Object describes a physical content (multimedia file or service), the corresponding `didl:Item` element encloses a `didl:Component` definition (lines 09-13) which introduces a `didl:Reference` to the essence of the content (line 11).

### 5.2 CAM Bundle Encoding

Figure 8 shows a part of the CAM Bundle specification using MPEG-21. CAM Bundles are encoded by means of `didl:Container` elements (lines 02-16) as they aggregate several descriptions of physical contents: the CAM Objects.

Regular metadata properties of CAM Bundles are encoded in a similar way to CAM Objects. However, special properties must describe the aggregation of CAM Objects into CAM Bundles. We distinguish between a CAM Object whose lifecycle is completely controlled by the CAM Bundle, and a CAM Object that is shared between several CAM Bundles, having its own life cycle. The latter case is modeled by the `didl:Descriptor` (lines 3-8). This descriptor introduces, through a sentence, a *core* specific property (`containsCAMObject` – lines 5-6) that retains the identifier of the (*freely*) aggregated CAM Object. The inclusion of (*hard*) bound CAM Objects is done through the inclusion of the `didl:Item` elements (line 10) defining the CAM Objects within the `didl:container` structure.

```
01: <didl:DIDL ...>
02:   <didl:Container>
03:    <didl:Descriptor>  <!-- referred CAM Objects -->
04:     <didl:Statement mimeType="text/xml"  xmlns:core="urn:c4h:core:2009">
05:      <core:containsCAMObject>urn:c4h:inst:0-39-36-1</core:containsCAMObject>
06:      <core:containsCAMObject>urn:c4h:inst:0-24-47-5</core:containsCAMObject>...
07:     </didl:Statement>
08:    </didl:Descriptor>
09:  ...
10:   <didl:Item> ... </didl:Item> <!—enclosed CAM Object definition -->
11:   ...
12:  </didl:Container>
13: </didl:DIDL>
```

**Figure 8 MPEG-21 Fragment describing CAM Bundle and CAM Bundle metadata**

### 5.3 Core metadata encoding

In the remaining of this section, we illustrate first some common CAM4Home properties that can be map straightforward onto MPEG21/7 constructs. Then, we illustrate the mechanism that we have employed for encoding metamodel concepts not included natively in MPEG-21.

Due to the number of requirements defined under the project, we have selected a few concepts that cover several multimedia metadata domain (e. g., creation information, semantic relationships, etc).

#### 5.3.1 Common metadata

CAM Objects and CAM Bundles descriptions are composed of some common metadata such as title, description, creation date and time, creator, etc. Some of this metadata are already enclosed in the `mpeg7:CreationInformationType` element. Figure 9 illustrates the encoding of some common metadata. For instance, the `core:creationDateTime` metadata is encoded (lines 17-21) by the `CreationCoordinate/CreationDate/TimePoint` element. The `core:description` metadata is encoded (lines 14-16) by the `Abstract/FreeTextAnnotation` text element. Information about the creator is brought by the `mpeg7:Creator` element (lines 4-12).

```
01:<didl:Descriptor>
02: <didl:Statement mimeType="text/xml; charset=UTF-8" >
03:  <mpeg7:Creation  xmlns:core="urn:c4h:core:2009">
04:  <mpeg7:Creator>                              <!-- encodes core:Creator -->
05:   <mpeg7:Role href="urn:mpeg:mpeg7:cs:RoleCS:AUTHOR">
06:    <mpeg7:Agent xsi:type="PersonType">
07:     <mpeg7:Name>
08:      <mpeg7:GivenName>Rob</mpeg7:GivenName>
09:     <mpeg7:FamilyName>Koenen< mpeg7:FamilyName>
10:    </mpeg7:Name>
11:   </mpeg7:Agent>
12:  </mpeg7:Creator>
13:  <mpeg7:Title xml:lang="en"> CAN 2010 football</mpeg7:Title>
14:  <mpeg7:Abstract>                             <!-- encodes core:description -->
15:    <mpeg7:FreeTextAnnotation> All matches played in CAN 2010</mpeg7:FreeTextAnnotation>
16:  </mpeg7:Abstract>
17:  <mpeg7:CreationCoordinates>                  <!-- encodes core:creationDateTime -->
18:   <mpeg7:CreationDate>
19:     <mpeg7:TimePoint>2008-07-13T14:10:00</mpeg7:TimePoint>
20:   </mpeg7:CreationDate>
21:  </mpeg7:CreationCoordinates>
22:       . . . . .
```

```
23: <mpeg7:Creation>
24:</didl:Descriptor>
```

**Figure 9 Common metadata encoding using MPEG-21/7**

### 5.3.2 Relationships inside the bundle

Within a CAM Bundle, we can express simple-qualitative or complex-quantitative relationships that apply for some of the contained CAM Object. A simple-qualitative relation (*Relationship*) only indicates its nature. The complex-quantitative relation introduce have some specific attributes in order to better characterize the relation between two CAM objects. The simple *Relationship* instances can be represented by a special-purpose `didl:Descriptor` that expresses the relationships between CAM Objects by using the predefined `mpeg7:relation` element.

Figure 10 illustrates the encoding of the "*SameEvent*" relationship between two CAM Objects referenced by the `mpeg7:source` (which corresponds to `core:sourceReference`) and `mpeg7:target` (which corresponds to `core:targetReference`) elements. The `mpeg7:typelist` attribute indicates the nature of the relation between the source and target CAM Objects (`core:relationType`).

```
01:<didl:DIDL xmlns:didl="urn:mpeg:mpeg21:2002:01-DIDL-NS">
02:  <didl:Container> ...
03:   <didl:Descriptor>
04:    <didl:Statement mimeType="text/xml; charset=UTF-8" >
05:     <mpeg7:relation typelist="urn:c4h:rel:SameEvent" source="#CAMObject1"
06:           target="#CAMObject2"/>
07:     </didl:Statement>
08:    </didl:Descriptor>
09:   </didl:Container>
10: </didl:DIDL>
```

**Figure 10 Relationship encoding between two CAM Objects within a CAM Bundle**

The concepts defined by MPEG-7 standard to describe the relationships between multimedia elements do not allow the description of complex relationships. Hence, we have defined a new extension of `mpeg7:RelationType` element in order to support additional descriptive parameters. Figure 11 shows the XML Schema definition of this new element (`ExtendedRelationship`). This new specification allows defining new

parameters for the relationships by introducing (name,value) pairs through the `Param` element.

```
01:<complexType name="ExtendedRelationship"> <!-- encodes subclasses of core:Relationship--
>
02: <complexContent>
03:  <extension base="mpeg7:RelationType">
04:   <element name="Param" minOccurs="0" maxOccurs="unbounded">
05:    <complexType>
06:     <attribute name="name" type="string" use="optional"/>
07:     <attribute name="value" type="anySimpleType" use="optional"/>
08:    </complexType>
09:   </element>
10:  </extension>
11: </complexContent>
12:</complexType>
```

**Figure 11: XML Schema for ExtendedRelationship element definition**

In Figure 12 we illustrate the encoding of the `AdvertiseRelationship` which is one of the relationships we have defined under CAM4Home project. It describe the fact that one object referenced by `#CAMObject1` includes some advertisement section that are seen as independent piece of content referenced by `#CAMObject2` (for instance, a movie trailer advertises the whole movie). `AdvertiseRelationship` contains attributes describing the advertisement characteristics (its position, duration, size, a direct buying link, etc.).

```
01:<didl:DIDL xmlns:didl="urn:mpeg:mpeg21:2002:01-DIDL-NS">
02: <didl:Container> ...
03:  <didl:Descriptor>
04:   <didl:Statement mimeType="text/xml; charset=UTF-8">
05:    <core:ExtendedRelationship  xmlns:core="urn:c4h:core:2009"
              source="#CAMObject1" target="#CAMObject2"
              typelist="urn:c4h:rel:AdvertiseRelationship">
06:      <core:Param>
07:       <core:name>rapidBuyURL</core:name>
08:       <core:value>http://store.apple.com/us</core:value>
09:      <core:Param>
10:    </core:ExtendedRelationship>
11:   </didl:Statement>
12:  </didl:Descriptor>
13: </didl:Container>
14:</didl:DIDL>
```

**Figure 12: Encoding of quantitative ExtendedRelationship instances**

For the community created metadata that is not directly supported by MPEG-21/7, we make extensive use of the `didl:Statement` construct in order to embed in MPEG-21 description specific metadata properties defined in CAM4Home.

### 5.3.3 Community Created Metadada

Community created metadata concept regroups all metadata created by users consuming the content (user comments, user ratings or social tags). Since there is no MPEG-21/7 concepts allowing the description of such data, we have created new concepts for user feedbacks description. Figure 13 illustrates the encoding of such metadata.

```
01:<didl:DIDL xmlns:didl="urn:mpeg:mpeg21:2002:01-DIDL-NS">
02: <didl:Container>
03:  <didl:Descriptor>
04:   <didl:Statement mimeType="text/xml; charset=UTF-8" >
05:    <core:CommunityCreatedMetadata  xmlns:core="urn:c4h:core:2009">
06:        . . . . .
07:     <core:SocialTag>
08:      <core:tagValue>foot, match, zidane<core:tagValue>
09:     </core:SocialTag>
10:     <core:UserComment>
11:      <mpeg7:CreationTime>2003-04-20T13:20:25+09:00</mpeg7:CreationTime>
12:      <core:Comment> your video is null<core:Comment>
13:     </core:UserComment>
14:     <core:UserRating>
15:      <mpeg7:CreationTime>2003-04-20T13:20:25+09:00</mpeg7:CreationTime>
16:      <core:Rate>1/5<core:Rate>
17:     </core:UserRating>
18:         . . . . .
19:    </core:CommunityCreatedMetadata >
20:   </didl:Statement>
21:  </didl:Descriptor>
22: </didl:Container> ...
23:</didl:DIDL>
```

**Figure 13 A fragment of Community Created Metadata instance encoding**

Even though in MPEG-7 some elements (such as `mepg7:KeywordAnnotation`) might have been used to encode social tag values, we have preferred to introduce new constructs in order to clearly state that such metadata are created by community lambda users and not necessarily by a clearly defined annotator. So, we had found it ambiguous to have homonymous elements carrying information with very different levels of confidence.

In the following section we discuss how the supplementary metadata can be encoded using MPEG-21 native constructs (such as DIA elements) or some extensions that we have designed.

## 5.4 Supplementary metadata encoding

The CAM4Home supplementary metadata model provides the information about the context in order to enable the interoperability of various platform and content delivery

services. The CAM4Home supplementary metadata provides the structures to profile users, communities, devices, networks and platform services. The MPEG-21 DIA natively covers a part of the CAM4Home supplementary metadata model, so naturally, we have tried to identify adequate mappings. MPEG-21 deals with most of the issues that we have defined under CAM4Home project. In this section we show only a part of CAM4Home supplementary user-related metadata specification using MPEG-21.

### 5.4. 2 Native constructs for user profile encoding

Figure 14 illustrates the use of MPEG-21 standard for encoding user profile. The example represents the encoding of metadata concerning user personal description (lines 4-9) which is a part of user profile.

```
01:<dia:Description xsi:type="UsageEnvironmentType">
                                    <!-- encodes suppl:C4HUserPersonalDescription -->
02: <dia:UsageEnvironment xsi:type="UserCharacteristicsType">
03:   <dia:UserCharacteristics xsi:type="UserInfoType">
04:    <dia:UserInfo xsi:type="mpeg7:PersonType">
05:     <mpeg7:Name>
06:      <mpeg7:GivenName>John</mpeg7:GivenName>
07:      <mpeg7:FamilyName>Doe</mpeg7:FamilyName>
08:     </mpeg7:Name>
09:    </dia:UserInfo>
10:   </dia:UserCharacteristics>
11:  </dia:UsageEnvironment>
12: </dia:Description>
```

**Figure 14 MPEG-21 Fragment describing User Profile**

### 5.4. 2 New constructs for user profile encoding

In Figure 15, we present some user related metadata (about the user physiological state – lines 4-8 in Figure 15) that is not enclosed in the MPEG-21/7. We have specialized the `dia:UserCharacteristicsBaseType` into `suppl:PhysiologicalStateType` in order to introduce specific information such as (blood pressure, fatigue level, etc.).

```
01:<dia:Description xsi:type="UsageEnvironmentType">
02: <dia:UsageEnvironment xsi:type="UserCharacteristicsType">
03:   <dia:UserCharacteristics xsi:type="suppl:PhysiologicalStateType">
04:    <suppl:PhysiologicalState  xmlns:suppl="urn:c4h:suppl:2009">
05:     <suppl:bloodPressure> 12 cmHg</suppl:bloodPressure>
06:     <suppl:fatigue>false</suppl:fatigue>
07:      ......
08:    </suppl:PhysiologicalState>
09:   </dia:UserCharacteristics>
10:  </dia:UsageEnvironment>
11: </dia:Description>
```

**Figure 15 PhysiologicalState instance encoding**

## 6. Discussion

In this section, we discuss some of our choices to encode the metadata model. We have adopted an XML-based approach to encode it as the processing of XML files is rather simple and widely spread. The data overhead resulting from the XML encoding is compensated by the flexibility of interpreting XML documents. In order to acquire significant compression rates and reduce the structure overhead, generic solutions exist [22][23] and they can be applied directly to encodings before the transmission as well as the interpretation.

If our proposed metadata model is made using plain XML (i.e. encoding based on XML Schema [11] only) it would have contained smaller data overhead. All the metadata related to a given element would have to be encoded together within the given element. However, we considered the extension of existing comprehensive multimedia description standards such as MPEG-21, as a well-established standard gives more visibility to our solution. The MPEG-21 covers similar topics as CAM4Home regarding the content (core part of CAM4Home) and the deployment context and environment (supplementary part of CAM4Home) descriptions. The main inconvenient of this solution is that the client would need to be capable of interpreting a very large set of encoding structures.

For covering parts of the metadata metamodel that are not fully supported by MPEG-21/7 we have applied two different methods. The first one consists in introducing specific metadata by means of `didl:Statement` elements. This is applied when no similar concepts exist in MPEG-21/7. The second one which corresponds to the extension of the existing description schemes and constructs of MPEG-21/7. This is applied when the existing MPEG-21/7 offer only restricted encoding possibilities (such as the `mpeg7:Relation` element or the `dia:UserCharacteristics`). We made this choice in

order to keep the description as compatible as possible with the native MPEG-21/7 structures. Effectively, in this way, even applications not able to fully interpret the extended versions, they can still use the information in a restricted manner. One can rely on the fact that a given structure (like a complex relationship) is basically an extension of a well-known structure and treat it as such.

## 7. Conclusion and Perspectives

In this chapter, we have exposed an MPEG-21 encoding of a metadata model that we have designed collaboratively within the CAM4Home project. In the future, new needs and new use contexts (intelligent networks, more powerful devices) will emerge. It is important to be able to respond to these two challenges by providing on one hand: means for extending the descriptions and, on the other hand, tools and encodings for adapting the descriptions to the interpretation capabilities of new devices or new networks.

The first part is ensured by the flexibility features of the encoding, as we have presented them in the previous section. The second part might require the automatic mapping between actual encodings (in MPEG-21) and future encodings. This work, in some way similar to ontology matching techniques, is currently explored in our research team [3][4]. In the future, we expect to free the client from any consideration related to a specific encoding. The client's queries are formulated using model terms, and the results are provided in terms of model instances without any reference to a specific encoding used to represent and transfer the information from one end to the other within the system.

## References


[1] M. Amielh and S. Devillers, "Bitstream syntax description language:Application of XML schema to multimedia content adaptation," in *Proc. 11th Int. World Wide Web Conf. (WWW2002)*, Honolulu, HI, May 2002

[2] M. Amielh and S. Devillers, "Multimedia Content Adaptation with XML," in *Proc. Int. Conf. on MultiMedia Modeling*, 2001.

[3] S. Amir, I.M. Bilasco, Md. H. Sharif, and C. Djeraba, "Towards a unified multimedia metadata management solution," *Intelligent Multimedia Databases and Information Retrieval: Advancing Applications and Technologies*, IGI Global, February 2010.

[4] S. Amir, "Un système d'intégration de métadonnées dédiées au multimédia," *INFORSID'2009*, Toulouse, France, Mai 26-29th, 2009.

[5] I. M. Bilasco , S. Amir, P. Blandin , C. Djeraba, J. Laitakari, J. Martinet, E. M. Gracia , D. Pakkala, M. Rautainen, M. Ylianttila, J. Zhou, "Semantics for intelligent delivery of multimedia content," in *Proc. SAC ACM 2010*, Sierre, Switzerland 22 – 26 march, 2010

[6] J. Bekaert, E. De Kooning, and H. de Sompel, "Representing digital assets using MPEG-21 Digital Item Declaration," *Int. J. Digit. Libr.* 6, 2 (Apr. 2006), 159-173.

[7] I. Burnett, R. Van de Walle, K. Hill, J. Bormans, and F. Pereira, "MPEG-21: Goals and achievements," *IEEE MultiMedia*, 10(4), 60-70

[8] E. Christensen, F. Curbera, G. Meredith, and S. Weerawarana, "Web Services Description Language (WSDL) 1.1," http://www.w3.org/TR/wsdl.

[9] Sharable Content Object Reference Model (SCORM). Advanced Distributed Learning. [Online] Available: http://www.adlnet.org/

[10] IMS Content Packaging XML Binding. IMS Global Learning consortium. [Online] Available: http://www.imsproject.org/

[11] D.C. Fallside (Ed.). 2002. "XML Schema Part0: Primer (W3C Recommendation)," Available at http://www.w3.org/TR/xmlschema-0/

[12] Metadata Encoding & Transmission Standard. The Library of Contress. [Online]. Available: http://www.loc.gov/standards/mets.

[13] E. Gustafsson, A. Jonsson "Always best connected," *Wireless Communications IEEE*, volume 10, issue 1, Feb. 2003.

[14] J. Hunter, "An overview of the MPEG-7 Description Definition Language (DDL)," *IEEE Trans. on Circuits and Systems for Video Technology*, 11(6), 765-772.

[15] International Organization for Standardization. (2003). *ISO/IEC 15938-5:2003: Information technology -- Multimedia content description interface -- Part 5: Multimedia description schemes* (1st ed.) Geneva, Switzerland

[16] International Organization for Standardization. (2004). *ISO/IEC 21000-2:2003. Information technology -- Multimedia framework (MPEG-21)-- Part 2: Digital Item Declaration* (1st ed.) Geneva, Switzerland

[17] International Organization for Standardization. (2003). *ISO/IEC 21000-3:2003: Information technology -- Multimedia framework (MPEG-21) -- Part 3: Digital Item Identification* (1st ed.) Geneva, Switzerland

[19] International Organization for Standardization. (2004). *ISO/IEC 21000-7:2004: Information technology -- Multimedia framework (MPEG-21)-- Part 7: Digital Item Adaptation* (1st ed.) Geneva, Switzerland

[20] G. Klyne, F. Reynolds, C. Woodrow, H. Ohto, "Composite Capability/Preference Profiles (CC/PP): Structure and Vocabularies", *W3C Recommandation*, http://www.w3.org/TR/CCPP-struct-vocab/

[21] D. Martin (Ed.), "OWL-S *Semantic markup for web services",* Member Submission, W3C, November. http://www.w3.org/Submission/OWL-S

[22] F. Pereira, A. Vetro, T. Sikora. "Multimedia Retrieval and Delivery: Essential Metadata Challenges and Standards," in *Proc. of the IEEE*, vol. 96, issue 4, April 2008.

[23] J. Schneider, T. Kamiya, "Efficient XML Interchange (EXI) format," Available at http://www.w3.org/TR/exi/

[24] H. Schulzrinne, E. Wedlund, "Application-layer mobility using SIP," *ACM SIGMOBILE Mobile Computing and Communications Review*, volume 4, issue 3, July 2000.

[25] TV-Anytime Forum. [Online]. Available: http://www.tv-anytime.org.

[26] J. Zhou, M. Rautiainen, M. Ylianttila, M. Foulonneau, P. Blandin, "Metamodeling for Community Coordinated Multimedia and Experience on Metamodel-Driven Content Annotation Service Prototype," in *Proc. IEEE Congress on Services Part II*, September 23-26, Beijing, China, 88-95.