The Semantics of MPEG-21 Digital Items Revisited

Christian Timmerer
Klagenfurt University, Austria
Universitätsstrasse 65-67
A-9020 Klagenfurt
+43 (0)463 2700 3621
christian.timmerer@itec.uni-klu.ac.at

Maria Teresa Andrade and Pedro Carvalho
INESC Porto, Porto, Portugal
Rua Dr. Roberto Frias 378
P-4200-465 Porto
+35 (0)122 2094000
{maria.andrade,pedro.carvalho}@inescporto.pt

Davide Rogai
Comm.it
Via M. De' Bernardi, 65
50145, Firenze. Italy
+39 (0)55 3434277
davide.rogai@comm-it.it

Giovanni Cordara
Telecom Italia Lab
Via Reiss Romoli, 274
10148, Torino. Italy
+39 (0)11 2286389
giovanni.cordara@telecomitalia.it

ABSTRACT
The MPEG-21 standard forms a comprehensive multimedia framework covering the entire multimedia distribution chain. In particular, it provides a flexible approach to represent, process, and transact complex multimedia objects which are referred to as Digital Items (DIs). DIs can be quite generic, independent of the application domain, and can encompass a diversity of media resources and metadata. This flexibility has an impact on the level of interoperability between systems and applications, since not all the functionality needs to be implemented. Furthermore, additional semantic rules may be implemented through the processing of the Digital Item which is possibly driven by proprietary metadata. This jeopardizes interoperability and consequently raises barriers to the successful achievement of augmented and transparent use of multimedia resources. In this context, we have investigated and evaluated the interoperability at the semantic level of Digital Items throughout the automated production, delivery and consumption of complex multimedia resources in heterogeneous environments. This paper describes the studies conducted, the experiments performed, and the conclusions reached towards that goal.

Categories and Subject Descriptors
H.5.1 [**Information Interfaces and Presentation (e.g., HCI)**]: Multimedia Information Systems – *audio input/output, evaluation/methodology, video (e.g., tape, disk, DVI).*

General Terms
Experimentation, Standardization, Languages.

MSi08, October 31, 2008, Vancouver, British Columbia, Canada.

Keywords
MPEG-21, Digital Items, Semantics, Evaluation, Metadata.

1.INTRODUCTION
The aim of the MPEG-21 Multimedia Framework is to enable transparent and augmented use of multimedia resources across a wide range of networks, devices, user preferences, and communities, notably for trading (of bits). As such, it provides the next step in MPEG's standards evolution, i.e., the transaction of Digital Items among Users [1].

A *Digital Item (DI)* is a structured digital object with a standard representation and metadata, i.e., it is the fundamental unit of transaction and distribution within the MPEG-21 multimedia framework. In other words, it enables the aggregation of multimedia resources together with metadata, licenses, identifiers, intellectual property management and protection (IPMP) information, and methods within a standardized structure. With this definition in mind one can create Digital Items that are very generic and independent of the application domain. Furthermore, it is possible to represent semantically similar (or equal) Digital Items in different ways in terms of structure and syntax elements. This has an impact on the level of interoperability that can be achieved within and across application domains. As interoperability across domains is somewhat difficult to achieve, we will focus our work on the automated production, delivery, and consumption of Digital Items in heterogeneous environments.

In this paper we investigated two types of interoperability with respect to the semantics of Digital Items:

— Interoperability at the level of the Digital Item itself, i.e., access/browse media resources and associated metadata, process/consume the Digital Item as such, etc.

— Interoperability at the level of the declaration of a Digital Item, i.e., processing of its representation in various ways (e.g., parsing, validation).

Therefore, we have investigated use cases both in the automated production of Digital Items and in their delivery within heterogeneous environments to different types of end user terminals. Based on the data models defined by these use cases we have defined evaluation criteria for the two types of

interoperability as introduced above. Finally, we have conducted a thorough analysis and evaluated the criteria against the use cases.

The innovation brought in by this work, include the identification of interoperability problems among compliant implementations of the same standard specification, more specifically the MPEG-21 standard, the definition of evaluation criteria and consequently the specification of a procedure to evaluate interoperability of MPEG-21 DIs produced by those different implementations.

The remainder of this paper is organized as follows. Section 2 reviews the major building blocks that may constitute a Digital Item and references recent (industry) adoptions. The inputs for the experiments conducted are described in Section 3 and their results are discussed in Section 4. Section 5 concludes the paper and provides recommendations for the construction of interoperable Digital Items.

2.MPEG-21 DIGITAL ITEMS

The standard representation of Digital Items is defined by a model – in Extended Backus–Naur Form (EBNF) [2] – which describes a set of abstract terms and concepts and may be expressed by the XML Schema based Digital Item Declaration Language (DIDL) [3]. The resulting XML document conformant to DIDL is called a DIDL document representing a Digital Item Declaration (DID). The DID may contain several building blocks as defined in the DIDL which defines the structure of the Digital Item. Some of these buildings blocks are reviewed in the following but for detailed information the interested reader is referred to [1][3]. Note that the building blocks described here are defined both in the abstract model and DIDL.

The *Item* element comprises a grouping of sub-items or components. In general, an item can be considered as a declarative representation of a Digital Item. Note that an item without sub-items can be considered as a logically indivisible work and an item that does contain sub-items can be considered a compilation.

The *Component* element defines a binding of a (multi-)media resource to a set of descriptors which provides information related to all or parts of the resource. These descriptors will typically contain control or structural information about the resource such as bit rate, character set, start points, or encryption information.

A *Descriptor* element associates information with the enclosing element, i.e., its parent (e.g., item) or following sibling (e.g., component). The information can itself be a component (e.g., thumbnail of an image) or a textual statement (e.g., metadata such as MPEG-7 [4] or Dublin Core [5]).

A *Resource* element is defined as an individually identifiable asset such as a video, audio clip, image, or textual asset. Note that the resource must be locatable via an unambiguous address or directly embedded within the Resource element.

Digital Items are configurable through the so-called choice/selection mechanism. A *Choice* element describes a set of related *Selection* elements which can affect the configuration of a Digital Item. As such it provides a generic and flexible way for multimedia content selection based on certain criteria defined by the Digital Item author. Such criteria may include rights expressions and/or usage environment constraints.

An example Digital Item describing the Lord of the Rings trilogy is provided below. Note that instead of the verbose XML representation a text-based notation has been chosen for better readability and space constraints (annotations are provided in italics).

```
DIDL
 Item
  Descriptor
   Statement (containing an identifier)
  Descriptor
   Statement (containing the title of the DI)
  Choice (minSelection=0, maxSelection=1)
   Selection (select_id=lotr1)
    Descriptor
     Statement (metadata about the selection)
   Selection (select_id=lotr2 + further meta)
   Selection (select_id=lotr3 + further meta)
  Item
   Condition (require=lotr1)
   Component
    Resource (mimeType=video/mpeg, ref=http://…)
  Item
   Condition (require=lotr2)
   Descriptor
    Statement (info about this item, e.g., title)
   Component
    Descriptor
     Statement (info about this component, e.g.
                coding params, bit rate, etc.)
    Resource (mimeType=video/mpeg, ref=http://…)
  Item
   Condition (require=lotr3)
   Component
    Resource (mimeType=video/mpeg, ref=http://…)
```

This Digital Item (DI) comprises one item with three sub-items which constitutes this DI as a collection of DIs, i.e., each part of the Lord of the Rings trilogy. The main Item element has two Descriptor elements, the first identifies the DI and the second provides metadata such as title, credits, etc. possibly expressed in yet another description format (e.g., MPEG-7, Dublin Core). The Choice element prescribes the possible sections, each having an identifier (select_id) and possibly further metadata using a Descriptor element. An instantiation of such a selection is referred to as the configuration of the Digital Item. The identifier is used by the Condition elements within the subsequent item elements in order to declare the relationship between the selections and the sub-Items of the overall Digital Item collection. For example, the second sub-item describes the second part of the trilogy (require=lotr2) providing further metadata about this item, and comprises a Component element that references the actual movie by using the Resource element. The difference between the two Descriptor elements of this item is that the one at the item level contains descriptive information about this item (i.e., about the second part of the Lord of the Rings trilogy) whereas the one at the component level provides metadata about the actual media resource.

In the following we will review some adoptions of the Digital Item concept in both (other) standardization bodies and industry:

— The first adoption of Digital Items was within the Universal Plug and Play (UPnP) forum as DIDL-Lite [6] which is derived from a subset of MPEG-21 DIDL. It is basically used as a container format within UPnP's content directory and enhanced with UPnP-specific data (e.g., media resource attributes such as bitrate, resolution, size, etc.) and Dublin Core metadata. As it is a derivation of a subset of DIDL, it is

not compatible with MPEG-21 DIDL and, thus, interoperability between the two formats is not guaranteed.

— Part 4 of MPEG-21 [7] – Intellectual Property Management and Protection (IPMP) Components – defines its own syntax (i.e., IPMPDIDL) enabling the declaration of protected Digital Items while maintaining interoperability with unprotected Digital Items following DIDL through the DI model.

— Some of the MPEG Multimedia Application Formats (MAFs) [8] restrict the usage of DIDL by defining means how to instantiate Digital Items for a given application format. For example, the Photo Player MAF, that provides an interoperable solution for digital photo library applications carrying JPEG images and their associated MPEG-7 metadata in an MPEG-4 file, restricts the structure of the Digital Item by defining specific semantics for the individual elements such as item, descriptor, component, etc. [9]. As MAFs just select a subset of DIDL without modifying syntax or semantics, they will preserve interoperability with DIDL.

— The abstract Digital Item model has been adopted within Microsoft's Interactive Media Manager (IMM) [10] and implemented using the Web Ontology Language (OWL) [11]. It uses Dublin Core but also allows for the inclusion of domain-specific metadata (e.g., IPTC, EXIF, XMP, SMPTE, etc) or custom ontology predicates. However, as it just adopts the abstract model, interoperability at this level may be provided but not at the actual implementation level. Interestingly, IMM also adopts part 3 of MPEG-21 – Digital Item Identification (DII) – which allows for uniquely identifying Digital Items and parts thereof [12]. Note that MPEG-21 DII does not define a new identification scheme but enables the inclusion of existing ones (e.g., ISBN, ISSN, ISAN, ISRC).

The information here provided plays an essential role to enable the reader to gain a better understanding of the DID nature, mechanisms and usage. It clearly shows that, although having a formal structure with well defined elements which offer specific functionality or serve specific purposes, it still constitutes a very flexible framework for the representation of complex multimedia objects. It allows different combination of the distinct elements and the insertion of multiple types of data, and metadata. The latter may follow specific and proprietary formats often conveying information intended for the processing or presentation of the DI itself.

3. DESCRIPTION OF INPUTS AND EXPECTED OUTPUTS

3.1 Use Case Scenarios

In order to evaluate the interoperability at the level of Digital Items and its declaration we have investigated two use case scenarios in the area of *automated production of cross-media Digital Items (UCS-1)* and the *provisioning of end-to-end Quality of Service (QoS) for Digital Items across heterogeneous terminals and networks (UCS-2)*.

Such use cases represent the referred scenarios of two projects funded by the European Commission, namely AXMEDIS [17] and ENTHRONE [18], adopting the MPEG-21 technology in the context of production, delivery, and consumption of Digital Items in heterogeneous and distributed environments. The members of those projects deeply cooperated to conduct such experiments: in particular, the DIs being used in each project have been exchanged and consumed by players implemented in both projects, in order to evaluate them against a list of criteria previously identified, as reported in next section.

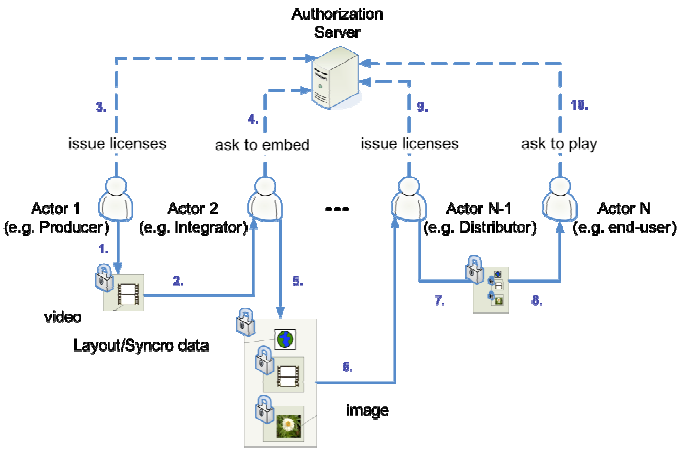

Figure 1. Digital Items workflow for a cross-media content value-chain.

The choice of those scenarios seems to fit with purpose of those experiments, since the DIs adopted in the projects are conformant to the MPEG-21, and they have been applied to different contexts in similar application scenarios. The DIs have been tailored to the specific use cases: they support a subset of all the functionalities/elements of the standard and include some proprietary metadata. Such DIs, therefore, represent generic examples of DIs, suitable for being tested and validated through interoperability tests.

Automated Digital Item creation (UCS-1) aims to realize Digital Item creation flows where these are created, integrated, and finally delivered to the end users. These flows can include automatic creation/processing of Digital Items, which can embed a presentation logic, in order to be flexibly rendered within the end-user devices. The Digital Item manipulation involves different actors:

— The Content Provider inserts in the value chain new Digital Items containing raw media files, encoded in any format;

— More Content Integrators, in different steps, annotate Digital Items, compose them to create new ones, and embed presentation logic in composed Digital Items in order to define layout, formatting and behavior.

At each step the DIs can be protected.

In order to *enable the provisioning of end-to-end QoS for Digital Items across heterogeneous environments (UCS-2)*, several actors along the delivery path are involved and play a crucial role that will be briefly highlighted:

— The Content Provider (CP) prepares the actual multimedia content as MPEG-21 Digital Items facilitating scalable coding formats and metadata formats.

— The Service Provider (SP) provisions and offers multimedia services to end-users and enriches the multimedia content with additional metadata with respect to Service Level Agreements (SLAs) [13] taking into account constraints

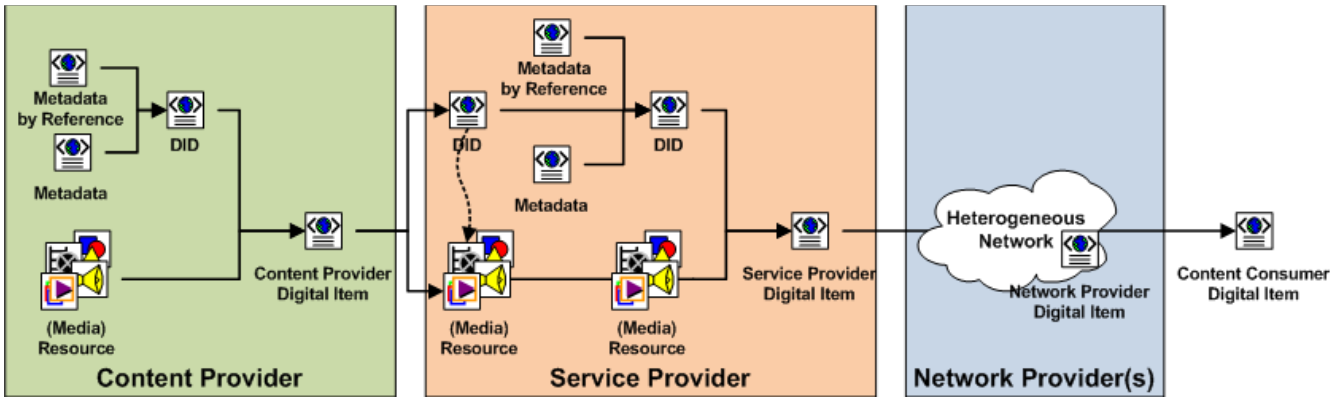

Figure 2. Provisioning of end-to-end Quality of Service (QoS) for Digital Items.

imposed by access networks for service provisioning towards the content consumer.

— The Adaptation Provider (AP) operates in close relationship with the SP and the NPs. Its goal is to provide improved QoS of content delivery while optimizing available system and network resources across the end-to-end chain. It takes content adaptation decisions according to the a-priori-known as well as dynamically received context information. Note that the actual content adaptation is done by the Content-/Service-/Network Providers.

— The Network Provider (NP) offers QoS-based network connectivity services at its autonomous domain level. Co-operation is needed among NPs for providing inter-domain QoS-based network connectivity services.

— The Content Consumer (CC) requests the services provided by the SP and consumes it through his end-device. The actual end device functions depend on the business model.

3.2Data Model of Inputs

3.2.1Use Case Scenario 1

The Use Case Scenario 1 (UCS-1) targets the delivery of a unique package along the value chain of the content production to the distribution and finally to the end-user. The "cross-media" package is created as the result of different composition/integration of existing DIs and it has to be consumed at the end-user site by a specific player. The player is able to coordinate the different resources included inside the DI by reading one or more presentation specific resources, e.g., LASeR [14], HTML [15], SMIL [16], etc.

Figure 1 depicts an example value chain for automated cross-media production in a governed environment. In this particular example, the Producer creates the initial Digital Item, in a protected form, including a video resource and associated metadata. An Integrator select the Producer's Digital Item to create a new one putting it together with an image; the Integrator adds to the Digital Item an HTML presentation layout which refers to the video and the image resources. The Integrator adds protection (information) to the Digital Item also. The Integrator could perform the required manipulation on the Digital Item on the basis of a specific license issued by the Producer. Further in the value chain, the Distributor selects the composed Digital Item to be distributed to her/his customers. The Distributor needs to issue licenses that grant his customer to play the composed Digital Item. Actually she/he could issue these licenses on the basis of a specific license issued by the previous actors in the value chain. The last actor that is involved in the value chain is the End-User: she/he plays the Digital Item provided by the Distributor, her/his player accesses and render the HTML presentation and the related resources. Also the End-user is allowed to play the Digital Item on the basis of the license issued by the Distributor.

This data model has been developed in the context of the AXMEDIS project [17] and differentiates between composite and basic objects. A composite object is structured as follows:

```
DIDL
 Item
  OBJECT_AXOID
  OBJECT_METADATA
  ITEM [0]
  ...
  ITEM [n]
```

Where:

— OBJECT_AXOID is a Descriptor/Statement combination containing a Digital Item Identifier;

— OBJECT_METADATA is a sequence of Descriptor elements containing the metadata of the object; and

— each ITEM describing a basic object which is introduced in the following.

The structure of a basic object:

```
DIDL
 Item
  OBJECT_AXOID
  OBJECT_METADATA
  CONTENT
```

OBJECT_AXOID and OBJECT_METADATA are defined in a similar way as above. CONTENT is a sequence of Component elements containing a resource or a reference to a resource without additional metadata.

The metadata defined within the OBJECT_METADATA can be clustered into domain-specific and domain-independent metadata. The former is defined in a proprietary way and allows for predefined usage within the framework where this structure has been specified. The latter comprises metadata assets based on

open standards such as MPEG-7 or Dublin Core in order to allow MPEG-21 terminals to access these metadata even if they are not compliant with this framework.

3.2.2Use Case Scenario 2

As this use case addresses end-to-end QoS (across heterogeneous terminals and networks), various requirements from all actors within the end-to-end delivery chain need to be taken into account. The actors involved can be categorized into content provider (CP), service provider (SP), network provider (NP) and customer/content consumer (CC). Thus, the Digital Item may undergo various – sometime significant – changes while being delivered from the CP towards the CC as illustrated in Figure 2. However, the main changes will happen within the CP/SP domain. Hence, we will focus on the requirements coming from CP and SP, e.g., system-wide identification, temporal availability, encoding characteristics, adaptation possibilities and expected resulting qualities, constraints to be considered, etc.

As shown in Figure 2, the content provider creates the initial Digital Item including the actual media resource and associated metadata. In this context, this Digital Item is referred to as the CP Digital Item. A service provider may enrich this CP Digital Item with additional information pertaining to a particular service offered to the content consumer. This may, for example, include further metadata or an adapted version of the original Digital Item which fits the a-priori known requirements of certain customers. The network providers are responsible for appropriate transportation of the so-called service provider Digital Items which may undergo well-defined modifications that optimize the transmission over heterogeneous networking infrastructures. Finally, the content consumer receives the desired Digital Item appropriate to her/his context, i.e., anywhere, anytime, and with any device.

This data model has been developed in the course of the ENTHRONE project [18] and differentiates between composed and final items. A composed item comprises sub-items which can be themselves composed or final whereas a final item does not contain any further sub-items but components. This differentiation is similar to the data model from UCS-1 from a semantic point of view but not from a syntactical point of view. Digital Items according to UCS-2 are structured as follows:

```
DIDL
 Declaration(s)  (referable descriptors)
 Container
  Descriptor(s)  (top-level container descriptors)
  Item           (composed item)
   Descriptor(s) (top-level item descriptors)
   Item          (final item)
    Descriptor(s) (item-level descriptors)
    Component(s)
     Descriptor(s) (component-level descriptors)
     Resource
   Item(s) (further composed or final items)
  Item(s)  (further composed or final items)
```

Where:

— Declaration elements may include descriptors that are used by reference (instead of duplicating them);

— Descriptor elements provide the metadata for the DI at different levels (i.e., container, composed item, final item, component); and

— one or more Component elements, each of them representing a variation of semantically equal media resources (e.g., different bit rates, resolutions, qualities).

The Descriptor elements may contain standardized metadata (e.g., MPEG-7, TVAnytime, MPEG-21) but also proprietary data, mainly required for the delivery as defined within this framework. This proprietary metadata is used for the coordination of various entities within the delivery chain (e.g., servers, proxies, adaptation gateways) in order to provide an agreed level of quality to the end user. It is important to note that this proprietary metadata does not hamper the consumption of media resources or open standards-based metadata at the receiving terminal.

3.3 Expected outputs

The main result to be expected shall provide an evaluation of the level of interoperability that can be achieved (based on the inputs) both at the level of the Digital Item and its declaration. Therefore, we have defined evaluation criteria which are presented and discussed in the next section.

In particular, the investigation at the level of the Digital Item itself requires a thorough analysis of the semantics of the Digital Items following the two data models that served as an input to this experiment. Additionally, we expected to derive some further requirements which need to be addressed in order to increase the semantic understanding of a Digital Item and, thus, its interoperability.

4.EXPERIMENTS CONDUCTED

4.1Evaluation Criteria

The evaluation criteria have been separated in those regarding the interoperability at the level of the Digital Item Declaration (DID), i.e., its XML representation, and the Digital Item itself. In particular, the latter shall give detailed information about the semantics of Digital Items.

4.1.1 Interoperability at the Level of the Digital Item Declaration

C 1. DID conformance according to the DID model as defined in ISO/IEC 21000-2? (yes – no, why?)

C 2. DID conformance according to the DIDL as defined in ISO/IEC 21000-2? (yes – no, why?)

Note: Valid according to DIDL schema and additional validation rules, i.e., we may use DID reference software for checking conformance.

C 3. Are the data models / DI models related to the use cases scenarios unambiguously interpretable from a syntactical and logical point of view? (yes, the DIs can be consumed – no, why?)

C 4. Are there issues in parsing a Digital Item Declaration? (no – yes, what kind of?)

Note: Loading into memory and resolving XINCLUDEs (version 2) or References (version 1).

C 5. Are all elements/features from ISO/IEC 21000-2 supported by the data models / DI models related to the CE use cases? (yes – no, which are not and why?)

C 6. Are there any syntax restrictions on top of ISO/IEC 21000-2 (e.g., min/max occurrence of elements) defined? (no – yes, why and is conformance still guaranteed?)

C 7. Are there any semantic restrictions on top of ISO/IEC 21000-2 (e.g., descriptor must contain a Digital Item Identifier) defined? (no – yes, why?)

4.1.2 Interoperability at the Level of the Digital Item

C 8. Are there issues in browsing the "content" (i.e., resources) of a Digital Item? (no – yes, which ones?)

C 9. Are there issues in displaying metadata in an appropriate way? (no – yes, which ones?)

C 10. Are there issues in retrieving and rendering of (media) resources? (no – yes, which ones?)

C 11. Are there issues in decoding/displaying the Digital Item as a whole (i.e., in case a presentation logic is embedded into the Digital Item)? (no – yes, which ones?)

4.2Evaluation and Discussion

In this section we evaluate and discuss the two use case scenarios and their data models against the criteria presented in the previous section.

The two data models conform both to the abstract digital item model and its declaration language (i.e., DIDL). The latter has been tested with XML tools and the available reference software [19]. Additionally, DIs composed according to the two data models can be consumed (i.e., displayed) by both players, i.e., DI according to UCS-1 can be consumed with the player from UCS-2 and vice-versa. Nonetheless, (meta-)data not understood by the respective player are ignored which means that in some cases only the media resources are displayed to the user, even if not in the intended manner. This was due to the fact that the player was not aware of the format (i.e., it was faced with proprietary metadata) or did not have installed the tools required for proper presentation of this data (i.e., for standardized metadata). As there were no issues in parsing the DID, criteria C 1 to C 4 have been considered as fulfilled.

Interestingly, both data models did not make use of the choice/selection mechanism which was either not required or provided through additional semantics defined on top of the MPEG-21 standard. For example, UCS-2 describes multiple variations of semantically equal media resources through multiple Component elements. It is assumed that a single but most appropriate component is delivered to the terminal, i.e., the one that is suitable given the available context (e.g., terminal and network conditions). However, for UCS-1 this would mean that only the first component (as defined in the hierarchy of the DID) is displayed to the user because in such a scenario – if multiple components are declared – a presentation logic must be present that describe how these components shall be presented to the user. Furthermore, both data models require the existence of an identifier for each item or sub-item. Thus, it becomes apparent that, for example, MPEG-21 Digital Item Identification (DII) is required for each application running on top of these data models. For criteria C 5 to C 7 we can conclude that both data models do not fulfil the criteria defined. This has an impact on the semantics but still allows for consumption of Digital Items in an interoperable way with the restriction that this may not be performed as originally intended.

The criteria C 8 to C 10 mainly target issues on whether appropriate processing tools (i.e., decoder, parser, interpreter, etc) for media resources and metadata have been installed at the targeted terminal. For media resources, solutions are already available that install appropriate decoders – if necessary – on-demand. However, although parsers for metadata are very easy to provide, the interpretation thereof is not (yet) possible in an unambiguous way. This can be explained due to the lack of a "decoder" specification for metadata, i.e., the counterpart of a decoder for the media resources. Furthermore, as already mentioned above, (meta-)data not recognizable by the receiving terminal are ignored, e.g., due to unknown namespace definition. Thus, it is advisable to use proprietary formats only for information that does not hamper the play-out of a Digital Item. In other words, use proprietary formats only for information that is not necessarily required or that provide only auxiliary information.

Finally, for C 11 – on consuming Digital Items as a whole – we investigated the usage of a presentation logic (embedded in the Digital Item) versus the implementation thereof within the specification of the application logic as part of the framework definition. For example, the existence of multiple Component elements calls for the need of a presentation logic which unambiguously defines how the various media resources – associated with these components – are presented to the end user. Examples of such presentation logics are HTML, SMIL, or LASeR. However, this requires the inclusion of this logic into the Digital Item (e.g., as a separate Resource) and an appropriate tagging thereof for unique identification by the receiving terminal. Furthermore, in cases multiple presentation logics of the same type are available, another discrimination is required either as part of the DI (e.g., priorities) or as part of the application parsing the DI (e.g., depending on the platform where the DI is consumed).

5. CONCLUSIONS

The DID structure and syntax enables the flexible description of the contents and structure of a DI. As intended, it does not define (by itself) how the DI should be presented. Nevertheless, each of the elements defined in the DID model is intended to represent a specific function.

Due to the flexibility of the DID, many application using MPEG-21 have chosen to implement only a subset of the functionalities/elements or added semantic restrictions to the structure to simplify the DI handling. Also, the use of proprietary metadata and application level rules have been used to drive the processing and presentation of the DIs.

The above mentioned factors have led to some interoperability issues between systems using MPEG-21. Although the implementation of only part of the standard does not imply an invalid DID, its processing may lead to a loss of functionality as some elements will be ignored. Also, the use of proprietary metadata or semantic processing rules can cause an unintended result when exchanging DIs between systems. Therefore, when a more tailored processing is required for the content of a DI, be it metadata or media resources, an additional layer seems necessary to homogenize the use of MPEG-21, even if only to make some processing operations transparent. A possible solution for this would be the use of Digital Item Processing [20] or the emerging MPEG standard with the working title Presentation of Structured Information (PSI) [21]. Although not experimented by the authors, as it was outside the scope of the proposed work, the use of one of these two approaches may prove to be an efficient solution to solve the identified interoperability problems.

Digital Item Processing (DIP) specifies a set of tools to allow the creators of the content to provide their suggestions of interaction with the DI, so that its consumption may be processed in a dynamic and controlled way. In particular, DIP allows to embed Digital Item Methods (DIMs) inside the DI Declaration like a standard library as for programming languages. DIMs, which are based on ECMAScript and use standardised Digital Item Base Operations (DIBOs), can be presented to the User for execution. The interaction of the User with the DIs, including its presentation on the User device, would then be achieved through the selection of methods to be executed upon the DI, thus ensuring an interoperable consumption of Digital Items as intended by the DIM creators. The use of DIP would therefore provide the means to convey the presentation logic inside the DID through the use of DIMS.

Presentation of Structured Information (PSI) represents a recent exploration activity [22] inside MPEG that will result in an extension of the LASeR specification [23]. PSI provides standard mechanisms to point to certain parts of the Digital Item Declaration for presentation purposes. The actual presentation of the DI is then performed according to the LASeR specification. For this reason, it is required that the declarative description of the LASeR presentation is embedded into the Digital Item as a resource.

The implementation of each of these two solutions however, requires that additional information is inserted in the Digital Item and in its respective declaration, as indicated above. This information is not currently mandatory for the generation of compliant MPEG-21 DIs. Accordingly, the effectiveness of any of them would necessarily require that somewhere else, for example, at the application domain level, the decision on whether to use DIP or LASeR+PSI on a mandatory basis, would have to be taken and then observed by the creators of the DIs.

6.ACKNOWLEDGMENTS

Part of this work is supported by the European Commission in the context of the AXMEDIS (contract no. 511299) and ENTHRONE (contract no. 038463) projects. Further information is available at http://www.axmedis.org and http://www.ist-enthrone.org respectively.

7.REFERENCES

[1] F. Pereira, J.R. Smith, A. Vetro, (eds.), "Special Section on MPEG-21", *IEEE Trans. on Multimedia*, vol. 7, no. 3, Jun. 2005.

[2] ISO/IEC 14977:1996, "Information technology -- Syntactic metalanguage -- Extended BNF".

[3] ISO/IEC 21000-2:2005, "Information technology -- Multimedia framework (MPEG-21) -- Part 2: Digital Item Declaration".

[4] B.S. Manjunath, P. Salembier, T. Sikora (eds.), *Introduction to MPEG-7: Multimedia Content Description Interface*. Wiley & Sons, Apr. 2002.

[5] J. Kunze, T. Baker, "The Dublin Core Metadata Element Set", *RFC 5013*, Aug. 2007. Web: http://www.ietf.org/rfc/rfc5013.txt, last accessed: Jul. 2008.

[6] UPnP™, *ContentDirectory:2 Service Template Version 1.01*, May 2006. Web: http://www.upnp.org/specs/av/UPnP-av-ContentDirectory-v2-Service-20060531.pdf, last access: Jul. 2008.

[7] S. Lauf, I. Burnett, "A protected Digital Item Declaration Language for MPEG-21", *Proceedings of AXMEDIS2006*, Nov./Dec. 2005.

[8] K. Diepold, F. Pereira, Wo Chang, "MPEG-A: Multimedia Application Formats", *IEEE Multimedia*, vol. 12, no. 4, pp. 34-41, Oct.-Dec. 2005.

[9] S. Yang, S. Lee, Y. M. Ro, S.-K. Kim, "Semantic Photo Album Based on MPEG-4 Compatible Application Format", *Proceedings of the International Conference on Consumer Electronics (ICCE 2007)*, Las Vegas, USA, Jan. 2007.

[10] Microsoft Interactive Media Manager (IMM), Web: http://www.codeplex.com/IMM, last accessed: Jul. 2008.

[11] D.L. McGuinness, F. van Harmelen (eds.), "OWL Web Ontology Language Overview", *W3C Recommendation*, Feb. 2004. Web: http://www.w3.org/TR/owl-features/, last accessed: Jul. 2008.

[12] I.S. Burnett, S.J. Davis, G.M. Drury, "MPEG-21 Digital Item Declaration and Identification – Principles and Compression", *IEEE Transactions on Multimedia*, vol. 7, no. 3, pp. 400-407, Jun. 2005.

[13] D. C. Verma, "Service Level Agreements on IP Networks", *Proceedings of the IEEE*, vol. 92, no. 9, pp. 1382-1388, Sep. 2004.

[14] J.-C. Dufourd, O. Avaro, C. Concolato, "An MPEG Standard for Rich Media Services", *IEEE Multimedia*, vol. 12, no. 4, pp. 60-68, Oct.-Dec. 2005.

[15] D. Raggett, A. Le Hors, I. Jacobs (eds.), "HTML 4.01 Specification", *W3C Recommendation*, Dec. 1999. Web: http://www.w3.org/TR/html401/, last accessed: Jul. 2008.

[16] D. Bulterman, et.al. (eds.), "Synchronized Multimedia Integration Language (SMIL 2.1)", *W3C Recommendation*, Dec. 2005. Web: http://www.w3.org/TR/SMIL2/, last accessed: Jul. 2008.

[17] AXMEDIS project. Web: http://www.axmedis.org/, last accessed: Jul. 2008.

[18] ENTHRONE project. Web: http://www.ist-enthrone.org/, last accessed: Jul. 2008.

[19] ISO/IEC 21000-8:2008, "Information technology -- Multimedia framework (MPEG-21) -- Part 8: Reference software". Web: http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_IEC_21000-8_2008_Reference_Software/, last accessed: Jul. 2008.

[20] F. De Keukelaere, F, S. De Zutter, R. Van de Walle, "MPEG-21 Digital Item Processing", *IEEE Transactions on Multimedia*, vol. 7, no. 3, pp. 427-434, Jun. 2005.

[21] ISO/IEC 14496-20:2005, "Information Technology - LASeR or Lightweight Application Scene Representation".

[22] Systems Subgroup, "Requirements for Presentation of Structured Information", *ISO/IEC JTC 1/SC 29/WG 11/N9904*, Archamps, France, Apr. 2008. Web: http://www.chiariglione.org/MPEG/working_documents/explorations/psi/PSI-CfP.zip, last accessed: Jul. 2008.