Position Paper on Relevant MPEG Metadata Technologies for the W3C Video on the Web Workshop December 2007

Edited by Ian S Burnett, Stephen Davis, Julie Lofton

The purpose of this position paper is not to deliver novel technologies from the authors but rather to remind, or make participants aware of technologies which have been developed by the ISO/IEC SC29WG11 MPEG standards body in the area of video description and usage in web connected environments. As such, the listed authors have edited together content available within the MPEG organisation and do not claim this information to be originally authored.

We start with a proposition that there are several fundamental elements which are required for Video to be delivered in a rich form over the web:

1. A container technology which allows video(s), metadata and time descriptions to be held in a structured format
2. Mechanisms by which text/XML descriptors ('tags') can be attached to timepoints and other structural areas in video content
3. Mechanisms by which small sections of XML documents might be requested and transferred alongside streaming video
4. Technologies that express the DRM control applied to content and metadata

To this end, MPEG has developed four technologies which are summarised in the following sections: Digital Items, Fragment identification, Fragment request Units and IPMP Components. We would note that MPEG has also developed the complete MPEG-7 Multimedia Description Framework which provides a wide range of descriptive tools for video (and other media). However, the focus of this position paper is on the metadata technologies surrounding the media and providing for its delivery and association of appropriate metadata.

1 Digital Items

The Digital Item is defined in ISO/IEC 21000-2 (freely available at the ITTF site) which is the second part of the MPEG-21 set of standards. A Digital Item is a structured digital object with a standard representation, identification and metadata within the MPEG-21 framework. This entity is the fundamental unit of distribution and transaction within this framework but it can be generally used to give a structured container for multimedia content and metadata.

1.1 Digital Item Declaration Model

Declaring a Digital Item involves specifying the resources, metadata, and their interrelationships for a Digital Item.

Part 2 of ISO/IEC 21000 defines a set of abstract terms and concepts to form a useful model for declaring Digital Items. The goal of this model is to be as flexible and general as possible, while providing for the "hooks" that enable higher level functionality. This, in turn, allows the model to

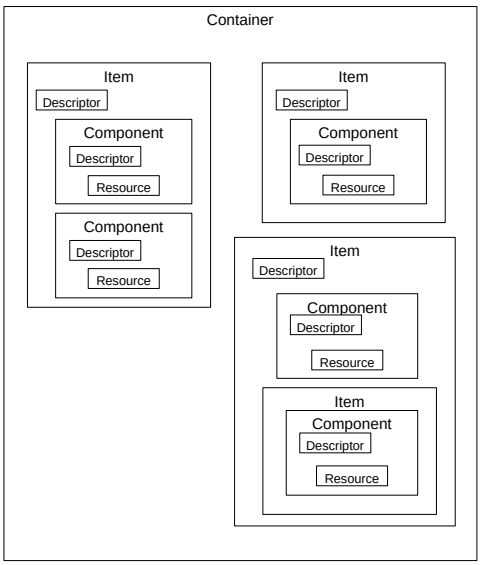

Figure 1 — Example Digital Item Declaration model

serve as a key foundation in the building of higher level models in other MPEG-21 elements. This model specifically does not define a language in and of itself. Instead, the model helps to provide a common set of abstract concepts and terms that can be used to define such a scheme, or to perform mappings.

1.2 Digital Item Declaration Language

Part 2 of ISO/IEC 21000 also specifies the XML based Digital Item Declaration Language (DIDL). This provides for the standard representation in XML of a Digital Item in the MPEG-21 Multimedia framework. The entities of the DID Model are represented in XML by the elements and attributes of DIDL. In addition a W3C XML Schema definition of DIDL is provided. The schema is also freely available at the ITTF web site. The schema can be used for schema validation of a DIDL document, however it is not sufficient for full DIDL validation since DIDL includes validation rules that are not able to be expressed in W3C XML Schema. The complete syntax and semantics, including the additional validation rules, are specified in the specification itself.

2 Fragment Identification

MPEG provides two mechanisms for identifying sections and timepoints in video. The first, from MPEG-7 uses a purely XML instantiation, while the second provides and extension of the fragment syntax in URIs to provide a richer fragment identification framework.

2.1 MPEG-7 – using an XML approach

The `TemporalSegmentLocator` which is included in the MPEG-7 Multimedia Descriptions Schemes (ISO/IEC 15938-5) standard describes the location of temporal media data such as video and audio. The localization within a target temporal media data can be described using either the media time information or the byte location information.

In the following example the location of a video segment is specified by the URI of a video file and the relative start time with respect to the beginning of the file and the duration of the segment.

```
<MyTemporalSegmentLocator>
        <MediaUri>http://www.mpeg7.org/demo.mpg</MediaUri>
        <MediaTime>
                <MediaRelTimePoint mediaTimeBase="../../MediaUri">PT3S</MediaRelTimePoint>
                <MediaDuration>PT10S</MediaDuration>
        </MediaTime>
</MyTemporalSegmentLocator>
```

MPEG-7 also provides for spatial locators and a detailed descriptive mechanism for the structure of video (amongst other media) content. This is illustrated in Figure 2

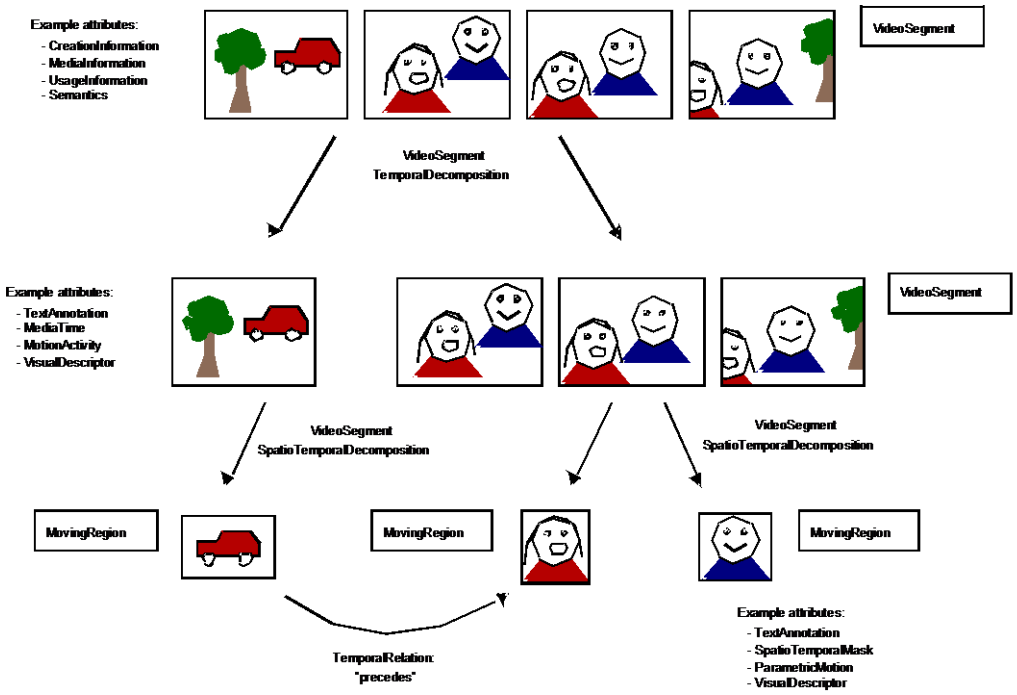

Figure 2: Example of the MPEG-7 Structure Description Tools functionality

The MPEG -7 mechanisms can be incorporated into Digital Items to describe relationships of descriptors and resources.

2.2 Fragment Identification – using a URI approach

Part 17 of ISO/IEC 21000 is entitled Fragment Identification for MPEG Resources and specifies a normative syntax for URI Fragment Identifiers to be used for addressing parts of any Resource whose Internet Media Type is one of: audio/mpeg, video/mpeg, video/mp4, audio/mp4, application/mp4, video/MPEG4-visual, application/mp21.

The tools specified in this part of ISO/IEC 21000 allow identification of a part of a Resource by providing a format for the reference to the part using a Fragment Identifier. MPEG URI Fragment Identifier schemes offer comprehensive and flexible mechanisms for addressing fragments of audiovisual content. Therefore, their use may potentially be extended to other audiovisual MIME types. The MPEG URI Fragment Identifiers are used after the '#' character of a URI reference. The syntax for URI Fragment Identifiers defined in this specification is based on the W3C XPointer Framework Recommendation and adds the ability to address: Temporal, spatial and spatio-temporal locations; Logical unit(s) of a resource according to a given Logical Model; A byte range of a resource; Items or Tracks of an ISO Base Media File.

ISO/IEC 21000-17 consists of three schemes:

1. The ffp() pointer scheme applies to files formats conforming to ISO/IEC 14496-12 and 14496-12/Amd.1 and allows the identification of an item or a track as defined in these specifications.
2. The offset() pointer scheme applies to any binary resource and identifies a byte range in a binary data stream.
3. The mp() pointer scheme provides two complementary mechanisms for identifying fragments in a multimedia resource:

    - Firstly, the mp() pointer scheme defines a set of so-called axes, which allow the identification of temporal, spatial or spatio-temporal fragments in a multimedia Resource, e.g. an audio, an image or a video, independent of the coding format.
    - Secondly, the mp() pointer scheme allows the addressing of fragments of a multimedia resource via a given hierarchical Logical Model of the resource, e.g. a track in an audio CD. Such Logical Models may be standardised or proprietary. This specification defines a syntax based on XPath's location paths for locating Logical Units in a hierarchical Logical Model.

For example, the following URI identifies a time point of 50 seconds from the beginning of a track in an MPEG-21 file. This Track is located in the file via its track_ID, which equals 101.

```
http://example.com/myFile.mp21#ffp(track_ID=101)*mp(/~time('npt','50'))
```

This example creates a moving region with a context restricted to the time interval between 10 and 30 seconds NPT:

```
mp(/~time('npt','10','30')/~moving-region(rect(0,0,5,5),pt(10,10,t(5)),pt(20,20)))
```

3 Fragment Request Units

Fragment Request Units (FRUs), which are part of the MPEG-B standard (ISO/IEC 23001-2 Fragment Request Units), can be used to request fragments of XML from a remote XML document. The FRUs are created in XML (valid to the FRU schema) from a selection of FRU commands. Briefly, basic FRU commands are as follows: Src, Query, and XMLPull. These commands allow a client to select a document, query a document, and issue XML Pull commands on a document. FRUs are capable of requesting any fragment of the XML document (based on fragment size and location), thus providing clients with random access. This allows a client to jump into any node in an XML document, or to simply, "navigate backwards" (such an operation may be entirely client side if previous XML fragments have been cached locally).

Related to FRUs, MPEG-7 Part 1 specifies a textual delivery method for delivering XML multimedia descriptors. This method is known as Textual Encoding format for MPEG-7 (TeM) and relies on the principle that an XML document can be divided into smaller XML fragments, which can be reassembled at the client. The local version of the XML document on the client side is manipulated through TeM specific commands sent by the server. TeM is capable of fragmenting an XML document and delivering these XML fragments to a client to recreate the original XML document structure. The fragments are encapsulated in XML using Fragment Update Unit (FUU) elements ad FRUs and FUUs can therefore work together, with the client sending FRU request using a query such as XPath and receiving the requested XML fragment via a FUU. The client is then able to reconstruct a partial XML instantiation in the document area of interest using only relevant information provided

by the FUU, instead of receiving the whole XML from the server, which would be prohibitive for e.g. mobile devices.

4 IPMP DIDL

The fourth part of MPEG-21, IPMP (Intellectual Property Management and Protection) Components (ISO/IEC 21000-4), provides an alternative normative Representation for parts of Digital Items that require protection through IPMP governance. This Representation is termed the IPMP Digital Item Declaration Language (IPMP DIDL), and defines governed XML elements corresponding to entities in the DID model. Each of these IPMP DIDL elements is intended to link a corresponding DIDL element (which may be encrypted) with information about the governance, so that the Digital Item or part of Digital Item thus represented is used in accordance with the Digital Item author's wishes.

For each entity in the DID model, an IPMP DIDL element is provided as a protected Representation of that entity, derived from the abstract DID model types as defined in the DID model schema in ISO/IEC 21000-2. The relationship between the schemas for IPMP DIDL, DIDL and the DID model is shown in Figure 3 below.

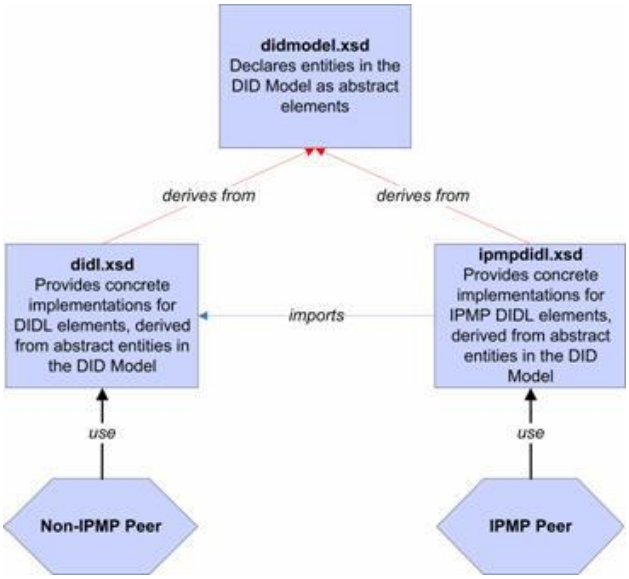

Figure 3: Schema relationship between DID model, DIDL and IPMP DIDL

5 References

[1] ISO/IEC 21000-2:2005 Digital Item Declaration (second edition)
[2] ISO/IEC 21000-17:2006 Fragment Identifiers for MPEG Resources
[3] ISO/IEC 21000-4:2005 IPMP Components
[4] ISO/IEC 15938-5:2003 Multimedia Description Schemes
[5] ISO/IEC 23001-2:2007 Fragment Request Units