M

MPEG-21 DIGITAL ITEM ADAPTATION

Christian Timmerer[a], Anthony Vetro[b], and Hermann Hellwagner[a]
[a]Department of Information Technology, Klagenfurt University, Austria
[b]Mitsubishi Electric Research Laboratories (MERL), USA

Synonym: *Overview of Part 7 of the MPEG-21 Multimedia Framework; Interoperable description formats facilitating the adaptation of multimedia content; ISO standard enabling device and coding format independence; ISO/IEC 21000-7:2007*

Definition: *MPEG-21 Digital Item Adaptation (DIA) enables interoperable access to (distributed) advanced multimedia content by shielding users from network and terminal installation, management, and implementation issues.*

Context and Objectives

Universal Multimedia Access (UMA) [1] refers to the ability to seamlessly access multimedia content from anywhere, anytime, and with any device. Due to the heterogeneity of terminals and networks and the existence of various coding formats, the adaptation of the multimedia content may be required in order to suit the needs and requirements of the consuming user and his/her environment. For enabling interoperability among different vendors, Part 7 of the MPEG-21 Multimedia Framework [2], referred to as Digital Item Adaptation (DIA) [3], specifies *description formats* (also known as *tools*) to assist with the adaptation of Digital Items. Digital Items are defined as structured digital objects, including standard representation, identification and metadata, and are the fundamental units of distribution and transaction within the MPEG-21 framework. The concept of MPEG-21 DIA is illustrated in Figure 1. As shown in this figure, only tools used to guide the adaptation engine are specified by the standard. The adaptation engines themselves are left open to various implementations. However, description formats and format-independent mechanisms that provide support for the adaptation in terms of media resource adaptation, description adaptation, and/or Quality of Service (QoS) management are within the scope of the standardization, and are collectively referred to in Figure 1 as DIA tools.

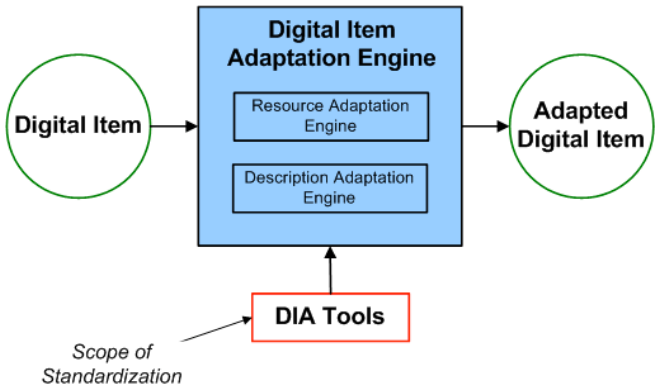

Figure 1. Concept of MPEG-21 Digital Item Adaptation [3].

The aim of this article is to provide an overview of the various DIA tools and their specifications. Key aspects of each tool will be reviewed and some details regarding their application and use will be provided.

Usage Environment Descriptions

The usage environment description includes the description of *terminal capabilities* and *network characteristics*, as well as *user characteristics* and *characteristics of the natural environment*. Such descriptions provide a fundamental input to any adaptation engine and enable device independence. The parameters that have been specified by the standard are reviewed below. Some example applications are also discussed.

Terminal Capabilities

The notion of a terminal in DIA is rather generic and represents all devices regardless of their location in the delivery chain. So, in addition to the typical consumer electronic devices such as mobile phones, televisions, audio players and computers, a terminal could also be a server, proxy or any intermediate network node. With this definition, terminal capabilities must be described in terms of both receiving and transmitting capabilities. Such a description is used to satisfy consumption and processing constraints of a particular terminal. The major categories are briefly outlined below:

- **Codec Capabilities**: specifies the coding formats a particular terminal is capable of encoding or decoding, e.g., an MPEG profile/level. Furthermore, codec parameters can be explicitly described, e.g., the maximum bit-rate of a decoder. With such information, an adaptation engine could transcode a video from MPEG-2 to MPEG-4 in case the source video is on a DVD and the terminal can only handle MPEG-4.
- **Input-Output Characteristics**: Display capabilities (e.g., resolution, color capabilities, rendering format), audio output capabilities (e.g., frequency range, power output, signal-to-noise ratio, the number of output channels) and user interaction inputs (e.g., keypad, touch screen, microphone) are the key items considered in this category.
- **Device Properties**: There are a variety of properties specified under this umbrella, but due to space limitations only power and storage characteristics are covered. The power characteristics tool is intended to provide information pertaining to the consumption, battery capacity remaining, and battery time

remaining. With such attributes, a sending device may adapt its transmission strategy in an effort to maximize the battery lifetime. Storage characteristics are defined by the input and output transfer rates, the size of the storage, and an indication of whether the device can be written to or not. Such attributes may influence the way that a Digital Item is consumed, e.g., whether it needs to be streamed or could be stored locally.

Network Characteristics

Two main categories are considered in the description of network characteristics: capabilities and conditions. The capabilities define static attributes of a network, while the conditions describe dynamic behavior. These descriptions primarily enable multimedia adaptation for improved transmission efficiency [4].

- **Network Capabilities**: These descriptions include attributes that describe the maximum capacity of a network and the minimum guaranteed bandwidth that a network can provide. Also specified are attributes that indicate if the network can provide in-sequence packet delivery and how the network deals with erroneous packets, i.e., does it forward, correct or discard them.
- **Network Conditions**: These tools specify attributes that describe the available bandwidth, error and delay. The error is specified in terms of packet loss rate and bit error rate. Several types of delay are considered, including one-way and two-way packet delay, as well as delay variation. Available bandwidth includes attributes that describe the minimum, maximum, and average available bandwidth of a network. Since these conditions are dynamic, time stamp information is also needed. Consequently, the start time and duration of all measurements pertaining to network conditions are also specified. However, the end points of these measurements are left open to the application performing the measurements, e.g., two routers, a network domain, end-to-end, etc.

User Characteristics

The descriptions concerning the characteristics of a user as specified by DIA may be used for a number of purposes, including the adaptive selection or personalization of content.

- **User Info**: General information about a user is specified in DIA by importing the Agent DS specified by MPEG-7 [5] providing information such as name and contact information, where a user can be a person, a group of persons, or an organization.
- **Usage Preference and History**: As with User Info, corresponding tools specified by MPEG-7 define the usage preference (e.g., in terms of genre) and usage history tools (e.g., recording a video program, playing back a piece of music) in DIA.
- **Presentation Preferences**: This class of tools defines a set of preferences related to the means by which Digital Items and their associated resources are presented or rendered for the user. That are, preferences related to audio-visual rendering, to the format or modality a user prefers to receive, to the priority of the presentation, as well as the preferences that direct the focus of a user's attention with respect to audio-visual and textual media.
- **Accessibility Characteristics**: These tools provide descriptions that would enable one to adapt content according to certain auditory or visual impairments of the user. For audio, an audiogram is specified for the left and right ears, which

specifies the hearing thresholds for a person at various frequencies in the respective ears. For visual related impairments, color vision deficiencies are specified, i.e., the type and degree of the deficiency. For example, given that a user has a severe green-deficiency, an image or chart containing green colors or shades may be adapted accordingly so that the user can distinguish certain markings. Such descriptions would also be very useful to simply determine the modality of a media resource to be consumed, but may also be used for more sophisticated adaptations, such as the adaptation of color in an image [6].

- **Location Characteristics**: This set of tools covers mobility characteristics and destination. The first of these tools aims to provide a concise description of the movement of a user over time which can be used to classify users, e.g., as pedestrians, highway vehicles, etc. Destination is a tool to indicate, as the name implies, the destination of a user, i.e., an intended location at a future time. Both tools could be used for adaptive location-aware services.

Natural Environment Characteristics

The natural environment pertains to the physical environmental conditions around a user such as lighting condition or noise level, or a circumstance such as the time and location that Digital Items are consumed or processed.

- **Location** and **Time:** These tools refer to the location and time of usage of content, respectively. Both are specified by MPEG-7 description tools, namely the Place DS for Location and the Time DS for Time [5]. These tools are referenced by both the mobility characteristics and destination tools and may also be used to support adaptive location-based services.
- **Audio-Visual Environment**: This set of tools describes audio-visual attributes that can be measured from the natural environment and affect the way content is delivered and/or consumed by a user in this environment. With respect to the visual environment, *illumination characteristics* that may affect the perceived display of visual information are specified. It has been observed that the overall illumination around a display device affects the perceived color of images on the display device and contributes to the distortion or variation of perceived color [7]. By compensating the estimated distortion, actual distortion caused by the overall illumination can be decreased or removed. For *audio*, the description of the noise levels and a noise frequency spectrum are specified. The noise level is represented as a sound pressure level in decibels, while the noise spectrum are the noise levels for 33 frequency bands of 1/3 octave covering the range of human audible bandwidth. An adaptation engine would enhance the perceived quality of the adapted audio signal by modifying the frequency attenuation of the original audio signal according to the noise characteristics. Interested readers may refer to [8] for further details.

Bitstream Syntax Descriptions

Due to the large diversity of competing or complementary scalable coding formats, a device needs to maintain the corresponding number of software modules in order to facilitate the manipulation of bitstreams based on all these formats. Thus, to solve this limitation and leverage the use of scalable multimedia formats, DIA defines a generic framework based on XML for bitstream adaptation [9] which offers coding format independence. In particular, DIA uses XML to describe the high-level structure of a

bitstream, i.e., how it is organized in packets or layers of data. The resulting XML document is called Bitstream Syntax Description (BSD). In most cases, the BSD does not describe the bitstream on a bit-by-bit basis and may describe the bitstream at different syntactic layers, e.g., finer or coarser levels of detail such as frames or scenes for a video resource, depending on the application. The actual bitstream adaptation is performed by transforming the BSD corresponding to an input bitstream, and then generating the adapted version of the bitstream from the transformed BSD without knowing the details of the coding format. This process is illustrated in Figure 2.

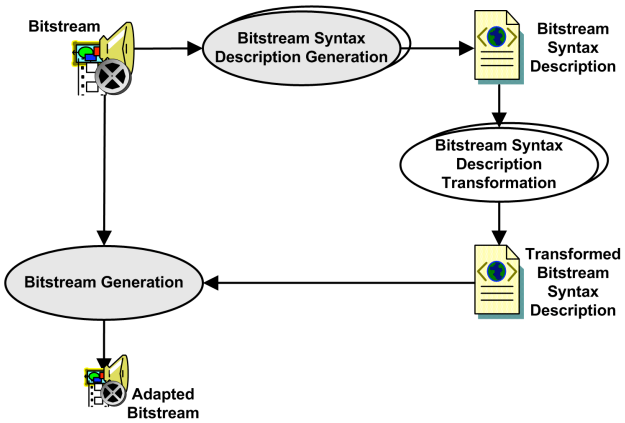

Figure 2. High-Level Architecture of Bitstream Syntax Description-based Adaptation [2].

The behavior of both the *BSD generation* and the *bitstream generation* processes are normatively specified by the DIA standard whereas the *BSD transformation* process is not normative and, thus, open for industry competition.

Terminal and Network QoS

The goal of terminal and network QoS is to select optimal parameter settings for media resource adaptation engines that satisfy constraints imposed by terminals and/or networks while maximizing QoS. To facilitate this, the *AdaptationQoS tool* specifies the relationship between constraints, feasible adaptation operations satisfying these constraints, and associated utilities (qualities) [10].

Three types of modules are specified by the standard to define this relationship, which allows an optimal choice for representing data according to the actual application requirements:

- **Look-Up Table (LUT)**: MxN matrix suitable for dense and discrete data.
- **Utility Function (UF)**: N-dimensional vector suitable for sparse and discrete data.
- **Stack Function (SF)**: Expressed using the Reverse Polish Notation (RPN) suitable for functional and continuous data.

Although each of the modules operates differently, a generic interface to the modules is provided through IOPins (Input/Output Pins). Each IOPin is a uniquely identifiable variable that is globally declared and referenced from within a module. An IOPin can be interpreted as input, output, or both, which also allows the interconnection of different

modules. The values of the IOPins can be either directly declared as continuous or discrete within the IOPin or as an external parameter specified by its semantics.

Furthermore, IOPin values can be further constrained by using the *Universal Constraints Description (UCD)* tool, also specified within DIA. The UCD allows further constraining the usage and usage environment of a Digital Item by means of limitation and optimization constraints. Both types of constraints are formulated using the stack function syntax as mentioned above. For instance, the usage of a Digital Item containing an image resource can be constrained in such a way that it should not be adapted to smaller than 20% of the receiving terminal's display resolution – this may be a restriction set by the content provider. However, the dimension of the image should be maximized if adaptation is required. On the other hand, the usage environment may define a similar constraint where the image should be smaller than 75% of the rendering terminal's display resolution due to specific application requirements, e.g., the rendering application should not run in full screen mode to conserve resources.

Metadata Adaptability

With regard to metadata adaptation, there are several important aspects covered by DIA. First, as the content is adapted, the associated metadata must also change accordingly. Second, if the metadata is transmitted and consumed, it may need to be scaled in order to meet terminal and network constraints. Third, given a very rich and detailed description of a piece of content, filtering is often necessary to obtain only the necessary or interesting parts of the metadata for a user. Finally, for the case when multiple sources of metadata for the same resource exist, an efficient means of integrating these different sources of metadata into a single description could be very useful. DIA specifies tools to assist with all of the above issues.

Session Mobility

In DIA, session mobility refers to the transfer of *configuration-state information* that pertains to the consumption of a Digital Item on one device to a second device. This enables the Digital Item to be consumed on the second device in an adapted way. The configuration-state of a Digital Item is referred to as an instantiation of Choices/ Selections that are part of the Digital Item Declaration (DID) [2]. During this transfer, *application-state information*, which pertains to information specific to the application currently rendering the Digital Item, may also be transferred.

To make the session mobility concepts more concrete, consider the following example of an electronic music album. The different songs of this album are expressed as individual Selections within a particular Choice of the DID. There may also be other Choices within the DID that configure the platform, acceptable media formats, and so on, for a given user. In this example, assume that a first user is listening to the second song on a certain device, e.g., an MP3 player, and this user would like to share this particular song with a another user. Assuming that they have the right to do this, the configuration-state of the DID on the first device, i.e., the Selection that pertains to the second song, would be captured and transferred to the second device, e.g., a mobile/smart phone. Since the platform and acceptable media formats may be different on the second device, potential Choices in the original DID concerning those Selections would still need to be made

there. Supplemental application-state information may also be transferred such as timing information related to songs or layout information if images and video are involved.

As shown by the above example, the user's current state of interaction with a Digital Item is completely described by both the configuration-state and application-state information.

DIA Configuration

The DIA Configuration tools are used to help guide the adaptation process considering the intentions of a DID author. There are essentially two types of tools that have been specified. The first is used to suggest the means by which Choice/Selections in a DID should be processed, e.g., displayed to users or configured automatically according to DIA descriptions, while the second is used to suggest particular DIA descriptions that should be used for resource adaptation.

As an example, consider a DID with two Choices: one pertaining to the video quality and another pertaining to the media format. The Descriptor of each Choice in the DID may include a DIA tag that indicates whether the Choice should be manually selected by the user or automatically configured according to relevant metadata. Assume that the user selects the video quality and the media format is configured automatically. Given that the media format should be configured automatically, the DIA Configuration tool may further describe the particular DIA description that would be required to perform the configuration; in this case, the supported media formats on the terminal.

DIA Conversions and Permissions

DIA also defines rights expressions to govern adaptations in an interoperable way [11]. More specifically, DIA provides a means to describe the adaptation capabilities that a terminal is capable of performing. The standard additionally specifies the relationship of arbitrary adaptation operations (e.g., transcoding, image cropping) and specific adaptation parameters (e.g., target transcoding format, width/height of cropped region). The AdaptationQoS tool is extended with the so-called *CrossConversionQoS* tool that takes into account different modality conversions such as video-to-image or image-to-speech conversions.

Finally, and most importantly, the *permitted DIA changes* and *change conditions* allows one to express the rights associated with an adaptation as an extension of the MPEG-21 Rights Expression Language (REL) and MPEG-21 Rights Data Dictionary (RDD) [12]. It should be noted that the REL and RDD already provide tools to permit playing, modifying, and adapting; however, only with coarse control. DIA essentially enables finer-grained control over the changes that can occur when playing, modifying, or adapting Digital Items and their component resources. In this way, a flexible copy protection system that allows only certain changes to the content could be designed. It should be emphasized that although changes to the content are governed, the standard does not specify the actual implementation of the adaptation engine.

Dynamic and Distributed Adaptation

The DIA standard also adds support for dynamic and distributed adaptation (DDA) [13]:

- **Dynamic adaptation** refers to the adaptation of Digital Items according to dynamically changing usage environments, e.g., the available bandwidth may drop during a streaming session and the Digital Item is consequently adapted to this new usage environment.
- **Distributed adaptation** refers to multiple adaptation steps successively performed on different MPEG-21 peers, e.g., a same media resource may be successively adapted on a server, network node, and/or terminal.

In order to comply with dynamic and distributed adaptation scenarios as indicated above, DIA introduces the concept of process units where a *process unit* is a well-formed XML fragment that can be consumed as such by the MPEG-21 peer, and to which a time information may be attached, indicating the point in time when it becomes available for consumption. As such it is more a processing-oriented concept rather than a delivery-oriented concept. Note that it does not depend on any encoding method used for delivering it.

The actual tools defined within MPEG-21 that support DDA are threefold and briefly outlined in the following:

- **Media resource streaming instructions**: tool enabling the streaming of arbitrary bitstreams described by a Bitstream Syntax Description.
- **XML streaming instructions**: tool for describing process units and assigning time information to them for streamed processing thereof.
- **Properties style sheet**: tool for dynamically applying media resource and XML streaming instructions to an XML document.

Concluding Remarks

The MPEG-21 Digital Item Adaptation standard offers a comprehensive and rich set of interoperable description formats (i.e., tools) that assist in the adaptation of Digital Items in various ways. In particular, Part 7 of the MPEG-21 Multimedia Framework is an important contribution towards fulfilling the challenges which are collectively referred to as Universal Multimedia Access (UMA).

See: Universal Multimedia Access, Multimedia Content Adaptation in MPEG-21, Multimedia Content Modeling and Personalization, MPEG-21 Multimedia Framework, Multimedia Adaptation Decision-Taking, Generic Multimedia Content Adaptation

References

1. A. Vetro, C. Christopoulos and T. Ebrahami, Eds., "Special Issue on Universal Multimedia Access", *IEEE Signal Processing Magazine*, 20(2), March 2003.
2. I. Burnett, F. Pereira, R. Van de Walle, R. Koenen, Eds., *The MPEG-21 Book*, John Wiley & Sons, 2006.
3. A. Vetro, C. Timmerer, "Digital Item Adaptation: Overview of Standardization and Research Activities", *IEEE Transactions on Multimedia*, 7(3):418–426, June 2005.
4. M. van der Schaar and Y. Andreopoulos, "Rate-Distortion-Complexity Modeling for Network and Receiver Aware Adaptation", *IEEE Transactions on Multimedia*, 7(3):471-479, June 2005.

5. B.S. Manjunath, P. Salembier, T. Sikora, Eds., *Introduction to MPEG 7: Multimedia Content Description Language*, John Wiley & Sons, 2002.
6. J. Song, S. Yang, C. Kim, J. Nam, J.W. Hong and Y.M. Ro, "Digital Item Adaptation for Color Vision Variations", *Proceedings of the SPIE Conference on Human Vision Electronic Imaging VIII*, Santa Clara, CA, January 2003.
7. Y. Huh and D.S. Park, "Illumination environment description for adaptation of visual contents", *ISO/IEC JTC1/SC29/WG11/M8341*, Fairfax, USA, May 2001.
8. B. Feiten, I. Wolf, E. Oh, J. Seo and H.-K. Kim, "Audio Adaptation According to Usage Environment and Perceptual Quality Metrics", *IEEE Transactions on Multimedia*, 7(3):446-453, June 2005.
9. S. Devillers, C. Timmerer, J. Heuer, and H. Hellwagner, "Bitstream Syntax Description-Based Adaptation in Streaming and Constrained Environments", *IEEE Transactions on Multimedia*, 7(3):463-470, June 2005.
10. D. Mukherjee, E. Delfosse, J.G. Kim and Y. Wang, "Optimal Adaptation Decision-Taking for Terminal and Network Quality of Service", *IEEE Transactions on Multimedia*, 7(3):454-462, June 2005.
11. C. Timmerer, T. DeMartini, and H. Hellwagner, "The MPEG-21 Multimedia Framework: Conversions and Permissions", *Proceedings of D·A·CH Security 2006*, Düsseldorf, Germany, March 2006.
12. X. Wang, T. DeMartini, B. Wragg, M. Paramasivam, and C. Barlas, "The MPEG-21 Rights Expression Language and Rights Data Dictionary", *IEEE Transactions on Multimedia*, 7(3):408-417, June 2005.
13. M. Ransburg, C. Timmerer, and H. Hellwagner, "Dynamic and Distributed Multimedia Content Adaptation based on the MPEG-21 Multimedia Framework", in: Mathias Lux *et al.*, Eds., *Multimedia Semantics: The Role of Metadata*, Springer, 2008.