Feature Article

Everything You Wanted to Know About MPEG-7: Part 2

Frank Nack
GMD-IPSI

Adam T. Lindsay
Starlab

Audio-visual information must allow some degree of interpretation, which can be passed onto or accessed by a device or a computer code. MPEG-7 aims to create a standard for describing these operational requirements. This article describes MPEG-7's concepts, terminology, and requirements. We also describe the influence of other approaches to multimedia content description on the development of the MPEG-7 standard and conclude with a short estimate of the time schedule for the standard.

Part 1 of this article provided an overview of the development, functionality, and applicability of MPEG-7. In Part 2 we discuss the description of MPEG-7's concepts, terminology, and requirements. We then compare MPEG-7 with other approaches to multimedia content description.

The increasing availability of audio-visual data in digital form impelled the work within the Moving Pictures Expert Group's MPEG-7. This data may include still pictures, graphics, 3D models, audio, speech, video, and information about how these elements combine in a multimedia presentation such as scenarios or personal characteristics.

The three main goals for MPEG-7 regarding this data include

■ description of multimedia content,

■ flexibility in data management, and

■ globalization and interoperability of data resources.

Keeping these goals in mind, we'll describe a few relevant terms. MPEG-7 will not define a monolithic system for content description but rather a set of methods and tools for the different aspects of multimedia description. Thus, MPEG-7 aims to standardize

■ a set of Description Schemes and Descriptors to describe data,

■ a language to specify Description Schemes, such as the Description Definition Language (DDL), and

■ a scheme for coding the Description.

Terminology and requirements

The following sections should clarify what the different terms mean within MPEG-7. We took the definitions from the *MPEG-7 Requirements Document*.[1] However, we added some explanation to provide a better understanding.

Data

Definition: Data is audio-visual information that will be described using MPEG-7 regardless of storage, coding, display, transmission, medium, or technology.

This broad definition encompasses graphics, still images, video, film, music, speech, sounds, text, and any other relevant audio-visual medium. Examples for MPEG-7 data include an MPEG-4 stream; a videotape; a CD containing music, sound, or speech; a picture printed on paper; or an interactive multimedia installation on the Web.

Feature

Definition: A Feature is a distinctive characteristic of the data that signifies something to somebody.

A Feature isn't a traditional signal processing term, but something "in the mind." Thus, you can't compare Features without a meaningful Feature representation (Descriptor) and its instantiation (Descriptor Value) for a given data set. Some examples of Features include the color of an image, fundamental frequency of a speech segment, rhythm of an audio segment, camera motion in a video, genre of a piece of music, title of a movie, and actors in a movie.

Descriptor and Descriptor Value

Definition: A Descriptor is a representation of a Feature. A Descriptor defines the syntax and semantics of the Feature representation.

For a Descriptor to function in MPEG-7, it must precisely define the semantics of the Feature, the associated data type, legal values, and an interpretation of the Descriptor Values (see below). An example might be `Color: string`. The data type may be composite, meaning that it can be formed by concatenating multiple instances of a data type. An example of this would be `RGB-Color: [integer, integer, integer]`.

It's possible to have several Descriptors representing a single Feature—that is, address different relevant requirements. Examples of multiple Descriptors for a single Feature include enumerated lists, color moments, and histograms that represent color.

The following example of a Descriptor's syntax is based on a Standard Generalized Markup Language (SGML)-like style. Please be aware that this is just our suggestion. The current state of the MPEG-7 standard hasn't defined a format yet.

Thus, the syntax of a Descriptor might look like this:

```
< name, typekind, spec> value </name>
```

Table 1 summarizes what the items represent.

A Descriptor Value is an instantiation of a Descriptor for a given data set (or subset thereof). Descriptor Values are combined via a Description Scheme to form a Description.

Description Scheme

Definition: A Description Scheme specifies the structure and semantics of the relationships between its components, which may be both Descriptors and Description Schemes.

Figure 1 presents an abstract representation of possible relations between Descriptors and Description Schemes. The arrows from DDL to Description Scheme signify that the Description Schemes are generated using the DDL.

The distinction between a Description Scheme and a Descriptor is that a Descriptor is concerned with the representation of a Feature whereas the Description Scheme deals with the structure of a Description. Thus, these are two very different concepts.

A simple Description Scheme for describing technical aspects of a shot might look like this, where those elements written in bold represent other Description Schemes:

Table 1. Representation of items in a Descriptor.

Item	Description
Name	Descriptor ID
Typekind	Data type of Descriptor Value: free text (plus language identifier), structured text (plus structure identification), integer, real, date, time/time index, version, enumeration (such as RGB values, vector), relation (Descriptor-Descriptor, Descriptor-Description Scheme, Description-Description), complex (a bitmap, histogram)*, function*, and possibly trigger (executable code)*
Spec	Contains a specification for the data type. For example, "free text" might specify the language the text is written in (Dutch, English, Japanese, and so on).
Value	An instantiation of the Descriptor is assigned to the Feature as pertaining to the data. Descriptor Values are combined via the mechanism of a Description Scheme to form a Description.

*Value could contain the actual data or a link

Figure 1. An abstract representation of possible relations between Descriptors and Description Schemes.
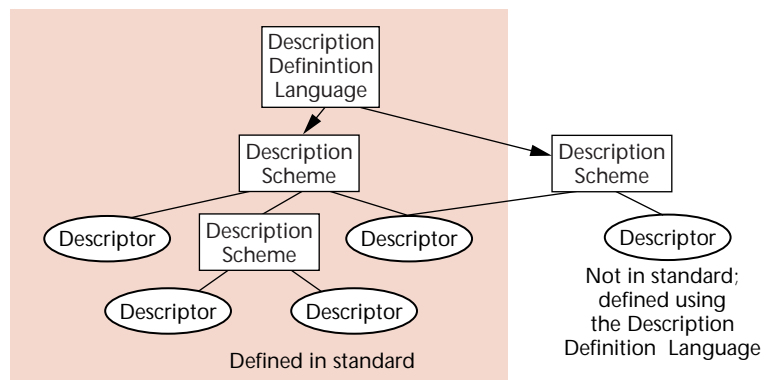
```
Shot_Technical_Aspects
  Lens     providing information about type
           (wide-angle), movement (zooms),
           state (deep focus), masking, and so on.
  Camera   providing information about distance
           (close-up), angle (overhead), move-
           ment (pan_left), position (viewpoint
           of the frame), and so on.
  Speed
  Color
  Granularity
  Contrast
```

Description

Definition: A Description consists of a Description Scheme (structure) and the set of Descriptor Values (instantiations) that describe the data.

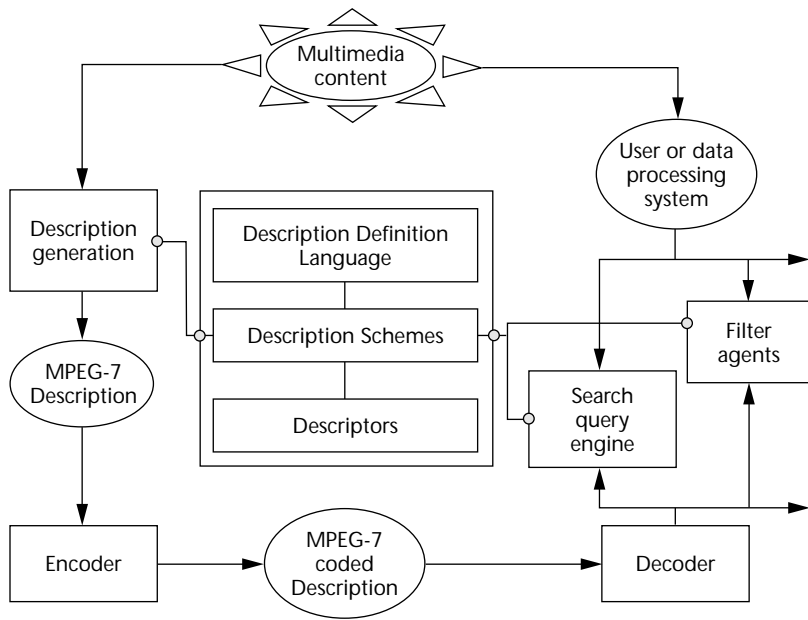A Description contains or refers to a fully or

Figure 2. An abstract representation of possible applications using MPEG-7.

partially instantiated Description Scheme. In other words, a Description is an instance of a description schema.

Coded description

Definition: A coded description is a Description that has been encoded to fulfill relevant requirements such as compression efficiency, error resilience, random access, and so on.

Description Definition Language

Definition: The DDL allows the creation of new Description Schemes and, possibly, Descriptors. It also allows the extension and modification of existing Description Schemes.

We discuss further aspects of the DDL development in the section "Influence of other approaches."

To provide a better understanding of the terminology described above, please see Figure 1. Note that the presence of a box or ellipse in one of these drawings doesn't imply that the corresponding element must appear in all MPEG-7 applications.

Figure 2 explains how MPEG-7 would work in practice. The left side portrays how data is annotated, whereas the right side demonstrates how described data can be retrieved. Square boxes describe tools doing things, such as encoding or decoding. Circular boxes describe static elements, such as a Description. For example, the "description generation" box is a description generation engine that produces an "MPEG-7 description" as output, shown here as a rectangular box. Note, there can be other streams from content to user, though they're not depicted here.

Sometimes there's no need for a binary efficient representation of a Description, which means the textual description will be used directly. In some cases the encoder and decoder are dispensable.

Requirements for major MPEG-7 components

With reference to the above terminology we'll now take a second look at the technical work of MPEG-7.

Based on the analysis of a wide range of potential applications[2] that could use MPEG-7 Descriptions, the MPEG-7 working group has compiled a list of requirements for Descriptors, Description Schemes, and the DDL, described in detail in the *MPEG-7 Requirement Document*.[1] Note that while the MPEG-7 standard should satisfy all requirements, not all requirements must be satisfied by each individual Descriptor or Description Scheme. The following section provides a summary of the most important of these requirements to give you an idea of what MPEG-7 hopes to achieve.

We understand that the DDL's design forms a core part of the work within MPEG-7. Therefore, we assume that the DDL will provide a solid descriptive (for example, SGML-based) underpinning for users to create their own Description Schemes and Descriptors, which provides great flexibility and ensures that the standard endures. To achieve this, MPEG-7 has identified a number of requirements the DDL should cover, of which the most important are

∎ *Compositional capabilities*. The DDL shall supply the ability to compose new Description Schemes and Descriptors, where a Description Scheme might be composed from multiple Description Schemes. A newly created Description Scheme must allow the creation of MPEG-7 compliant Descriptions.

∎ *Transformational capabilities*. The DDL shall allow the reuse, extension, and inheritance of existing Descriptors and Description Schemes.

∎ *Unique identification*. The DDL shall provide

mechanisms to uniquely identify Description Schemes and Descriptors so that they can be referred to unambiguously.

■ *Data types.* The DDL shall provide a set of primitive data types, such as text, integer, real, date, time/time index, version, and so on to succinctly describe composite data types that might arise from processing digital signals (such as histograms, graphs, RGB values). Also, the DDL must provide a mechanism to relate Descriptors to data of multiple media types of inherent structure, particularly audio, video, audio-visual presentations, the interface to textual descriptions, and any combinations of these.

■ *Relationships within a Description Scheme and between Description Schemes.* The DDL provides the capability to express relationships between Description Schemes and among elements of a Description Scheme. The DDL expresses the semantics of these relations, such as spatial, temporal, structural, and conceptual relations.

■ *Relationship between Description and data.* The DDL shall supply a rich model for links and/or references between one or several Descriptions and the described data.

Even though the set of requirements expresses a certain agreement about the DDL's scope, a number of unresolved problems still exist, where the working group has to conduct further studies. For example, some discussion concerns the problem of the excecutional functionality of the DDL, where composition operations on Description Schemes, partial extraction, and subsequent extension are key issues. For detailed information about open issues in the DDL development, see the *DDL Development Document, Version 1.*[3]

The requirements for Descriptors and Description Schemes establish general, functional, and coding requirements common for audio and visual media as well as those only relevant for each individual medium. These requirements are derived from analyzing a wide range of potential applications that could use MPEG-7 Descriptions. MPEG-7 doesn't target any one application in particular. Rather, the elements that MPEG-7 standardizes shall support as broad a range of applications as possible.

The general requirements for Descriptors and Description Schemes deal with types of Features (such as objective, subjective, production, and so on), Feature hierarchies, priorities, scalability, cross modality, and abstraction levels and their hierarchization. Functional requirements cover aspects such as content- and similarity-based retrieval, streaming, referencing (analog) data, browsing, copyright information, and so forth. Coding requirements handle problems such as extraction and efficient representation.

The requirements for the individual medium types—for example, audio and video—cover aspects such as

■ *Feature types* such as color, visual objects, texture, shape, motion, and so on for video, and frequency contour, audio objects, timbre, harmony, amplitude envelope, and so on for audio.

■ *Data formats* for video such as MPEG-1, MPEG-2, or MPEG-4; analog video and film; Joint Photographic Experts Group (JPEG); Virtual Reality Modeling Language (VRML); and so on; and digital audio formats for audio such as MPEG-1 Audio, Compact Disc, vinyl records, magnetic tape media, MPEG-4 Structured Audio Orchestra Language (SAOL), and so on.

■ *Data classes* such as natural video, still pictures, graphics, animation (2D), and so on for video, and soundtrack (natural audio scene), music, atomic sound effects (such as clapping), speech, and so on for audio.

Besides the requirements for Descriptors and Description Schemes, MPEG-7 authors have already identified a number of system requirements. However, these requirements represent just the beginning of clarifying the scope of MPEG-7 systems. The general question that must be answered is whether MPEG-7 systems should encompass other areas outside of the normative parts of MPEG-7 (DDL, Description Schemes, Descriptors) such as the presentation of MPEG-7 Descriptions, the presentation of the media elementary streams, transportation protocols, and so on.

For more details regarding the terminology and the requirements for the DDL, Descriptors, and Description Schemes, we refer you to the *MPEG-7 Requirements Document.*[1]

Other approaches exist to describe multimedia content. The MPEG-7 working group recognizes the fact that other organizations such as the W3 Consortium address similar problems while MPEG-7 is being developed. Next we'll briefly

address the most relevant approaches and their differences from and their influences on MPEG-7.

The influence of other approaches

Like previous MPEGs, MPEG-7 explicitly considers other standards and functionalities for content description, such as SGML, Extensible Markup Language (XML), HyTime, and the work of the Multimedia and Hypermedia Experts Group (MHEG). The following section describes the influence of these descriptive languages, particularly on the development of the DDL.

SGML

Based on his work on the Generalized Markup Language (GML),[4] Goldfarb developed—under the support of the International Organization for Standardization (ISO)—the formal definition of SGML. The ISO adopted it as standard 8879[5] in October 1986.

SGML formalizes an open, nonproprietary language to describe a document's structure, thus providing a way to store, index, and search for archival information, and to easily construct subsets of the information. Furthermore, SGML tackles the problem of how to transfer documents from computer to computer by providing techniques that let computers with different character sets and encoding schemes communicate in an internationally agreed way.[6]

SGML documents consist of a number of interrelated document components, called elements (such as title, author, abstract, and so on). Each element found in the document is given a generic identifier—a markup—such as `<author>`, which presents the beginning, and `</author>`, which presents the end of the element "author." The Document Type Definition (DTD), with which each SGML document starts, defines the document class structure in terms of the elements it contains. To generate large documents efficiently, they can be built up from a series of subdocuments, each existing as a document in its own right. SGML represents tree structures in a document hierarchy that can be as strict (or loose) as necessary. However, SGML's document structure provided by the DTD requires that authors and designers work within this structure. Inflexibility may be avoided by implementing a comprehensive document analysis phase that identifies all reasonable structures. SGML's strength lies in identifying and representing logical elements for a document, but it shows weaknesses in the flexible extension and maintenance of a description. Since SGML was developed for describing the structure and content of text documents, it lacks sufficient support for describing multimedia data (such as audio-visual data types), primitive data types, and composite data types that may arise from processing digital signals. Furthermore, SGML doesn't define the metadata describing access mechanisms and the temporal- and spatial-oriented links to data.

Nevertheless, SGML provides a good example for developing MPEG-7's DDL in respect to structuring and describing content. In particular, the experiences of developing DTDs are most useful in respect of standardization. The ability to create classes is of interest, as some industries have defined a standard DTD to let organizations easily exchange information. For example, the aeronautical industry uses the ATA2100 standard mandatorily. However, other organizations use an adapted version of the ATA2100 internally and convert to standard code when exchanging data.

XML

Developed by the World Wide Web Consortium (W3C), XML gives developers the power to deliver structured data from a variety of applications to the desktop for local computation and presentation. XML is also a format for server-to-server transfer of structured data. XML itself is a simplified version of SGML and was designed to maintain SGML's most useful parts.

While SGML requires that structured documents reference a DTD to be valid, XML allows for "well-formed" data and can be delivered without a DTD. The addition of well-formed XML remains one of the strong points of XML. It's one of the fundamental differences between XML and SGML. Thus, XML was designed so that SGML can be delivered, as XML, over the Web.

The concept of name spaces is XML's second important strength. Name spaces let developers uniquely qualify the element names and relationships. Name spaces also make these names recognizable to avoid name collisions on elements that have the same name but are defined in different vocabularies. They allow mixing tags from multiple name spaces, which proves essential if data comes from multiple sources.

For example, a bookstore may define the `<TITLE>` tag to mean the title of a book, contained only within the `<BOOK>` element. A directory of people, however, might define `<TITLE>` to indicate a person's position, for instance: `<TITLE>President</TITLE>`. Name spaces help define this distinction.

Due to XML's structural and declarative strengths, the MPEG-7 Requirements Group identified XML during the evaluation of submissions for the Call for Proposals (see the section "MPEG-7 schedule" below) as the basis for developing the DDL. However, the group also recognized that XML was designed with the requirements of the Web in mind, which mainly focuses on text documents. Audio-visual media, on the other hand, are much more dynamic and unpredictable. As a result, the DDL requires additional techniques and features from the required parsers such as object orientation, inheritance mechanisms, conditional instantiations of elements, temporal and spatial linking mechanisms from Descriptions into the data and between elements within a Description, and so on.

XML doesn't provide most of these features, but they're addressed by other conceptual frameworks under development within W3C, such as Synchronized Multimedia Integrated Language, or SMIL (see http://www.w3.org/TR/WD-smil), Schema for Object-Oriented XML (SOX), Resource Description Framework (RDF), and the XML Schema. We now give a brief overview of these approaches with respect to the DDL development.

XML 1.0 (see http://www.w3.org/TR/REC-xml) provides some linking mechanisms such as

▌ the Linking Specification (XLink; see http://www.w3.org/TR/1998/WD-xlink-19980303) that connects a source (that is, the actual document or a uniform resource identifier) with a target given by a uniform resource locator (URL) or URI,

▌ the Extended Pointer Specification (XPointer; see http://www.w3.org/TR/WD-xptr) that addresses a particular subsource, such as an area of or a particular point in a different XML document, and

▌ the XML Path Language (XPath; see http://www.w3.org/TR/xpath), another recent development in the world of links regarding addressing within a document.

However, these can't address the temporal needs required by MPEG-7 audio-visual data (such as the description of content for an analog video). SMIL (see http://www.w3.org/AudioVideo/#SMIL) attempts to overcome this problem within XML. SMIL is a style sheet that holds all the stylistic information about a Web multimedia presenta-

> XML Schema combines the work of W3C schema approches, providing most of their strengths.

tion for real-time delivery over the Web. Though SMIL is thought of as a layout language, the proposed specification describes mechanisms for linking to media files such as Audio Visual Interchange (AVI), Wave sound files (WAV), QuickTime movies (MOV), RealAudio (RA), RealVideo (RV), RealMedia (RM), Musical Instrument Digital Interface (MIDI), Shockwave, and Flash. Though not fully implemented yet, these might fulfill the MPEG-7 requirements for temporal and spatial linking mechanisms.

SOX (see http://www.w3.org/tr/note-sox) provides facilities for defining the structure, content, and semantics of XML documents to enable XML validation and automated content checking. The strengths of SOX compared to XML are an extensible data typing mechanism (which features three varieties of data types: scalar, enumerated, and format), a content model, and attribute inheritance. The major disadvantage of SOX is the design purpose, that is, the validation of business documents in e-commerce applications. This means SOX is designed for static forms rather than complex multimedia structures.

The W3C RDF working group aims to establish a Description Scheme Language (DSL) that provides the basis for creating generic tools for authoring, manipulating, and searching machine-understandable data on the Web (see http://www.w3.org/RDF/). On the technical level, RDF will provide XML name spaces to allow the use of a unique vocabulary for developed schemas. Its class (typically authored for a specific purpose or domain) and property-oriented structures (attributes) provide inheritance mechanisms (subclasses) and thus support object-oriented content description, the incremental modification of a base schema, and multiple descriptions (such as views) on data. This makes RDF extensible and shareable. However, RDF has major limitations with respect to MPEG-7 requirements. For example, RDF has no linking mechanisms to spatio-temporal

extents of a digital signal (such as a link to a spatially localized object in an MPEG-2 stream), limited data typing, no cardinality, no support for range constraints, and so on.

The latest approach of a schema for content description within the W3C is the XML Schema (see http://www.w3.org/tr/xmlschema-1/). The introduction of the XML Schema Structure Description states that the purpose of the XML Schema is

> to define and describe a class of XML documents by using their constructs to constrain and document the meaning, usage, and relationships of their constituent parts: data types, elements and their content, attributes and their values, entities and their contents, and notations. Schema constructs may also provide for the specification of implicit information, such as default values. Schemas are intended to document their own meaning, usage, and function through a common documentation vocabulary.

The XML Schema combines the work of different W3C schema approaches such as SOX and the Document Definition Markup Language (DDML), thus providing most of their strengths. Moreover, the W3C envisages that XML Schema and RDF syntax and schema will continue to co-exist but there will be ways to map between them.

Based on a thorough analysis and use of XML Schema for the encoding of preliminary Description Schemes, the MPEG-7 DDL development group has decided to use the XML Schema language as the basis for the DDL. However, certain reservations were raised at the Vancouver MPEG meeting in July 1999[3] concerning this approach. The major concerns were:

■ XML Schema's versioning is unstable,

■ it's still too declarative,

■ it needs better reference mechanisms than XLink and XPointer,

■ it doesn't support the definition of user-defined data types,

■ MPEG-7's dependency on the output and time schedule of W3C XML Schema WG, and

■ the effect of W3C's copyright of XML Schema

language on the ability to add MPEG-7-specific extensions.

As a result of these concerns, further discussions at the Vancouver meeting led to the decision to develop an MPEG-7-specific language in parallel with the XML Schema development being carried out within W3C. A new grammar based on a proposal by Australia's Distributed Systems Development Center (DSTC)[7] with modifications to ensure simple mapping to XML Schema, was recently developed. Based on this grammar, the following tasks are currently being performed:

■ specify the Extended Backus-Naur Form (EBNF) and an XML DTD for the new grammar,

■ specify the validation mechanisms that a parser must provide, and

■ develop a validating parser for this DDL.

To ensure synchronized development efforts, an official liaison between MPEG and the W3C was established. For detailed information about the status of the DDL development please look at the current DDL Development Document, available at the URLs provided at the end of the section "MPEG-7 schedule."

HyTime

The Hypermedia/Time-Based Structuring Language (HyTime),[8] is an SGML-based standard framework for integrated open hypermedia. By definition, HyTime documents are SGML documents and they entirely conform to SGML. In terms of functionality, however, HyTime extends the power of SGML in many ways.[9]

HyTime offers a list of definitions for element types, called architectural forms. These forms, or classes, allow modeling components for hypermedia documents such as hyperlinks or event schedules. These element classes permit multiple inheritance, which means that elements can inherit semantic and syntactic features not only from the governing DTD, but also from any number of other DTDs. This leads to greater flexibility compared to standardized DTDs for hypermedia documents.

HyTime's strongest asset is the support for linking and scheduling of documents in time and space. HyTime provides a standardized interface for the use of links such as clink (contextual link), ilink (independent link), agglink (aggregate traversal), and query link.

Although HyTime is one of the most powerful standards for general information management, the MPEG Requirements Group sees it as too generic and thus difficult to understand and handle. Nevertheless, parts of the HyTime standard, in particular those related to linking, are important for the MPEG-7 DDL.

MHEG-5

MHEG is developing within ISO several standards that deal with the coded representation of multimedia and hypermedia information.

The MHEG series of standards specify the coded representation of multimedia and hypermedia information objects (MHEG objects) for interchanging applications or systems as final-form units. These standards also specify the means of interchange from storage devices to local networks to telecommunication or broadcast networks.

MHEG-5 is the fifth part of the MHEG standard suite. It was developed to support the distribution of multimedia and hypermedia applications in a client-server architecture across platforms of different types. MHEG-5 defines a coded representation of multimedia and hypermedia interactive applications—that is, the syntax and the associated semantics allowing an author to build applications with the following features:

❚ text, graphic, and audio-visual output components,

❚ input components such as entry fields, buttons, and sliders, and

❚ behavior based on events that trigger actions applied to these components.

Though MHEG-5 shares common ground with MPEG-7 with respect to the description of content, MHEG-5 mainly focuses on providing solutions for multimedia applications. It addresses problems such as synchronization and control of information streams, which is why MHEG-5 suits applications such as video on demand and interactive TV well. However, these types of applications only form a fraction of the spectrum within MPEG-7.

Thus, MHEG-5 is of paramount importance for developing Descriptors and Description Schemes in the domain of interactive TV and video on demand, since here the sectional plane between the two standards can be found.

MPEG-7 schedule

Having discussed a number of issues about MPEG-7, the obvious question to ask is: Where are we in the process of MPEG-7 standardization?

At the time of writing, MPEG-7 has left the definition phase, where its scope, objectives, and requirements were defined. Since October 1998 MPEG-7 has entered its collaborative process, by launching a Call for Proposals (CFP).[10]

A CFP asks for technology for the standard. It's addressed to all interested parties, no matter whether they have participated in MPEG.

MPEG work usually occurs in two stages, competitive and collaborative. In the competitive stage, participants work on their technology by themselves. In answer to the CFP, people submit their technology to MPEG, after which MPEG makes a fair comparison between the submissions. MPEG experts conduct the proposal evaluation and development of the standard that follows.

Evaluation starts with assessing the scope and technical merits of each proposal as follows.

For the normative part, MPEG requires

❚ Descriptors

❚ Description Schemes

❚ A DDL

❚ Coding methods for compact representation of Descriptions

❚ Systems tools addressing the MPEG-7 Systems requirements specified in the MPEG-7 requirements document[1]

While Descriptor, Description Scheme, DDL, and coding scheme proposals were evaluated in February 1999, the proposals addressing MPEG-7 systems tools weren't part of that process. These tools were considered at the Seoul MPEG meeting, where the proposers presented and showed demos in March 1999. Further definition and review of systems requirements are ongoing.

To develop MPEG-7 for the eXperimentation Model (XM)—the continually evolving implementation set that will become a reference implementation—MPEG will also need technical contributions of

❚ Extraction methods

❚ Search methods

■ Evaluation and validation techniques

The proposals from the development group are needed to design and improve the XM. They were discussed during the Lancaster, UK meeting and in the MPEG-7 Evaluation Ad Hoc Group(s) between Lancaster and Seoul. The conclusion of these discussions was submitted as input to Seoul for the design of XM 1.0.

Note that it's not necessary for a proposal to address all of the elements and requirements listed above. A proposal can, for example, only propose Descriptors for some set of Features, or Description Schemes, or parts of Description Schemes. Selected aspects of different proposals will be incorporated into a common model (the XM) during the collaborative phase of the standard with the goal of building the best possible model. In that respect, the requirements document consists of requirements that MPEG sets in creating the MPEG-7 standard.

More than 600 people responded to the CFP by December 1998, from more than 90 countries. Within these preproposals, more than 500 included Descriptors, 100 included Description Schemes, and approximately 15 were DDLs.

During the collaborative phase, the XM gets updated and improved iteratively until MPEG-7 reaches the Working Draft stage. Improvements to the XM occur through core experiments, which are prompted by the contribution of new elements for the standard. Core experiments will be defined to test the contributed tools within XM's framework according to well-defined test conditions and criteria. The goal is to develop the best possible XM. Finally, those parts of the XM (or of the Working Draft) that correspond to the normative elements of MPEG-7 will be standardized.

Note that although the time for entering the initial CFP has passed, the core experiments provide an ideal opportunity for new organizations to join in the MPEG-7 efforts by proposing technology that fulfills still needed functionalities or provides a significant improvement over existing MPEG-7 technologies.

You can find a detailed description of the evaluation process in the *MPEG-7 Evaluation Document*.[11] The *Description of MPEG-7 Content Set*[12] describes the available audio-video material for these tests. For distribution and relevant licensing issues, see the documents *Distribution of MPEG-7 Content Set*[13] and *Licensing Agreement for the MPEG-7 Content Set*.[14] The *MPEG-7 Proposal Package Description* document[15] discusses the XM's test phase.

For more detailed information on current events and news, please see http://www.darmstadt.gmd.de/mobile/MPEG7/, http://www.mpeg-7.com, and http://www.cselt.it/mpeg/.

Conclusion

MPEG-7 is an ambitious standardization effort from the Motion Pictures Expert Group. A number of open questions still exist, but the established results point to a promising future. However, the most important question still needs to be answered, that is: What is the balance between flexibility and compatibility within MPEG-7?

The MPEG-7 working group has to decide whether they follow a specific, bottom-up approach for a few individual domains, or if the intention is to let anyone create their own MPEG-7 solution. The group's decision will have a clear influence on the option of standardizing only the DDL or a DDL and a core set of Descriptors and Description Schemes.

We believe that MPEG-7 should make a strong showing in some core applications by establishing Description Schemes and variants that would serve the video, image, music, speech, and sound indexing communities well, allowing a number of initial products to target those basic standards. MPEG-7 should provide a level of genericity (in the Descriptors) and power (in the DDL) that will let specialized communities (such as biomedical or remote sensing imaging) adapt the standard to their uses.

Furthermore, we believe that MPEG-7's core goal is to provide interoperability. At the end of MPEG-7, whether version 1 or 2, there should exist a single DDL, a generic set of Descriptors for audio and visual features, and a specific Description Scheme that serves specific applications. However, even the authors are divided on the question of how to handle cases where a Feature cannot be captured by simply structuring existing Descriptors into a novel Description Scheme. The problem is that a Descriptor built using the DDL might allow the novel Description Scheme to be perfectly parsable, but the new defined Descriptor at the bottom of whatever structure might provide semantic information that other computers can't understand. On the other side, introducing a registration body seems most problematic, especially since this might also lead to forced incompatibilities due to a variety of competing but incompatible Descriptors. Ultimately, struggling with these sorts of questions makes the MPEG-7 process intellectually stimulating and rewarding.

We have faith that we will see a standard that provides the compatibility of content descriptions, allowing a given community to adopt it early. MPEG-7 should also offer the flexibility for that community to grow and include other special interests.   **MM**

References

1. *ISO/MPEG N2859, MPEG-7 Requirements*, MPEG Requirements Group, Int'l Organization for Standardization, Geneva, July 1999.

2. *ISO/MPEG N2860, Applications for MPEG-7*, MPEG Requirements Group, Int'l Organization for Standardization, Geneva, July 1999.

3. *ISO/IEC JTC 1/SC29/WG11 N2862, MPEG-7 Description Definition Language Document, v 1.0,* Int'l Organization for Standardization, Geneva, July 1999.

4. C.F. Goldfarb, E.J. Mosher, and T.I. Peterson, "An Online System for Integrated Text Processing," *Proc. American Society for Information Science*, Vol. 7, American Society for Information Science Press, Silver Spring, Md., 1970, pp. 147-150

5. *ISO 8879, Standard Generalized Markup Language (SGML)*, Int'l Organization for Standardization, Geneva, Oct. 1986.

6. M. Bryan, *SGML*, Addison-Wesley, Reading, Mass., 1992.

7. J. Hunter, *A Revised Proposal for an MPEG-7 DDL*, M4518, Distributed Systems Development Center (DSTC), Brisbane, Australia, 1999.

8. *ISO/IEC JTC1/SC18 WG8, W1920 rev., Hypermedia/Time-Based Structuring Language (HyTime)*, 2d ed., Int'l Organization for Standardization, Geneva, May 1997.

9. S. DeRose and D. Durand, *Making Hypermedia Work—A User's Guide to HyTime*, Kluwer Academic Publishers, Boston, 1994.

10. *ISO/MPEG N2469, Call for Proposals for MPEG-7 Technology*, MPEG Requirements Group, Int'l Organization for Standardization, Geneva, October 1998.

11. *ISO/MPEG N2463, MPEG-7 Evaluation Process Document*, MPEG Requirements Group, Int'l Organization for Standardization, Geneva, October 1998.

12. *ISO/MPEG N2467, Description of MPEG-7 Content Set*, MPEG Requirements Group, Int'l Organization for Standardization, Geneva, October 1998, http://www.darmstadt.gmd.de/mobile/MPEG7/Documents/N2467.html.

13. *ISO/MPEG N2468, Distribution of MPEG-7 Content Set*, MPEG Requirements Group, Int'l Organization for Standardization, Geneva, October 1998, http://www.darmstadt.gmd.de/mobile/MPEG7/Documents/N2468.html.

14. *ISO/MPEG N2466, Licensing Agreement for the MPEG-7 Content Set*, MPEG Requirements Group, Int'l Organization for Standardization, Geneva, October 1998, http://www.darmstadt.gmd.de/MPEG7/Documents/N2466.html.

15. *ISO/MPEG N2464, MPEG-7 Proposal Package Description (PPD)*, MPEG Requirements Group, Int'l Organization for Standardization, Geneva, October 1998.

Frank Nack is a member of the Mobile Interactive Media (Mobile) division at the German National Research Center for Information Technology-Integrated Publication and Information Systems Institute (GMD-IPSI). His research interests include video representation, digital video production, interactive storytelling, and media-networked-oriented agent technology. He earned his PhD in computing at Lancaster University, UK. He is an active member of the MPEG-7 standardization group, where he served as editor of the *Context and Objectives Document* and the *Requirements Document*. He now chairs the MPEG-7 Description Definition Language (DDL) development group.

Adam Lindsay is the principal investigator of the multimedia research division at Starlab, a research firm based in Brussels, Belgium. He joined Starlab after earning his MS in mid-level representation of musical melody at the Massachusetts Institute of Technology Media Lab, Cambridge, Massachusetts. His current research is on applying MPEG-7-style metadata to multimedia to make it more intelligent about itself. He is an active member of the MPEG-7 standardization group, where he served as the editor of the MPEG-7 *Applications Document* and as the leader in MPEG-7 Audio activities.

Readers may contact Nack at GMD-IPSI Dolivostr. 15, 64293 Darmstadt, Germany, e-mail nack@darmstadt.gmd.de. Contact Lindsay at Starlab, Excelsiorlaan 40-42, B-1930, Zaventem, Belgium, e-mail adam@starlab.net.