Title: Introduction to MPEG-7 (version 1.0)
Status: approved
Source: Requirements
Editor: José M. Martínez (UPM-GTI, ES)

TABLE OF CONTENTS

THE CONTEXT OF MPEG-7 ................................................................. 1
MPEG-7 OBJECTIVES........................................................................ 2
CREATING MPEG-7 APPLICATIONS .............................................. 3
METHOD OF WORK, WORK PLAN, AND CURRENT STATUS........... 4
THE UNIQUENESS OF MPEG-7 IN THE 21ST CENTURY MEDIA LANDSCAPE .... 5
REFERENCES .................................................................................. 6
CONTACT POINTS ............................................................................ 6
ANNEX A: MPEG-7 TERMINOLOGY .............................................. 7
ANNEX B: MPEG-7 FAQS............................................................... 8
ANNEX C: MPEG-7 AND OTHER MPEG STANDARDS ...................... 10
Introduction to MPEG-7

Accessing audio and video used to be a simple matter - simple because of the simplicity of the access mechanisms and because of the poverty of the sources. The transition between two millennia abounds with new ways to produce, offer, filter, search, and manage digitized multimedia information. Broadband is being offered with increasing audio and video quality and speed of access. The trend is clear. In the next few years, users will be confronted with such a large number of content provided by multiple sources that efficient and accurate access to this almost infinite amount of content will seem to be unimaginable. This challenging situation demands a timely solution to the problem. MPEG-7 is the answer to this need.

MPEG-7 aims at offering a comprehensive set of audiovisual description tools to create descriptions, which will form the basis for applications enabling the needed quality access to content, which implies good storage solutions, high-performance content identification, proprietary assignation, and fast, ergonomic, accurate and personalized filtering, searching and retrieval. This is a challenging task given the broad spectrum of requirements and targeted multimedia applications, and the broad number of audiovisual features of importance in such context. The question of identifying and managing content is not just restricted to database retrieval applications such as digital libraries, but extends to areas like broadcast channel selection, multimedia editing, and multimedia directory services.

MPEG-7 is an ISO/IEC standard being developed by MPEG (Moving Picture Experts Group), the committee that also developed the Emmy Award winning standards known as MPEG-1 and MPEG-2, and the MPEG-4 standard. The MPEG-1 and MPEG-2 standards are used in many applications, including DVD and digital television. MPEG-4 provides the standardized technological elements enabling the integration of the production, distribution and content access paradigms of the fields of interactive multimedia, mobile multimedia, interactive graphics and enhanced digital television. This is a kind of repeated in the section “MPEG-7 and other MPEG standards”.

This document gives an introductory overview of the MPEG-7 standard. More information about MPEG-7 can be found at the MPEG-7 website http://drogo.cselt.it/mpeg/ and the MPEG-7 Industry Focus Group website http://www.mpeg-7.com. These web pages contain links to a wealth of information about MPEG, including much about MPEG-7, many publicly available documents, several lists of ‘Frequently Asked Questions’ and links to other MPEG-7 web pages.

The Context of MPEG-7

More and more audiovisual information is available, from many sources around the world. The information may be represented in various forms of media, such as still pictures, graphics, 3D models, audio, speech, video. Audiovisual information plays an important role in our society, be it recorded in such media as film or magnetic tape or originating, in real time, from some audio or visual sensors and be it analogue or, increasingly, digital. While audio and visual information used to be consumed directly by the human being, there is an increasing number of cases where the audiovisual information is created, exchanged, retrieved, and re-used by computational systems. This may be the case for such scenarios as image understanding (surveillance, intelligent vision, smart cameras, etc.) and media conversion (speech to text, picture to speech, speech to picture, etc.). Other scenarios are information retrieval (quickly and efficiently searching for various types of multimedia documents of interest to the user) and filtering in a stream of audiovisual content description (to receive only those multimedia data items which satisfy the user’s preferences). For example, a code in a television program triggers a suitably programmed VCR to record that program, or an image sensor triggers an alarm when a certain visual event happens. Automatic transcoding may be performed from a string of characters to audible information or a search may be performed in a stream of audio or video data. In all these examples, the audiovisual information has been suitably “encoded” to enable a device or a computer code to take some action.

Audiovisual sources will play an increasingly pervasive role in our lives, and there will be a growing need to have these sources processed further. This makes it necessary to develop forms of audiovisual information representation that go beyond the simple waveform or sample-based, compression-based (such as MPEG-1 and MPEG-2) or even objects-based (such as MPEG-4) representations. Forms of representation that allow some degree of interpretation of the information’s meaning are necessary. These forms can be passed onto, or accessed by, a device or a computer code. In the examples given above an image sensor may produce visual data not in the form of PCM samples (pixels values) but in the form of objects with associated physical measures and time information. These could then be stored and processed to verify if certain programmed conditions are met. A video recording device could receive descriptions of the audiovisual information associated to a program that would enable it to record, for example, only news with the exclusion of sport. Products from a company could be described in such a way that a machine could respond to unstructured queries from customers making inquiries.
MPEG-7 will be standard for describing the multimedia content data that will support these operational requirements. The requirements apply, in principle, to both real-time and non real-time as well as push and pull applications. MPEG will not standardize or evaluate applications. MPEG may, however, use applications for understanding the requirements and evaluation of technology. It must be made clear that the requirements in this document are derived from analyzing a wide range of potential applications that could use MPEG-7 descriptions. MPEG-7 is not aimed at any one application in particular; rather, the elements that MPEG-7 standardizes shall support as broad a range of applications as possible.

**MPEG-7 Objectives**

The MPEG-7 standard aims at providing standardized core technologies allowing description of audiovisual data content in multimedia environments. It will extend the limited capabilities of proprietary solutions in identifying content that exist today, notably by including more data types. Audiovisual data content that has MPEG-7 data associated with it, may include: still pictures, graphics, 3D models, audio, speech, video, and composition information about how these elements are combined in a multimedia presentation (scenarios). Special cases of these general data types may include facial expressions and personal characteristics. MPEG-7 description tools do, however, not depend on the ways the described content is coded or stored. It is possible to create an MPEG-7 description of an analogue movie or of a picture that is printed on paper, in the same way as of digitised content.

MPEG-7 Description tools allow to create descriptions (the result of using the MPEG-7 description tools at the users will) of content that may include:

- Information describing the creation and production processes of the content (director, title, short feature movie)
- Information related to the usage of the content (copyright pointers, usage history, broadcast schedule)
- Information of the storage features of the content (storage format, encoding)
- Structural information on spatial, temporal or spatio-temporal components of the content (scene cuts, segmentation in regions, region motion tracking)
- Information about low level features in the content (colors, textures, sound timbres, melody description)
- Conceptual information of the reality captured by the content (objects and events, interactions among objects)

All these descriptions are of course coded in an efficient way for searching, filtering, etc.

To accommodate this variety of complementary content descriptions, MPEG-7 approaches the description of content from several viewpoints. Currently five viewpoints are defined: Creation & Production, Media, Usage, Structural aspects and Conceptual aspects. The five sets of description elements developed on those viewpoints are presented here as separate entities. However, they are interrelated and can be combined in many ways. Depending on the application, some will present and others can be absent or only partly present.

A description generated using MPEG-7 description tools will be associated with the content itself, to allow fast and efficient searching for, and filtering of material that is of interest to the user. The type of content and the query do not have to be the same; for example, visual material may be queried using visual content, music, speech, etc. It is the responsibility of the search engine and filter agent to match the query data to the MPEG-7 description.

Figure 1 explains a hypothetical MPEG-7 chain in practice. The circular boxes depict tools that are doing things, such as encoding or decoding, whereas the square boxes represent static elements, such as a description. The grayed boxes in the figure encompass the normative elements of the MPEG-7 standard. The standard does not describe the process of (automatic) extraction of descriptions/features, nor does it specify the search engine, filter agent, or any other program that can make use of the descriptions.
Note: There can be other streams from content to user; these are not depicted here. Furthermore, it is understood that there might be cases where a binary efficient representation of the description is not needed, and a textual representation would suffice. Thus, the use for the encoder and decoder is optional.

Figure 1: An abstract representation of possible applications using MPEG-7.

MPEG-7 addresses many different applications in many different environments, which means that it needs to provide a flexible and extensible framework for describing audiovisual data. Therefore, MPEG-7 does not define a monolithic system for content description but rather a set of methods and tools for the different viewpoints of the description of audiovisual content. Having this in mind, MPEG-7 is designed to take into account all the viewpoints under consideration by other leading standards such as, SMPTE Metadata Dictionary, Dublin Core, EBU P/Meta, and TV Anytime, which are focused to more specific applications or application domains, whilst MPEG-7 tries to be as generic as possible. MPEG-7 uses also XML Schema as the language of choice for the textual representation of content description and for allowing extensibility of description tools. Considering the popularity of XML, usage of it will facilitate interoperability in the future.

The main elements of the MPEG-7’s standard are:

- Descriptors (D): representations of Features, that define the syntax and the semantics of each feature representation,
- Description Schemes (DS), that specify the structure and semantics of the relationships between their components. These components may be both Descriptors and Description Schemes,
- A Description Definition Language (DDL) to allow the creation of new Description Schemes and, possibly, Descriptors and to allows the extension and modification of existing Description Schemes,
- System tools, to support multiplexing of descriptions, synchronization issues, transmission mechanisms, coded representations (both textual and binary formats) for efficient storage and transmission, management and protection of intellectual property in MPEG-7 descriptions, etc.

Creating MPEG-7 Applications

The elements that MPEG-7 standardizes will support a broad range of applications (for example, multimedia digital libraries, broadcast media selection, multimedia editing, home entertainment devices, etc.). MPEG-7 will also make the web as searchable for multimedia content as it is searchable for text today. This would apply especially to large content archives, which are being made accessible to the public, as well as to multimedia catalogues enabling people to identify content for purchase. The information used for content retrieval may also be used by agents, for the selection and filtering of broadcasted "push" material or for personalized advertising. Additionally, MPEG-7 descriptions will allow fast and cost-effective usage of the underlying data, by enabling semi-automatic multimedia presentation and editing.
All domains making use of multimedia will benefit from MPEG-7. Considering that at present day it is hard to find one not using multimedia, please extend the list of the examples below using your imagination:

- Digital libraries, Education (image catalogue, musical dictionary, Bio-medical imaging catalogues, …)
- Multimedia editing (personalised electronic news service, media authoring)
- Cultural services (history museums, art galleries, etc.),
- Multimedia directory services (e.g. yellow pages, Tourist information, Geographical information systems)
- Broadcast media selection (radio channel, TV channel, …)
- Journalism (e.g. searching speeches of a certain politician using his name, his voice or his face),
- E-Commerce (personalised advertising, on-line catalogues, directories of e-shops, …)
- Surveillance (traffic control, surface transportation, non-destructive testing in hostile environments, etc.),
- Investigation services (human characteristics recognition, forensics),
- Home Entertainment (systems for the management of personal multimedia collections, including manipulation of content, e.g. home video editing, searching a game, karaoke, …)
- Social (e.g. dating services),

Imagine the things you’ll be able to do having MPEG-7 enabled technology. You’ll be able to:

- Play a few notes on a keyboard and retrieve a list of musical pieces similar to the required tune, or images matching the notes in a certain way, e.g. in terms of emotions.
- Draw a few lines on a screen and find a set of images containing similar graphics, logos, ideograms, …
- Define objects, including colour patches or textures and retrieve examples among which you select the interesting objects to compose your design.
- On a given set of multimedia objects, describe movements and relations between objects and so search for animations fulfilling the described temporal and spatial relations.
- Describe actions and get a list of scenarios containing such actions.
- Using an excerpt of Pavarotti’s voice, obtaining a list of Pavarotti’s records, video clips where Pavarotti is singing and photographic material portraying Pavarotti.

**Method of Work, Work Plan, and current status**

The method of development is comparable to that of the previous MPEG standards. MPEG work is usually carried out in three stages: definition, competition, and collaboration. In the definition phase, the scope, objectives and requirements for MPEG-7 were defined. In the competitive stage, participants worked on their technology by themselves. The end of this stage was marked by the MPEG-7 Evaluation following an open Call for Proposals (CfP). The Call asked for relevant technology fitting the requirements. In answer to the Call, all interested parties, no matter whether they participate or have participated in MPEG, were invited to submit their technology to MPEG. Some 60 parties submitted, in total, almost 400 proposals, after which MPEG made a fair expert comparison between these submissions. Selected elements of different proposals will be incorporated into a common model (the eXperimentation Model, or XM) during the collaborative phase of the standard. The goal is building the best possible model, which is in essence a draft of the standard itself. During the collaborative phase, the XM is updated and improved in an iterative fashion, until MPEG-7 reaches the Committee Draft (CD) stage in October 2000, after several versions of the Working Draft. Improvements to the XM are made through Core Experiments (CEs). CEs are defined to test the existing tools against new contributions and proposals, within the framework of the XM, according to well-defined test conditions and criteria. Finally, those parts of the XM (or of the Working Draft) that correspond to the normative elements of MPEG-7 will be standardized.

The current work plan for MPEG-7 is shown below:

<table>
<thead>
<tr>
<th>Call for Proposals</th>
<th>Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>October 1998</td>
<td>February 1999</td>
</tr>
</tbody>
</table>
Currently MPEG-7 concentrates on the specification of description tools (Descriptors and Description Schemes), together with the development of the MPEG-7 reference software, known as XM (eXperimentation Model). The XML Schema Language was chosen as the base for the Description Definition Language (DDL), but with some further developments since XML Schema is not enough to fulfill all the DDL requirements.

The MPEG-7 Audio group develops a range of Description Tools, from generic audio descriptors (e.g., waveform and spectrum envelopes, fundamental frequency) to more sophisticated description tools like Spoken Content and Timbre. Generic Audio Description tools will allow the search for similar voices, by searching similar envelopes and fundamental frequencies of a voice sample against a database of voices. The Spoken Content Description Scheme (DS) is designed to represent the output of a great number of state of the art Automatic Speech Recognition systems, containing both words and phonemes representations and transition likelihoods. This alleviates the problem of out-of-vocabulary words, allowing retrieval even when the original word was wrongly decoded. The Timbre descriptors (Ds) describe the perceptual features of instrument sound, that make two sounds having the same pitch and loudness appear different to the human ear. These descriptors allow searching for melodies independently of the instruments.

The MPEG-7 Visual group is developing four groups of description tools: Color, Texture, Shape and Motion. Color and Texture Description Tools will allow the search and filtering of visual content (images, graphics, video) by dominant color or textures in some (arbitrarily shaped) regions or the whole image. Shape Description Tools will facilitate “query by sketch” or by contour similarity in image databases, or, for example, searching trademarks in registration databases. Motion Description Tools will allow searching of videos with similar motion patterns that can be applicable to news (e.g. similar movements in a soccer or football game) or to surveillance applications (e.g., detect intrusion as a movement towards the safe zone).

The Multimedia Description Schemes group is developing the description tools dealing with generic (basic structures, common DSs), audiovisual (structure of video and audio) and archival features (collections, streaming). Its central tools deal with content management and content description. Content Management description tools cover the viewpoints of Media, Creation and Production, and Usage. Media description tools allows searching for preferred storage formats, compression qualities, and aspect ratios among others. Creation and Production descriptions tools cover the typical archival and credits information (e.g., title, creators, classification). Usage description tools deal with description related to the use of the described content (e.g. rights, broadcasting dates and places, availability, audience, financial data). The Content Description ones cover both structural and conceptual viewpoints. Structural description tools provide segmentation, both spatial and temporal, of the content. This allows, among other functionalities, assigning descriptions to different regions and segments (e.g., to provide the means for a segment annotation instead of only a global one) and providing importance rating of temporal segments and regions (e.g., allowing to differentiate among regions of the content for adaptive coding with different quality). Conceptual description tools allow providing semantic based description (e.g., linguistic annotations of the content, and object and event description from a knowledge viewpoint). Besides the Content Description and Content Management description tools, there are others targeted to content organization (e.g., to organize an archive of image’s descriptions in a repository), navigation and access (e.g., to display a summary of videos through relevant short sequences or keyframes for quick browsing), and user preferences (e.g. for agent based selection or filtering of favorite programs).

The Uniqueness of MPEG-7 in the 21st century Media landscape

How many times have you seen science fiction movies such as 2001, A Space Odyssey or Star Trek and think, ‘Wow, we are so far away from having some of the fancy gadgets depicted in these movies!’ In 2001, Hal, the talking computer intelligently navigates and retrieves information or runs complex operations instigated by spoken input. What about the communicator in Star Trek? Surely, today’s mobile phones are the first signs of a ‘Star Trek’ communicator where AV content can be broadcasted, filtered, searched, navigated and retrieved.
MPEG-7 is, at last, the beginning of the road to the realization of dreams of so many imaginative minds of the 20th century. MPEG-7 will indeed play a key role to that often heard refrain, ‘this is what I thought computers were supposed to do!’ MPEG-7 will enable applications that mould computers around human requirements and not humans around computer requirements. Unlike today’s state-of-the-art technology, MPEG-7 allows for objective description of features so as to enable content disclosure based on facts, rather than on (unpredictable) human annotations. Finding information by rich spoken queries, hand-drawn images, and humming will improve the user-friendliness of computer systems and finally address what most people expect computers to be able to do.

For professionals, a new generation of applications providing tools for high-quality information search and retrieval will be possible. For example, TV program producers will be able to search with ‘laser-precision’ for occurrences of famous entities, stored in thousands of hours of audiovisual records, in order to create material for a program about that entity. Program production time will reduce and the quality of program content will increase.

MPEG-7 is a multimedia content description standard, which closely addresses how humans expect to interact with computer systems because it develops rich descriptions that reflect those expectations. MPEG-7 is about the future of media in the 21st century. This is not an overstatement. MPEG-7 provides a comprehensive and flexible framework for describing the content of multimedia. To describe content implies knowledge of elements it consists of, as well as, knowledge of interrelations between those elements. The most straightforward application is multimedia management, where such knowledge is prerequisite for efficiency and accuracy. However, there are other serious implications. Imprinted knowledge of content and structure, so far elitarian knowledge possessed by content creators only, is made public here, allowing content manipulation, and ultimately content reuse – new content creation. Copyrights issues are not banal here. Other issues and concerns arise, but they are balanced by incredible economical, educational, and ergonomic benefits that will be brought by MPEG-7 technology. Potential concerns will be resolved, and in some years we will not be able to imagine media without MPEG-7 technologies.

References

There are a number of documents available at the MPEG Home Page at http://drogo.cselt.it/mpeg/, including:

- MPEG-7 Requirements
- MPEG-7 Applications
- MPEG-7 Context, Objectives and Technical Roadmap
- MPEG-7 Principal Concepts List
- MPEG-7 Overview
- MPEG-7 DDL WD
- MPEG-7 Visual WD
- MPEG-7 Audio WD
- MPEG-7 MDS WD and XM

Information more focused to industry is also available at the MPEG-7 Industry Focus Group Web site at http://www.mpeg-7.com.

Contact points

- Neil Day (neil@garage.co.jp)
  For Technical Issues
  - Requirements: Rob Koenen (r.h.koenen@research.kpn.com)
  - Audio: Adam Lindsay (atl@comp.lancs.ac.uk)
  - Visual: Thomas Sikora (sikora@hhi.de)
  - Multimedia DS: Philippe Salembier (philippe@gps.tsc.upc.es)
  - Systems: Olivier Avaro (olivier.avaro@francetelecom.fr)
  - XM Software Implementation: Stephan Herrmann (stephanh@lis.e-technik.tu-muenchen.de)

Editor:

- José M. Martinez (jms@gti.ssr.upm.es)
Annex A: MPEG-7 Terminology

1. **Data**

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

   **Notes**
   This definition is intended to be sufficiently broad to encompass graphics, still images, video, film, music, speech, sounds, text and any other relevant AV medium.

   **Examples**
   Examples for MPEG-7 data are an MPEG-4 stream, a video tape, a CD containing music, sound or speech, a picture printed on paper, and an interactive multimedia installation on the web.

2. **Feature**

   **Definition**
   A Feature is a distinctive characteristic of the data which signifies something to somebody.

   **Notes**
   Features themselves cannot be compared without a meaningful feature representation (descriptor) and its instantiation (descriptor value) for a given data set.

   **Examples**
   Some examples are: color of an image, pitch of a speech segment, rhythm of an audio segment, camera motion in a video, style of a video, the title of a movie, the actors in a movie etc.

3. **Descriptor**

   **Definition**
   A Descriptor (D) is a representation of a Feature. A Descriptor defines the syntax and the semantics of the Feature representation.

   **Notes**
   A descriptor allows an evaluation of the corresponding feature via the descriptor value. It is possible to have several descriptors representing a single feature, i.e. to address different relevant requirements.

   **Examples**
   For example for the color feature, possible descriptors are: the color histogram, the average of the frequency components, the motion field, the text of the title, etc. More examples of Features and their associated Descriptors are provided in Table 1.

4. **Descriptor Value**

   **Definition**
   A Descriptor Value is an instantiation of a Descriptor for a given data set (or subset thereof).

   **Notes**
   Descriptor Values are combined via the mechanism of a Description Scheme (see point 5) to form a Description (see point 6).

   **Examples**

5. **Description Scheme**

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

   **Notes**
   The distinction between a DS and a D is, that a D contains only basic data types, as provided by the DDL (see point 8), and does not refer to another D or (sub)DS.

   **Examples**

---

Introduction to MPEG-7
A movie, temporally structured as scenes and shots, including some textual descriptors at the scene level, and color, motion and some audio descriptors at the shot level.

6. Description

   Definition
   A Description consists of a DS (structure) and the set of Descriptor Values (instantiations) that describe the Data.

   Notes
   Depending on the completeness of the set of Descriptor Values, the DS may be fully or partially instantiated. Whether or not the DS is actually present in the Description depends on technical solutions still to be provided.

   Examples

7. Coded Description

   Definition
   A Coded Description is a Description that has been encoded to fulfil relevant requirements such as compression efficiency, error resilience, random access, etc.

   Notes

   Examples

8. Description Definition Language

   Definition
   The Description Definition Language (DDL) is a language that allows the creation of new Description Schemes and, possibly, Descriptors. It also allows the extension and modification of existing Description Schemes.

   Notes
   It is not yet clear to which extend the DDL will allow the creation of new descriptors.

   Examples

Annex B: MPEG-7 FAQs

1. What is MPEG-7?

   MPEG-7 will be a standardised description of various types of multimedia information. This description will be associated with the content itself, to allow fast and efficient searching for material that is of interest to the user. MPEG-7 is formally called ‘Multimedia Content Description Interface’. The standard does not comprise the (automatic) extraction of descriptions/features. Nor does it specify the search engine (or any other program) that can make use of the description.

2. From whom or where did the demand for MPEG-7 come?

   The demand logically follows the increasing availability of digital audiovisual content. MPEG members recognised this demand, and initiated a new work item. The work on the definition of MPEG-7 has already started to attract new people to MPEG.

3. Why is MPEG-7 needed?

   Nowadays, more and more audiovisual information is available, from many sources around the world. Also, there are people who want to use this audiovisual information for various purposes. However, before the information can be used, it must be located. At the same time, the increasing availability of potentially interesting material makes this search more difficult. This challenging situation led to the need of a solution to the problem of quickly and efficiently searching for various types of multimedia material interesting to the user. MPEG-7 wants to answer to this need, providing this solution.
4. Who is currently participating in the development of the MPEG-7 standard?

The people taking part in defining MPEG-7 represent broadcasters, equipment manufacturers, digital content creators and managers, transmission providers, publishers and intellectual property rights managers, as well as university researchers.

5. Where are you in the process of specifying the MPEG-7 standard?

We are in the collaborative phase of the standardisation process. This means that we have passed the Call for Proposals and the evaluation of the submissions to that CfP. We are currently performing experiments (so-called Core Experiments) to continuously improve the technology on the table for standardization. This testing is carried out in a common environment, called the experimentation Model (XM). Experiments are carried out in well-defined test conditions and according to pre-defined criteria. The goal is to develop the best possible standard.

6. Will MPEG-7 include audio or video content recognition?

The standardisation of audiovisual content recognition tools is beyond the scope of MPEG-7. Following its principle ‘specifying the minimum for maximum usability, MPEG-7 will concentrate on standardising a representation that can be used for description. Development of audiovisual content recognition tools will be a task for industries which will build and sell MPEG-7 enabled products. In developing the standard, however, MPEG might build some coding tools, just as it did with the predecessors of MPEG-7, namely MPEG-1, -2 and -4. Also for these standards, coding tools were built for research purposes, but they did not become part of the standard itself.

7. Will MPEG-7 support audio or video content retrieval?

In the same way that MPEG will not standardise the tools to generate the description, MPEG-7 will also not standardise the tools that use the description. It might however be necessary to address the interface between the description and the search engine.

8. What form will the "descriptions" of multimedia content in MPEG-7 take?

The words ‘descriptions’ or ‘features’ represent a rich concept, that can be related to several levels of abstraction. Descriptions vary according to the types of data. Furthermore, different types of descriptions are necessary for different purposes of the categorisation.

9. Will the standard allow automatic extraction of descriptions as well as manual entry?

The descriptions that conform to the MPEG-7 standard could be entered by hand, but they could also be automatically extracted. Some features can be best extracted automatically (colour, texture), but for some other features (‘this scene contains three shoes and that music was recorded in 1995’) this is very hard or even impossible.

10. A 'Call for Proposals', how does that work?

A Call for Proposals (CfP) asks for technology for inclusion in the standard. It is addressed at all interested parties, no matter whether they participate or have participated in MPEG. MPEG work is usually carried out in two stages, a competitive and a collaborative one. 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. Based on the outcome of the evaluation, MPEG decided which proposals to use for the collaborative stage. In this stage, members of the Experts Group work together on improving and expanding the standard under construction, building on the selected proposals.

11. What is the relationship between MPEG-7 and other MPEG activities?

MPEG-7 can be used independently of the other MPEG standards - the description might even be attached to an analog movie. The representation that is defined within MPEG-4, i.e. the representation of audiovisual data in terms of objects, is however very well suited to what will be built on the MPEG-7 standard. This representation is basic to the process of categorisation. In addition, MPEG-7 descriptions could be used to improve the functionality of previous MPEG standards.

12. If I want to get involved in MPEG-7, what do I need to know about the other MPEG standards?

In principle, knowledge about the other three MPEG standards is not required for taking part in the MPEG-7 work. However, since some of MPEG-7’s tools may be close to those of MPEG-4, some knowledge about them could be useful.
13. If I want to know more about the other MPEG standards, where do I look?

You can start by taking a look at MPEG’s home page (http://www.cselt.it/mpeg/) which contains many useful references, including more lists with "Frequently Asked Questions” about MPEG activities.

14. So what happened to MPEG-5 and -6? (And how about 3?)

MPEG-3 existed once upon a time, but its goal, enabling HDTV, could be accomplished using the tools of MPEG-2, and hence the work item was abandoned. So after 1,2 and 4, there was much speculation about the next number. Should it be 5 (the next) or 8 (creating an obvious binary pattern)? MPEG, however, decided not to follow either logical expansion of the sequence, but chose the number of 7 instead. So MPEG-5 and MPEG-6 are, just like MPEG-3, not defined.

15. When will MPEG-7 replace the existing MPEG-1 and MPEG-2 standards?

MPEG-7 will not replace MPEG-1, MPEG-2 or MPEG-4. It is intended to provide complementary functionality to these other MPEG standards: representing information about the content, not the content itself (“the bits about the bits”. This functionality is the standardisation of multimedia content descriptions.

17. If I want to know more about, be involved in, or give an input to the MPEG-7 development process, whom should I contact?

You can contact any of the people listed below with their email addresses and telephone numbers. To visit MPEG meetings you need to be on your national delegation, but the people listed in the contact points can explain how this works.

Annex C: MPEG-7 and other MPEG Standards

Currently there are three MPEG standards dealing with compression, decompression, processing, and coded representation of moving pictures, audio and their combination.

MPEG-1 is a standard for storage and retrieval of moving pictures and audio on storage media, that is very successful. It is the de-facto form of storing moving pictures and audio on the World Wide Web and is used in millions of Video CDs. Digital Audio Broadcasting (DAB) is a new consumer market that makes use of MPEG-1 audio coding.

MPEG-2 is a standard for digital television, that has been the timely response for the satellite broadcasting and cable television industries in their transition from analogue to digital formats. Millions of set-top boxes incorporating MPEG-2 decoders have been sold in the last 3 years.

MPEG-4 is a standard for multimedia applications that supports the creation of rich, reusable, and interactive multimedia content that can be used by different distribution networks (broadcasting, internet, CDs,...) and terminals (PCs with Web browsers, TV sets, Set-Top-Boxes, DVD players, ...). MPEG-4 is the first real multimedia representation standard, allowing interactivity and a combination of natural and synthetic materials, coded in the form of objects that are integrated to compose multimedia presentations (scenarios).

In principle, MPEG-1, -2, and -4 are designed to represent the information itself, while MPEG-7 is meant to represent information about the information. Looking from another perspective: MPEG-1, -2, and -4 make content available, while MPEG-7 allows you to find the content you need.

MPEG-7 can be used independently of the other MPEG standards - the description might even be attached to an analog movie. MPEG-7 descriptions could be used to improve the functionalities of previous MPEG standards, but will not replace MPEG-1, MPEG-2 or MPEG-4. It is intended to provide complementary functionality to these other MPEG standards: representing information about the content, not the content itself ("the bits about the bits").

Besides these standards, MPEG started recently the development of MPEG-21, a standard that aims at creating a Multimedia Framework taking into consideration the different components involved in the delivery of content from the creator to the user.