WP9: A Review of Data and Metadata Standards and Techniques for Representation of Multimedia Content

Maria Grazia Di Bono, Gabriele Pieri, Ovidio Salvetti

ISTI-CNR Pisa

Deadline August 31st, 2004

Contents

1. Introduction 4

2. State of the Art on Multimedia data and metadata standards 7

2.1 Multimedia data standards overview 7

2.1.1 Video standards 7

2.1.1.1 MPEG family 7

2.1.1.2 Other Video Standards 8

2.1.2 Audio standards 9

2.1.2.1 MP3 is not MPEG3 9

2.1.2.2 Some other audio standards 9

2.1.3 Image standards 10

2.1.4 Multimedia presentation standards: a brief overview 11

2.2 Multimedia metadata standards overview 12

3. Multimedia data and metadata standards overview in the NoE 34

3.1 Analysis of the results 35

4. Standardised Metadata frameworks 36

4.1 MPEG-21: a brief overview 36

4.2 XML technologies and metadata, semantic web and interoperability 41

4.2.1 Extensible Markup Language (XML) and metadata 41

4.2.2 Semantic web and interoperability 43

5. Conclusions 47

A. Questionnaire 49

B. Standardization Bodies 55

C. Reference Projects 56

D. Contributing Partners 58

E. Next steps within Muscle 59

References 60

Bibliography 64

Chapter 1

Introduction

In the last years multimedia resources have been used in a lot of applications paying attention not only on the traditional highly professional markets and the gaming enhancement field. Multimedia (MM) data, in the form of still pictures, graphics, 3D models, audio, speech, video and such combination of these (e.g. MM presentations) are going to play a gradually more important role in our life. So the need to enable computational interpretation and processing of such data and resources and also share and exchange them across the network is widely growing on.
Internet MM communications, such as video and audio on demand, video conferencing and distance e-Learning, give us an idea of the growing diffusion of MM data. Related to this situation, much effort has been put in developing standards for *co*ding and *dec*oding MM data. In fact, understanding that most of the MM data are redundant, MM codecs use compression algorithms to identify and use redundancy, offering the possibility to efficiently exchange data across networks.
Moreover, the rapid expansion of Internet has caused a growing demand for systems and tools which can satisfy the more sophisticated requirements for storing, managing, searching, accessing, retrieving and sharing complex resources having many different formats and available on several media types.

A multimedia system is generally composed of different components (Fig. 1): database, multimedia storage server, network and client systems in a more and more mobile environment. Moreover, new standardised initiatives try to bind these components together. For example, there are the new and emerging standards by ISO/IEC JTC 1/SC 29/WG 11 MPEG (Moving Picture Experts Group), that is MPEG-4, MPEG-7 and MPEG-21 standards. They offer standardized technology for coding MM data, natural and synthetic (e.g., photography, face animation), continuous and static (e.g., video, image), as well as for describing content (*metadata*) and for open MM frameworks for a reasonable and interoperable use of MM data in a distributed environment [18].

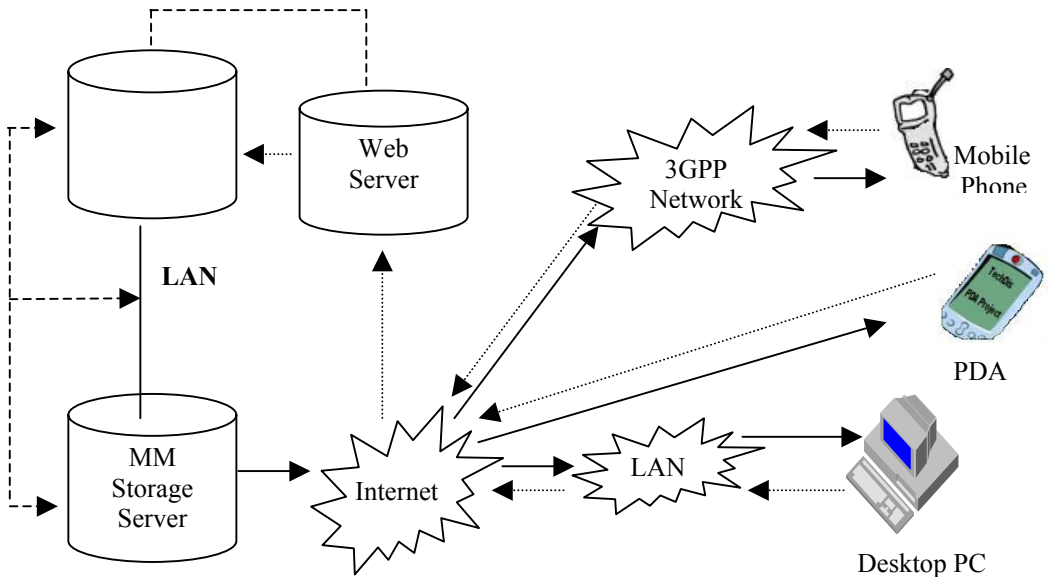

Fig. 1 Structure and information flow of a distributed multimedia system

Metadata have a relevant role in this context, representing the value-added information that describes the administrative, descriptive, preservation and technical characteristics associated with MM resources.

The use of metadata in MM distributed systems provides many advantages, as the possibility to allow searching for MM data by content. Finding multimedia objects by their content in a distributed database means searching them on the base of content descriptions and similarity measures. For example, it could be possible to list all videos from an on-line database in which there is a specified actor or song.
Another use of metadata is oriented to describe environment characteristics (usage or representation preferences) and network constraints. For instance, we can consider the situation in which a user is looking for all the soccer events of a specific weekend but the available bandwidth for his terminal is limited. The database has to search not only videos related to the specified events but also it should consider the bandwidth constraints. In this case, it will be more efficient offering alternatives like to show only key-images extracted from the video.
Metadata is also used to describe intellectual properties that may guarantee a reasonable use of data above all in commercial applications.

Metadata information can be automatically or manually extracted from MM documents (video, audio, audio-visual documents, MM presentations, etc) also considering annotations.
Because of the high cost and subjectivity associated with human-generated metadata, a large number of research initiatives is focusing on technologies to enable automatic classification and segmentation of digital resources (e.g., automatic generation of metadata for textual documents, images, audio and video resources).

Many consortia are working on a number of projects in order to define new MM (meta-)data standards. A reference list is shown in Appendix B.

One of the more recent approach is to combine a specific MM metadata standard with other standards useful to describe other application domains, in order to have a more complete characterisation of the specific problem without creating a new standard.
New metadata initiatives such as TV-Anytime [7], MPEG-21 [23], NewsML [9], and several communities (museums, education, medicine and others), want to combine MPEG-7 MM descriptions with new and existing metadata standards for simple resource discovery (Dublin Core [11]), rights management (INDECS [12]), geo-spatial (FGDC [13]), educational (GEM [14], IEEE LOM [15]) and museum (CIDOC CRM [16]) content, to satisfy their domain-specific requirements.
In order to do this, it is necessary to have a common understanding of the semantic relationships between metadata terms from different domains. To this purpose, XML schema provides a first support for expressing semantic knowledge and RDF schema can provide a way to do this [17], even if there are also other frameworks able to realise the same task.
Among these new initiatives, the perspective of MPEG-21 standard has been analysed. The vision for MPEG-21 is to define a MM framework to enable transparent and augmented use of MM resources across a wide range of networks and devices used by different communities. Considering the *Digital Item* as the fundamental unit of distribution and transaction (*structured digital objects, including a standard representation and identification, and metadata*), MPEG-21 allows users to exchange, access, consume, trade and otherwise manipulate *Digital Items* in efficient, transparent and interoperable way.

This document gives an overview of the most significant part of the most important MM data and metadata standards considering both the context inside and outside the *Network of Excellence* (NoE).

In chapter 2, Section 2.1 analyses the state of the art on MM data standards, subdividing them in four categories: video, audio, image and MM presentation standards.
Section 2.2 presents an overview of the most well known MM metadata standards. Also the general concepts of MM metadata are discussed, focusing our attention mainly on a comparative analysis rather then on a complete and detailed description. Among a wide range of standards, particular attention is put on MPEG family (MPEG-7 and MPEG-21) because of their description and representation power of the MM objects, jointly with their large diffusion, the possibility of extensions for different specific domains and the availability of some tools for automatic metadata extraction.

Chapter 3 is focused on the description of data and metadata standards used inside the NoE. These data have been collected distributing to all the partners a questionnaire specifically designed (see Appendix A). The questionnaire was set up as a fundamental tool to acquire information useful for defining a future strategy for the construction of a common representative model of the MM objects preserved by each partner of the NoE. Besides, we considered this census as a first step to reach the objectives pointed out, covering the interoperability needs of the NoE.

Chapter 4 gives an overview of the main web standard technologies, like XML and RDF, able to describe metadata and define a way to achieve semantic web and interoperability goals, also providing a synthetic overview about ontologies.

Finally, Chapter 5 takes conclusions from all the above mentioned and analysed subjects, trying to suggest a possible way to follow in order to satisfy the NoE sharing, exchanging and interoperability needs.

The questionnaire sent to all the organizations of the NoE is available in Appendix A.

A list of reference consortia working on MM (meta-)data standards is shown in Appendix B.

Appendix C shows a list of international reference projects regarding MM metadata, also including initiatives oriented to interoperability aspects.

Appendix D presents a list of partners who have given specific contributions to the state of the art inside the NoE.

Appendix E gives a brief overview of the next steps that we consider fundamental to carry on successfully this activity in the next future.

Finally, Bibliographic references are listed in the References and Bibliography Sections.

Chapter 2

State of the Art on Multimedia data and metadata standards

2.1 Multimedia data standards overview

Standardization bodies continue to work on media standards in order to provide a common approach to enable interoperability, better quality and efficiency under specific constraints.
The result of recent advances in hardware and networking is that multimedia applications are becoming mainstream, spanning a large spectrum of consumer applications. Examples of technologies used in such applications are" image post-processing, video processing and indexing, speech recognition, speech synthesis, and music authoring. For this new trend, applications should support a wide spectrum of commonly used media formats to succeed.
In recent years there has been a wide proliferation of MM standards, the major part of which can be grouped as follows:

- *Video:* in this category we can remember the MPEG-1, MPEG-2, MPEG-4, QuickTime, Sony DV, AVI, ASF, Real-Media, …
- *Audio* : among the most known standards we can remember Raw PCM, WAV, MPEG-1, MP3, GSM, G.723, ADPCM
- *Image:* the most diffuse image standards are JPEG, TIFF, BMP, GIF
- *MM Presentations:* among these standard types we can cite SMIL and MHEG

A brief overview of reference standards is analysed in the sections below, subdividing them into the three categories aforementioned.

2.1.1 Video standards

2.1.1.1 MPEG family

MPEG-1

In development for years, MPEG-1 became an official standard for encoding audio and video in 1993. It can be described as the simplest of the MPEG standards, it describes a way to encode audio and video data streams, along with a way to decode them. The default size for an MPEG1 video is 352x240 at 30fps for NTSC (352x288 at 25fps for PAL sources). These were designed to give the correct 4:3 aspect ratio when displayed on the rectangular pixels of TV screens. For a computer-based viewing audience, 320x240 square pixels gives the same aspect ratio. Good up to about 1.5Mbps, MPEG1 delivers roughly VHS quality at 30 frames per second. You can scale up or down in size or bit-rate, but from 1.2-1.5Mbps is the sweet spot where you'll get the most bang for your bit-rate buck.

MPEG-2

The MPEG-2 standard builds upon MPEG-1 to extend it to handle the highest-quality video applications. It is a common standard for digital video transmission at all parts of the distribution chain. Broadcast distribution equipment, digital cable head-ends, video DVDs, and satellite television all employ MPEG-2.
You'll need special capture cards to encode MPEG-2 in real-time on a PC. But in the streaming world, MPEG-2 is a great source from which to trans-code the various Real, WindowsMedia and Quicktime formats we serve to our viewers.
MPEG-2 needs about 6Mbps to provide the quality you're used to seeing on movie DVDs, although data rates up to 15Mbps are supported. 720X480 is the typical 4:3 default resolution, while 1920x1080 provides support for 16:9 high-definition television.

MPEG-4 : Internet Streaming and Synchronized Multimedia

Where MPEG-2 was designed to scale up to broadcast and high-definition quality and operating requirements, MPEG-4 goes the other way. It's designed to scale down to dial-up internet bandwidths and to tiny devices like cell phones and PDAs; as well as still remain viable for high-quality desktop streaming up to 1Mbps.
But MPEG-4 is much more than just an audio and video compression/decompression scheme. It's a container for all kinds of media objects (images, text, video, animation, interactive elements like buttons and image maps, etc) and a way to choreograph them into a synchronized, interactive presentation. MPEG-4 also has standard interfaces to allow plugging in a DRM scheme called Intellectual Property Management and Protection (IPMP).
MPEG-4 is still at the frontier of media technologies. The specification is extensive, and each vendor implements it in their own way. Try a variety of MPEG-4 tools and you'll find lots of incompatibilities. But some are working to smooth the landscape. The Internet Streaming Media Association (ISMA) [71] is an industry consortium dedicated to interoperability among MPEG-4 products and services. Essentially, any implementation that's ISMA-compliant will work with any other.

2.1.1.2 Other Video Standards

AVI

A format developed by Microsoft Corporation for storing video and audio information is *AVI format* (Audio Video Interleave). It is limited to 320 x 240 resolution and 30 frames per second, neither of which is adequate for full-screen, full-motion video. However, AVI video does not require any special hardware, making it the lowest common denominator for MM applications. Many MM producers use this format because it allows them to sell their products to the largest base of users.

Quicktime

A competing video format is *QuickTime*, which is a video and animation system developed by Apple Computer. QuickTime is built into the Macintosh operating system and is used by most Mac applications that include video or animation. PCs can also run files in QuickTime format, but they require a special QuickTime driver. In February 1998, the ISO standards body gave Quicktime a boost by deciding to use it as the basis for the new MPEG-4 standard.

2.1.2 Audio standards

2.1.2.1 MP3 is not MPEG3

It's the magical ability to squeeze the 1.4 Mbps audio stream from a standard audio CD down to a sweet-sounding 128kbps that has made MP3 the de facto standard for digital music distribution. You can find MP3 support in every major media player on every computer platform, and dozens of consumer electronic devices can play MP3s. It's as close to a universal format for audio as you'll find. MP3 is actually part of the MPEG1 standard. The audio portion of the MPEG1 spec contains three different compression schemes called layers. Of the three, Layer 3 provides the greatest audio quality and the greatest compression. At 8kbps, MP3 will sound like a phone call intelligible, but nothing you'd ever call high-fidelity. Good-quality music starts at about 96kbps, but generally you'll want 128 or 160kbps to get "CD quality" reproduction.

2.1.2.2 Some other audio standards

PCM

Short for Pulse Code Modulation, PCM is a sampling technique for digitising analogue signals, especially audio signals. PCM samples the signal 8000 times a second; each sample is represented by 8 bits for a total of 64 Kbps. Since it is a generic format, it can be read by most audio applications. Similar to the way a plain text file can be read by any word-processing program. PCM is used by Audio CDs and digital audio tapes (DATs). It is also a very common format for AIFF and WAV files.

ADPCM

Short for Adaptive Differential Pulse Code Modulation, ADPCM is a form of pulse code modulation (PCM) that produces a digital signal with a lower bit rate than standard PCM. It produces a lower bit rate by recording only the difference between samples and adjusting the coding scale dynamically to accommodate large and small differences. It works by analysing a succession of samples and predicting the value of the next sample. It then stores the difference between the calculated value and the actual value. Some applications use ADPCM to digitise a voice signal so voice and data can be transmitted simultaneously over a digital facility normally used only for one or the other.

WAV (RIFF)

WAVE format, the Microsoft WAV sound file format, is derived from the RIFF (Resource Interchange File Format). The WAV files can be recorded at 11kHz, 22kHz, and 44kHz, in 8 or 16-bit mono and stereo formats. A WAV file consists of three elements: a header, audio data, and a footer. The header is mandatory and contains the specifications for the file (information on interpreting the audio data) and optional material including copyright. The audio data are in the format specified by the header. The footer is optional and, if present, contains other annotation. Usually, the data in a WAV file take the form of PCM bit streams.

AIF (AIFF)

The Audio Interchange File Format (AIFF) was developed by Apple computer to store high-quality sampled sound and musical instrument information. AIF is a popular file format for transferring files between the Mac and the PC. This format supports 8-bit files only; mono up to 44.1 KHz, and stereo up to 22 KHz.

2.1.3 Image standards

JPEG : Joint Photographic Experts Group

In general, what people usually mean when they use the term "JPEG" is the image compression standard they developed. JPEG was developed to compress still images, such as photographs, a single video frame, something scanned into the computer, and so on. You can run JPEG at any speed that the application requires. For a still picture database, the algorithm doesn't have to be very fast. If you run JPEG fast enough, you can compress motion video -- which means that JPEG would have to run at 50 or 60 fields per second. This is called motion JPEG or M-JPEG. You might want to do this if you were designing a video editing system. Now, M-JPEG running at 60 fields per second is not as efficient as MPEG-2 running at 60 fields per second because MPEG was designed to take advantage of certain aspects of motion video.

BMP

A representation, consisting of rows and columns of dots, of a graphics image in computer memory. The value of each dot (whether it is filled in or not) is stored in one or more bits of data. For simple monochrome images, one bit is sufficient to represent each dot, but for colours and shades of grey, each dot requires more than one bit of data. The more bits used to represent a dot, the more colours and shades of grey that can be represented.
The density of the dots, known as the resolution, determines how sharply the image is represented. This is often expressed in dots per inch (dpi) or simply by the number of rows and columns, such as 640 x 480.
Bit-mapped graphics are often referred to as raster graphics. The other method for representing images is known as vector graphics or object-oriented graphics. With vector graphics, images are represented as mathematical formulas that define all the shapes in the image. Vector graphics are more flexible than bit-mapped graphics because they look the same even when you scale them to different sizes [http://www.webopedia.com].

GIF

Short for Graphics Interchange Format, another of the graphics formats supported by the Web. Unlike JPG, the GIF format is a loss less compression technique and it supports only 256 colours. GIF is better than JPG for images with only a few distinct colours, such as line drawings, black and white images and small text that is only a few pixels high. With an animation editor, GIF images can be put together for animated images. The compression algorithm used in the GIF format is owned by Unisys, and companies that use the algorithm are supposed to license the use from Unisys [http://www.webopedia.com].

PNG

Short for Portable Network Graphics, the third graphics standard supported by the Web (though not supported by all browsers). PNG was developed as a patent-free answer to the GIF format but is also an improvement on the GIF technique. An image in a loss less PNG file can be 5%-25% more compressed than a GIF file of the same image. PNG builds on the idea of transparency in GIF images and allows the control of the degree of transparency, known as opacity. Saving, restoring and re-saving a PNG image will not degrade its quality. PNG does not support animation like GIF does [http://www.webopedia.com].

TIFF

Acronym for Tagged Image File Format, one of the most widely supported file formats for storing bit-mapped images on personal computers (both PCs and Macintosh computers). What made the TIFF so different was its tag-based file structure. Where the BMP is built on a fixed header with fixed fields followed by sequential data, the TIFF has a much more flexible structure. At the beginning of each TIFF is a simple 8-byte header that points to the position of the first Image File Directory (IFD) tag. This IFD can be of any length and contain any number of other tags enabling completely customised headers to be produced. The IFD also acts as a road map to where image data is stored in the file as the tagged nature of the format means that this needn't be stored sequentially. Finally the IFD can also point to another IFD as each TIFF can contain multiple sub files [http://www.webopedia.com].

2.1.4 Multimedia presentation standards: a brief overview

Several cross platform video and audio standards have been established including still and motion JPEG, and a number of different MPEG standards. So far, there has been no standard method of bringing all these formats together to produce MM presentations. Several models aim to solve this by providing a system independent presentation standard for hardware and software engineers and presentation authors to conform to. In this way, a presentation created on one hardware platform should be viewable on others.

SMIL (pronounced smile) stands for Synchronized Multimedia Integration Language [20] It is a mark-up language, like HTML and is designed to be very easy to learn and deploy on Web sites. Recommended from the World Wide Web Consortium (W3C) it allows developers to create time-based multimedia documents on the web. Based on XML, it is able to mix many types of media, text, video, graphics, audio and vector based animation together and to synchronize them according to a timeline.
Some of the main features of this standard can be listed as follows:

- The presentation is composed from several components that are accessible via URIs, e.g. files stored on a Web server.

- The components have different media types, such as audio, video, image or text. The begin and the end time of different components are specified according to events in other media components. For example, in a slide show, a particular slide is displayed when the narrator in the audio starts talking about it.

- Familiar looking control buttons such as stop, fast-forward and rewind allow the user to interrupt the presentation and to move forwards or backwards to another point in the presentation.

- Additional functions are "random access", i.e. the presentation can be started anywhere, and "slow motion", i.e. the presentation is played slower than at its original speed.

- The user can follow hyperlinks embedded in the presentation.

MHEG is an abbreviation for the Multimedia and Hypermedia Experts Group [19]. This is another group of specialists, eminent in their field which has been set up by ISO, the International Standards Organisation. This group had the task of creating a standard method of storage, exchange and display of MM presentations. In particular we can distinguish between MHEG-5 and MHEG-4.
The first one allows us to manage MM applications across computer networks; for each MM object it doesn't defines a compression scheme, each object has an own compression standard, while the last one doesn't defines tools to create a multimedia structure but it is able to combine a multimedia information stream in time by integrating it with different components as text, video, images each compressed in a specific way according to the media which it represents.
The aim of the standard MHEG-5 consists in defining an object-oriented model to codify the synchronization of the multimedia objects in a standard way. The synchronization regards not only the objects themselves (the activation of a musical piece together with the end of a video film) but also events generated by users (the press of such a bottom using mouse) and temporal events (as an example after one minute from the visualization of an image an audio comment will be activated).

Its basic goals are:

- To provide a simple but useful, easy to implement framework for multimedia applications using the minimum system resources

- To define a digital final form for presentations, which may be used for exchange of the presentations between different machines no matter what make or platform

- To provide extensibility i.e. the system should be expandable and customisable with additional application specific code, though this may make the presentation platform dependent

2.2 Multimedia metadata standards overview

Metadata is the value-added information which documents the administrative, descriptive, preservation, technical and usage history characteristics associated with resources. It provides the underlying foundation upon which digital resources management systems are based to provide fast, precise access to relevant resources across networks and between organizations.

Multimedia content analysis refers to understanding semantic meanings of MM documents through metadata extracted using common techniques of image and signal processing and image analysis and understanding.
Metadata is an important aspect of the creation and management of digital images and other MM files. The information contained in the metadata standards can regards the following aspects:

- the technical format of the image file
- the process by which the image was created
- the content of the image

Following these standards helps organizations to consistently record information about their MM documents in a way that facilitates retrieval and sharing in a networked environment.

Metadata for MM documents can be classify according the following three types:

1.*Descriptive or Content metadata*: is information about the object captured in the document (the object's name, title, materials, dates, physical description, etc..). Content metadata is very important, as it is the main way by which people can search and retrieve the MM documents from a database. There are standards available to assist in determining what information should be recorded about the object, and how to record it.

2.*Technical metadata*: is also essential to properly manage digital images. Technical metadata is data about the MM document itself (not about an object in the document). For example, for a digital image, it can include information about: the technical processes used in image capture or manipulation or colour or file formats, and some of the technical information that is recorded about the image, such as the image file type, must be machine-readable (following specific technical formats) in order for a computer system to be able to properly display the image.

3.*Administrative metadata* includes information related to the management of MM documents (such as rights management).

MM Metadata can be also classified according to other criteria considering the level of data description, the producibility and the domain dependence [4]. They can be classified by:

- *Level*: we can distinguish between a *technical level* in which lower level aspects of the multimedia content is described and a *semantic level* in which aspects of higher level of abstraction on the multimedia content are taken into account.

- *Producibility*: the production of metadata can either be automatic which is a very desirable property from the economic point of view and regards more frequently the low level technical metadata; for semantic metadata describing the information covered by multimedia content it is typically required human knowledge. So in these cases the metadata production is manually performed.

- *Dependencies*: metadata can be domain-dependent, for instance the position of a tumour can be interesting for medical applications, while the colour distribution of an image can be useful for many application domains. Metadata can also be media type-dependent considering for instance the colour distribution as applicable only to visual media while the creation date applicable to any media.

Metadata represents surely a gain in terms of benefits produced for MM data descriptions but most of all for MM applications (content analysis).
There are also disadvantages related to metadata. Some of them are its cost, its unreliability, its subjectivity, its lack of authentication and its lack of interoperability with respect to syntax, semantics, vocabularies and languages. However, there are many researchers currently investigating strategies to overcome different aspects of these limitations in an effort to provide more efficient means of organizing content on the Internet.

The main reference topics related to techniques and projects developed for multimedia content analysis can be briefly summarised as follows [3]:

- *Automatic document Indexing/Classification*: there is a variety of techniques to classify documents in subject categories. These techniques include: Bayesian analysis of the patterns of words in the document, clustering of sets of documents according to similarity measures, neural networks, sophisticated linguistic inferences, the use of pre-existing sets of categories and seeding categories with keywords. The most common methods used by auto-categorization software are based on scanning every word in a document and analysing the frequencies of patterns of words and, based on a comparison with an existing taxonomy, assigning the document to a particular category in the taxonomy. Other approaches use "clustering" and "taxonomy building" techniques searching through all combinations of words to find cluster of documents that appear to be together. Some systems are capable of automatically generating a summary of a document by scanning through the document and finding important sentences using rules like: the first sentence of the first paragraph is often important.

- New researches are focusing on *Semantics-Sensitive Matching* [1] and A*utomatic Linguistic Indexing* in which the system is capable of recognizing real-world objects or concepts [2]. Image retrieval research has moved on from the IBM QBIC (query by image content) system (QBIC, 2001) which uses colours, textures, and shapes to search for images [8]. In particular the IBM CUEVideo project [51] combines video and audio analysis, speech recognition, information retrieval and artificial intelligence.

- *Speech recognition* is increasingly being applied to the indexing and retrieval of digitised speech archives. Speech recognition systems can generate searchable text that is indexed to time code on the recorded media, so users can both call up text and jump right to the audio clip containing the keyword. Normally, running a speech recogniser on audio recordings does not produce a highly accurate transcript because speech-recognition systems have difficulty if they haven't been trained for a particular speaker or if the speech is continuous. However the latest speech recognition systems will work even in noisy environments, are speaker-independent, work on continuous speech and are able to separate two speakers talking at once.

- *Video Indexing and retrieval*: the latest video indexing systems combine a number of indexing methods, embedded textual data, scene change detection, visual clues and continuous-speech recognition to convert spoken words into text. Some systems [52] can automatically analyse videos and extract named entities from transcripts which can be used to produce time and location metadata. This metadata can then be used to explore archives dynamically using temporal and spatial graphical user interfaces.

- The *Annotation* systems also represent an useful tool for metadata extraction. The motivation behind annotation systems is related to the problem of metadata trust and authentication. Users can attach their own metadata, opinions, comments, ratings and recommendations to particular resources or documents on the Web, which can be read and shared with others. The basic philosophy is that we give more probably value and trust to the opinions of people we respect than metadata of unknown origin. The W3C's Annotea system [55] and DARPA's Web Annotation Service [50] are two web-based annotation systems which have been developed. Other annotation tools for film/video and multimedia content (IBM VideoAnnEx, 2001) [53], (Ricoh MovieTool, 2002) [54], (DSTC's FilmEd,

2003) [57] and tools to enable the attachment of spoken annotations to digital resources (PAXit, 2003) [56] such as images or photographs have been developed.

- *Metadata for Preservation* : a lot of initiatives are focusing on metadata pursuing the goal of the multimedia resource preservation. Such initiatives include: Reference Model for an Open Archival Information System (OAIS, 2002) [58], the CURL Exemplars in Digital Archives project (CEDARS, 2002) [59], the National Library of Australia (NLA) PANDORA project (PANDORA, 2002) [60]. These initiatives rely on the preservation of both digital objects and associated metadata for an easy interpretation in the future. The preservation metadata provides sufficient technical information about the resources and can facilitate the long-term access of the digital resources by providing a complete description of the technical environment needed to view the work, the applications and version numbers needed, decompression schemes as well as any other files that need to be linked to it.

A large number of metadata standardisation initiatives has been developed in recent years, in order to describe multimedia contents in so many different domains and to grant sharing, exchanging and interoperability across wide range networks.
We can distinguish between two different standard typology, according to what each of them represents in terms of its functionalities.
The first typology is directly related to the representation of multimedia content for a specific domain and each of these standards can be referred as a *standardised description scheme*, while the second one considers the possibility of integrating more metadata standards mapped on different application domains, providing rich metadata models for media descriptions together with languages allowing one to define other description schemes for arbitrary domains. These last standards can be referred as *standardised metadata frameworks* and have been shortly described in the Chapter 4.

Table 1 represents a selection of several metadata standards description schemes, which can be considered the most frequently cited and representative for a quite wide range of different application domains. So, in table 1 is illustrated a list of descriptive characteristics of each reference standard taken in account. In particular, information about standardisation bodies, last version dates, described MM data types, application domains, description semantic levels and the way by which metadata has been produced (manually or automatically) have been considered.

A short quick overview of the meaningful metadata standards, schematically described in the table 1, is proposed in the next sections.

	MARC	Dublin Core	CDWA	VRA Core	CSDGM	Z39.87	LOM	DIG35	METS	JPX	SMPTE Metadata Dictionary
Standardization Body	Library of Congress	Dublin Core Metadata Initiative (DCMI)	Art Information Task Force (AITF)	Visual Resource Association	Federal Geographic Data Committee (FGDC)	National Information Standard Organization (NISO)	IEE (LTSC)	Digital Imaging Group (DIG of I3A)	Digital Library federation (DLF)	Joint Photographic Experts Group (JPEG)	Society of Motion Picture and Television Engineers (SMPTE)
Year	Current version MARC 21 since 1999	Current version 1.1 since 1999	Current version 2.0 since 2000	Current version 3.0 since 2002	Update version since from 1998	2002	2002	Current version 1.1 April 2001	Last review 2001	2000	Last review 2004
MM Type	Any	Any	Any	Images	Any	Images	Any	Images	Any	Images	Any
Domain	Bibliographic media description	Bibliographic media description	Description of Art works	Description of images of Art works	Description of Geographic media	Description of still images	Description of educational media	Description of digital images	Description of digital objects	Description of digital images	Description of audio/video documents
Level	Largely semantic	Largely semantic	Largely semantic	Largely semantic	Semantic and technical	Technical	Largely semantic	Semantic and technical	Semantic and technical	Semantic and technical	Semantic and technical
Producibility	Mainly manual	Mainly manual	Mainly manual	Mainly manual	Manual and Automatic	Mainly automatic	Mainly manual	Mainly manual	Mainly manual	Mainly manual	Manual and Automatic

Table 1. Selection of several standardised description schemes.

MARC

MARC 21 [63] is an implementation of the American national standard, Information Interchange Format (ANSI Z39.2) and its international counterpart, Format for Information Exchange (ISO 2709). These standards specify the requirements for a generalized interchange format that will accommodate data describing all forms of materials susceptible to bibliographic description, as well as related information such as authority, classification, community information, and holdings data. The standards present a generalized structure for records, but do not specify the content of the record and do not, in general, assign meaning to tags, indicators, or data element identifiers. Specification of these elements are provided by particular implementations of the standards. The MARC formats are defined for five type of data, in particular for data listed below:

- *Bibliographic Data*: contain format specifications for encoding data elements needed to describe, retrieve, and control various forms of bibliographic material. It is an integrated format defined for the identification and description of different forms of bibliographic material. MARC 21 specifications are defined for books, serials, computer files, maps, music, visual materials, and mixed material.

- *Holdings Data:* contain format specifications for encoding data elements pertinent to holdings and location data for all forms of material.

- *Authority Data*: contain format specifications for encoding data elements that identify or control the content and content designation of those portions of a bibliographic record that may be subject to authority control.

- *Classification Data:* contain format specifications for encoding data elements related to classification numbers and the captions associated with them. Classification records are used for the maintenance and development of classification schemes.

- *Community Information:* provide format specifications for records containing information about events, programs, services, etc. so that this information can be integrated into the same public access catalogues as data in other record types

The MARC 21 formats are communication formats, primarily designed to provide specifications for the exchange of bibliographic and related information between systems. They are widely used in a variety of exchange and processing environments. As communication formats, they do not consent internal storage or display formats to be used by individual systems.

Dublin Core

The Dublin Core Metadata Initiative (DCMI) [11] began in 1995 with an invitational workshop in Dublin, Ohio that brought together librarians, digital library researchers, content providers and text-mark-up experts to improve discovery standards for information resources. The original Dublin Core merged as a small set of descriptors that quickly drew global interest from a wide variety of information providers in the arts, sciences, education, business, and government sectors.
The Dublin Core is not intended to displace any other metadata standard. Rather it is intended to co-exist, often in the same resource description, with metadata standards that offer other semantics. In fact, on one hand simplicity allows the cost of creating metadata to reduce and promotes interoperability, while on the other hand, simplicity does not accommodate the semantic and functional richness supported by complex metadata schemes.
The Dublin Core metadata element set is a set of 15 descriptors which can be briefly listed below:

- *Name* : The label assigned to the data element
- *Identifier* : The unique identifier assigned to the data element
- *Version* : The version of the data element
- *Registration Authority* : The entity authorised to register the data element
- *Language* : The language in which the data element is specified
- *Definition* : A statement that clearly represents the concept and essential nature of the data element
- *Obligation* : Indicates if the data element is required to always or sometimes be present (contain a value)
- *Data-type* : Indicates the type of data that can be represented in the value of the data element
- *Maximum Occurrence* : Indicates any limit to the repeatability of the data element
- *Comment* : A remark concerning the application of the data element

The design of Dublin Core consists in encouraging the use of richer metadata schemes in combination with itself. Richer schemes can also be mapped to Dublin Core for export or for cross-system searching. On the other hand, simple Dublin Core records can be used as a starting point for the creation of more complex descriptions.

CDWA

CDWA stands for Categories for the Description of Works of Art [64], a metadata schema designed by of the Art Information Task Force (AITF), to "describe the content of art databases by articulating a conceptual framework for describing and accessing information about objects and images." It was released in February 1996 and its last version dates back to September 2000.
This Metadata schema is very extensive and developed for use by art specialists. There are 26 main categories, and each category has its own set of subcategories. All the categories are fit into two groups:

- "Object, Architecture, or Group" as the information intrinsic to the work
- "Authorities/Vocabulary Control" as the information extrinsic to the work.

It was formulated with needs of academic researchers and represents the minimum information required to completely describe a particular work of art or museum object.

VRA Core

VRA Core Categories is the Visual Resources Association's approach of categorizing visual documents that represent objects of art or architecture [66]. The VRA Core Categories provides a template designed but not limited to visual works collections. As VRA Data Standard Committee pointed, "CDWA was exhaustive in its list of elements needed to describe museum objects, it was not entirely satisfactory for the description of images, and in particular, did not cover all of the elements needed for the description of architecture and other site-specific works", there is a need to expand the concept to non-art objects and visual document for which VRA was created. Compared with and Benefit from CDWA, VRA Core categories is designed to cover most visual materials. It does not have such comprehensive categories as CDWA. Similar to Dublin Core, It provides a core set of elements, which could be expanded by adding new elements as needed. It contains 17 categories that can be used to describe both work and representations of the work, defined as images.
These categories are listed below:

- *Record Type*: identifies the record as being either a WORK record, for the physical or created object, or an IMAGE record, for the visual surrogates of such objects.
- *Type:* identifies the specific type of Work or Image being described in the record.

- *Title:* the title or identifying phrase given to a Work or an Image. For an Image record this category describes the specific view of the depicted Work.

- *Measurements:* the size, shape, scale, dimensions, format, or storage configuration of the Work or Image. Dimensions may include such measurements as volume, weight, area or running time. The unit used in the measurement must be specified.

- *Material:* The substance of which a work or an image is composed.

- *Technique:* the production or manufacturing processes, techniques, and methods incorporated in the fabrication or alteration of the work or image.

- *Creator*: the names, appellations, or other identifiers assigned to an individual, group, corporate body, or other entity that has contributed to the design, creation, production, manufacture, or alteration of the work or image

- *Date:* date or range of dates associated with the creation, design, production, presentation, performance, construction, or alteration, etc. of the work or image.

- *Location:* the geographic location and/or name of the repository, building, or site-specific work or other entity whose boundaries include the Work or Image.

- *ID number*: The unique identifiers assigned to a Work or an Image.

- *Style/ Period:* A defined style, historical period, group, school, dynasty, movement, etc. whose characteristics are represented in the Work or Image.

- *Culture:* the name of the culture, people, or adjectival form of a country name from which a Work or Image originates or with which the Work or Image has been associated.

- *Subject:* terms or phrases that describe, identify, or interpret the Work or Image and what it depicts or expresses. These may include proper names (e.g., people or events), geographic designations (places), generic terms describing the material world, or topics (e.g., iconography, concepts, themes, or issues).

- *Relation:* terms or phrases describing the identity of the related work and the relationship between the Work being catalogued and the related work. Note: If the relationship is essential (i.e. when the described work includes the referenced works, either physically or logically within a larger or smaller context), use the Title.Larger Entity element.

- *Description:* a free-text note about the Work or Image, including comments, description, or interpretation, that gives additional information not recorded in other categories.

- *Source:* a reference to the source of the information recorded about the work or the image. For a work record, this may be a citation to the authority for the information provided. For an image, it can be used to provide information about the supplying Agency, Vendor or

Individual; or in the case of copy photography, a bibliographic citation or other description of the image source. In both cases, names, locations, and source identification numbers can be included.

- *Rights:* information about rights management; may include copyright and other intellectual property statements required for use.

CSDGM

The objectives of the standard are to provide a common set of terminology and definitions for the documentation of digital geo-spatial data [13].
This standard is intended to support the collection and processing of geo-spatial metadata. It is intended to be useable by all levels of government and the private sector.
The standard establishes the names of data elements and compound elements (groups of data elements) to be used for these purposes, the definitions of these compound elements and data elements, and information about the values that are to be provided for the data elements.

The main classes representing the elements can be briefly described as follows:

- Identification Information
- Data Quality Information
- Spatial Data Organization Information
- Spatial Reference Information
- Entity and Attribute Information
- Distribution Information
- Metadata Reference Information
- Citation Information
- Time period Information
- Contact Information

NISO Z39.87

The purpose of NISO Z39.87 is to define a standard set of metadata elements for digital images [65]. Standardizing the information allows users to develop, exchange, and interpret digital image files. It has been designed to facilitate interoperability between systems, services, and software, as well as to support the long-term management of and continuing access to digital image collections.
The design objectives of this NISO initiative are to define a metadata set that interoperates with and meets the goal outlined by the DIG35 metadata standard. To that end, the NISO group has adapted the original DIG35 goals as follows:

1. *Interchangeable*: The NISO metadata set is based on a sound conceptual model that is both generally applicable to many applications and assured to be consistent over time.

2. *Extensible and scalable*: The NISO metadata set enables application developers and hardware manufacturers to utilize additional metadata fields. This allows future needs for metadata to be fulfilled with limited disruption of current solutions.

3. *Image file format independent*: The NISO metadata set does not rely on any specific file format and can therefore be supported by many current and future file formats and compression mechanisms.

4. *Consistent*: The NISO metadata set works well with existing standards and it is usable in a variety of application domains and user situations.

5. *Network-ready*: The NISO metadata set provides seamless integration with a broad variety of systems and services. Integration options include database products and the utilization of XML schemas (the recommended implementation method).

The main categories of metadata elements can be summarised and organised as follows:

- *Basic Image Parameters:* Format, File Size, Compression, Colour Space, etc.

- *Image Creation:* Source Type, Producer, Scanning System for Capture, Digital Camera Set-up for Capture.

- *Imaging Performance Assessment:* Spatial metrics (sampling frequency, image dimensions, source dimensions), Energetics (bits per sample, colour map, white point, etc), target data (type, image data, performance data).

- *Change History:* Image processing (date, source data, software), previous image metadata.

LOM

This Standard is a multi-part standard that specifies Learning Object Metadata, trough a conceptual data schema that defines the structure of a metadata instance for a learning object [15].
For this Standard, a learning object is defined as any digital or non-digital entity that may be used for learning, education or training. A metadata instance for a learning object describes relevant characteristics of the learning object to which it applies. Data elements describe a learning object and are grouped into *categories*. The LOM data model is a hierarchy of data elements, including aggregate data elements and simple data elements, the leaf nodes of the hierarchy.
Its schema consists of nine such categories, each of them structured following a hierarchic architecture:

1. The *General* category groups the general information that describes the learning object as a whole.

2. The *Life cycle* category groups the features related to the history and current state of this learning object and those who have affected this learning object during its evolution.

3. The *Meta-Metadata* category groups information about the metadata instance itself (rather than the learning object that the metadata instance describes).

4. The *Technical* category groups the technical requirements and technical characteristics of the learning object.

5. The *Educational* category groups the educational and pedagogic characteristics of the learning object.

6. The *Rights* category groups the intellectual property rights and conditions of use for the learning object.

7. The *Relation* category groups features that define the relationship between the learning object and other related learning objects.

8. The *Annotation* category provides comments on the educational use of the learning object and provides information on when and by whom the comments were created.

9. The *Classification* category describes this learning object in relation to a particular classification system.

DIG35

The focus of the DIG35 Initiative Group is on defining metadata standards [67].
By establishing standards, the Initiative Group seeks to overcome a variety of challenges that have arisen as the sheer volume of digital images being used has increased. Among these are efficiently archiving, indexing, cataloguing, reviewing, and retrieving individual images, whenever and wherever needed.
Formed in April of 1999, the vision of the DIG35 Initiative Group is to "provide a standardized mechanism which allows end-users to see digital image use as being equally as easy, as convenient and as flexible as the traditional photographic methods while enabling additional benefits that are possible only with a digital format."
The main categories in which metadata elements can be classified are listed below:

- *Image Creation metadata:* camera capture, scanner capture, image source, creator, capture settings, scanner capture, captured item.

- *Content Description metadata:* caption, capture time, location, person, thing, organization, event, audio, dictionary reference.

- *Metadata History metadata:* processing summary, processing hints (cropped, transformed, retouched).

- *Intellectual Property Rights metadata:* Names, description, dates, exploitation, identification, contact point, history.

METS

The Metadata Encoding and Transmission Standard (METS) is another recently emergent standard designed to encode metadata for electronic texts, still images, digitised video, sound files and other digital materials within electronic library collections [30]. Written in XML schema, METS offers a coherent overall structure for encoding all relevant types of metadata (descriptive, administrative, and structural) used to describe digital library objects.
The organisation of the standard can be described by the following categories:

- *METS Header* : the METS Header contains metadata describing the METS document itself, including such information as creator, editor, etc.

- *Descriptive Metadata*: the descriptive metadata section may point to descriptive metadata external to the METS document (e.g., a MARC record in an OPAC or an EAD finding aid maintained on a WWW server), or contain internally embedded descriptive metadata, or both. Multiple instances of both external and internal descriptive metadata may be included in the descriptive metadata section.

- *Administrative Metadata:* the administrative metadata section provides information regarding how the files were created and stored, intellectual property rights, metadata regarding the original source object from which the digital library object derives, and information regarding the provenance of the files comprising the digital library object (i.e., master/derivative file relationships, and migration/transformation information). As with descriptive metadata, administrative metadata may be either external to the METS document, or encoded internally.

- *File Section*: the file section lists all files containing content which comprise the electronic versions of the digital object. <file> elements may be grouped within <fileGrp> elements, to provide for subdividing the files by object version.

- *Structural Map:* the structural map is the heart of a METS document. It outlines a hierarchical structure for the digital library object, and links the elements of that structure to content files and metadata that pertain to each element.

- *Structural Links:* the Structural Links section of METS allows METS creators to record the existence of hyperlinks between nodes in the hierarchy outlined in the Structural Map. This is of particular value in using METS to archive Websites.

- *Behaviour:* a behaviour section can be used to associate executable behaviours with content in the METS object. Each behaviour within a behaviour section has an interface definition element that represents an abstract definition of the set of behaviours represented by a particular behaviour section. Each behaviour also has a mechanism element which identifies a module of executable code that implements and runs the behaviours defined abstractly by the interface definition.

JPEG2000 and JPX

JPX is an extension of the JPEG2000 image standard file format (jp2) and can be defined as the container which allows to describe the jp2 image and the associated metadata.
The objective of this format can be described in the following list [37]:

- Specify extended decoding processes for converting compressed image data to reconstructed image data
- Specify an extended code-stream syntax containing information for interpreting the compressed image data.
- Specify an extended file format.
- Specify a container to store image metadata.
- Define a standard set of image metadata.
- Provide guidance on extended encoding processes for converting source image data to compressed image data.
- Provide guidance on how to implement these processes in practice.

The interesting part for our goals is the definition of a set of metadata for image description and the specification of a container able to store image metadata.
Starting from the DIG35 standard, on which it is largely based, it defines an XML language that allows to represent a complete set of metadata related to the image.
The metadata elements specify information such as how the image was created, captured or digitised, or how the image has been edited since it was originally created, including the intellectual

property rights information, as well as the content of the image, such as the names of the people and places in the image.
They can be grouped into four different sections and a common section for the types definition:

- *Image Creation Metadata*: the Image Creation metadata defines the how metadata that specifies the source of which the image was created. For example, the camera and lens information and capture condition are useful technical information for professional and serious amateur photographers as well as advanced imaging applications.

- *Content Description Metadata*: the Content Description metadata defines the descriptive information of who , what , when and where aspect of the image. Often this metadata takes the form of extensive words, phrases, or sentences to describe a particular event or location that the image illustrates. Typically, this metadata consists of text that the user enters, either when the images are taken or scanned or later in the process during manipulation or use of the images.

- *History Metadata*: The Metadata History is used to provide partial information about how the image got to the present state. For example, history may include certain processing steps that have been applied to an image. Another example of a history would be the image creation events including digital capture, exposure of negative or reversal films, creation of prints, or reflective scans of prints. All of these metadata are important for some applications. To permit flexibility in construction of the image history metadata, two alternate representations of the history are permitted. In the first, the history metadata is embedded in the image metadata. In the second, the previous versions of the image, represented as a URL/URI, are included in the history metadata as pointers to the location of the actual history. The history metadata for a composite image (i.e., created from two or more previous images) may also be represented through a hierarchical metadata structure. While this specification does not define the how or how much part of the processing aspect, it does enable logging of certain processing steps applied to an image as hints for future use.

- *Intellectual Property Rights Metadata*: The Intellectual Property Rights (IPR) metadata defines metadata to either protect the rights of the owner of the image or provide further information to request permission to use it. It is important for developers and users to understand the implications of intellectual property and copyright information on digital images to properly protect the rights of the owner of the image data.

- *Fundamental Metadata Types and Elements*: The Fundamental metadata types define common data types that may be used within each metadata groups. Those include an address type or a person type which is a collection of other primitive data types. The Fundamental metadata elements define elements the are commonly referenced within other metadata groups. These include a definition for language specification and a timestamp.

JPX is composed of several boxes, among which two are very interesting:

1. the MPEG-7 Binary box.
2. XML box.

This first box contains metadata in MPEG-7 binary format (BiM) as described in the MPEG-7 section, while the second box, already defined within the jp2 file format specifications allows to create metadata description schemes able to represent the considered MM data.

An interesting initiative is represented by the mapping between Dublin Core element set and the XML metadata used in JPX files [38].

SMPTE Metadata Dictionary

SMPTE Metadata Dictionary (SMPTE 335M-2001) [62] has been developed by the Society of Motion Picture and Television Engineer (SMPTE) [61]. It is composed by a set of metadata elements describing audio/video documents that can be grouped into the following categories:

- *Identification*: this class is reserved for abstract identifiers and locators.
- *Administration*: reserved for administrative and business related metadata.
- *Interpretation*: reserved for information on interpreting the data.
- *Parametric*: reserved for parametric and configuration metadata.
- *Process*: reserved for information about the essence or metadata processing.
- *Relational*: is reserved for information about the relationships between data.
- *Spatio-temporal*: reserved for information about space and time.
- *Organisationally Registered Metadata*: this class contains two sub-class, from which the first one is reserved for metadata registered by any organisation for private use and the second one is reserved for metadata registered by any organisation for public use.
- *Experimental Metadata*: in this part users may create their own structure consistent with the Encoding standard.

At the present time the standard definitions have been published in the form of a spreadsheet, but the first step in making the Metadata Dictionary more accessible is to convert the dictionary to a more agreeable format like XML.
Progress is now also being made in providing standardized translation of SMPTE metadata to and from web-friendly XML, and in reconciling SMPTE metadata to MPEG-7 and other descriptive metadata schemes.

MPEG-7: a brief overview

MPEG-7 is an ISO/IEC standard for descriptions of multimedia content. It can be classified into the group of standardised description schemes, but in contrast with other standardised description schemes aforementioned, it has not been developed in a restricted application domain but it has been intended to be applicable to a wide range of application domains.

In a world of more and more content stored in more places, the ability to identify, search, index, and publish information about content is key. MPEG-7 provides the tools needed for managing the exponential growth and distribution of multimedia content over the Internet, in digital broadcast networks, and in home and remote databases. Additionally, it enables highly sophisticated management, search, and filtering of the content.

The range of applications is extensive and includes the following ones:

- *Audio*: Searching for songs by humming or whistling a melody.
- *Graphics*: Sketching a few lines and getting a set of images containing similar graphics, logos, and ideograms.
- *Image*: Checking whether your company logo was advertised on a TV channel as contracted.
- *Visual*: Allowing mobile phone access to video clips of goals scored in a soccer game.
- *Multimedia*: Describing actions and receiving lists of corresponding scenarios.

Motivated by the need of efficient search and retrieval of such content, MPEG-7 was intended to provide a wide coverage audio, visual and more general aspects of multimedia content description.
It is not a coding standard and is not used to store compressed multimedia content. It is addressed to many different applications in many different environments and has a set of methods and tools to describe multimedia content from different view points.
Complex and customized metadata structures can be defined using the XML-based Description Definition Language (DDL).
Metadata schemes can include descriptions of semantic elements (i.e. shapes, colours, people, objects, motion, musical notation); catalogue elements (copyright and access rules, parental ratings, title, location, date, etc); or structural elements (technical stats about the media). So, search engines, live broadcasts and content management systems all can benefit from a standard, human- and machine-readable way to describe and identify content.
As aforementioned, MPEG-7 uses the Extensible Mark-up Language (XML) as a language for textual representation of the multimedia content.
The XML schema is the base for the Description Definition Language (DDL) used for defining the syntax of MPEG-7 description tools [21]. It also allows the extensibility of these tools.
The main elements of the MPEG-7 standard can be listed as below [18]:

- *Description Tools*: Descriptors (Ds) that define the syntax and the semantic of each feature or metadata element; and Description Schemes (DSs) that describe the structure and the semantic of the relationship among their components which can be both Descriptors and Description Schemes.
- *DDL*: the language that defines the syntax of the MPEG-7 description tools, it allows to create new description schemes and also eventually to modify a previous one in order to extend it for a better representation of world of interest.
- *Classification schema*: it defines a list of typical words belonging to a specific application world and their corresponding meanings. For instance, it allows the definition of the file formats in a standard way.
- *Extensibility*: it is supported through the extensibility mechanism of the description tools.
- *System tools*: which support a binary coded representation in order to have an efficient storage and transmission and provide necessary and appropriate transmission mechanisms.

The major functionalities of all part of MPEG-7 can be grouped as follows [22]:

- MPEG-7 Systems includes the binary format for encoding MPEG-7 descriptions and the terminal architecture.

- The DDL is based on XML Schema Language, but it has not been designed specifically for audiovisual content description, then there are certain MPEG-7 extensions which have been added. As a consequence, the DDL can be broken down into the following logical normative components:
  - The XML Schema structural language components
  - The XML Schema data-type language components
  - The MPEG-7 specific extensions

- MPEG-7 Visual Description tools consist of basic structures and Descriptors that cover the following basic visual features: colour, texture, shape, motion, localization and face recognition. Each category consists of elementary and sophisticated Descriptors.

- MPEG-7 Audio provides structures, in conjunction with the Multimedia Description Schemes part of the standard, for describing audio content. These structures are a set of low-level Descriptors, for audio features across many applications (e.g., spectral, parametric, and temporal features of a signal), and high-level Description Tools that are more specific to a set of applications. Those high-level tools include general sound recognition and indexing Description Tools, instrumental timbre Description Tools, spoken content Description Tools, an audio signature Description Scheme, and melodic Description Tools to facilitate query-by-humming.

- MPEG-7 Multimedia Description Schemes (MDS) comprises the set of Description Tools (Descriptors and Description Schemes) dealing with generic as well as multimedia entities. They are used whenever more than one medium needs to be described (e.g. audio and video). They can be grouped into five categories according to their functionalities:
  - Content description: representation of perceivable information.
  - Content management: information about the media features, the creation and the usage of the AV content.
  - Content organization: representation the analysis and classification of several AV contents.
  - Navigation and access: specification of summaries and variations of the AV content.
  - User interaction: description of user preferences and usage history pertaining to the consumption of the multimedia material.

- The eXperimentation Model (XM) software is the simulation platform for the MPEG-7 Descriptors (Ds), Description Schemes (DSs), Coding Schemes (CSs), and Description Definition Language (DDL). The XM applications are divided in two types: the server (extraction) applications and the client (search, filtering) applications.

- MPEG-7 Conformance includes the guidelines and procedures for testing conformance of MPEG-7 implementations.

Using metadata element descriptions, their attribute, the definition of simplex or complex types and the representation of the relations among that elements it is possible to completely describe the multimedia content.
The MPEG-7 descriptions of content can 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 spatial-temporal components of the content (scene cuts, segmentation in regions, region motion tracking).
- Information about low level features in the content (colours, textures, sound timbres, melody description).
- Conceptual information of the reality captured by the content (objects and events, interactions among objects).
- Information about how to browse the content in an efficient way (summaries, variations, spatial and frequency sub bands, etc.).
- Information about collections of objects.
- Information about the interaction of the user with the content (user preferences, usage history)

Table 2 shows some of the features which can be represented using MPEG-7.

Visual	Audio	Multimedia
Colour: • Colour space • Dominant colours • Colour quantisation • …	**Audio framework:** • audio waveform • audio power • …	**Content management:** • creation information • creation tool • creator • …
Texture: • Edge histogram • Homogeneous texture • …	**Timbre:** • harmonic instrument timbre • percussive instrument timbre • …	**Content semantic:** • classification scheme • text annotation • graph
Shape: • object region-based shape • contour-based shape •3D shape • …	**Sound recognition and indexing:** • sound model • sound classification model • sound model state path • …	**Navigation and summarization:** • hierarchical summary • visual summary component • audio summary component • …
Motion: • camera motion • object motion trajectory • …	**Melody:** • melody contour • melody sequence • …	**Content organization:** • collection • classification model • …
Localization: • region-locator • space-temporal locator • …	**Spoken content:** • spoken content lattice • spoken content header • …	**User :** • usage history • user preferences • …

Table 2. Multimedia data features describable using MPEG-7.

Some details about visual descriptor features can be highlighted as follows, other details are available in [22].

Visual Descriptors

Colour:

- *Colour Space:* four colour space are defined (RGB, YcrCb, HSV, HMMD). Alternatively one can specify an arbitrary linear transformation matrix from RGB coordinate.

- *Colour Quantization:* this descriptor is used to specify the quantization method which can be linear, non linear (in MPEG-7 uniform-quantization is referred as linear quantization and non-uniform quantiser as non-linear).

- *Dominant Colour:* this feature describes the dominant colours in the underlying segment, including the number of dominant colours, a confidence measure on the calculated dominant colours, and for each dominant colour the value of each colour component and its percentage.

- *Colour Histogram:* several types of histograms can be specified:
    - *The common colour histogram,* which includes the percentage of each quantized colour among all pixels in a segment or in a region.
    - *The GoF/GoP histogram,* which can be the average, median or intersection of conventional histograms over a group of frames or pictures.
    - *Colour-structure histogram,* which is indented to capture some spatial coherence of pixels with the same colour.

- *Compact Colour Descriptor:* instead to specify the entire colour histogram, it will be possible to specify the first two coefficients of the Haar transform of the colour histogram.

- *Colour Layout:* this is used to describe in a coarse level the colour pattern of the image. So an image can be reduced to an 8x8 blocks with each block represented by its dominant colour.

Shape:
- *Object Bounding box:* this descriptor specifies the rectangular box enclosing two- or three-dimensional object. In addition to the size, centre and orientation of the box, the occupancy of the object in the box is also specified by the ratio of the object area (volume) to the box area.

- *Contour-Based Descriptor:* this descriptor is applicable to a 2-D region with a closed boundary. MPEG-7 has chosen the use of the peaks in the curvature scale space representation to describe a boundary, which has been found to reflect human perception of shape.

- *Region-Based Shape descriptor*: it can be used to describe the shape of any 2-D region, which may consists of several disconnected sub-regions. MPEG-7 has chosen to use the Zernike moments to describe the geometry of a region. The number of moments and the value of each of them are specified.

Texture:
- *Homogeneous texture:* this is used to specify the energy distribution in different orientations and frequency bands (scales). This can be obtained using Gabor transform with six orientation zones and five scale bands.

- *Texture Browsing:* this descriptor specifies the texture appearance in terms of regularity, coarseness and directionality.

- *Edge Histogram:* it is used to describe the edge orientation distribution in an image. Three types of edge histograms can be specified, each with five entries, describing the percentages of directional edges in four possible orientations and non-directional edges. The global edge histogram is accumulated over every pixel in an image; the local histogram consists of 16 sub-histograms, one for each block in an image; the semi-global histogram consists of eight sub-histograms, one for each group of rows and columns in an image.

Motion:

- *Camera motion:* Seven possible camera motion are considered: panning, tracking (horizontal translation), tilting, booming (vertical translation), zooming, translation along the optical axis and rolling (rotation around the optical axis). For each motion two moving directions are possible. For each motion type and direction the presence (i.e., duration), speed and the amount of motion are specified.

- *Motion Trajectory:* it is used to specify the trajectory of the non rigid moving object in terms of 2D or 3D coordinates of certain key point selected. For each key point the trajectory between adjacent sampling times is interpolated by a specific interpolation function (either linear or parabolic).

- *Parametric Object motion:* this is used to specify the 2D motion of rigid objects. Five types of motion are included: translation, rotation/scaling, affine, planar prospective and parabolic. In addition, the coordinate origin and time duration need to be specified.

- *Motion Activity:* it is used to describe the intensity and the spread of activity on a video segment. Four attributes are considered: intensity activity, measured by the standard deviation of the motion vector magnitudes; direction of activity, determined from the average of the motion vector directions; spatial distribution of activity, derived from the run lengths of blocks with motion magnitudes; the temporal distribution of activity, described by the histogram of the quantized activity levels over individual frame in a shot.

Audio Descriptors

The new Audio Description Tools specified for MPEG-7 Audio version 2 are:

- *Spoken Content*: a modification to the version 1 Description Tools for Spoekn Content is specified.

- *Audio Signal Quality*: If an AudioSegment DS contains a piece of music, several features describing the signal's quality can be computed to describe the quality attributes. The AudioSignalQualityType contains these quality attributes and uses the ErrorEventType to handle typical errors that occur in audio data and in the transfer process from analogue audio to the digital domain. However, note that this DS is not applicable to describe the subjective sound quality of audio signals resulting from sophisticated digital signal processing, including the use of noise shaping or other techniques based on perceptual/psychoacoustic considerations. For example, in the case of searching an audio file on the Internet, quality information could be used to determine which one should be downloaded among several search results. Another application area would be an archiving system. There, it would be possible to browse through the archive using quality information, and also the information could be used to decide if a file is of sufficient quality to be used e.g. for broadcasting.

- *Audio Tempo*: The musical tempo is a higher level semantic concept to characterize the underlying temporal structure of musical material. Musical tempo information may be used as an efficient search criterion to find musical content for various purposes (e.g. dancing) or belonging to certain musical genres. Audio Tempo describes the tempo of a musical item according to standard musical notation. Its scope is limited to describing musical material with a dominant musical tempo and only one tempo at a time. The tempo information consists of two components: The frequency of beats is expressed in units of beats per minute (bpm) by AudioBPMType; and the meter that defines the unit of measurement of beats (whole note, half-note, quarter-note, dotted quarter note etc.) and is described using MeterType. Please note that, although MeterType has been initially defined in a different context, it is used here to represent the unit of measurement of beats in a more flexible way, thus allowing to also express non-elementary values (e.g. dotted half-note). By combining Bpm and Meter the information about the musical tempo is expressed in terms of standard musical notation.

Currently there are additional proposed tools for enhancing MPEG-7 Audio functionality, which may be developed to be part of Amendment 2 of MPEG-7 Audio:

- Low Level Descriptor for Audio Intensity.
- Low Level Descriptor for Audio Spectrum Envelope Evolution.
- Generic mechanism for data representation based on 'modulation decomposition.
- MPEG-7 Audio-specific binary representation of descriptors.

MPEG-7 is addressed to applications that can be stored (on-line or off-line) or streamed (e.g. broadcast), and can operate in both real-time and non real-time environments. A 'real-time environment' in this context means that the description is generated while the content is being captured.

An MPEG-7 processing chain includes three principal steps:

- Feature extraction step, that can be considered the analysis phase.
- The description itself, made using the standard.
- The search engine, that corresponds to the application level.

To fully exploit the possibilities of MPEG-7 descriptions, automatic extraction of features will be extremely useful.
Automatic extraction is not always possible. However, neither automatic nor semi-automatic feature extraction algorithms are inside the scope of the standard. The main reason is that their standardization is not required to allow interoperability, leaving space for industry competition. Another reason not to standardize analysis is to allow making good use of the expected improvements in these technical areas.
Figure 2 shows a possible representation of the use of the standard [22].
A multimedia content description is obtained via manual or semi-automatic feature-extraction process.
The audio-visual (AV) description may be stored (as depicted in the figure) or streamed directly.
If we consider a pull scenario, client applications will submit queries to the descriptions repository and will receive a set of descriptions matching the query for browsing (just for inspecting the description, for manipulating it, for retrieving the described content, etc.).
In a push scenario, a filter (e.g., an intelligent agent) will select descriptions from the available ones and perform the programmed actions afterwards.

However, audio-visual content may include or refer to text in addition to its audio-visual information.

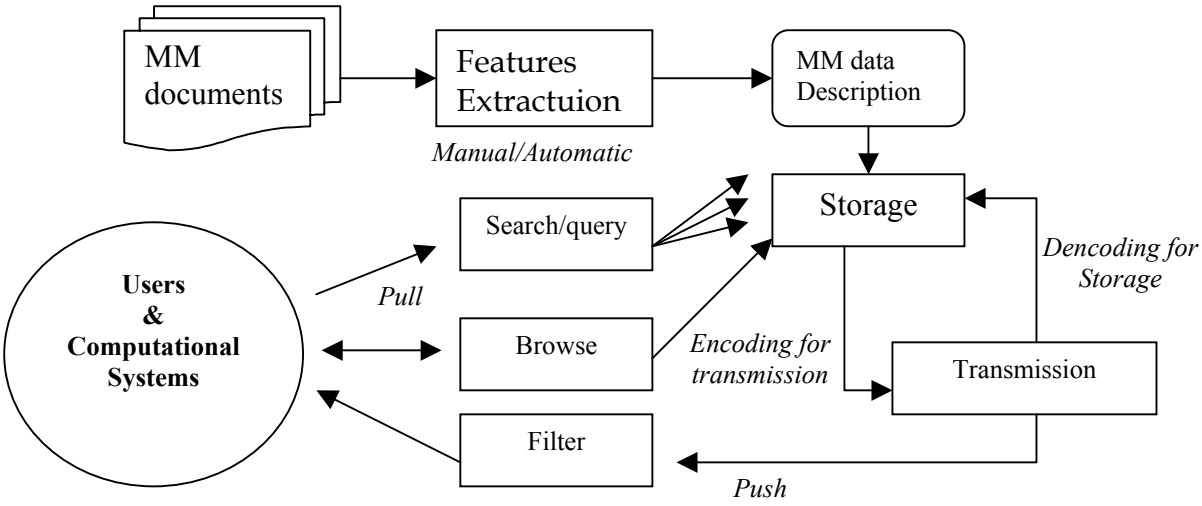

Figure 2. Representation of a possible use of MPEG-7.

To this purpose MPEG-7 has standardized different Description Tools for textual annotation and controlled vocabularies, taking into account existing standards and practices.
The standard provides a number of different basic constructs for textual annotation. The most flexible text annotation construct is the data type for free text. Free text allows the formation of an arbitrary string of text, which optionally includes information about the language of the text.
However, the standard provides also a tool for more structured textual annotation by including specific fields corresponding to the questions "Who? What object? What action? Where? When? Why? and How?".
Moreover, more complex textual annotations can also be defined by describing explicitly the syntactic dependency between the grammatical elements forming sentences, for example relation between a verb and a subject.

Regarding formats for coding MM description information represented by MPEG-7, the language used is without doubts XML. MPEG-7 standardizes an XML language for audiovisual metadata and uses XML to model this rich and structured data.
XML has not been designed to deal ideally in a real-time, constrained and streamed environment like in the multimedia or mobile industry.
To overcome the lack of efficiency of textual XML, MPEG-7 System defines a generic framework to facilitate the processing of MPEG-7 descriptions: BiM (Binary Format for MPEG-7) that enables the streaming and the compression of any XML documents.
BiM coders and decoders can deal with any XML language. Technically, the schema definition (DTD or XML Schema) of the XML document is processed and used to generate a binary format that has three main properties:

- Is connected to Schema knowledge
- Structural redundancy (element name, attribute names, and so on) is removed from the document, so the document structure is highly compressed (98% in average).

- Unlike a zipped XML document, a BiM file can be processed directly at binary level, allowing fast parser and filtering.

There is also another format used by the standard. It is based on a textual encoding called TeM, which enables the dynamic and/or progressive transmission of descriptions using only text.

BiM and TeM are two similar methods to fragment and convey descriptions as a description stream. Both methods allow one to convey arbitrary descriptions conformant to MPEG-7 MDS, Visual, and Audio, but structural differences in the TeM- and BiM-encoded representation of the description as well as in the decoding process exist.

A more detailed description of MPEG-7 and each own part can be found in the web site listed in references section [22].

Chapter 3

Multimedia data and metadata standards overview in the NoE

Pursuing the goal to have a wider knowledge over (MM) data and metadata used by the scientific community, we elaborated a questionnaire to be proposed to the members of the NoE, taking into account a previous initiative of the ECHO (European Chronicles On-line) project [34]. The aims of this questionnaire were multiple: to understand the level of familiarity inside the NoE community with metadata and its management, to know the typologies of data and metadata the NoE community is using, to be aware of the eventual needs of the NoE community when processing MM data.

In order to have some reliable guidelines for the design of the metadata model, two kinds of information were necessary to acquire. On the one hand we had to understand what kind of MM data are used within the NoE and which services are required in the NoE; on the other hand, which services could be implemented. This questionnaire was addressed to the NoE partners. It aimed at obtaining a panorama on the existing situation about the cataloguing of MM documents in each archive. Moreover, it aimed at learning the needs of the partners when implementing efficient speech recognition, digital video abstracting, images and signals processing depending on their interests, and also which metadata fields could be useful to accomplish their tasks.
In particular, then, to provide suitable services, the questionnaire contains few sections to ask for information about the future needs of the NoE partners. We also needed to fully understand whether and how the MM data are grouped together and structured, which MM data components would be interesting to describe and identify, and, finally, how the metadata set will be structured. This is the reason why the questionnaire was logically divided in two parts: the first one, in which general information was required in order to focus on the way in which content providers operate, and the second part in which the metadata description used by content providers were explicitly required.
The questionnaire was formulated following specific criteria relating to the typology of the community to whom it is addressed and the kind of information we are interested in.
The main characteristics taken into account are summarized in the following:

- standard models eventually used
- how many document sets each data providers has defined to group the audio visual documents
- which relationships are defined between the different document sets
- which (if any) different components of the same MM document are stored separately
- to which components of the MM document the metadata fields are associated
- if different support are used to store the same MM document
- in which formats the MM document are stored
- if the same MM document can be stored in different formats, quality, resolution

- if they use controlled vocabularies, Thesauri enumerated lists and, if so, if they are fixed "a priori" or they can be subject to dynamic modification

- which metadata fields are required to implement an efficient speech recognition or digital video abstracting

- which metadata fields are provided by the academic partners making speech recognition or digital video abstracting.

3.1 Analysis of the results

The questionnaire, as it was formulated, was distributed to all members of the NoE community and the answers were collected and analysed in order to purse the aforementioned objectives.

From the results collected we wanted also to understand whether or not a common metadata standard was already in use by the community. If already a common metadata standard was in use by a large part of the community this could give us hints for the directions to follow toward establishing a common standard for the NoE.

Questionnaire results have shown that the types of MM data used by the NoE members are above all images and still images (JPG, TIFF, Sun raster, BMP, GIFF), video and video sequences (YUV, MPEG-1, MPEG-2, AVI, QuickTime). Sometimes the same MM data are stored in different formats (e.g. AVI and MPEG-1), using as storage supports principally hard disks, but also CDs or DVDs and giving in some cases different space-time resolutions of the same MM document (e.g. CIF or QCIF formats for video data).
About the models currently used to describe the MM data used, the results have shown that the main part of the NoE members doesn't use any kind of metadata standard to represent the MM data elaborated except for one case in which the MPEG-7 standard is adopted and another case in which the model used is a Video Database Management System (BilVideo) [35], developed by the group itself to provide support for spatial-temporal and semantic querying using a web-based graphical user interface. The last model is absolutely not based on metadata.
It has been also verified that the main use of metadata is not oriented to group or classify large amount of MM documents, but rather to get a low-to-high level description of individual pieces of content for subsequent specific applications.
For example, we checked that in one case video segments based on spatial-temporal relationships can be recovered in semi-automatic way (BilVideo) with the aim to implement the same recovering process also taking into account semantic predicates. In another case, still images or visual part of video sequences can be also recovered even if no details are given to understand the process followed. Moreover, in a third case, shots and frames can be extracted automatically and scenes in a semi-automatic way from visual documents, aiming at create a "table of content" or an "index" of a MM document.
In the case in which the MPEG-7 metadata standard is used, MM data elaboration is oriented to process images and videos in order to extract low-level information (colour, shape, texture, motion), structural information (table of content, indexes) and a limited number of semantic features (face, body, objects, etc.).

Chapter 4

Standardised Metadata Frameworks

The main problem related to the use of the *standardised metadata description schemes*, as we have seen before, is the fixed number of describing attributes and their specialisation on fixed domains with the consequent lesser flexibility in representing needs of particular application domains.

In the light of these difficulties another standard class has been evolved, considering the possibility of integrating more metadata standards mapped on different application domains, providing rich metadata models for media descriptions together with languages allowing one to define other description schemes for arbitrary domains.
These last standards can be referred as *standardised metadata frameworks*. They promised simpler interoperability among different application domains and also the possibility to simpler create generic tools able to process multimedia descriptions.

We can summarise the main characteristics of the standardised metadata frameworks in the following points:

- No domain-dependence description scheme.
- Rich generic data-model for multimedia content description.
- Schema definition languages that allow defining description schemes for arbitrary application domains.

Table 3 shows some of reference metadata standard frameworks described classifying them according to relevant characteristics [4].

During the research phase particular attention has been given to one of the latest initiative following this tendency which is represented by the MPEG-21, a new MPEG metadata standard framework still under development.
It has been described in more details in the next section.

An overview on XML technologies applied to metadata, techniques for semantic web and interoperability, in particular the RDF framework, has also been taken into account in the following sections.

4.1 MPEG-21: a brief overview

MPEG-21 is the ISO/IEC 21000 standard from the Moving Picture Experts Group (MPEG) which define an open multimedia framework.
The vision of MPEG-21 is to define a multimedia framework to enable augmented and transparent use of multimedia resources across a wide range of networks and devices used by different communities.
The intent is that this framework will cover all the entire multimedia content delivery chain including creation, production, delivery, personalisation, presentation and trade [18].

MPEG-21 is based on two essential concepts [18][23]:

	PICS [24]	**MCF** [25]	**RDF** [26]	**Topic Maps** [27]	**XTM** [27]
Standardisation body	World Wide Web Consortium (W3C)	World Wide Web Consortium (W3C)	World Wide Web Consortium (W3C)	International Organization for Standardization (ISO)	TopicMaps.org Consortium
Year	1996	1997	1999	2000	2001
Data Model	Attribute based	Semantic network-based/ directed labelled graphs	Semantic network-based/ directed labelled graphs	Semantic network-based/ labelled graphs	Semantic network-based/ labelled graphs
Schema Definition Language	Rating Systems	Standard vocabulary	RDF Schema	Topic Map Constraint Language (TMCL)	Topic Map Constraint Language (TMCL)
Domain	Qualitative rating of web resources	Machine understandable, semantic description of web resources	Machine understandable semantic description of web resources	Semantic mapping of information resources	Semantic mapping of information resources
Level	Semantic	Semantic	Semantic	Semantic	Semantic
MM Type	Any	Any	Any	Any	Any

Table 3. Some characteristics of some reference metadata standard frameworks.

- the definition of a fundamental unit of distribution and transaction (*the Digital Item*) which can be considered the "what" of the Multimedia Framework (e.g., a video collection, a music album).
- the concept of *Users* interacting with Digital Items considering them as the "who" of the Multimedia Framework.

The goal of MPEG-21 can thus be express as defining the technology needed to support Users to exchange, access, consume, trade and otherwise manipulate Digital Items in an efficient, transparent and interoperable way.

An User is any entity that interacts in the MPEG-21 environment or makes use of a Digital Item. Such Users include individuals, consumers, communities, organisations, corporations, consortia, governments and other standards bodies and initiatives around the world. Users are identified specifically by their relationship to another User for a certain interaction. MPEG-21 provides a framework in which one User interacts with another User and the object of that interaction is a Digital Item commonly called content.
Some such interactions are creating content, providing content, archiving content, rating content, enhancing and delivering content, aggregating content, delivering content, syndicating content, retail selling of content, consuming content, subscribing to content, regulating content, facilitating transactions that occur from any of the above, and regulating transactions that occur from any of the above. Any of these are "uses" of MPEG-21, and the parties involved are Users.

The digital item is the fundamental unit of distribution and transaction, composed by structured digital objects including standard representation and identification and metadata.
To better understand the meaning of a digital item, we can consider the example of a multimedia family book. It can be composed by photos, videos and related text comments. We would like to have a way to represent each digital component as an individual entity and describe it and its relationships with other represented entity. Moreover, we would like to have also technical information about our entities and to grant that who is not related to our family has not the possibility to view the book, in other words we have to define a consumption rights.

Starting from these basic definitions, twelve parts of MPEG-21 have currently been defined [18][23]:

- *Vision, technologies and strategies*: this part is a technical report describing the vision of MPEG-21, giving an overview of the remaining parts, describing strategies for further developments and finally focusing on the relationships or collaborations with other standardisation bodies or consortia.

- *Digital Item Declaration (DID)*: this part is described by three sections:
    - *Model*: describes a set of abstract terms and concepts to form a useful model for defining Digital Items. Within this model, a Digital Item is the digital representation of "a work", and as such, it is the thing that is acted upon (managed, described, exchanged, collected, etc.) within the model.
    - *Representation*: is a normative description of the syntax and semantics of each of the Digital Item Declaration elements, as represented in XML.
    - *Schema*: Normative XML schema comprising the entire grammar of the Digital Item Declaration representation in XML.

- *Digital Item Identification (DII):* this part introduces tools to define unique identifiers used to name and address DIs and their parts. The specifications includes:

    - How to uniquely identify Digital Items and parts thereof (including resources).
    - How to uniquely identify IP related to the Digital Items (and parts thereof), for example abstractions.
    - How to uniquely identify Description Schemes.
    - How to use identifiers to link Digital Items with related information such as descriptive metadata.
    - How to identify different types if Digital Items.

- *Intellectual Property Management and Protection (IPMP)*: MPEG-21 project will define an interoperable framework for Intellectual Property Management and Protection. The project includes standardized ways of retrieving IPMP tools from remote locations, exchanging messages between IPMP tools and between these tools and the terminal. It also addresses authentication of IPMP tools, and has provisions for integrating Rights Expressions according to the Rights Data Dictionary (RDD) and the Rights Expression Language (REL).

- *Rights Expression Language (REL)*: this part is seen as a machine-readable language that can declare rights and permissions using the terms as defined in the Rights Data Dictionary. The REL is intended to provide flexible, interoperable mechanisms to support transparent and augmented use of digital resources in publishing, distributing, and consuming of digital movies, digital music, electronic books, broadcasting, interactive games, computer software and other creations in digital form, in a way that protects digital content and honours the rights and conditions specified for digital contents.

- *Rights Data Dictionary (RDD)*: The Rights Data Dictionary (RDD) comprises a set of clear, consistent, structured, integrated and uniquely identified Terms to support the MPEG-21 Rights Expression Language. The RDD specification is designed to support the mapping and transformation of metadata from the terminology of one namespace (or Authority) into that of another namespace (or Authority) in an automated or partially-automated way, with the minimum ambiguity or loss of semantic integrity.

- *Digital Item Adaptation (DIA)*: it refers to the terminals and networks and defines tools to support the adaptation process of a DI with respect to the environment description. In this context a Digital Item is subject to a resource adaptation engine, as well as a descriptor adaptation engine, which produce together the adapted Digital Item. The set of tools available are referred to the follow topics: User Characteristics, Terminal Capabilities, Network Characteristics, Natural Environment Characteristics, Resource Adaptability, Session Mobility.

- *Reference Software*: is a set of reference software collecting all software for descriptor generation and use as well as the implementation of the MPEG-21 tools.

- *File Format*: An MPEG-21 Digital Item can be a complex collection of information. Both still and dynamic media (e.g. images and movies) can be included, as well as Digital Item information, meta-data, layout information, and so on. It can include both textual data (e.g. XML) and binary data (e.g. an MPEG-4 presentation or a still picture). For this reason, the MPEG-21 file format has been defined in order to enable the DI to be efficiently stored and accessed.

- *DI Processing (DIP)*: it specifies a processing architecture and a set of operations which can be used to process Dis. These operations are called Digital Item Methods or DIMs. A DIM defines an intended method for configuring, manipulating or validating a DI.

- *Persistent association tools*: this part consists in a set of tools introducing techniques that link information to identify and describe content with the content itself.

- *Resource Delivery Test bed*: this is a reference platform to provide a flexible text environment to evaluate MPEG streaming technologies over IP networks. As a result, new scalable video codecs will be evaluated and developed.

The following table sets out the current timetable for MPEG-21standardisation.

Part	Title	Current Status	Completion Date
1	Vision, Technologies and Strategies	International Standard	Dec 2002
2	Digital Item declaration (DID)	International Standard	Dec 2002
3	Digital Item Identification (DII)	International Standard	Dec 2002
4	Intellectual Property Management and Protection (IPMP)	Working Draft	Sept 2004
5	Rights Expression Language (REL)	Final Committee Draft	Sept 2003
6	Rights Data Dictionary (RDD)	Final Committee Draft	Sept 2003
7	Digital Item Adaptation (DIA)	Final Committee Draft	Sept 2003
8	Reference Software	Final Committee Draft	Jul 2004
9	File Format	Final Committee Draft	Sept 2003

Table 4. Current timetable for MPEG-21standardisation.

Finally, we can say that MPEG-21 multimedia framework gives a "big picture" of interoperability [18], that is it offers a framework to build an infrastructure for the delivery and consumption of multimedia content. The components proposed are Users who interact with Dis. A DI can be an elemental piece of content, for instance a simple picture or an audio track, or a complete collection of audio-visual material. But the DI is only a container, in which it is indispensable to put the description of the DI itself using an existing metadata standard, in order to facilitate the integration in the whole distributed multimedia system.

More details on MPEG-21 standard framework can be found on the web site of the Moving Picture Expert Group (MPEG) in [23].

The following sections give a brief overview on the XML technologies applied to metadata description, the semantic web and interoperability way, given more emphasis to one of the aforementioned multimedia metadata standard frameworks, the Resource Description Framework (RDF) that is one of the most used frameworks.

4.2 XML technologies and metadata, semantic web and interoperability

Metadata must be expressed in a standardized way in order to be read, searched, exchanged by computer systems, and understood by humans. It can be recorded in a table or a database, or expressed in an HTML (Hypertext Mark-up Language) or XML (eXtensible Mark-up Language) document.

XML and its associated technologies, XML Namespaces, XML Query languages, XML Databases are enabling implementers to develop metadata schemas, application profiles, large repositories of XML metadata, and search interfaces using XML Query Language.
These technologies are the key to enabling the automated computer-processing, integration and exchange of information over internet [28][29].

4.2.1 Extensible Markup Language (XML) and metadata

XML (the Extensible Markup Language, W3C XML, 2003) is a simple, very flexible text format derived from the Standard Generalised Markup Language (SGML, ISO 8879).
Originally designed to meet the challenges of large-scale electronic publishing, XML will have a profound impact on the way data is exchanged on the Internet.
An important feature of this language is the separation of content from presentation, which makes it easier to select and/or reformat the data.
A DTD (document type definition) can be used to ensure that XML documents conform to a common grammar. Thus a DTD provides a syntax for an XML document, but the semantics of a DTD are implicit. That is, the meaning of an element in a DTD is either inferred by a human due to the name assigned to it, is described in a natural-language comment within the DTD, or is described in a document separate from the DTD.

If the growing proliferation of DTDs is indicative of the large use of XML on the web, then web developers will still be faced with the problem of semantic interoperability, i.e., the difficulty in integrating resources that were developed using different vocabularies and different perspectives on the data.
To achieve semantic interoperability, systems must be able to exchange data in such a way that the precise meaning of the data is readily accessible and the data itself can be translated by any system into a form that it understands.

Because XML makes it possible to exchange data in a standard format, independent of storage, it has become the de-facto standard for representing metadata descriptions of resources on the Internet, but while XML can be viewed as a little support for expressing semantic knowledge, RDF represents a valid way to do this.

XML Schema Language

XML Schema Language (W3C XML Schema, 2003) [21] provides a means for defining the structure, the content and semantics of XML documents.
It provides a list of XML markup constructs which can constrain and document the meaning, usage and relationships of the constituents of a class of XML documents: data types, elements and their content, attributes and their values, entities and their contents and notations. Thus, the XML

Schema language can be used to define, describe and catalogue XML vocabularies for classes of XML documents, such as metadata descriptions of web resources or digital objects.
XML Schemas have been used to define metadata schemas for a number of specific domains or applications, such as METS [30], MPEG-7 [22], MPEG-21 [23], NewsML [9].
An additional major metadata development has been the employment of W3C's XML Schemas and XML Namespaces to combine metadata elements from different domains/namespaces into "application profiles" or metadata schemas which have been optimised for a particular application.

XML Query

The mission of the XML Query Working Group (W3C XML Query, 2003) is to provide flexible query facilities to extract data from real and virtual documents on the Web, thereby providing the needed interaction between the web world and the database world.

Ultimately, collections of XML files will be accessed like databases.
The new query language, Xquery (Working draft 23 July 2004, [31]) is still evolving but it will provide a functional language comprised of several kinds of expressions that can be nested or composed with full generality.

XML Databases

There is a large amount of research and development going on in the area of XML databases.
Ronald Bourret [5] provides an excellent overview of the current state of this work and a comparison of current XML database technologies.
Bourret divides XML Database solutions into the following categories [29]:

- *Middleware*: software you call from your application to transfer data between XML documents and databases.

- *XML-Enabled Databases*: Databases with extensions for transferring data between XML documents and themselves.

- *Native XML Databases*: Databases that store XML in "native" form, generally as some variant of the DOM mapped to an underlying data store. This includes the category formerly known as persistent DOM (PDOM) implementations.

- *XML Servers*: XML-aware J2EE servers, Web application servers, integration engines, and custom servers. Some of these are used to build distributed applications while others are used simply to publish XML documents to the Web. Includes the category formerly known as XML application servers.

- *Content Management Systems*: Applications built on top of native XML databases and/or the file system for content/document management and which include features such as check-in/check-out, versioning, and editors.

- *XML Query Engines*: Standalone engines that can query XML documents.

- *XML Data Binding*: Products that can bind XML documents to objects. Some of these can also store/retrieve objects from the database.

4.2.2 Semantic web and interoperability

According to the World Wide Web Consortium (W3C), “The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation".

The Semantic Web has two main motivators:

- The first is *data integration*, which is a significant bottleneck in many IT applications. Current solutions to this problem are mostly ad hoc. In fact, each time a specific mapping is made between the data models (schemas) of the data sources involved. If the semantics of data sources were described in a machine-interpretable way, the mappings could be constructed at least in semi-automatic way.
- The second motivator is *more intelligent support for end users*. If computer programs can infer consequences of information on the web, they can give better support in finding information, selecting information sources, personalizing information, combining information from different sources, and so on.

The W3C has defined such open standards for metadata syntax as RDF (Resource Description Framework) [33] and OWL (Web Ontology Language) [44], and support for these standards from both industry and academia is rapidly increasing.

Professional groups increasingly are building metadata vocabularies (or *ontologies*). Large ontologies exist for medical terminology, genomics, geographic information systems, and law, just to mention a few.
These terminologies are typically hand built, but systems are rapidly getting better at learning them in a semi-automatic way from large volumes of text.

An important open problem is that of automatically finding translations between different terminologies that have been designed for the same domain, the so called *"ontology mapping problem"*.

RDF (Resource Description Framework)

The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries. It is a collaborative effort led by W3C with participation from a large number of researchers and industrial partners. It is based on the Resource Description Framework (RDF), which integrates a variety of applications using XML for syntax and URIs for naming [32].
RDF is *"to specify semantics for data based on XML in a standardized interoperable manner."* [28].
Technically, RDF is not a language, but a data model of metadata instances. The basic data model is very simple; it consists of nodes connected by labelled arcs, where the nodes represent web resources and the arcs represent properties of these resources.
The development of RDF has been motivated by the following uses [26]:

- Web metadata: providing information about Web resources and the systems that use them (e.g. content rating, capability descriptions, privacy preferences, etc.)

- Applications that require open rather than constrained information models (e.g. scheduling activities, describing organizational processes, annotation of Web resources, etc.)

- To do for machine processable information (application data) what the World Wide Web has done for hypertext: to allow data to be processed outside the particular environment in which it was created, in a fashion that can work at Internet scale.

- Inter-working among applications: combining data from several applications to arrive at new information.

- Automated processing of Web information by software agents: the Web is moving from having just human-readable information to being a world-wide network of cooperating processes.

RDF is designed to represent information in a minimally constraining, flexible way. It can be used in isolated applications, where individually designed formats might be more direct and easily understood, but RDF's generality offers greater value from sharing. The value of information thus increases as it becomes accessible to more applications across the entire Internet.
RDF uses the key concepts exposed briefly in the follow [26]:

- *Graph data model*: the structure of any expression in RDF is a collection of triples, each consisting of a subject, a predicate and an object. A set of such triples is called an RDF graph. This can be illustrated by a node and directed-arc diagram, in which each triple is represented as a node-arc-node link, hence the term "graph". Each triple represents a statement of a relationship between the things denoted by the nodes that it links.

- *URI-based vocabulary*: A node may be a URI (Uniform Resource Identifiers) with optional fragment identifier (URI reference, or URIref), a literal, or blank. A URI reference or literal used as a node identifies what that node represents. A URI reference used as a predicate identifies a relationship between the things represented by the nodes it connects. A predicate URI reference may also be a node in the graph. In particular, a blank node is just a unique node that can be used in one or more RDF statements, but has no intrinsic name.

- *Data types*: Data types are used by RDF in the representation of values such as integers, floating point numbers and dates. A data type is defined by a lexical space, a value space and a lexical-to-value mapping. RDF predefines just one data type “rdf:XMLLiteral”, used for embedding XML in RDF.

- *Literals*: literals are used to identify values such as numbers and dates by means of a lexical representation. Anything represented by a literal could also be represented by a URI, but it is often more convenient or intuitive to use literals. They can be “plain literals” meaning that the string can be combined with other language tags, or “typed literals” meaning that the string can be combined with data type URI.

- *Expression of simple facts*: some simple facts indicate a relationship between two things. Such a fact may be represented as an RDF triple in which the predicate names the relationship, and the subject and object denote the two things. Thus, a more complex fact is expressed in RDF using a conjunction (logical-AND) of simple binary relationships. RDF does not provide means to express negation (NOT) or disjunction (OR).

- *Entailment*: in brief, an RDF expression A is said to entail another RDF expression B if every possible arrangement of things in the world that makes A true also makes B true. On this basis, if the truth of A is presumed or demonstrated then the truth of B can be inferred .

An RDF document is written in XML. The XML language used by RDF is called RDF/XML. By using XML, RDF information can easily be exchanged between different types of computers using different types of operating system and application languages.

The RDF data model defines a simple model for describing interrelationships among resources in terms of named properties and values. RDF properties may be thought of as attributes of resources and they also can represent relationships between resources. As such, the RDF data model can therefore be similar to an entity-relationship diagram.
The RDF data model, however, provides no mechanisms for declaring these properties, nor does it provide any mechanisms for defining the relationships between these properties and other resources. That is the role of RDF Schema.
To allow the creation of controlled, sharable, and extensible vocabularies, the RDF working group has developed the RDF Schema Specification [33]. This specification defines a number of properties that have specific semantics.
A schema defines not only the properties of the resource (e.g. Title, Author, Subject, Size, Colour, etc.) but may also define the kinds of resources being described (e.g. books, Web pages, people, companies, etc.). It specifies the mechanisms needed to define such elements, to define the classes of resources they may be used with, to restrict possible combinations of classes and relationships, and to detect violations of those restrictions.
Since it is possible that different schemas may use the same strings to represent different conceptual meanings, RDF uses XML namespaces to assign a separate namespace to each schema. In RDF, schemas are extended by simply referring to objects from that schema as resources in a new schema. Since schemas are assigned to unique URIs (Uniform Resource Identifiers), the use of XML namespaces guarantees that exactly one object is being referenced.
More details on RDF specifications can be found in [33].

Ontologies

The W3C Web Ontology Working Group is building upon the RDF Core work to develop a language for defining structured web based ontologies which will provide richer integration and interoperability of data among descriptive communities. This is the Web Ontology Language (OWL) [44] which in turn is building upon the DAML+OIL specification [43] developed by DARPA [50].

An ontology consists of a set of concepts, axioms, and relationships that describes a domain of interest. An ontology is similar to a dictionary or glossary, but with greater detail and structure and expressed in a formal language (e.g., OWL) that enables computers to process its content.
Ontologies can enhance the functioning of the web to improve the accuracy of Web searches and to relate the information in a resource to the associated knowledge structures and inference rules defined in the ontology.
Ontologies can differ in respect to the scope and purpose of their content. The most prominent distinction is between the:

- *domain ontologies* describing specific fields, like medicine.
- and *upper level ontologies* describing the basic concepts and relationships considered when information about any domain is expressed in natural language.

The synergy among ontologies, springs from the cross-referencing between upper level ontologies and various domain ontologies.
Upper ontologies provide a structure and a set of general concepts upon which domain-specific ontologies (e.g. medical, financial, engineering, sports etc.) could be constructed. An upper

ontology is limited to concepts that are abstract and generic enough to address a broad range of domain areas at a high level.
Computers utilize upper ontologies for applications such as data interoperability, information search and retrieval, automated inferencing, and natural language processing.

Ontologies should be publicly available and different data sources should be able to commit to the same ontology for shared meaning. Also, they should be able to extend other ontologies in order to provide additional definitions. In fact, often shared ontologies are not sufficient. An organization may find that an existing ontology provides 90% of what it needs, but the remaining 10% is critical. In such cases, the organization should not have to create a new ontology, but an ontology which extends an existing one and adds any desired identifiers and definitions.

Different ontologies may model the same concepts in different ways. The language should provide primitives for relating different representations, thus allowing data to be converted to different ontologies and enabling a "web of ontologies".

Many communities are developing domain-specific or application-specific ontologies. Some examples include biomedical ontologies such as OpenGALEN [46] and SNOMED CT [47].

A large number of research efforts are focussing on the development of tools for building and editing ontologies [48], these are moving towards collaborative tools such as OntoEdit [49].

A number of research and standards groups are working on the development of common conceptual models (or upper ontologies) to facilitate interoperability between metadata vocabularies and the integration of information from different domains. For example the Harmony project developed the ABC Ontology/Model [45], a top-level ontology to facilitate interoperability between metadata schemas within the digital library domain. Other interesting projects are reported in Appendix C.

Chapter 5

Conclusions

A review of data and metadata standards and tools used to represent MM information is presented in this document, as the preliminary step necessary to achieve the goals that WP9 is pursuing.
These goals can be briefly resumed in the following items:

- To grant interaction between different communities in the NoE (vision, speech, text, etc.).
- To define a strategy for the NoE to develop, maintain and provide an integrated (meta-)data service able to support multiple MM data standards, multiple users and the management of distributed and heterogeneous data, metadata and connected methodologies.
- To act as a facilitator in the communication, exchange and interoperability by advising on common formats, creating exchange interfaces, establishing additional standards, where needed, and formats for metadata.
- To add content-based representation, protection & rights issues.
- To co-operate with other organization bodies.

In particular, this report analyses the current data and metadata standards used within the NoE and other MM communities with the aim at exchanging MM information.
In the light of the existing state of the art, inside and outside the NoE, it has been also pointed out how the interoperability needs of the NoE can be covered by these standards.
A possible strategy for defining a representative metadata model of heterogeneous MM data can also be evicted, taking into account the existing network technologies able to grant semantic mappings and network resource interoperability.

From our analysis, the main metadata features individuated are:

- *metadata description schemes*, which represent singular metadata specifications
- *metadata frameworks*, which are intended as high level containers. These frameworks are thought to implement interoperability, especially between different description schemes. They mainly contain metadata description schemes and the relationships among the different schemes.

Then, an overview of the main and mostly used metadata schemes and frameworks is presented. Each scheme has its own MM data types (i.e., images, sounds, videos,…) and domain of application (i.e., bibliographic data, geographical data, digital images,…). Each framework aims at combining different specific metadata standards expanding the domain of the applications covered.

In order to have a wider knowledge over MM data and metadata used by the scientific community, a specific questionnaire was elaborated and proposed to the members of the NoE, aiming at understanding what kind of MM data are used and which services could be implemented to answer the needs of the partners when working on their applications of interest.
The questionnaire was formulated following specific criteria relating to the typology of the community to whom it is addressed and the kind of information we are interested in [34].

The collected results have shown that the MM metadata standards are not commonly used within the NoE, even if interoperability needs come out clearly.

Therefore, future work might be oriented to consider the possibility to individuate a mature metadata standard in conjunction with a metadata framework (RDF in particular) to allow the use of different standards describing heterogeneous resources of specific application domains.

Besides, semantic mappings between specific-domain metadata vocabularies and the integration of information from different domains could be also taken into account by constructing extensible ontologies independent of any domain metadata vocabulary, aiming at the definition of an unambiguous machine-understandable formal representation of the semantics associated with multimedia descriptions (e.g., RDF, OWL).

At the same time it could be opportune to interact more deeply with MPEG-21 evolution.

As a final consideration, we like to highlight that analysing the state of the art of MM (meta-)data standards, only a part of them, having a clear international relevance, has been taken into account. Other current initiatives have not been included in this report since they have not been considered particularly attractive for our purpose, but surely we will pay attention to all the new emerging standards in the following of the Muscle activity, also searching an active interaction with standardisation bodies.

Appendix A

Questionnaire

A proposal to collect information within the NoE

To enable resource sharing and exchanging of the MUSCLE multimedia (MM) data collections over the Web, it is necessary to define content description standards or metadata standards for complex, multi-layered, time-dependent, information rich data streams.

In order to have some reliable guidelines for the design of the metadata model, two kinds of information are necessary to acquire. On the one hand we have to understand what kind of MM data are used within the NoE and which services are required in the NoE; on the other hand, which services could be implemented. This questionnaire is addressed to the NoE partners. It aims at obtaining a panorama on the existing situation about the cataloguing of MM documents in each archive. Moreover it aims at learning the needs of the partners to implement an efficient speech recognition, digital video abstracting, images and signals processing depending on their interests, and learning which metadata fields they could provide.

In particular, to provide suitable services, the questionnaire aims at acquiring further information to understand the future needs of the NoE partners. We need to fully understand not only a possible metadata set of the MM data, but whether and how the MM data are grouped together and structured, which MM data components would be interesting to describe and identify, and, finally, how the metadata set will be structured. This is the reason why the questionnaire is logically divided in two parts: the first one, in which general information is required in order to focus on the way in which content providers operate, and the second part in which the metadata used by content providers are explicitly required.
The questionnaire is formulated following specific criteria relating to the typology of the persons to whom it is addressed and the kind of information we are interested in:

- standard models eventually used
- how many document sets each data providers has defined to group the audio visual documents
- which relationships are defined between the different document sets
- which (if any) different components of the same MM document are stored separately
- with which components of the MM document the metadata fields are associated
- if different support are used to store the same MM document
- in which formats the MM document are stored
- if the same MM document can be stored in different formats, quality, resolution
- if they use controlled vocabularies, Thesauri enumerated lists and, if so, if they are fixed “a priori” or they can be subject to dynamic modification
- which metadata fields are required to implement an efficient speech recognition or digital video abstracting
- which metadata fields are provided by the academic partners making speech recognition or digital video abstracting

Kind	**Question**	**Answer**		**Comment**	
		The existing situation	**The desired situation**	**The existing situation**	**The desired situation**
Model	**1)** Do you use some cataloguing criteria corresponding to a standard for your MM data? If so, which one(s)?				
Model	**2)** Describe your actual situation in terms of the model currently being used (answer the question providing information only about the existing situation) .				
MM Data Types	**3)** Which kind of MM data do you deal with ? For example audio-visual documents, audio files, images etc…				
Sets	**4)** How many sets have you defined to group your MM data? For each set, supply us with the criterion that characterises it (e.g. images and/or audio data extracted from an audio-visual document containing persons and/or human voices)				
Sets	**5)** If you have more than one set, are there any relationships between the different sets? If so, describe the relationships? (e.g. a set is a subset of another, or a MM data might belong to two or more sets or two sets are partially or totally described through the same metadata set)				

Sets	**6)**If you have more than one set, is the same metadata set used to describe different multimedia data sets? That is, do you have a common metadata model for all the sets? If not, is the model common for only some sets? If so, which ones?				
Documents Structure	**7)**Which component(s) of the same MM data can be recovered? For example, the audio, the video track or text, audio, images in a SMIL presentation. What are the relationships between them? For example, the audio track is the audio of a MM data extracted from the (MPGx or Real Video) file and not stored separately.				
Documents Structure	**8)** On the inside of the same MM data, are the sequences, scenes, shots, frames in an audio-visual document or text, images in a MM presentation, or objects in an image identifiable? If so, which one(s) and how are they identifiable?				
Documents Structure	**9)** Are there other types of decompositions (physical or maybe only logical ones) of the MM data? If so, which ones? (e.g. an audio-visual document regarding a tennis match is composed of parts each related to the sets and games of the match)				
Documents Format	**10)** Do you have the same MM data stored on different supports? If so, which ones? For example, on tape, CD-ROM, …				

Documents Format	**11)** In which formats do you have your MM data? For example, in AVI, Real Video, JPEG, BMP, MPEGX, SVHS, …				
Documents Format	**12)** Do you have the same MM in different formats? If so, specify which ones. For example, the same MM data can be available in AVI, MPEGX or in JPEG, BMP and so on.				
Documents Format	**13)** Do you have the same MM data stored with different resolutions and quality? If so, specify.				
Metadata Sets	**14)** For each MM data set, provide the information indicated in Table1.				
Metadata Sets	**15)** Which kind of elaborations on your MM data do you perform? For example speech recognition, image and signal processing, digital video abstracting, …				
Metadata Sets	**16)** In your opinion, which are the metadata fields that are required to implement an efficient speech recognition, digital video abstracting or other kind of elaborations on the MM data? List those metadata fields providing them with a description as indicated in Table 2.				
Metadata Sets	**17)** Which are the metadata fields that you are able to provide? List those metadata fields providing them with a description as indicated in Table 2.				
Documents Format	**18)** Which are the aspects of the MM data that you consider interesting to describe and that are not dealt within this questionnaire?				

Table 1: Required information

Set Name		**Insert the name of the MM data set related to the metadata set below.**					
Name	**Name Object to which the metadata field is associated**	**Description**	**Type**	**Mandatory**	**Constraints**	**Other**	**Automatic Extraction**
Insert the name of each metadata field in the described MM data set	Indicate whether the metadata field is relative to the object belonging to the set or whether it is relative to one of its components. If it is relative to a component indicate a kind of component. The term component means the decomposition of a MM data in video tack, audio track, scenes, frames, text etc… and also its decomposition into possible logical components.	Insert the description of the metadata field.	Indicate the type of value assumed by the metadata field. In the case in which the metadata field type is date or time, indicate the specific format used.	Indicate whether the metadata field is mandatory (Yes) or not (No).	Indicate whether constraints are defined on the metadata field. For example specify whether the assumed value is restricted by the value of another metadata field or other specific conditions.	Insert other informatio n that you think is interesting.	Indicate whether the metadata field is automatically (Yes) or manually (No) extracted.

Table 2: Required information

Name	Name Object to which the metadata field is associated	Description	Type	Mandatory	Constraints	Other	Automatic Extraction
Insert the name of each metadata field in the described MM data set	Indicate whether the metadata field is relative to the object belonging to the set or whether it is relative to one of its components. If it is relative to a component indicate a kind of component. The term component means the decomposition of a MM data in video tack, audio track, scenes, frames, text etc… and also its decomposition into possible logical components.	Insert the description of the metadata field.	Indicate the type of value assumed by the metadata field. In the case in which the metadata field type is date or time, indicate the specific format used.	Indicate whether the metadata field is mandatory (Yes) or not (No).	Indicate whether constraints are defined on the metadata field. For example specify whether the assumed value is restricted by the value of another metadata field or other specific conditions.	Insert other information that you think is interesting.	Indicate whether the metadata field is automatically (Yes) or manually (No) extracted.

Appendix B

Standardization Bodies

Reference list

- W3C (http://www.w3.org)
- DSTC *Distributed Systems Technology Centre* (http://www.dstc.eu.au)
- AES, *Audio Engineering Society* (http://www.aes.org)
- EBU, *Engineering Broadcasting Union* (http://www.ebu.ch)
- SMPTE, *Society of Motion Picture and Television Engineering* (http://www.smpte.org)
- CEN/ISSS, *European Committee for Standardization / Information Society Standardization System* (http://www.cenorm.be)
- ISO, *International Organization for Standardization* (http://www.iso.org)MS, *Global Learning Consortium* (http://www.imsglobal.org)
- LTSC, *IEEE Learning Technology Standard Committee* (http://ltsc.ieee.org)
- MPEG, *Moving Picture Expert Group* (http://mpeg.org)
- I3A, *International Imaging Industry Association* (http://www.i3a.org)
- NISO, *National Information Standards Organisation* (http://www.niso.org)
- DCMI working groups (http://dublincore.org/groups)
- VRA, Visual Resources Association Data Standards Committee (http://www.vraweb.org)
- AITF, *Art Information Task Force* promoted by the Getty Standards and Digital Resource Management Program (http://www.getty.edu)
- DLF, *Digital Library Foundation working group (*http://www.diglib.org/standards.htm)
- SWG, Standards Working Group promoted by the Federal Geographic Data Committee (http://www.fgdc.gov/metadata/contstan.html)
- Network Development MARC Standards Office (http://www.loc.gov/marc/ndmso.html)

Appendix C
Reference Projects

A brief list of recent reference projects of international importance, also together with other interesting international initiatives about MM metadata are listed below.

Some of these initiatives use tools like XML/RDF [21][33] or ontologies (DAML+OIL or OWL) [43] to realise semantic mappings among metadata description schemes of different domains in order to grant interoperability.

- *MAENAD* (Multimedia Access across Enterprise Networks and Domains). The objectives of this project are to develop underlying data models, ontologies, metadata schemas (RDF, XML), metadata generators, annotation tools, repositories, querying and inferencing languages, search and presentation interfaces and search engines which can provide solutions to the problems of finding, assimilating, presenting, preserving, and managing MM content across domains (cultural, educational, scientific), metadata schemes and media types (e.g. text, image, audio, video, music, cartographic, HTML, SMIL, XML). It is composed by sub-projects, some of which are listed below:

    - *Sunago*: the aim of this project is to develop ontologies, standards, tools and techniques to enable the incorporation and exploitation of multimedia resources with the semantic web. This will potentially involve construction of efficient knowledge-based multimedia systems that automatically extract, from a multimedia input, semantic information (objects, events, properties, relations) described in ontologies of specific domains enriched with audiovisual data. The extracted semantic metadata can be used for classification, summarisation, indexing, searching and efficient retrieval of multimedia content [39][40][41][42].

    - *Indigenous Collection Management*: the objective of this project is to investigate how information technology tools and standards can be refined and extended to enable indigenous communities to preserve and protect their unique indigenous cultures, knowledge and artefacts whilst supporting traditional protocols and facilitating better cross-cultural communication and understanding.

    - *FilmEd*: it aims to define collaborative tools for annotating, summarizing, deconstructing, and discussing moving image titles and for generating multimedia educational programs about films or videos; effective tools for storing, searching, querying, browsing and presenting structured multimedia analyses of films and videos; software tools, workflows and models for handling the rights management of film and video resources; a technical and financial model for the effective delivery and reuse of moving image metadata and segments within the education sector at the tertiary level.

    - *Harmony* (International Digital Library Project): it is an international project investigating the key issues in describing complex multimedia resources in digital libraries through the study of a conceptual model for interoperability among

community-specific metadata vocabularies, its related technologies and to develop mechanisms to map between community specific vocabularies using such a conceptual model. A resulting work is the ABC ontology model [45].

- *DICEMAN* (Distributed Internet Content Exchange with MPEG-7 and Agent Negotiations) (http://www.cordis.lu/infowin/acts/analysys/products/thematic/agents/ch3/diceman.htm): The main objective of the project has been to develop an end-to-end chain of technologies for indexing, storage, search and trading of digital audio-visual content. The targeted application domain for this chain of technologies is the current audio-visual content archiving industry and its professional end users. The technical work focus on: MPEG-7 indexing through a COntent Provider's Application (COPA); the use of FIPA Agents to search and locate the best content; and support for electronic commerce and rights management.

- *MusicBrainz Metadata Initiative 2.1* (http://musicbrainz.org/MM/index.html): this initiative is designed to create a portable and flexible means of storing and exchanging metadata related to digital audio and video tracks. It is a content description model for audio and video tracks on the Internet, using RDF/XML to facilitate the exchange of audio/video related metadata.

- *ROADS* (Resource Organisation and Discovery in Subject-based services) funded by the Joint Information Systems Committee (JISC) [69]: the overall objective of the ROADS project is to design and implement a user-oriented resource discovery system. It investigate the creation, collection and distribution of resource descriptions, to provide a transparent means of searching for, and using resources. The object is not to create an individual and idiosyncratic system but to draw on, and help create, standards of good practice which can be widely adopted by subject communities to aid and automate the process of resource organisation and discovery [68].

- *Warwick Framework* [70]: this framework is a container architecture for diverse sets of metadata. It provides an architecture for the interchange of distinct metadata packages. The architecture has two fundamental components, containers and packages. Containers are the unit for aggregating metadata packages. A container may be transient, existing only to transfer packages between systems, or persistent. In its persistent form a container is stored on one or more servers and is accessible using a global identifier (URI). It should be noted that a container may be wrapped within another object, i.e. one that is a wrapper for both data and metadata. Each package is a typed object of one of the following kinds:
    - metadata set, for example a Dublin Core or MARC record
    - indirect, i.e. a reference to an external object using a URI
    - container, these can be nested to any level of complexity

This list represents only a brief overview of some reference projects working on multimedia metadata standards, with the aim to give an idea of the increasing emphasis on this field and try to motivate our initiatives in this direction.

Appendix D

Contributing Partners

Enis Cetin	Bilkent University (Turkey)	cetin@ee.bilkent.edu.tr
Matthieu Cord	ENSEA	cord@ensea.fr
Montse Pardas	UPC	montse@gps.tsc.upc.es

Appendix E

Next steps within Muscle

Next steps in the project:

- *Webcast /helpdesk*: increase familiarity with and adherence to standards through seminars, helpdesk (e.g. a mailing list or a dedicated web site allowing users to exchange opinions and know-how using a forum or a database as a node of the Virtual Lab).

- *Create a "repository"* made up of all documents, articles, papers and tools to stimulate the NoE to complete the "internal" state of the art on the multimedia (meta-)data standards and tools.

- *Interactions with standardisation bodies*

Next reports:

- Guidelines on standards for representation and exchange of data, metadata and semantic.

- *A*nalysis of the open problems and interoperability limits of the existing standard and tools.

- Structuring tools for standardisation and exchange of data, metadata and semantic.

- *Interaction with standardisation bodies*: report on the liaison with other standardisation bodies.

References

[1] Wang J.Z., Jia Li, Wiederhold, G., *"SIMPLIcity: semantics-sensitive integrated matching for picture libraries"*, Pattern Analysis and Machine Intelligence, IEEE Transactions on, Vol. 23, Issue 9, pp. 947-963, September 2001.

[2] Jia Li, Wang J.Z.,*" Automatic Linguistic Indexing of Pictures by a statistical modeling approach "*, Pattern Analysis and Machine Intelligence, IEEE Transactions on, Vol. 25, Issue 9, pp. 1075-1088, September 2003.

[3] Hunter J., *"Working towards MetaUtopia - A Survey of Current Metadata Research"*, Library Trends, Organizing the Internet, Edited by Andrew Torok, 52(2), Fall 2003.

[4] Utz Westermann, Wolfgan Klas, *"An Analysis of XML Database Solutions for the Management of MPEG-7 Media Descriptions",* ACM Computing Surveys, Vol. 35, No. 4, December 2003, pp. 331-373Bourret R. (2003, January). XML and Databases http://www.rpbourret.com/xml/XMLAndDatabases.htm

[6] Hunter J., *"Adding Multimedia to the Semantic Web - Building an MPEG-7 Ontology"*, International Semantic Web Working Symposium (SWWS), Stanford, July 30 - August 1, 2001.

[7] TV-Anytime Forum, http://www.tv-anytime.org/

[8] IBM QBIC System, 2001. http://wwwqbic.almaden.ibm.com

[9] NewsML http://www.newsml.org/

[10] ISO/IEC 15938-2 FCD Information Technology - Multimedia Content Description Interface - Part 2: Description Definition Language, March 2001, Singapore

[11] Dublin Core Metadata Element Set, Version 1.1, 2 July, 1999. http://www.purl.org/dc/documents/rec-dces-19990702.htm

[12] Indecs Metadata Model, November, 1999 (http://www.indecs.org/)

[13] Content Standard for Digital Geospatial Metadata (CSDGM), http://www.fgdc.gov/metadata/contstan.html

[14] GEM, The Gateway to Educational Materials http://www.geminfo.org/

[15] IEEE Learning Technology Standards Committee's Learning Object Meta-data Working Group. Version 3.5 Learning Object Meta-data Scheme.

[16] CIDOC Documentation Standards Group, Revised Definition of the CIDOC Conceptual Reference Model, September 1999 (http://cidoc.ics.forth.gr/index.html)

[17] RDF Model and Syntax Specification, W3C Recommendation 22 February 1999. http://www.w3.org/TR/REC-rdf-syntax/

[18] Kosch H., "Distributed Multimedia database technologies supported by MPEG-7 and MPEG-21", Boca Raton CRC Press 2004.

[19] MHEG www.mheg.org/

[20] SMILE http://www.w3.org/AudioVideo/

[21] XML Schema a Recommendation of W3C, May 2001, http://www.w3.org/XML/Schema.

[22] MPEG-7 http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm

[23] MPEG-21 http://www.chiariglione.org/mpeg/standards/mpeg-21/mpeg-21.htm

[24] Platform for Internet Content selection (PICS) http://www.w3.org/PICS/

[25] Meta Content Framework (MCF) using XML http://www.w3.org/TR/NOTE-MCF-XML-970624/

[26] Resource Description Framework (RDF) http://www.w3.org/RDF/

[27] XML Topic Maps (XTM) http://www.topicmaps.org/xtm/index.html

[28] Heflin, J., Hendler, J. *"Semantic Interoperability on the Web"*. In Proceedings of Extreme Markup Languages 2000. Graphic Communications Association, 2000. pp. 111-120.

[29] Hunter J., *"Working Towards MetaUtopia - A Survey of Current Metadata Research"*, Library Trends, Organizing the Internet, Edited by Andrew Torok, 52(2), Fall 2003.

[30] METS (Metadata Encoding and Transmission Standard) http://www.jisc.ac.uk/index.cfm?name=techwatch_report_0205

[31] XQuery http://www.w3.org/TR/xquery/

[32] Semantic Web, http://www.w3.org/2001/sw/

[33] RDF Schema Specification, http://www.w3.org/TR/1999/PR-rdf-schema-19990303/

[34] European Chronicles On-line (ECHO), http://pc-erato2.iei.pi.cnr.it/echo/

[35] Ozgur Ulusoy, Ugur Gudukbay, Mehmet Emin Donderler, Ediz Saykol. BilVideo: A Video Database Management System, by Bilkent University Multimedia Database Group, http://www.cs.bilkent.edu.tr/~ediz/bilmdg/bilvideo/

[36] IBM CUEVideo project, 2000, http://www.almaden.ibm.com/projects/cuevideo.shtml

[37] JPEG2000 JPX, http://www.jpeg.org

[38] G. Colyer, K.Ishii, J. Hunter, "Mapping between Dublin Core and JPX (JPEG 2000) Metadata", ISO/IEC JTC1/SC29/WG1 N2736, 15 October, 2002. http://www.jpeg.org/jpeg2000/metadata.html

[39] Hunter J., Drennan J., Little S., "Realizing the Hydrogen Economy through Semantic Web Technologies", *IEEE Intelligent Systems Journal - Special Issue on eScience*, January 2004.

[40] Doerr M., Hunter J., Lagoze C., "Towards a Core Ontology for Information Integration", *Journal of Digital Information, Volume 4 Issue 1*, April 2003.

[41] Hunter J., "Enhancing the Semantic Interoperability of Multimedia through a Core Ontology", *IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on Conceptual and Dynamical Aspects of Multimedia Content Description*, January 2003.

[42] Little S., Geurts J., Hunter J., "Dynamic Generation of Intelligent Multimedia Presentations through Semantic Inferencing", ECDL '02, Rome, September 2002.

[43] DAML+OIL, March 2001. (http://www.daml.org/2001/03/daml+oil-index)

[44] W3C Web Ontology Language (OWL) (2003, March 31). Guide, Version 1.0, W3C Working Draft (http://www.w3.org/TR/owl-guide)

[45] Lagoze C., Hunter J., "The ABC Ontology and Model" (v3.0), *Journal of Digital Information, Vol 2 , Issue 2*, November 2001.

[46] OpenGALEN (2002). (http://www.opengalen.org/)

[47] SNOMED CT (2003). (http://www.snomed.org/)

[48] Denny M., (2002 November 6). Ontology Building: A Survey of Editing Tools, Nov 2002. (http://www.xml.com/pub/a/2002/11/06/ontologies.html)

[49] Sure Y., Erdmann M., Angele J., Staab S., Studer R. and Wenke D., (2002, June). "OntoEdit: Collaborative Ontology Engineering for the Semantic Web", In Proceedings of the first International Semantic Web Conference 2002 (ISWC 2002), Sardinia, Italy. (http://link.springer.de/link/service/series/0558/papers/2342/23420221.pdf)

[50] DARPA Object Service Architecture Web Annotation Service (1998). (http://www.icc3.com/ec/architecture/webannotations.html)

[51] CueVideo, IBM Almaden Research Center (http://www.almaden.ibm.com/projects/cuevideo.shtml)

[52] ENVIE: Extensible News Video Information Exploitation, Informedia (2003) (http://www.informedia.cs.cmu.edu/)

[53] VideoAnnEx Annotation Tool, IBM (http://www.research.ibm.com/VideoAnnEx/)

[54] Ricoh MovieTool, (http://www.ricoh.co.jp/src/multimedia/MovieTool/)

[55] W3C Annotea Project, (http://www.w3.org/2001/Annotea/)

[56] PAX-it Image Annotation Tool, (http://www.paxit.com/paxit/image_annotation.asp)

[57] FilmEd project, DSTC (http://metadata.net/filmed/)

[58] Open Archival Information System (OAIS) Resources, 2002 (http://www.rlg.org/longterm/oais.html)

[59] CEDARS curl exemplars in digital archives,2002 (http://www.leeds.ac.uk/cedars/)

[60] PANDORA Project, National Library of Australia, 2002 (http://pandora.nla.gov.au/index.html)

[61] Society of Motion Picture and Television Engineers (SMPTE) (http://www.smpte.org/)

[62] SMPTE Metadata Dictionary, (http://www.smpte-ra.org/mdd/)

[63] MARC Standard (http://www.loc.gov/marc/)

[64] CDWA Standard (http://www.getty.edu/research/conducting_research/standards/cdwa/)

[65] NISO Z39.87 Standard (http://www.niso.org/standards/)

[66] VRA Core (http://www.vraweb.org/vracore3.htm)

[67] DIG35 Standard (http://www.i3a.org/i_dig35.html)

[68] ROADS project (http://www.ukoln.ac.uk/metadata/roads)

[69] Joint Information Systems Committee (JISC) (http://www.jisc.ac.uk/)

[70] Warwick Framework (http://www.ukoln.ac.uk/metadata/resources/wf.html)

[71] ISMA, Internet Streaming Media Alliance (http://www.isma.tv/home)

Bibliography

Ching-Yung Lin, Belle L. Tseng, Milind Naphade, Apostol Natsev and John R. Smith, “VideoAL: A Novel End-to-End MPEG-7 Automatic Labeling System” , *IEEE Intl. Conf. on Image processing, Barcelona, Sep. 2003.*

Ching-Yung Lin, Belle L. Tseng and John R. Smith, “VideoAnnEx: IBM MPEG-7 Annotation Tool for Multimedia Indexing and Concept Learning,”, *IEEE Intl. Conf. on Multimedia & Expo, Baltimore, July 2003.*

Carlo Meghini, Fabrizio Sebastiani and Umberto Straccia. “A model of multimedia information retrieval”. *Journal of the ACM, 48(5):909-970, 2001.*

SMITH, J. R., AND CHANG, S.-F. 1997. Visually searching the Web for content. *IEEE Multimedia 4*, 3, 12–20.

J. Magalhães, F. Pereira, "Using MPEG standards for multimedia customization", *Signal Processing: Image Communication, 2004.*

Eyvind Fossbakk, Pilar Manzanares, Jose L. Yago, Andrew Perkis, “An MPEG-21 framework for streaming media ”, *in Proc. 2001 IEEE Workshop on Multimedia Signal Processing, Cannes 3-5 October, 2001.*

Y. Wang, Z. Liu and J. Huang, "Multimedia content analysis using audio and visual information," *IEEE Signal Processing Magazine, Vol. 17 n. 6,* November 2000.

van Beek P., Smith J.R., Ebrahimi T., Suzuki T., Askelof J. ,“Metadata-driven multimedia access”, *IEEE Signal Processing Magazine, Vol. 20, n. 2, p. 40-52*, March 2003.

Bormans J., Gelissen J., Perkis A., “MPEG-21: The 21st century multimedia framework”, *IEEE Signal Processing Magazine, Vol. 20, n. 2, p. 53-62*, March 2003.

Gal Ashour, Arnon Amir, Dulce Ponceleon, Savitha Srinivasan, “Architecture for Varying Multimedia Formats”, *ACM Multimedia Workshop*, Marina Del Rey CA USA, 2000.

Zheng Yin, Zhengfang Xu, and Abdulmotaleb El Saddik, “Study of Metadata for Advanced Multimedia Learning Objects”, CCECE 2003 – CCGEI 2003, Montreal, May 2003.

Martin Doerr, Jane Hunter, Carl Lagoze, “Towards a Core Ontology for Information Integration”, *Journal of Digital Information, Vol. 4 Issue 1*, April 2003.

C. Lagoze, J. Hunter, "The ABC Ontology and Model" (v3.0), *Journal of Digital Information, Vol 2 , Issue 2*, November 2001.

J. Hunter, "MetaNet - A Metadata Term Thesaurus to Enable Semantic Interoperability Between Metadata Domains", *Journal of Digital Information, Special Issue on Networked Knowledge Organisation Systems, Volume 1, Issue 8*, April 2001.

J. Hunter, F. Nack, "An Overview of the MPEG-7 DDL Proposals", *Signal Processing: Image Communication Journal, Special Issue on MPEG-7, Vol 16 pp 271-293*, 2000.

van Harmelen, F., " The semantic web: what, why, how, and when", *IEEE Distributed Systems Online, Vol. 5 n. 3 p. 1-4* , March 2004.

Greg Goth, "Multimedia Search: Ready or Not?", *IEEE Distributed Systems Online, Vol. 5, n.7*, July 2004.

Catarci, T.; Donderler, M.E.; Saykol, E.; Ulusoy, O.; Gudukbay, U., "BilVideo: a video database management system", *IEEE Multimedia Vol. 10, n. 1, p. 66-70*, March 2003.