An Overview on Current Emerging Multimedia Standards

Content List

Introduction 3
Image Compression Standards 4
JPEG 2000 4
PNG 5
Audio-Visual Broadcasting Standards 6
UMIDS 6
MHP 7
P/Meta 7
TV Anytime 8
DC 8
Content-oriented Standards 9
MPEG-4 9
XMT 11
MPEG-7 12
MPEG-21 13
MHEG-5 / MHEG 6 15
Web-oriented Standards 15
HyTime 15
SMIL 16
SVG 18
XHTML 18
XSL 19
CSS 19
Additional Web-oriented Standards for the semantic Web 19
RDF 19
OIL 20
References 22

Introduction

The world of digital media is changing swiftly. A couple of years ago the digitisation of media meant rapid developments with respect to storage, reproduction and transportation of information, mainly served by advances in hardware technology (such as networkable and storage-intensive computers, CD-ROMs, DVD, video and camcorders, IP-telephony, Webcams, synthesizers, Musical Instrument Digital Interface (MIDI), Digital Audio Workstations (DAW), e-books, and so on). As a result, our social ways to exchange information have already deeply changed. More and more people are acquainted with the creative process of producing and the technical intricacies of receiving audio-visual information and, as exemplified by the popularity of the Internet, use these new skills extensively.

Moreover, the result of the ubiquitous World Wide Web together with high-speed networks is that multimedia is rapidly moving into real-world applications such as eBusiness, training, education and communications. Media content description, indexing and retrieval, and content delivery are among the component technologies required to build such applications. Also the convergence of television, computers and the Internet brings out media delivery scenarios that were not feasible before. Hence, today's service/content providers will need to embrace new business models in order to attract and sustain consumers.

However, the flood of mixed media data can transform the web, like any large information space, into a dismal place. If that should be avoided, we have to understand and handle context. That requires more than the characterisations of audio-visual information on a perceptual level using objective measurements such as those based on image or sound processing or pattern recognition, which are mainly provided by current research.

Multimedia standards aim to define open specifications of various types of multimedia information. MPEG-7, TV-Anytime, and Open Cable are examples of different multimedia standardization efforts. Such open standards facilitate fast and efficient retrieval, and promote interoperability among applications. While that is the Holy Grail, development of media applications in a distributed environment with competing emerging standards offers its share of challenges.

This report aims at providing an overview on current multimedia standards by describing the environments they emerged from and the problems they are addressing. However, this overview will mainly cover the latest developments, which means it will not describe already established standards such as GIF, JPEG, Quicktime, MPEG1 and MPEG2, MP3, etc.

Image Compression Standards

JPEG 2000

In March 1997 a new call for contributions were launched for the development of a new standard for the compression of still pictures, the JPEG2000. This project, JTC 1.29.14 (15444), was intended to create a new image coding system for different types of still pictures (bi-level, grey-level, colour, multi-component), with different characteristics (natural images, scientific, medical, remote sensing, text, rendered graphics, etc) allowing different imaging models (client/server, real-time transmission, image library archival, limited buffer and bandwidth resources, etc) preferably within a unified system. This coding system should provide low bit-rate operation with rate-distortion and subjective image quality performance superior to existing standards.
The standard is intended to complement and not to replace the current JPEG standards. The standard is expected to reach the International Standard (IS) stage in December 2000 The standardisation process, which is co-ordinated by the JTC1/SC29/WG1 of ISO/IEC has already (as of May 2000) produced the Working Draft (WD) and the Committee Draft (CD) [1] documents. JPEG 2000 refers to all parts of the standard - current proposals are for 6 parts, with part 1 (the core) to be delivered and agreed as a full ISO International Standard by the end of the year 2000 [2] The parts are:

Part 1, JPEG 2000 Image Coding System
Part 2, Extensions (adds more features and sophistication to the core)
Part 3, Motion JPEG 2000
Part 4, Conformance
Part 5, Reference software (currently Java and C implementations are scoped)
Part 6, Compound Image file format (for pre-press and fax like applications)

Some of the features that this standard possesses are the following:

- Superior low bit-rate performance: This standard offers performance superior to the current standards at low bit-rates (e.g. below 0.25 bpp for highly detailed grey-scale images). This significantly improved low bit-rate performance is achieved without sacrificing performance on the rest of the rate-distortion spectrum.
- Continuous-tone and bi-level compression: It is desired to have a coding standard that is capable of compressing both continuous-tone and bi-level images. The standard strives to achieve this with similar system resources.
- Lossless and lossy compression: It is desired to provide lossless compression naturally in the course of progressive decoding. Examples of applications that can use this feature include medical images, image archival applications, network applications and pre-press imagery.
- Progressive transmission by pixel accuracy and resolution: Progressive transmission that allows pictures to Be reconstructed with increasing pixel accuracy or spatial resolution is essential for many applications. This feature allows the reconstruction of images with different resolutions and pixel accuracy, as needed or desired, for different target devices.

- Region of Interest Coding: Often there are parts of a picture that are more important than other parts of it. This feature allows user-defined Regions-Of-Interest (ROI) in the image to be compressed with better quality than the rest of the image.
- Robustness to bit-errors: It is desirable to consider robustness to bit-errors while designing the code stream. This is an important feature for wireless communication applications.

PNG

The PNG (Portable Network Graphics) format provides a portable, legally unencumbered, well-compressed, well-specified standard for lossless bitmapped image files (raster images). Although the initial motivation for developing PNG was to replace GIF but can also replace many common uses of TIFF. The design provides indexed-colour, greyscale, and truecolour support, plus an optional alpha channel. It provides both full file integrity checking and simple detection of common transmission errors. Also, PNG can store gamma and chromaticity data for improved colour matching on heterogeneous platforms. It is designed to work well in online viewing applications, such as the World Wide Web, so it is fully streamable with a progressive display option., with minimal cost to developers.

GIF features retained in PNG include:

- Indexed-colour images of up to 256 colours.
- Streamability: files can be read and written serially, thus allowing the file format to be used as a communications protocol for on-the-fly generation and display of images.
- Progressive display: a suitably prepared image file can be displayed as it is received over a communications link, yielding a low-resolution image very quickly followed by gradual improvement of detail.
- Transparency: portions of the image can be marked as transparent, creating the effect of a non-rectangular image.
- Ancillary information: textual comments and other data can be stored within the image file.
- Complete hardware and platform independence.
- Effective, 100% lossless compression.

Important new features of PNG, not available in GIF, includ**e:**

- Truecolour images of up to 48 bits per pixel.
- Greyscale images of up to 16 bits per pixel.
- Full alpha channel (general transparency masks).
- Image gamma information, which supports automatic display of images with correct brightness/contrast regardless of the machines used to originate and display the image.
- Reliable, straightforward detection of file corruption. Faster initial presentation in progressive display mode.

PNG is designed to be:

- Simple and portable: developers should be able to implement PNG easily.
- Legally unencumbered: to the best knowledge of the PNG authors, no algorithms under legal challenge are used. (Some considerable effort has been spent to verify this.)
- Well compressed: both indexed-colour and truecolour images are compressed as effectively as in any other widely used lossless format, and in most cases more effectively.
- Interchangeable: any standard-conforming PNG decoder must read all conforming PNG files.
- Flexible: the format allows for future extensions and private add-ons, without compromising interchangeability of basic PNG.
- Robust: the design supports full file integrity checking as well as simple, quick detection of common transmission errors.

The W3C does not currently have plans to work on any future versions of PNG, though were the necessity to arise, and were an activity in that area to receive the support of Members, the Consortium could in principle support some future activity. Further detailed information about PNG can be obtained from [3].

Audio-Visual Broadcasting Standards

UMIDS

The Dynamic Metadata Dictionary—Unique Material Identifiers (UMIDs) is a development of the Society of Motion Picture and Television Engineers (SMPTE).

A Unique material Identifier provides a reference for all captured audio-visual content units in a clip or shot, so that particular a content unit can always be located either locally, remotely or on a archive storage medium. The UMIDS provides for the link between the essence (video, audio, graphics, stills etc.) and the metadata and generates a time code and date (time-axis) for synchronising this data. UMIDs can also be used to track copyright information about the essence. UMIDs are automatically generated by the systems and support granularity up to the level of shots and video and audio frames. The UMID structure contains a set of components that each represents a key-aspect of the audio- or video essence. There are two sorts of UMIDs: the basic UMID, with the minimal components necessary for the unique identification (the essential metadata) and the extended UMID that provides information on the creation time and date, recording location and the name of the organisation and the maker (the signature metadata). Together these basic and extended UMID have a defined length of 64 bytes.

UMID in itself includes some of the most important metadata about a content clip. And is an important identification standard for the audio-visual production environment. Supports granularity (important in the process of creation and authoring). Provides a link between metadata and essence. There is high industrial support by some of the most important manufacturers of audio-visual equipment and media-management systems.

Already included in most of the (forthcoming) metadata standards for the audio-visual production environment, such as MPEG-7. For detailed information please see [4].

MHP

The Multimedia Home Platform (MHP) as part of the Digital Video Broadcasting (DVB) Project.

MHP is a series of measures designed to promote the harmonised transition from analogue TV to a digital interactive multimedia future. Based around a series of Java APIs (Application Programming Interfaces) for DVB set-top-boxes, MHP promises to provide a domestic platform, which will facilitate convergence like no other DVB specification.

Typical MHP resources are MPEG processing, I/O devices, CPU, memory and a graphics system. The core of the MHP is based around a platform known as DVB-J. This includes the Java virtual machine as originally specified by Sun Microsystems. A number of software packages provide generic application program interfaces (APIs) to a wide range of features of the platform. MHP applications access the platform only via these specified APIs. MHP implementations are required to perform a mapping between these specified APIs and the underlying resources and system software.

Some 1000 pages long, MHP defines the application lifecycles, security and data download mechanisms for enhanced broadcast, interactive and indeed full Internet applications. For detailed information or the MHP standard document please see [5].

P/Meta

The P/Meta Standard developed by the Production Technology Management Committee (PMC) of the European Broadcasting Union (EBU), shall provide a common exchange framework and a format between members (and others) that builds on SMPTE outputs and the additional insights provided by the British Broadcasting Corporation's (BBC) Standard Media Exchange Framework (SMEF).

Tasks

1.To establish understanding between EBU members of the media-related data interchange requirements of media commissioner/publishers (broadcasters), suppliers (producers) and consumers, using the BBC Standard Media Exchange Framework (SMEF) as the core information architecture.

2.To validate and extend the SMEF model as appropriate against members' requirements in terms of data and process, noting local synonyms (or translations), to create an 'E-SMEF'. This would extend the thinking to the development of a commercial process framework for exchange of media between EBU members.

3.Using E-SMEF, to apply emerging SMPTE metadata standards to the production and broadcast or distribution process, and study the feasibility of creating and adopting common exchange formats for essence and metadata.

4.To establish understanding of the use of unique identifiers in metadata e.g. the SMPTE UMID, as a crucial linkage tool between unwrapped data (metadata) and wrapped or embedded metadata in media files or streams, and develop protocols for their management between members.

5.As an aid to commercial and system interoperability between members, and in co-operation with standards bodies in related industries such as music and print publishing, to collate all relevant unique identifier schemes and map them against each other.

This could be in collaboration with the EU INDECS project and the DOI Foundation, and extend to cover their data models too. For further information, not only on the other production but also broadcasting and networking activities, look at [6].

TV Anytime

The TV Anytime Forum is an association of organizations that develops specifications to enable audio-visual and other services based on mass-market, high-volume digital storage.

Digital broadcasting offers the opportunity to provide value-added interactive services, that allow end users to personalize and control the material of interest, an evolution of TV into an integrated entertainment / information gateway. Anticipating this evolution, a group of organizations has formed the TV Anytime Forum to direct development of standards, tools and technologies towards a common framework. In their call for contributions [7] they identified metadata as one of the key technologies enabling their vision.

The role of metadata within TV Anytime is crucial: it will enable selection of content - e.g. via searching or filtering - capturing of content items for later viewing or further processing, navigation through stored and remote content, and access of content items. In addition, it may enable content referencing, user profiling and preferences, and even support the management of rights.

TV-Anytime is currently the most important test application within MPEG-7 (see below). Further detailed information about TV Anytime can be obtained from [8].

DC

The Dublin Core Metadata Initiative (DC) defined a metadata element set intended to facilitate the discovery of electronic resources. This standard supports a number of description communities mainly from the digital library world, in particular due to its link to the Resource Description Framework (RDF), but also proves influential in the P/Meta standard.

Each Dublin Core element is defined using a set of ten attributes from the ISO/IEC 11179 [9] standard for the description of data elements. These include:

- 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)

- Datatype - 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 definitions provided here include both the conceptual and representational form of the Dublin Core elements. The Definition attribute captures the semantic concept and the Datatype and Comment attributes capture the data representation.
Each Dublin Core definition refers to the resource being described. A resource is defined in [10] as "anything that has identity". For the purposes of Dublin Core metadata, a resource will typically be an information or service resource, but may be applied more broadly.
Further information about DC can be retrieved from [11].

Content-oriented Standards

MPEG-4

MPEG-4 is a new standard providing core technologies for efficient storage, transmission and manipulation of video data in multimedia environments. Outstanding novelties of MPEG-4 are the philosophy of considering scenes as compositions of audio-visual objects (AVO's), the support of hybrid coding of natural video and 2D/3D graphics in a common context (e.g. virtual 3D worlds) and the provision of advanced system and interoperability capabilities. Especially interactive computer and communication systems will benefit from these particular advantages of MPEG-4.
Though the results of MPEG-1 and MPEG-2 served well for wide-ranging developments in such fields as interactive video, CD-ROM, and digital TV, it became soon apparent that the requirements of the now spreading multimedia applications aimed beyond the established achievements. Thus, MPEG started in 1993 with work to provide the standardised technological elements enabling the integration of the production, distribution, and content access paradigms of such fields as digital television, interactive graphics applications (synthetic content) and interactive multimedia (World Wide Web, distribution of and access to content). MPEG-4 version 1, formally called ISO/IEC 14496 [12], is available as international standard since December 1998. The fully backward compatible extensions under the title of MPEG-4 Version 2 were frozen at the end of 1999, to acquire the formal International Standard Status early 2000. Some work, on extensions in specific domains, is still in progress.
The aim of MPEG-4 is to provide a set of technologies to satisfy the needs of authors, service providers and end users alike, by avoiding the emergence of a multitude of proprietary, non-interworking formats and players.
The standard should be used to allow the development of systems, which can be configured for a vast number of applications (among others, real time communications, surveillance and mobile multimedia). This is achieved by providing standardised ways to:

- interact with the material based on encoding units of aural, visual or audio-visual content, called "media objects". These media objects can be of natural or synthetic

origin; this means they could be recorded with a camera or microphone, or generated with a computer;
- interact with the content, based on the description of the composition of these objects to create compound media objects that form audio-visual scenes; in a way one can understand the composition of audio-visual MPEG-4 objects as the representation of the real world, where spatial and temporal relations between objects allow the user to interact with these objects in a way similar to everyday usage,
- integrate different data types, allowing the harmonisation of natural and synthetic objects, the usage of 2D and 3D respectively mono and stereo video or multiview video, and mono, stereo and multi channel audio, etc.;
- multiplex and synchronise the data associated with media objects, so that they can be transported over network channels providing a QoS
- interact with the audio-visual scene generated at the receiver's end, e.g. to manipulate some characteristics of an object, to access selected parts of a scene, to remove an object from one scene and to integrate it into another, etc.

To meet the swift technological progress within multimedia, MPEG-4 developed the 'Syntactic Description Language' (MSDL). This approach not only aimed at the integration of new algorithms through defined tools, but also that at any time new tools, techniques, and concepts could be adopted which should have provided improved or new functionality. In other words MSDL was the way to guarantee the flexibility of the standard, preventing eventual obsolescence and narrowness of scope. However, MSDL did not become part of the standard and was replaced by the 'Binary Format for Scene Description (BIFS)', which filled a limited part of the role.
The major extensions of MPEG-4 over MPEG-2, with respect to the three main foci (i.e. content/interaction, flexibility/extensibility, and integration) are:

- A standard functionality, e.g. synchronisation of audio and video, modi for short delay and usage via networks
- Scalability, where in particular the content based scalability is of importance, since this is the mechanism for prioritisation of objects within a scene
- Content based manipulation of a bitstream without transcoding
- Content based access (indication, hyperlink, request, up- and downloading, deleting, pre-viewing, etc.)
- Content combination (e.g. text and graphic overlay, mixing of synthetic and video and/or audio data, etc.)
- Efficient coding of several streams at the same time (e.g. stereo video or several views of the same event, etc.)
- An improved efficiency in coding (improvement of data quality with low bitrates compared to existing standards – e.g. H.263)
- A robustness in error susceptible environments due to an elasticity towards remaining errors (a selective look-ahead error correction, error control, error masking, etc.)
- An improved random access on parts of an audio-visual sequence.

Like its predecessors, MPEG-4 is concerned with streams. Since MPEG-4 subdivides audio-visual content into objects, the standardised characteristics of a stream are concerned with multiplexing, demultiplexing and the synchronisation of multiple steams. Figure 2 describes the relationship between different streams based on the MPEG-4 ‘System Layer Model.’

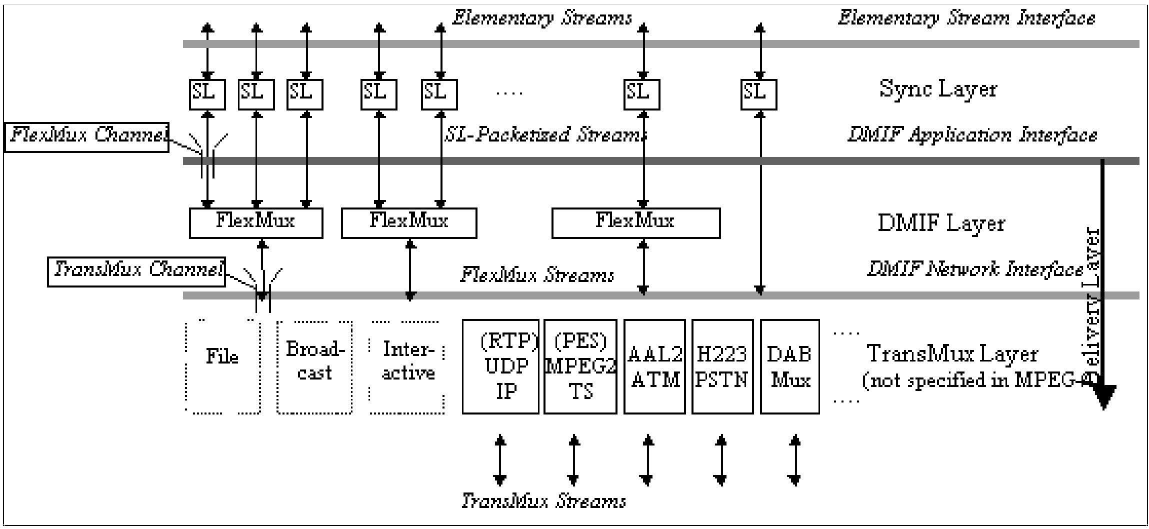

Figure 2 The MPEG4 System Layer Model [12]

MPEG-4 allows attaching meta-data to objects about their content. Users of the standard can use this ‘Object Content Information (OCI)’ datastream to send textual information along with MPEG-4 content. It is also possible to classify content according to pre-defined tables, which will be defined outside of MPEG. However, there is no structure and format defined for the meta-data.

The abandonment of MPEG-3 caused much speculation about the next number. Should it be 5 (the next) or 8 (creating an obvious binary pattern)? MPEG, however, decided not to follow either logical expansion of the sequence, but chose the number of 7 instead. So MPEG-5 and MPEG-6 are, just like MPEG-3, not defined. For further details about the different MPWEG standards please see [13].

XMT

The MPEG-4 specification, in its binary form, basically provides a conformance point between the sender and the receiver of the content. As such, the coded form cannot be "reverse-engineered" in a consistent manner to represent the content author's original intentions.

The XMT has been designed to provide an exchangeable format between content authors whilst preserving the author's intentions in a high-level textual format. In addition to providing a suitable, author-friendly abstraction of the underlying MPEG-4 technologies, another important consideration for the XMT design was to respect existing practices of content authors such as the Web3D X3D, W3C SMIL and HTML.

The XMT framework consists of two levels of textual syntax and semantics: the XMT-A format and the XMT-Ω format. The XMT-A is an XML-based version of MPEG-4 content which closely mirrors its binary representation.

The goal of the XMT-A format is to provide a deterministic one-to-one mapping to ISO/IEC 14496:1999 binary representations and to be interoperable with the X3D [14], which is being developed for VRML 200x (X3D) by the Web 3D Consortium. It contains a subset of the X3D, as well as the X3D-like representations of MPEG-4 specific features such as Object descriptors (OD), BIFS update commands and 2D composition. The ODs are used to associate scene description components to the actual elementary streams that contain the corresponding coded data while the BIFS update commands serve as a mechanism that allows a scene to be remotely manipulated, and portions of the scene to be progressively streamed in order to reduce bandwidth requirements.

The XMT-Ω is a high-level abstraction of MPEG-4 features designed based on the W3C SMIL [15], an XML-based language that allows authors to create dynamic, interactive multimedia presentations. Using SMIL, authors can describe the temporal behaviour and layout of multimedia presentations, as well as associate hyperlinks with the media objects in the presentation. For every XMT-Ω element, there is a mapping to a sequence of XMT-A elements.
Moreover, for those authors who wish to control the implementation of certain portions of their presentation, the XMT provides an escape mechanism from XMT-Ω to XMT-A. The escape mechanism enables content authors to mix and match the two formats, XMT-Ω and XMT-A, overriding the default, standard mapping the XMT provides.

MPEG-7

The MPEG-7 standard, also known as "Multimedia Content Description Interface", aims at standardizing tools for efficient, interoperable, and extensible description of multimedia material in order to enable and facilitate applications that process, manage, and extract useful knowledge from the pervasively available multimedia material. MPEG-7 is facing this challenge by standardizing Descriptors (Ds) - features of the multimedia data, Description Schemes (DSs) - semantic groupings of Ds and DSs, a Description Definition Language (DDL) to define and extend Ds and DSs, and a coded representation to enable efficient transmission, storage, and access of descriptions [16]. MPEG-7 has recently entered the Committee Draft phase and it is expected to become an International Standard by September 2001.
The key contributions of MPEG-7, which set it apart from other multimedia metadata efforts, are the following:

- support of the description of the content depicted on the multimedia material and its management, such as creation, usage, and media storage
- description of other facets of multimedia material, such as summaries for efficient browsing; variations for personalized access; multimedia collections, classification schemes, and content models for content organization and knowledge representation; and user preferences in consumption of content for user interaction
- an integrated framework to describe the spatio-temporal structure and the conceptual aspects associated with multimedia material
- a comprehensive set of descriptors for features of audio-visual data such as colour, texture, motion, shape, speech recognition, instrument timbre, and sound effects, among others

- a mechanism to extend and specify new descriptors and description schemes using the MPEG-7 DDL, which is based on XML-Schema [17].

Existing multimedia applications will greatly benefit from adopting the MPEG-7 standard. The standard set of MPEG-7 descriptors and description schemes provides the base for interoperability and communication among the different components of the same applications and applications from different vendors launching the complete distribution of application tasks and enhancing the interaction among different applications (e.g. distributed editing and description of multimedia). The generic-purpose descriptors and description schemes covering a wide range of aspects in the description of multimedia permit expansive functionality and capabilities for multimedia applications (e.g. user-personalized applications for multimedia browsing and retrieval). The mechanism for specifying descriptors and description schemes from scratch or extending standard ones satisfies the need of applications for specialization to achieve current and future objectives and requirements. This is important since it is impossible for MPEG-7 to standardize all possible multimedia description tools needed for current and future multimedia applications.

Furthermore, it opens the horizon for new and exciting multimedia applications towards the universal and personalized access, retrieval, filtering, and browsing of multimedia information. For detailed documentation about MPEG-7, please refer to [18, 19].

MPEG-21

The aim of starting MPEG-21 [20] has been:

- to understand if and how these various components of a multimedia infrastructure fit together
- to discuss which new standards may be required, if gaps in the infrastructure exist and, once the above two points have been reached and
- to actually accomplish the integration of different standards.

The purpose of this effort is to globalise consumer choice. There are very few if any standard technologies that would allow the implementation of consumer/device capability. There is a need to produce specifications of standardized interfaces and protocols which allow consumers to access the widest possible variety of content providers, both commercial and non-commercial. Putting the consumer as the focal point will expand the business opportunities on a global basis. Content and service providers of all sizes will have the opportunity to reach a previously unreachable and/or fragmented consumer market.

A multimedia framework is required to support this new type of commerce. Such a framework requires that a shared vision or roadmap is understood by its architects to ensure that the systems that deliver e-content are interoperable and that transactions are simplified and, if possible, automated. This should apply to the infrastructure requirements for content delivery, content security, rights management, secure payment, and the technologies enabling them – and the list is probably not exhaustive.

The scope of MPEG-21 could therefore be described as the integration of the critical technologies enabling transparent and augmented use of multimedia resources across a wide range of networks and devices to support the following functions:

- Content creation
- Content production
- Content distribution
- Content consumption and usage
- Content representation
- Intellectual property management and protection
- Content identification and description
- Financial management
- User privacy
- Terminals and network resource abstraction
- Event reporting

From its background in key technology and information management standards related to the management and delivery of multimedia content, MPEG is uniquely positioned to initiate such an activity. However, it is recognized that the integration of such disparate technologies can only be achieved by working in collaboration with other bodies.
This may look like a very complex project to address, but it is believed that the enabling factor is the integration of a set of critical technologies implementing the functionalities mentioned above, that can be brought to bear on the problem.
MPEG-21 takes the following statement to describe its overall vision:

- To enable transparent and augmented use of multimedia resources across a wide range of networks and devices.

Its goal is to create an interoperable multimedia framework by:

- Understanding how the components of the framework are related and identify where gaps in the framework exist;
- Achieving the integration of standards to support harmonized technologies for the management of multimedia content;
- Developing new specifications which allow:
- access, (re)use of and interaction with multimedia objects across networks and/or capable devices
- the implementation of multiple business models including those requiring the management of automated rights and payments transactions throughout the value chain
- the privacy of content users to be respected.

For a detailed description of MPEG-21 see [20].

MHEG-5 / MHEG 6

SVGMHEG is an acronym for 'Multimedia and Hypermedia information coding Experts Group'. This group is developing within ISO several standards, which deal with the coded representation of multimedia/hypermedia information.
The MHEG series of standards specify the coded representation of multimedia/hypermedia information objects (MHEG objects) for the interchange of applications or systems as final-form units, and by any means of interchange from storage devices, local networks, to telecommunication or broadcast networks.
MHEG-5 is the fifth part of the MHEG standard suite. It was developed to support the distribution of multimedia/hypermedia applications in a client/server architecture across platforms of different types and brands. MHEG-5 defines a coded representation of multimedia/hypermedia interactive applications, i.e. the syntax and the associated semantics allowing an author to build applications with the following features:

- text, graphic and audio-visual output components,
- input components such as entry fields, buttons, sliders,

and behaviour based on events that trigger actions applied to these components.
The part 6 (MHEG-6) was created to extend MHEG-5 by adding computing (data processing) and communication functions with the external environment, such as servers, local device drivers. For all that, MHEG-6 represents Program and Applet objects using the JavaTM bytecode representation as interchange format, the Java virtual machine to interpret this format, and a basic set of Java APIs (java.lang, java.util, java.io) to provide access to basic platform functionality. A programming interface enables programs to have access to MHEG-5 objects and to the services of the MHEG-5 interpreter: the MHEG-5 API which is defined as a Java specification and consists of one Java package, called iso.mheg5. MHEG-6 is mainly intended to support the distribution of interactive retrieval (client/server) applications running on limited resources set-top-boxes involving asymmetrical channels with audiovisual streams on the downward channel.
MHEG-5/6 is mainly focused on providing solutions for multimedia applications, addressing problems such as synchronisation and control of information streams, which is why MHEG-5/6 is well suited to applications such as "Video On Demand" and interactive TV. Thus, MHEG-5/6's vision was of paramount importance for the development of descriptors and description schemes in the domain of interactive TV and video on demand. However, the current influence of this technology might be doubted by looking at the developments in frameworks such as TV Anytime. See also [21].

Web-oriented Standards

HyTime

HyTime [22], is a SGML based standard framework for integrated open hypermedia. HyTime documents are by definition SGML documents and they entirely conform to SGML. In terms of functionality, however, HyTime extends the power of SGML in many ways.

HyTime offers a list of definitions for element types, so called ‘Architectural Forms‘. These forms or classes allow the modelling of components for hypermedia documents, i.e. hyperlinks or event schedules. These element classes allow multiple inheritance, which means that elements can inherit semantic and syntactic features not only from the governing DTD, but also from any number of other DTDs, which leads to a much greater flexibility compared to standardised DDTs for hypermedia documents..
The strongest asset of HyTime is the support of linking and scheduling of documents in time and space. HyTime provides a standardised interface for the use of links, i.e. clink (Contextual link), ilink (independent link), agglink (aggregate traversal), query link.
However, though HyTime is one of the most powerful standards for general information management, it is perceived as too generic and thus difficult to understand and handle.

SMIL

W3C's Synchronized Multimedia Activity has focused on the design of a new language for choreographing multimedia presentations where audio, video, text and graphics are combined in real-time. The language, the Synchronized Multimedia Integration Language (SMIL) is written as an XML application and is currently a W3C Recommendation [23].
Simply put, it enables authors to specify what should be presented when, enabling them to control the precise time that a sentence is spoken and make it coincide with the display of a given image appearing on the screen.
The basic idea is to name media components for text, images, audio and video with URLs and to schedule their presentation either in parallel or in sequence. A typical SMIL presentation has the following characteristics:

- The presentation is composed from several components that are accessible via URL's, e.g. files stored on a Web server.
- The components have different media types, such as audio, video, image or text. The begin and end times of different components are specified relative 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.

The following are examples of applications that can be created using SMIL:

- Photos taken with a digital camera can be coordinated with a commentary
- Training courses can be devised integrating voice and images.
- A Web site showing the items for sale, might show photos of the product range in turn on the screen, coupled with a voice talking about each as it appears.
- Slide presentations on the Web written in HTML might be timed so that bullet points come up in sequence at specified time intervals, changing colour as they become the focus of attention.

- On-screen controls might be used to stop and start music.

The W3C (World Wide Web Consortium) recently announced its intentions to make SMIL-2 a W3C recommendation. SMIL-2 is an attempt to repackage the SMIL-1.0 language as a set of modules that can be integrated into any other XML language where SMIL-related functionality is required and to extend the functionality of the existing SMIL-1.0 language.

According to the official W3C descriptions, SMIL 2.0 has the following design goals:

- Extending the functionality contained in SMIL 1.0. SMIL 1.0 has been a very popular format for describing web-based multimedia productions. (Over 150,000 SMIL-compatible players have been released during the past two years). SMIL 2.0 will build on SMIL 1.0, primarily by extending the temporal constructs available to the user, and by adding new functionality to support media transitions and object animations. SMIL 2.0 also provides a more formally defined timing and interaction specification.
- Reusable markup modules. In order to make timed-based manipulation objects available to other XML languages, SMIL 2.0 adopts a modularisation approach that define functionality across ten basic modules:
  - Animation: control of run-time attribute value manipulation, including object movement;
  - Content Control: control over the conditional evaluation of document parts, based on both static predefined attributes and user-defined custom test attributes;
  - Layout: facilities for the definition and management of visual and audio rendering spaces;
  - Linking: facilities for the definition and management of (temporal) anchors and links;
  - Media Objects: elements and attributes for controlling the (partial) activation of media objects;
  - Metadata: elements and attributes for defining XML-compliant meta data in a presentation;
  - Structure: elements for defining the over-all structure of a presentation, including basic content containers;
  - Timing and Synchronization: basic and advanced elements and attributes for defining the coordinated activation of objects;
  - Transitions: elements and attributes to define and control transitions on media objects; and
  - XML Language Integration: rules for integrating SMIL modules into other XML languages.
- SMIL Profiles. Where SMIL 1.0 defined a single language profile that was geared to a separate media player model, SMIL 2.0 allows the integration of SMIL modules in terms of various profiles. These profiles include a new language profile, an integration of SMIL timing into XHTML documents (XHTML+SMIL) and a profile geared to integrating basic timing for minimal environments (such as mobile telephones).

- Accessibility Support. Part of the SMIL mandate was to extend and improve the accessibility features of SMIL 1.0. In SMIL 2.0, authors may include closed captions closed audio in their SMIL presentations.

A detailed description of SMIL 2.0 can be found at [24].

SVG

W3C’s SVG (Scalable Vector Graphics) is a language for describing two-dimensional graphics in XML [25]. SVG allows for three types of graphic objects: vector graphic shapes (e.g., paths consisting of straight lines and curves), images and text. Graphical objects can be grouped, styled, transformed and composited into previously rendered objects. The feature set includes nested transformations, clipping paths, alpha masks, filter effects and template objects.

SVG drawings can be interactive and dynamic. Animations can be defined and triggered either declaratively (i.e., by embedding SVG animation elements in SVG content) or via scripting. Sophisticated applications of SVG are possible by use of supplemental scripting language with access to SVG's Document Object Model(DOM), which provides complete access to all elements, attributes and properties. A rich set of event handlers such as "onmouseover" and "onclick" can be assigned to any SVG graphical object. Because of its compatibility and leveraging of other Web standards, features like scripting can be done on XHTML and SVG elements simultaneously within the same Web page.

SVG is a language for rich graphical content. For accessibility reasons, if there is an original source document containing higher-level structure and semantics, it is recommended that the higher-level information be made available somehow, either by making the original source document available, or making an alternative version available in an alternative format which conveys the higher-level information, or by using SVG's facilities to include the higher-level information within the SVG content.

A detailed description of SVG is available at [26].

XHTML

W3C’s XHTML (The Extensible HyperText Markup Language) is a family of current and future document types and modules that reproduce, subset, and extend HTML 4 [HTML]. XHTML family document types are XML based, and ultimately are designed to work in conjunction with XML-based user agents. The details of this family and its evolution are discussed in more detail in the section on Future Directions.

XHTML 1.0 [27, 28] is the first document type in the XHTML family. It is a reformulation of the three HTML 4 document types as applications of XML 1.0 [XML]. It is intended to be used as a language for content that is both XML-conforming and, if some simple guidelines are followed, operates in HTML 4 conforming user agents. Developers who migrate their content to XHTML 1.0 will realize the following benefits:

- XHTML documents are XML conforming. As such, they are readily viewed, edited, and validated with standard XML tools.
- XHTML documents can be written to operate as well or better than they did before in existing HTML 4-conforming user agents as well as in new, XHTML 1.0 conforming user agents.

- XHTML documents can utilize applications (e.g. scripts and applets) that rely upon either the HTML Document Object Model or the XML Document Object Model [DOM].
- As the XHTML family evolves, documents conforming to XHTML 1.0 will be more likely to interoperate within and among various XHTML environments.

The XHTML family is the next step in the evolution of the Internet. By migrating to XHTML today, content developers can enter the XML world with all of its attendant benefits, while still remaining confident in their content's backward and future compatibility. For details, please see [27, 28].

XSL

XSL is a language for expressing stylesheets. It consists of two parts:

- XSL Transformations (XSLT): a language for transforming XML documents [29]
- An XML vocabulary for specifying formatting semantics (XSL Formatting Objects)

An XSL stylesheet specifies the presentation of a class of XML documents by describing how an instance of the class is transformed into an XML document that uses the formatting vocabulary. For background information on style sheets, see the Web style sheets resource page. XSL is developed by the W3C XSL Working Group whose charter is to develop the next version of XSL.

CSS

Cascading Style Sheets (CSS) is a simple mechanism for adding style (e.g. fonts, colours, spacing) to Web documents. Style sheets describe how documents are presented on screens, in print, or perhaps how they are pronounced. W3C has actively promoted the use of style sheets on the Web since the Consortium was founded in 1994. The Style Sheets Activity has produced two W3C Recommendations [30, 31], which are widely, although not consistently, implemented in browsers.
By attaching style sheets to structured documents on the Web (e.g. HTML), authors and readers can influence the presentation of documents without sacrificing device independence or adding new HTML tags.

Additional Web-oriented Standards for the semantic Web

RDF

The Resource Description Framework (RDF) encourages the view of "metadata as being data" by using XML (the extensible Markup Language) as its encoding syntax. The resources being described by RDF are, in general, anything that can be named via a URI (Uniform Resource Identifier). The broad goal of RDF is to define a mechanism for describing resources that makes no assumptions about a particular application domain, nor defines the semantics of any application domain. The definition of the mechanism should be domain neutral, yet the mechanism should be suitable for describing information about any domain.

At the core, RDF data consists of nodes and attached attribute/value pairs. Nodes can be any web resources (pages, servers, basically anything for which you can give a URI), even other instances of metadata. Attributes are named properties of the nodes, and their values are either atomic (text strings, numbers, etc.) or other resources or metadata instances. In short, this mechanism allows us to build labelled directed graphs.

The essence of RDF is the model of nodes, attributes, and their values. In order to store instances of this model into files or to communicate these instances from one agent to another, a graph serialization syntax is needed, i.e. XML. RDF and XML are complementary; there will be alternate ways to represent the same RDF data model, some more suitable for direct human authoring.

RDF in itself does not contain any predefined vocabularies for authoring metadata. However, resource description communities require the ability to say certain things about certain kinds of resources. For describing bibliographic resources, for example, descriptive attributes including "author", "title", and "subject" are common. For digital certification, attributes such as "checksum" and "authorization" are often required. The declaration of these properties (attributes) and their corresponding semantics are defined in the context of RDF as an RDF schema. 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 (books, Web pages, people, companies, etc.).

Anyone can design a new vocabulary; the only requirement for using it is that a designating URI is included in the metadata instances using this vocabulary. This use of URIs to name vocabularies is an important design feature of RDF: many previous metadata standardization efforts in other areas have foundered on the issue of establishing a central attribute registry. RDF permits a central registry but does not require one.

For a detailed description of RDF and RDF-Schema see [32, 33]

OIL

The second generation Web, that we could call the "Knowledgeable Web", aims at machine processable meaning of information. This coincides with the vision that Tim Berners-Lee calls the Semantic Web in his recent book "Weaving the Web", and for which he uses the slogan "Bringing the Web to its full potential" (www.semanticweb.org). The Knowledgeable Web will enable intelligent services such as information brokers, search agents, information filters etc. Ontologies will play a crucial role in enabling the processing and sharing of knowledge between programs on the Web. Ontologies are generally defined as a "representation of a shared conceptualisation of a particular domain". They provide a shared and common understanding of a domain that can be communicated across people and application systems.

One example of the use of ontologies on the Knowledgeable Web is in e-commerce sites where ontologies are needed

- to enable machine-based communication between buyer and seller,
- to enable vertical integration of markets (e.g. www.verticalnet.com)
- to leverage reusable descriptions between different marketplaces.

A further example of the use of ontologies can be found in search engines. By using ontologies the search engines can escape from the current keyword-based approach, and can find pages that contain syntactically different, but semantically similar words (e.g. www.hotbot.com).

A final example is the work provided in MPEG-7, where the results, i.e. the DSs and Ds, can be understood as an ontology for audio-visual descriptions, so far mainly for broadcasting.

However, a prerequisite for such a role ontologies may play, is the development of a joint standard for integrating ontologies with existing and arising web standards. The Ontology Inference Layer (OIL) is a proposal for such a standard. OIL [34] is a Web-based representation and inference layer for ontologies, which combines the widely used modelling primitives from frame-based languages with the formal semantics and reasoning services provided by description logics. Furthermore, OIL is the first ontology representation language that is properly grounded in W3C standards such as RDF/RDF-Schema and XML/XML-Schema.

References

[1] ISO/IEC JTC1/SC29/WG1 N1646: "JPEG2000 Final Committee Draft v1.0", March 16, 2000.

[2] http://www.jpeg.org/JPEG2000.htm

[3] http://www.w3.org/TR/REC-png.html

[4] Society of Motion Picture and Television Engineers (SMPTE), *Television—Unique Material Identifier (UMID)*, SMPTE330M-2000, SMPTE, White Plains, N.Y., 2000.

[5] http://www.dvb.org/standards/index.html

[6] http://www.ebu.ch/home_4.html

[7] The TV Anytime Forum: Call for Contributions on Metadata, Content Referencing, and Rights Management, Document: TV0114r3, 17th Dec. 1999. (Available from http://www.tv-anytime.org/).

[8] http://www.tv-anytime.org/

[9] ISO 11179 - Specification and Standardization of Data Elements, Parts 1-6.ftp://sdct-sunsrv1.ncsl.nist.gov/x3l8/11179

[10] RFC2396 - Uniform Resource Identifiers (URI): Generic Syntax, Internet RFC 2396. http://www.ietf.org/rfc/rfc2396.txt">http://www.ietf.org/rfc/rfc2396.txt

[11] http://purl.org/dc/

[12] Koenen, R. (2000). *MPEG-4 Overview – (La Baule Version)*. Dublin, October 2000, ISO/IEC JTC1/SC29/WG11 N3747 http://www.cselt.it/mpeg/standards/mpeg-4/mpeg-4.htm

[13] http://www.cselt.it/mpeg/

[14] ISO-IEC FDIS 14772:200x, Information Technology - Computer graphics and image processing - The Virtual Reality Modeling Language (VRML).

[15] "Synchronized Multimedia Integration Language (SMIL) 1.0 Specification" W3C Recommendation, http://www.w3.org/TR/REC-smil/ (June 1998).

[16] http://archive.dstc.edu.au/mpeg7-ddl/

[17] http://archive.dstc.edu.au/mpeg7-ddl/latest_primer.html , http://www.w3.org/TR/xmlschema-1/ , http://www.w3.org/TR/xmlschema-2/

[18] http://www.darmstadt.gmd.de/mobile/mpeg7/

[19] http://www.mpeg-7.com/

[20] MPEG-21, Multimedia Framework architecture, ISO/IEC JTC1/SC29/WG11/N3500, Beijing, July 2000,

[21] http://www.mhegcentre.com/info/about.htm

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

[23] Synchronized Multimedia Integration Language (SMIL) 2.0 Specification. Work in progress. W3C Working Drafts are available at http://www.w3.org/TR 21 September

[24] SMIL 2.0: http://www.w3.org/TR/smil20/

[25] Extensible Markup Language (XML) 1.0", 10 February 1998. Available at http://www.w3.org/TR/REC-xml/.

[26] SVG 1.0: http://www.w3.org/TR/SVG/

[27] XHTML 1.0: http://www.w3.org/TR/xhtml1/

[28] XHTML 1.1: http://www.w3.org/TR/xhtml11/

[29] XSL Transformations (XSLT) Version 1.0: http://www.w3.org/TR/xslt

[30] Cascading Style Sheets, level 1: http://www.w3.org/TR/REC-CSS1

[31] Cascading Style Sheets, level 2: http://www.w3.org/TR/REC-CSS2/

[32] Resource Description framework (RDF): http://www.w3.org/RDF/

[33] Resource Description Framework (RDF) Schema Specification 1.0: http://www.w3.org/TR/2000/CR-rdf-schema-20000327/

[34] Ontology Inference Layer(OIL): www.ontoknowledge.org/oil