



**ISO/TC 46/SC 9/Working Group 1
for ISO Project 15706: International Standard Audiovisual Number (ISAN)**

Web page: <http://www.nlc-bnc.ca/iso/tc46sc9/isan.htm>

ISO/TC 46/SC 9/WG 1 N 85
1999-02-25

Compound AV objects and IPMP – questions to be asked

This document was prepared by Didier J. MARY, FIAPF/AGICOA technical liaison for MPEG. It summarises the present situation of IPMP (of which ISAN will be part of) and questions which remains to be solved as regards 'hierarchy of ISANs'.

I) Useful background information

From « Overview of the MPEG-4 Standard » (N2459 October 1998/Atlantic City)

Scope and features of the MPEG-4 standard

The MPEG-4 standard under development will provide a set of technologies to satisfy the needs of authors, service providers and end users alike.

- For *authors*, MPEG-4 will enable the production of content that has far greater reusability, has greater flexibility than is possible today with individual technologies such as digital television, animated graphics, World Wide Web (WWW) pages and their extensions. Also, it will be possible to better manage and protect content owner rights.
- For *network service providers* MPEG-4 will offer transparent information which will be interpreted and translated into the appropriate native signaling messages of each network with the help of relevant standards bodies. However the foregoing excludes Quality of Service considerations, for which MPEG-4 will provide a generic QoS descriptor for different MPEG-4 media. The exact translations from the QoS parameters set for each media to the network QoS are beyond the scope of MPEG-4 and are left to be defined by network providers. Signaling of the MPEG-4 media QoS descriptors end-to-end, will enable transport optimization in heterogeneous networks.
- For *end users*, MPEG-4 will enable many functionalities which could potentially be accessed on a single compact terminal and higher levels of interaction with content, within the limits set by the author. An MPEG-4 applications document exists which describes many end user applications including, among others, real time communications, surveillance and mobile multimedia.

For all parties involved, MPEG wants to avoid the emergence of a multitude of proprietary, non-interworking formats and players.

MPEG-4 achieves these goals by providing standardized ways to:

1. represent units of aural, visual or audiovisual 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;
2. describe the composition of these objects to create compound media objects that form audiovisual scenes;
3. multiplex and synchronize the data associated with media objects, so that they can be transported over network channels providing a QoS appropriate for the nature of the specific media objects; and
4. interact with the audiovisual scene generated at the receiver's end.

1.1 Coded representation of media objects

Audiovisual scenes are composed of several media objects, organized in a hierarchical fashion. At the leaves of the hierarchy, we find primitive media objects, such as :

- still images (e.g. as a fixed background),
- video objects (e.g. a talking person - without the background)
- audio objects (e.g. the voice associated with that person);
- etc.

MPEG standardizes a number of such primitive media objects, capable of representing both natural and synthetic content types, which can be either 2- or 3-dimensional. In addition to the media objects mentioned above and shown in Figure 1, MPEG-4 defines the coded representation of objects such as:

- text and graphics;
- talking synthetic heads and associated text used to synthesize the speech and animate the head;
- synthetic sound

A media object in its coded form consists of descriptive elements that allow to handle the object in an audiovisual scene as well as of associated streaming data, if needed. It is important to note that in its coded form, each media object can be represented independent of its surroundings or background. The coded representation of media objects is as efficient as possible while taking into account the desired functionalities. Examples of such functionalities are error robustness, easy extraction and editing of an object, or having an object available in a scaleable form.

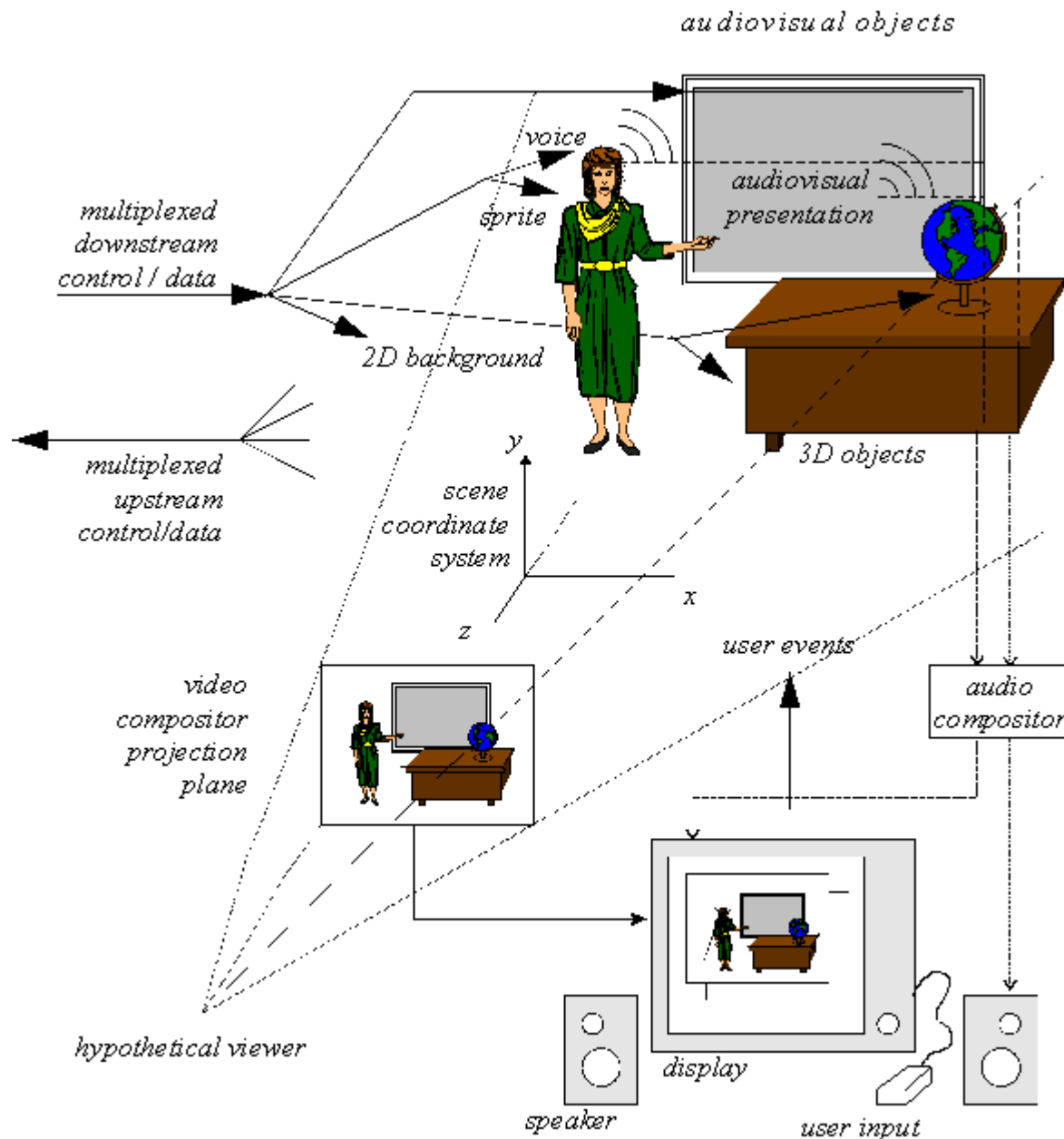
1.2 Composition of media objects

Figure 1 gives an example that highlights the way in which an audiovisual scene in MPEG-4 is described as composed of individual objects. The figure contains compound media objects that group primitive media objects together. Primitive media objects correspond to leaves in the descriptive tree while compound media objects encompass entire sub-trees. As an example: the visual object corresponding to the talking person and the corresponding voice are tied together to form a new compound media object, containing both the aural and visual components of a talking person. Such grouping allows authors to construct complex scenes, and enables consumers to manipulate meaningful (sets of) objects.

More generally, MPEG-4 provides a standardized way to describe a scene, allowing for example to:

- place media objects anywhere in a given coordinate system;
- apply transforms to change the geometrical or acoustical appearance of a media object;
- group primitive media objects in order to form compound media objects;
- apply streamed data to media objects, in order to modify their attributes (e.g. moving texture belonging to an object; animation parameters animating a moving head);
- change, interactively, the user's viewing and listening points anywhere in the scene.

The scene description builds on several concepts from VRML in terms of both its structure and the functionality of object composition nodes and extends it to fully enable the aforementioned features.



[...]

2.8 Content-related IPR identification and protection

MPEG-4 provides mechanisms for protection of intellectual property rights (IPR), as outlined in section 1.5. This is achieved by supplementing the coded media objects with an optional Intellectual Property Identification (IPI) data set, carrying information about the contents, type of content and (pointers to) rights holders. The data set, if present, is part of an elementary stream descriptor that describes the streaming data associated to a media object. The number of data sets to be associated with each media object is flexible; different media objects can share the same data sets or have separate data sets. The provision of the data sets allows the implementation of mechanisms for audit trail, monitoring, billing, and copy protection.

Next to identifying rights, each of the wide range of MPEG-4 applications has a set of requirements regarding protection of the information it manages. These applications can have different security requirements. For some applications, users exchange information that has no intrinsic value but that must still be protected to preserve various rights of privacy. For other applications, the managed information has great value to its creator and/or distributors requiring high-grade management and protection mechanisms. The implication is that the design of the IPMP framework must consider the complexity of the MPEG-4 standard and the diversity of its applications. This IPMP framework leaves

the details of IPMP systems designs in the hands of applications developers. The level and type of management and protection required depends on the content's value, complexity, and the sophistication of the associated business models.

The approach taken allows the design and use of domain-specific IPMP systems (IPMP-S). While MPEG-4 does not standardize IPMP systems themselves, it does standardize the MPEG-4 IPMP interface. This interface consists of IPMP-Descriptors (IPMP-Ds) and IPMP-Elementary Streams (IPMP-ES).

IPMP-Ds and IPMP-ESs provide a communication mechanism between IPMP systems and the MPEG-4 terminal. Certain applications may require multiple IPMP systems. When MPEG-4 objects require management and protection, they have IPMP-Ds associated with them. These IPMP-Ds indicate which IPMP systems are to be used and provide information to these systems about how to manage and protect the content. (See Figure 17)

Besides enabling owners of intellectual property to manage and protect their assets, MPEG-4 provides a mechanism to identify those assets via the Intellectual Property Identification Data Set (IPI Data Set). This information can be used by IPMP systems as input to the management and protection process.

2.9 Object Content Information

MPEG-4 will allow attaching information to objects about their content. Users of the standard can use this '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.

4.1.12 Object Content Information (OCI)

Requirement

MPEG-4 shall provide the possibility to associate content description information to the various audiovisual objects in the scene

Specification

1. MPEG-4 shall support normative Object Content Information (OCI) data (syntax and semantics.) Room for private description information data shall also be provided.
2. The amount of normative Object Content Information used for each specific case should depend on the content provider's needs and thus the OCI syntax should be flexible enough to accommodate very different needs, in an efficient way. The minimum amount of OCI possible to add should have an insignificant weight (null, if possible) in the bitstream budget. This means that no MPEG-4 application should be unnecessarily loaded with object content information, if it does not want to provide this type of data.
3. Taking into account the content-based nature of the MPEG-4 audio-visual representation, where a scene is composited by many objects, independently accessible and usable, it should be possible to associate Object Content Information to all levels of a scene hierarchy, down to each elementary object.
4. Object Content Information syntax and semantics should be as much as possible the same at all levels of a scene hierarchy, down to each elementary object.
5. Taking into account the MPEG-4 time schedule and the emerging MPEG-7 effort, Object Content Information (OCI) in MPEG-4 should be limited to textual descriptors and description classifiers.

Example

The following types of data are considered examples of Object Content Information (OCI) data: IPR data (following the specification in this document), data concerning content description and classification, e.g. movie, news, sports, music, children, game, etc. organized in one or more layers, data concerning parental rating, e.g. by ages, data concerning events, such as event name, event description, start time, duration, etc, data concerning language (for audio), content textual description.

Note

The provision of Object Content Information (OCI), is essential to allow the selection, retrieval, and access of services, scenes, events, etc. Although in principle OCI data can be regarded as MPEG-7 information, the fulfillment of this MPEG-4 requirement should not constrain, in any way, the MPEG-7 development.

AV Object = Audio Visual Object : a representation of a real or virtual object that can be manifested aurally and/or visually

AV Objects are generally hierarchical, in that they may be defined as composites of other AV objects, which are called *sub-objects*. AV objects that are composites of sub-objects are called *compound AV objects*. All other AV objects are called *primitives*. AV objects that cannot be decomposed into sub-objects are called *primitive AV objects*.

Identification of Intellectual Property

Requirement

1. The MPEG-4 standard shall provide for possibility to record, transmit and retrieve the identifiers of the copyrighted components that compose the MPEG 4 object, using existing identification systems, e.g. International Standard Book Number (ISBN number).
2. The MPEG-4 standard shall provide the capability to uniquely identify the registration authority. (e.g. a reference to the ISBN agency)
3. It shall be possible to identify a composed MPEG-4 object (i.e. a sum of other copyrighted objects) as a separate copyrighted object.

Note

- The ability to identify the registration authority is necessary in order to access information, specific to the particular component of content (object), in external databases.

Multiple objects can have dependent identifiers, but if the authority is the same for each object, there is no need to repeat the authority ID over and over. An example would be the case in which a movie is identified by an International Standard Audio-visual Number (ISAN), containing several songs identified by International Standard Work Code for Tunes (ISWC-T).

<h3>A few definitions</h3>

- **Visual Object**

Any elementary or composite AV object which presents visually (e.g. video object, 2D/3D synthetic object, text object, any meaningful combination of these objects).

Note:

Special cases of a Visual Objects are Video Object or a Facial Animation Object.

- **Audio Object**

A sound source, or a composition of sound sources or channels, specified by a model of the elementary behavior and layout of real-world or synthetic sound(s) approximated by the model (e.g. wave table, sample stream, TTS stream, control stream that drives algorithmic or sample-based synthesis of music, program defining a voice), localized within an explicit or implicit 2D or 3D aural environment.

- **Video Object**

A temporal stream of moving images, or a spatial and temporal composition of such streams within a scene, with any supporting data for temporal, resolution, and quality scalability. A degenerate video object is a texture map. Audio and video objects are potentially independent of the viewing conditions and frame rate of their presentation, and are interpreted by a presentation layer.

- **Scene Composition**

A hierarchical nesting of AV objects in space and time that supports the content-based access, manipulation, and scalability requirements of MPEG-4. Such a nesting may:

1. Compose spatial-temporal scenes from efficient shared representations of objects;
2. Provide alternative or subordinated representations of AV objects for scalability;
3. Provide space-time partitioning to manage large data sets and user access to portions of them;
4. Represent structural relationships that enable user manipulation or interaction between objects;
5. Provide granularity or priorities for selecting scene elements to present within terminal resources.

A scene composition may include local and/or remote, streaming and/or downloaded objects that are named, identified uniquely, spatially located, and synchronized to each other in a specific view.

[...]

From « Managing Intellectual Property Identification and Protection within MPEG-4 » (N1918 October 1997/Fribourg)
--

[...]

4 Specifying the Requirements

The requirements for managing information about the copyright contents of MPEG-4 objects can be expressed in the following terms:

- Identification of content
- The persistence of content identification in modified MPEG-4 objects
- Content Protection

It is proposed that requirements which relate specifically to the **identification of content** to be proposed in the first version of the MPEG-4 standard. Requirements for persistence of content identification and content protection will be proposed in version 2.

5 Identification of Content

As a minimum requirement, it must be possible to identify from an IP Identification Data Set the **identity** of content which is contained within an MPEG-4 object. In cases when more than one type of content resides in the object, multiple occurrences of IP Identification Data Sets will occur (i.e. for audiovisual works).

The Requirements for identification can be summarised as follows:

1. Identify all of the individual intellectual property components of digital content (i.e. an audiovisual production and related contents such as sound recordings, still pictures and video clips, etc.)

[...]

5.5 Revisions to the IP Identification Data Set

The IP Identification Data Set has been revised to include fields which simply relate to achieving the successful identification of content. The following changes have been introduced:

- To provide maximum flexibility, none of the fields in the IPI Data Set is mandatory.

[...]

7 Conclusion

It has been possible by distinguishing the requirements to define what is required solely for the **identification** of intellectual property, to enable an effective solution to be agreed between the representatives of the creative industries during the Fribourg meeting. The results which are documented provide a specification for inclusion within the MPEG-4 Committee Draft which meets the separate needs of the creative industries to support the interests of the people which they represent.

One important requirement which remains to be defined is the persistent association of IP Identification Data Sets with MPEG-4 objects and their resistance to removal due to object modification.

The new WIPO treaty establishes the principle that it is an offence for any person to attempt to alter or remove electronic rights management information. The scope of this legal protection is also understood to cover the distribution, broadcast and communication to the public of works or objects from which such electronic rights management information has been removed.

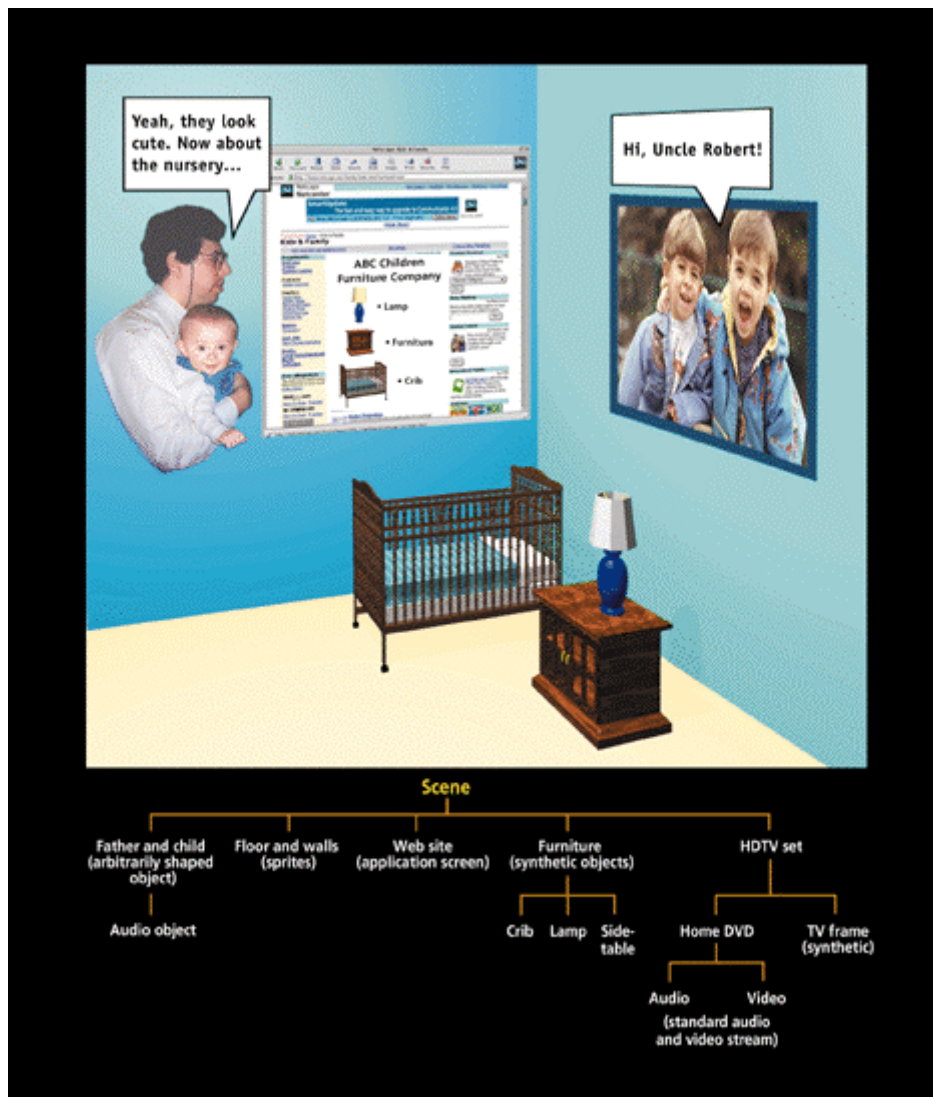
WG11 wishes to alert the individuals and organisations who intend to use the identification and protection mechanisms supported by MPEG-4 that these issues must be addressed in addition to the MPEG-4 technical standardisation process.

II) Questions, remarks and considerations to be further evaluated

Where MPEG-1 is a standard used to play out audio and video in linear streams, allowing the same type of access as a home VCR, and MPEG-2 was introduced for compression and transmission of digital television signals, MPEG-4, as described in the documents above is a « hierarchical nesting of AV objects in space and time », based on a radical object-oriented paradigm.

Here's how Rob Koenen explains this paradigm in his document « MPEG-4 Multimedia for our time - *IEEE Spectrum February 1999 Volume 36 Number 2* » :
« At the atomic level, to use a chemical analogy, the audio and video components of MPEG-4 are known as objects. These can exist independently, or multiple ones can be grouped together to form higher-level audiovisual bonds, to coin a phrase. The grouping is called composition, and the result is an MPEG-4 scene [see figure below]. The strength of this so-called object-oriented approach is that the audio and video can be easily manipulated.

Visual objects in a scene are described mathematically and given a position in a two- or three-dimensional space. Similarly, audio objects are placed in a sound space. When placed in 3-D space, the video or audio object need only be defined once; the viewer can change his vantage point, and the calculations to update the screen and sound are done locally, at the user's terminal. This is a critical feature if the response is to be fast and the available bit-rate is limited, or when no return channel is available, as in broadcast situations. »



Different types of multimedia that can be transmitted with MPEG-4 appear in the scene above, a man and his infant son on-line with his offstage wife. The tree chart below, called a scene graph, represents the media as independent or compound objects. One compound object comprises the father and child (an arbitrarily shaped video) and the audio track of his voice. Other objects are the floor and walls--"sprites," here used for easily changed backgrounds; the Web site of the furniture store--an application mapped as a screen texture; and the computer-generated (synthetic) furniture the father has chosen from the Web site for his wife to look at and interactively move around. Simultaneously playing on a synthetic HDTV set is a movie from the family's home digital versatile disk (DVD).

When considering the above presentation focusing on « movies¹ » (of any kind), the following considerations arise.

Movie conversion

A movie converted to a MPEG-4 object could follow one of these schemes :

- a primitive AV object, which means a movie is a « complete » object that cannot be decomposed in different sub-elements (video + audio [voice] + music),
- a compound AV object, which means video is a video object, audio [voice] is an audio object and music² is another audio object (this means 3 AV objects at least³)

¹ I'll use this term to differentiate from « Audio/Video objects ».

Then when analysing related IPMP information :

- in the first case, the AV object contains one and only one IPI data set, specifically an ISAN number.
- in the second case, the AV object would contain 4 IPI data sets : one for video, one for audio, one for music, and one for the whole AV object. The ISAN might be in one of these IPI data sets (probably the one for the global object).

Question : how should be/is the hierarchy organised ? Which IPI data set should be/is the « most important » one ?

Dubbing

Following the same scheme as above, one could consider :

- having one primitive AV object of a movie for each language it could be broadcasted in. This means each language generated version of the same movie, would have a different ISAN.
- having one compound AV object, containing multiple audio [voice] objects, one for each language it could be broadcasted in. This means the ISAN number would be same for any language. The only difference would remind in the sub-IPI data set relating to the different audio [voice] objects (considering the hierachy model of IPI data sets is consistent).

This issue should be further analyzed

AV objects modification

One of the goals for MPEG-4 version 2, is persistence and resitance to removal of IPMP data sets. MPEG-4 version 1 did not include such requirements.

MPEG-4, as any other standard is designed primarily to create or expand new business opportunities. This means any hardware or software designed to use MPEG-4 files or data, might be able to allow any kind of modification of any MPEG-4 object, should it be a primitive or compound object.

I sent the following questions to the IPMP mail reflector a few weeks ago :

Let's suppose I have 2 AVOs (pieces of multimedia of any sort = Audio, or video, or ...). Each one has its IP data set and IP system embedded. OK, to be more specific, let's say each AVO only contains its unique identification number embedded and no encryption scheme attached. Ok, now I cut a part of each AVO (eg : if a video, I cut 3 seconds of it). Then I create a new empty AVO of the same kind, and I paste each previous part to this new one. (see sample image)

Q1 : When I cut part of an AVO, are the IP data set and IP system retained ?

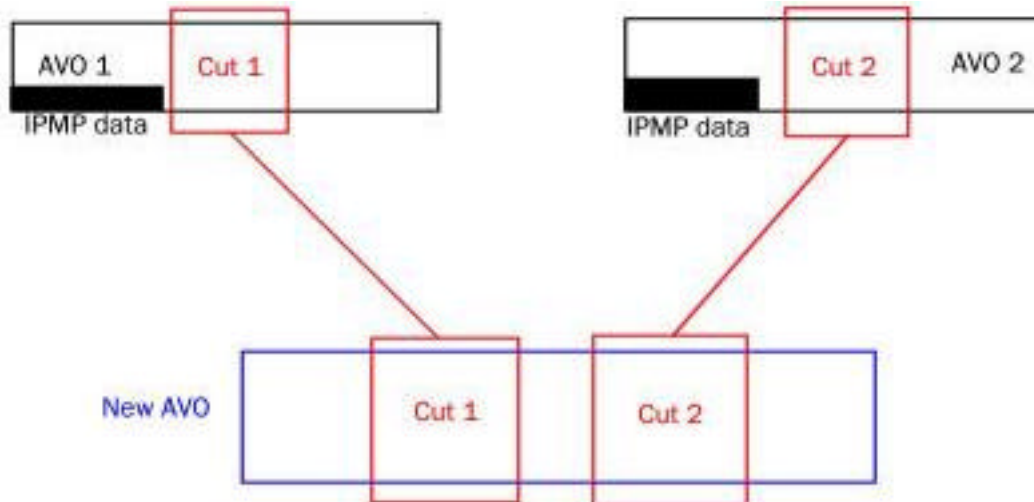
Q2 : When I paste it to a new AVO, are they retained too ?

² To simplify, I'll consider that music is composed of only one tune. If each tune is identified on its own, then of course each would be an independent object, having its own IPI data set.

³ See comments on dubbing

Q3 : If I merge parts of different AVOs are their respective IP information retained ? Can then each information be retrieved independently, e.g. for tracking purposes ?

Q4 : any other issues ?



and got the following answers :

In theory the answer is YES to all of the above questions. However, in practice depends on :

- 1) the time interval that the watermark is inserted (say that a watermark, an ISRC code, is inserted every 1.4 secs),
- 2) the duration of the watermark is approx 0.6 sec,
- 3) the content of the audio track which means that not necessarily a watermark is inserted every 1.4 secs but this interval from time to time can be longer and of course,
- 4) the duration of the music part that has been cut out

Thus if you have a 1 min piece of music A and you cut out the 0.5 min of this and paste with another 0.5 min which has been cut out from another track B then you will retain all the watermark information. However, if you think to do the same using only 3 secs of audio everything is possible from to extract whole ISRC codes, half codes or may be nothing.

Of course there are different watermark algorithms which will behave different under the same scenario. The above answer is based in CRL's audio watermarking algorithm.

Sincerely

Panos Kudumakis

(Dr Panos Kudumakis is Staff Scientist at CRL, a UK laboratory specialised in audio watermarking).

you have some interesting questions, that lie on the border of what MPEG must deal with and what is a question of legal systems. Let me try to explain.

> Q1 : When I cut part of an AVO, are the IP data set and IP system retained ?

Not automatically. The editing system in question must support this functionality. This is not a matter for MPEG to enforce. It is supported however, because the author/editor can find the data and copy it to the newly created object. If he strips it he might be doing

something illegal, but MPEG plays no role there. Basically, this is the answer to all your questions.

In principle, it is possible to protect objects so that IP information cannot be stripped. The counterpart of such content could be a 'safe' editing system that enforces some IP identification rules and also outputs protected content,

Only this editing system could read and modify the objects and use them for creating new content. This could even be rule based, such that some things are allowed and some things aren't.

The content (protected MPEG-4 objects) would not be editable in other editing systems. (Unless of course someone cracks the code and still does so, which might -again- constitute an illegal act)

It is important to point out that creating such environments is again NOT something that MPEG should or even can do - it is enough to provide the hooks, as the IPMP spec does.

> Q2 : When I paste it to a new AVO, are they retained too ?
Same answer as above.

> Q3 : If I merge parts of different AVOs are their respective IP information retained ?
> Can then each information be retrieved independently, e.g. for tracking purposes ?

If you keep the different objects as different objects you can retain the info. Alternatively you create some new objects, apply for an ISAN (unique number) and store the IP info in the ISAN database. Still alternatively, you use multiple IP identification sets, which can apply to one Elementary stream. Of course precise information will be lost if you collapse multiple objects into one, but you still know what it was composed of. If you are building an authoring system, this is something you'd better support (but I don't know what you're up to :-)

> Q4 : any other issues ?

Whatever you do, the standard does not tell you what is allowed and what is not, from a legal point of view. It just specifies which bitstreams are syntactically valid and how to decode them.

"Syntactically valid" does not equal "lawful" (if only because you can never tell how a syntactically valid bitstream was created).

Conditions will vary according to content owner, country, usage, etc. As I said above, industrial consortia may decide to build systems that enforce a chosen set of rules in some way or other, and MPEG-4 provides the hooks for building these systems.

hope this helps,
Rob

ps: something for the IPMP FAQ?

(Rob Koenen is a senior consultant and project manager with the research facility of KPN, the Dutch telecommunications operator. He has been actively involved in shaping the MPEG-4 and MPEG-7 standards, and has chaired the requirements group within the Moving Picture Experts Group since 1996.)

As far as I understand, the standard as defined today explains what can be done, but doesn't make the proper use of IPI data sets mandatory. Lawful and unlawful behaviors depend on the business and legal environments.

I then believe some enforcement should be put on MPEG-4 version 2, so that removing or modifying any IPI data set or any hierarchy of IPI data sets should be technically not feasible (this could be achieved by adding specifications for the proper conformance of hardware or software that could be used for MPEG-4 objects modification). Plus, some kind of information to manufacturers should be done, explaining that a « good » MPEG-4 object must contain all of its data. If it loses some (i.e. IPI data set), it's no more « good ».

Some more considerations on MPEG-4 editing
--

One « easy » modification to an MPEG-4 AV object could be removing the audio [voice] object (let's say French audio), to replace by one from an other language (dubbing).

Some other edits could be to removal/edit of some part of the music (replace a tune by another one, change orchestration ...), or even part of the movie (i.e. replace some images by new ones, add new special effects, colour the movie ...).

In all these cases, if we suppose working with a primitive AV object, should the IPI data set (ISAN) be kept unchanged or be upgraded to a new one ? If kept, how to differentiate the 2 versions ? If upgraded, should the previous IPI data set be kept somewhere (and where), for history tracking ?

In all these cases if working with compound objects, how should the IPI data set hierarchy evolve ?

If creating a new AV object, by merging 2 other AV objects (or part of them), should all the IPI data set hierarchy be kept in the final file ?

As a conclusion, there is quite a lot of complex data relating to MPEG-4 AV objects and there will be more and more if the processes generate unmanageable schemes, that tend to nothing.

Dropping some recommendations or needs could be in some cases the best choice.

DJM

2/25/99