

# Metadata Carriage in Digital Television

**Richard Chernock**

Triveni Digital  
Princeton Junction, NJ

**Michael A Dolan**

Consultant  
Alpine, CA

## ABSTRACT

Metadata (data about data) has always been an important aspect of broadcast television operations. For example, the title of a program is metadata about the contents of that program. As we move forward into the era of digital television (DTV) broadcasting, metadata assumes an even greater importance in management of the digital components. The increased complexity and inter-relationships of the components of the DTV signal is requiring that more information be available to the facility equipment and consumer receiver. This information is needed to intelligently understand how to fully present the television programming as well as to offer enhanced capabilities. Luckily, this same complexity (and the increased bandwidth in DTV) offers the capability of carrying this descriptive information.

Work has taken place within ATSC to begin to define the generalized carriage of metadata in an ATSC emission stream and provide a model from which to view the interface from the facility to the emission. In order to facilitate the work, a common set of terminology and timeline concepts usable in the facility was created. Special attention was paid to the interface between the emission domain and the facility domain in order to ensure that the necessary metadata defined by others, including the SMPTE metadata dictionary [1], can successfully cross this boundary.

This paper will summarize the work and findings of this activity.

## INTRODUCTION

The first thing we must understand when discussing metadata is how it is different from “data”, a term which is often used interchangeably in practice. Everyone understands video (when they see it), and everyone understands audio (when they hear it). But “data”, and especially its distinction from metadata, is a bit more elusive. SMPTE, in its Task Force Report [2] addresses this topic well, and the reader is encouraged to review that reference for a more detailed coverage of this topic. In a nutshell, elements of a broadcast that are intended to provide a direct visual or aural affect on the viewer are defined to be the “essence” of the broadcast. Elements that are to be used by the

receiving equipment to organize and present the essence or provide descriptive information to the viewer, is deemed metadata. For example, a computer graphics file found on the web is an obvious form of “data essence”. An identifier for that item that is never seen by the viewer but needed by a receiver to process the image is metadata about that data essence. Similarly, a channel number is (identifier) metadata about a programming service being provided to the viewer – it is metadata about the collection of video, audio and data essences.

As we move into the digital world, there is an increasing proliferation of metadata. This metadata is starting to take the form of camera angle information, digital rights information, as well as private information being inserted for specialized business models such as ratings measurement. As the volume of metadata information grows, it starts to push the limits of the traditional MPEG-2 transport mechanism for carrying metadata – the MPEG-2 descriptor [3]. With a little analysis, one sees that it is possible to quickly run out of *descriptor\_tag* values when assigning them to the individual metadata items, for example. Another issue is the size limitation of the descriptor structures themselves (at most 255 bytes of payload). Finally, there is the lack of ability of most table sections that contain descriptor loops to physically hold the volume of all possible descriptors for that table.

The obvious answer to these technical problems is to define a special MPEG-2 Program Element encapsulation that contains the metadata, in order to avoid the naming and sizing issues mentioned above. But, when one moves the metadata out from its implicit binding with other MPEG-2 Transport structures, then linkages and, in some cases, timing synchronization issues emerge and must be addressed.

This paper will address the current thinking in the area of generalized metadata carriage in the ATSC DTV transport.

## NEEDS FOR GENERALIZED METADATA CARRIAGE

As the systems for the production, distribution and emission of digital television content evolve, the amount of metadata continually increases. A disturbing recent trend is that new methods of carriage seem to be invented for each new form of metadata. While there are some categories of metadata that have special needs and greatly benefit from a specialized carriage mechanism (such as ATSC PSIP [4], whose carriage mechanism is designed to speed channel acquisition), most metadata does not really need special treatment.

Utilization of a generalized carriage mechanism would likely be beneficial, trading off the large advantage of increased efficiency at the receiver by using a common metadata decoder for a smaller loss in efficiency at emission which might have been realized with specialized forms of carriage. Considering that the receivers greatly outnumber the transmitters and that there is a very strong driving force for reducing receiver pricing, utilizing a generalized mechanism has some clear advantages in most circumstances.

As metadata moves through the distribution, from production to emission and finally consumption, it passes through a number of domains that determine the methods for carrying metadata and associating it with related essence. The ideal situation would be for a single common method for dealing with metadata, avoiding the need for conversions at the interfaces between domains. The most likely scenario is that there will continue to be different carriage mechanisms for metadata in the different domains, as the requirements and capabilities of the systems involved are considerably different. If, in this situation, there were only one method of carriage in each domain, the necessary conversion at the domain interfaces would be greatly simplified.

## TERMINOLOGY AND TIMELINE CONCEPTS

Before discussing the carriage of metadata, it is first helpful to define a set of terms that name certain somewhat well-known, but currently not well-defined, aspects of the elements that emerge at the facility/emission boundary.

**Work** – sequence of related images, audio and/or data, which is intended to be presented to a viewer; including metadata and a Timebase; identified by a unique Work identifier (such as ISAN [5]).

**Show** – collection of Works with its own Timebase

**Timebase** – continuous and monotonically increasing time relative to the first piece of a Work or Show (i.e. first frame of video); rate is the same as normal viewing.

**Work Segment** – a contiguous subset of a Work; identified with a single start time and end time pair referenced to the Work's Timebase; and a defined subset of the Elements of the Work.

**Show Segment** – a contiguous subset of a Show; identified with a single start time and end time pair referenced to the Show's Timebase; and a defined subset of the Elements of the Show.

**Element** – one instance of the video, audio, data or metadata.

**Virtual Channel** – sequential delivery of Shows.

**Event** – the announcement of a time-continuous part of a Virtual Channel aligned with Show time boundaries.

**Segment ID** – the metadata item that identifies a Segment; scoped to the Work or Show respectively.

Note that a Timebase may be implicit by counting frames or wall clock time during playback, or explicit by the use of SMPTE 12M (or SMPTE ATR) timecode. A Segment may be implicit through the insertion of segment ID metadata at certain point in time, or explicit by the use of Timebase bindings. Show Segments do not extend beyond the Show, and Work Segments do not extend beyond the Work. But, Segments can overlap and there may be gaps between Segments. A Show corresponds to, and is signaled by an Event in ATSC PSIP.

The possible elements of a Work are shown in Figure 1. There are fundamentally the three kinds of essence, plus metadata, plus a timeline.

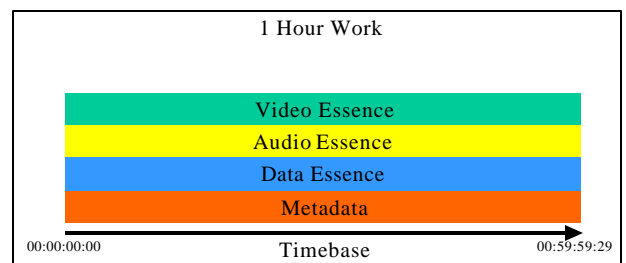
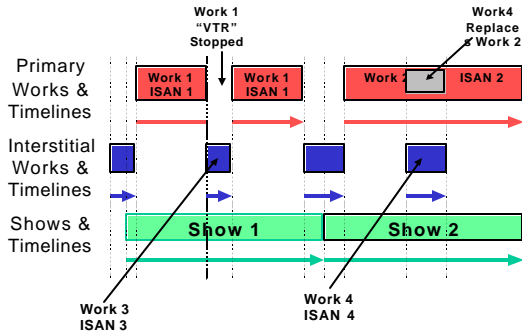


Figure 1. Elements of a Work.

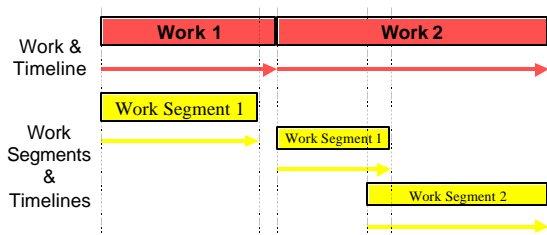
The mapping of Works to Shows is shown in Figure 2 below. Note that this diagram introduces the notion of “primary” Works and “interstitial” Works. The latter are primarily advertising insertions into the primary Work, traditionally described as the currently airing program. Also note that the use of the term, “program” is avoided as much as possible due to its overloaded use in MPEG-2 as a service or channel.



**Figure 2. Mapping of Works to Shows.**

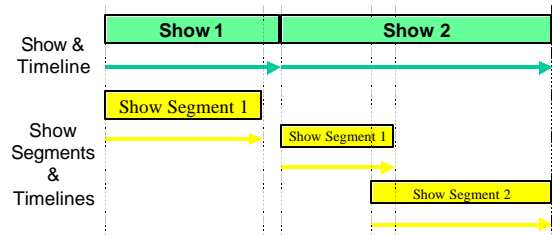
Note that Figure 2 also introduces the notion of a content identifier. In this case the use of ISAN [4] is shown, but any Work identifier is possible.

The application of Work Segments is shown in Figure 3 below with the relationship between Works and their Segments indicated. Work Segments, while they may overlap, are still contained within the Work they relate to.



**Figure 3. Relationship of a Work and its Segments.**

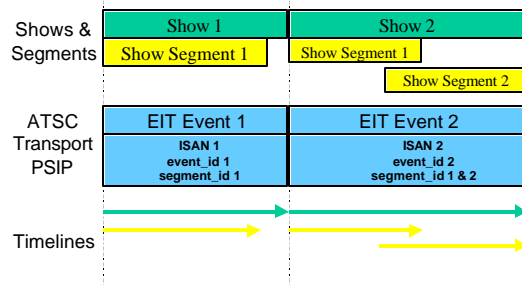
The relationship of Show Segments to Shows is shown in Figure 4.



**Figure 4. Relationship of a Show and its Segments.**

Like Work Segments, Show Segments are confined to the Show to which they refer. There is no pre-defined relationship between Work Segments and Show Segments (although there may be).

Finally, Figure 5 shows the mapping of the Shows to the ATSC transport. The mapping of Works to the transport can be inferred from the previous figures. Note that the Shows are constrained to effectively be an ATSC “event” as signaled in the PSIP EIT table.



**Figure 5. Mapping of Shows to the ATSC Transport**

## MPEG-2 METADATA AMENDMENT

One approach to a generalized method to carry metadata in an MPEG-2 emission stream revolves around an amendment to MPEG-2 Systems entitled "Amendment 1: Carriage of metadata over ITU-T Rec H.222.0 | ISO/IEC 13818-1 streams" [6] that is near finalization. As the title implies, this amendment is designed to provide mechanisms to carry metadata within an MPEG-2 Transport. While originally intended for the carriage of MPEG-7 [7] metadata within the MPEG-2 Transport, this amendment was broadened to encompass the carriage of any encoding of metadata, including SMPTE KLV [8] as well as TV Anytime.

As is typical at the MPEG-2 Systems level, the contents of the amendment describe a toolkit. The tools in the kit provide mechanisms for encapsulation, signaling and binding of metadata. Users of MPEG-2 (such as the ATSC) can select which tools to use from the toolkit and then constrain the way in which these tools are used in order to fit their environment.

These tools provide a means for carrying metadata within the MPEG-2 transport stream (encapsulation), letting a receiver know which are the different metadata components (signaling) and establishing connections between the metadata and the essence that it is associated with (binding).

### **Encapsulation methods**

The needs for metadata carriage in any particular transport and service are widely varied. In recognition of this, a number of encapsulations have been made available for metadata, with the method selected based on the characteristics of the metadata and its relationship to the essence. Additionally, the notion of a metadata access unit (AU) has been introduced, in a similar fashion to the access units for video and audio essence. A metadata access unit is a chunk of metadata that is useful to a decoder by itself.

The encapsulation types supported for metadata are:

- PES (stream type 0x15), which is used for synchronized / synchronous transport of metadata and may carry metadata objects or collections up to ~64KB in length.
- Metadata Sections (stream type 0x16), which is used for asynchronous transport of metadata without a carousel delivery mechanism (~4KB length limitation).
- Data Carousel (0x17), which is used for the carousel delivery of metadata objects without the need for hierarchical organization.
- Object Carousel (0x18), for the carousel delivery of metadata objects with need for a hierarchical organization, specifically including directories in the same notion as for traditional computer file systems.
- Synchronized Download Protocol (x19) for the synchronized delivery of metadata objects.

As can be seen from the list above, the encapsulation types cover just about all conceivable needs for conveying metadata in the MPEG-2 Transport Stream.

### **Signaling and Binding methods**

The mechanisms for signaling and binding metadata involve the use of a number of new MPEG-2 descriptors defined in the metadata amendment. These descriptors (each described in more detail below) allow the "labeling of content", association of a metadata service with audiovisual (or other) essence, identification of streams of metadata, and the description of metadata decoder parameters.

#### *Content Labeling Descriptor*

The content labeling descriptor is used to both assign a label to content which can be used for the association of metadata, as well as establishing the type of timeline to be used for the metadata (if any) and the relationship between the metadata timeline and the content timeline (as discussed below).

#### *Metadata Pointer Descriptor*

The metadata pointer descriptor is associated with essence (either a single component or an entire program) and provides the binding identifier for the related metadata. Its purpose is to provide the information necessary to locate the related metadata (which might be in the same MPEG-2 program, elsewhere in the same MPEG-2 Transport Stream or even in a location external to the MPEG bit stream). A typical application would be to place this descriptor in a data construct associated with the essence component (for example in the PMT inner loop describing a video bit stream).

#### *Metadata Descriptor*

The metadata descriptor is associated with a metadata bit stream (typically by being placed in the PMT inner loop associated with the metadata stream). It carries the identifier for the metadata which matches the identifier in the metadata pointer descriptor. This descriptor also carries information to allow the specific metadata object to be identified amongst a larger collection of metadata.

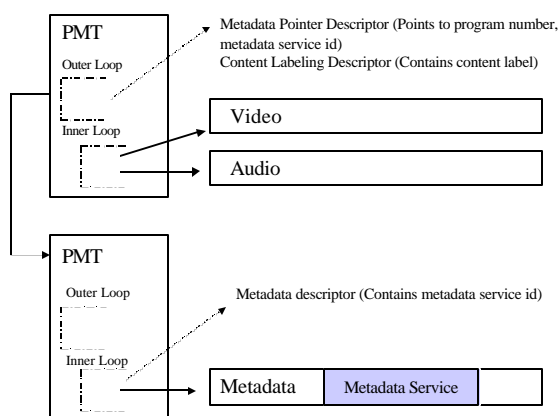
In the situation where information needs to be carried to provide hints to the metadata decoder (for example, identification of the specific decoder or parameters needed by the decoder), this descriptor either carries the information or provides the necessary binding information to find it.

### Metadata STD Descriptor

The metadata STD descriptor provides the parameters for the STD (System Target Decoder) model for the metadata, the buffer size and leak rates. These parameters are used by the decoder to ensure that resources are available to decode the metadata without loss.

### Putting it all Together

Figure 6 below shows an example of the linkages provided by the signaling and binding components described above. In this example, an MPEG program containing essence (content bit stream) has associated metadata carried in the same Transport Stream, but in a different MPEG program (metadata bit stream).



**Fig 6 Metadata linkages**

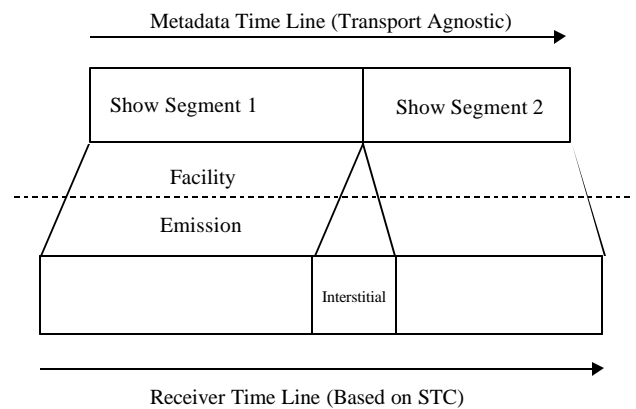
The PMT associated with the content bit stream carries two descriptors in the program (outer) loop, the content labeling descriptor and the metadata pointer descriptor. The content labeling descriptor associates a label with the content. The metadata pointer descriptor indicates that there is associated metadata and provides the information necessary to find and identify it. In this case, the metadata pointer descriptor points to the MPEG-2 program number for the bit stream that carries metadata stream (binding it to the PMT *program\_number* field). In this example, it also carries a *metadata\_service\_id* value that is used to match up to the individual metadata service (a metadata stream can carry multiple metadata services).

The PMT associated with the metadata stream carries a metadata descriptor in the program element (inner) loop associated with the metadata stream. The *metadata\_service\_id* field in this descriptor is used to match this metadata stream with the appropriate content. The form of encapsulation for the metadata is

signaled by the *stream\_type* field in the PMT for the Program Element associated with the individual metadata stream.

### Timeline and synchronization concepts

Metadata quite often has temporal relationships to essence that need to be preserved through the emission and presentation processes. These timing relationships may be loose (i.e. associations on the order of seconds) or tight (i.e. association to individual video field/frames). The former may be solved without any form of time signaling (emitting the metadata at the appropriate moment will suffice). Tight timing relationships typically require explicit signaling of the timing by associating a time code with the metadata itself, in a similar manner to the problem of synchronizing data essence to video in broadcast [9]. There are two necessary aspects to using timing relationships: establishing a relationship between timelines and signaling the synchronization points along the common timeline.



**Figure 7 Timing model for delivery of content and metadata**

One of the common problems encountered is the use of different timelines at different points in the distribution. A number of constraints have been made in the metadata amendment to address these problems:

- No time discontinuities shall occur in the metadata time line.
- The metadata time line shall be locked to the MPEG-2 sampling clock of the content (i.e. 90KHz clock).
- Each time reference in the metadata stream refers to the same metadata time line.

The content labeling descriptor allows signaling of the time base used for the content, as well as establishing the relationship (offset) between the metadata timeline and the content timeline. The allowed time bases are: MPEG-2 STC (System Time Clock), DSM-CC NPT (Normal Play Time) [10], and privately defined time bases. As mentioned above, in all cases a constraint is made that the relationship be expressed in units of a 90KHz clock (matching the STC "tick" rate). When a time base that does not utilize a 90KHz clock is used, then values along the timeline must be expressed as their 90KHz equivalents (expressed in a similar form to the PCR as defined in 13818-1 – i.e. a 33 bit value).

The synchronization points are signaled in the metadata encapsulation itself (either PES encapsulation or the synchronized download protocol), using PTS values. While this mechanism provides a basis for a synchronization system, there are a number of implementation details which must be accounted for, in a similar fashion to synchronizing data to video as discussed in [9], especially in the area of ensuring that the metadata is delivered sufficiently in advance of its use time to allow decoding to occur.

#### **Application to generalized carriage.**

The MPEG-2 Metadata amendment provides a good framework for a generalized scheme for the carriage of metadata in an emission bit stream. The tools defined in the amendment provide the necessary features to carry all forms of metadata and provide the necessary identification and associations to connect metadata and essence without the need to continue to define new MPEG-2 descriptors for every metadata item.

#### **RECOMMENDATIONS**

Some open issues that remain in ATSC are the signaling of Show (or even Work) segments. Right now there is no explicit signaling and some transport features such as Directed Channel Change (DCC) may need it to identify "splice points". Also, the issue of which synchronization mechanism to use in ATSC is still open except for the general framework defined in the ISO MPEG draft amendment. Finally, there is debate as to whether the Work timeline should be exposed in the emission. That is, a timeline that aligns with the original Work (such as the physical tape in the VTR as it stops for interstitials). This exposure may be needed to perform proper synchronization with data services, for example.

As the metadata requirements continue to grow, it will be critical to organize them into more general carriage mechanisms, rather than continue to add them piecemeal into the MPEG-2 descriptors and table sections.

The MPEG-2 Amendment provides an excellent framework upon which to build this general mechanism. Utilization of the tools in the MPEG metadata amendment by the ATSC will require further constraints to fit into the ATSC Transport environment.

#### **REFERENCES**

- 1) SMPTE RP210, SMPTE Metadata Dictionary, <http://www.smpite-ra.org/mdd/>
- 2) SMPTE/EBU Task Force for Harmonized Standards for the Exchange of Program Material as Bitstreams – Final Report: Analyses and Results, July 1998
- 3) ISO/IEC 13818-1 | ITU-T Rec. H.222.0:1996, Information Technology, "Generic coding of moving pictures and associated audio – Part 1: Systems"
- 4) ATSC Standard A/65A with Amendments 1, 2 and 3: Program and System Information Protocol for Terrestrial Broadcast and Cable
- 5) ISO 15706:2002, "Information and documentation -- International Standard Audiovisual Number (ISAN)"
- 6) ISO/IEC JTC1/SC29/WG11 N5270, "Information Technology – Generic Coding of Moving Pictures and Audio: Systems – Amendment 1: Carriage of metadata over ITU-T Rec H.222.0 | ISO/IEC 13818-1 streams"
- 7) ISO/IEC 15938 FDIS Information Technology Multimedia Content Description Interface
- 8) SMPTE 336M-2001 Television: Data Encoding Protocol using Key-Length-Value
- 9) "Issues in Tight Synchronization between Data and Video", R. Chernock, NAB Broadcast Engineering Conference Proceedings, 2001
- 10) ISO/IEC 13818-6, 1998 — MPEG-2 Digital Storage Media command & Control, Chapter 8