

# IRT Metadata Workshop

*February 2004*



Mike Cox

*Mirador Techniques Ltd.*

*Chair, SMPTE W25 Committee*

# A Quote ...



“The traditional things you do in Broadcasting aren’t at all like what goes on in a database!”

*J. Haine*



- SMPTE and the EBU have been looking at all the implications of Digital TV for some time
- In September 1998 they published a joint report on technical standards required for the cost-effective interchange of digital material between businesses.

*“EBU/SMPTE Task Force for Harmonized Standards for the Exchange of Programme Material as Bitstreams”*

# Metadata

## Semantics and Syntax

- A Dictionary lists “words” and their definitions
  - But words on their own convey no intelligent meaning
- Words need to be associated together into phrases
  - To keep together words that qualify each other
  - To make sure the correct syntax is conveyed
- The same words and phrases can convey different semantics
  - A film can be about a wedding or a tragedy
  - But both will have actors, scene descriptions etc

Words and phrases will only make proper sense if we identify exactly and unambiguously every part of what we are talking about

# Content



Raw Programme Material  
together with  
related Information of any kind

# Essence

The raw programme material itself.

*I.e. The Digital (or analogue) representation of the  
Pictures, Sound and Text*

*Examples include:*

*Captions, Graphics, Moving Image, Audio, Sub-  
titles*

# Metadata

Broadly, Data about the Essence

*Describes what the Essence is - everything from vital technical data to administrative data*

*Examples include:*

*Sampling rate, Aspect ratio, Archive synopsis,  
Supplier telephone number, Title*

# Therefore:

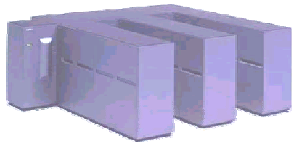
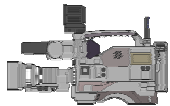
Content = Essence + Metadata

*and content is what is traded.*



# Metadata

## A rough Breakout



- Technical
  - *Physicality, Parametric, object structure*
- Production
  - *Conceptual, contextual, treatment etc*
- Content Interpretation
  - *Descriptions, genre, cataloguing*
- Business and Administration
  - *Branding, merchandising, contractual etc*
- Post production
  - *Edit decision, parametric data*
- Content repository
  - *Storage media, Rights*
- Publication and Distribution
  - *Publication media, slot data, audience data*



# Metadata Principles



# Principles

- Metadata is gathered piecemeal - only becomes information when there is enough data to give meaning.
  - Isolated data items tend to be ignored as meaningless
- Metadata must be captured at the earliest opportunity and preserved without re-keying.
  - Beware of focus, laziness and “nothing to do with me”
- Some Metadata will be liable to change and some sacrosanct.
  - E.g. Rights, Ownership, technical playback standards



# Principles

- Metadata which belongs together must be kept associated - e.g. roles and names
  - Can involve several layers of nesting
- Not all metadata will be required at all points in the process
  - The need to access it will skip stages
- When the metadata associated with Essence changes a new Version of Content is created
  - with implications for embedding



# Principles

- Atomise.
  - Always break information down into indivisible pieces
  - So, Name could be split into, say, 3 attributes: Forename, Middle Names, Surname
- Don't store computable attributes.
  - If something can be computed, only store the information needed to compute it
  - So, Age would be better replaced by Date of Birth



# The Digital Standards for Metadata

A horizontal bar with a rainbow gradient, transitioning from green on the left to red in the center and blue on the right.

## *The Metadata Dictionary*

*- one of a family of SMPTE Registries*

# SMPTTE Metadata Dictionary

## *SMPTTE 335M/RP210*

- Lists each metadata element
  - With a Name and a Definition
  - With a data Type (ASCII? Boolean? Integer?)
  - With its expected Length (One byte or a whole script?)
  - Against a unique ISO/SMPTTE Label (the Key)
- Is grouped for management purposes
  - Into Classes
  - Under Nodes and Sub-Nodes
- Is intended to be backwards compatible
  - Easy to make additions to
  - Deletions are never permitted

# Dictionary Groupings



- Identification
- Administration
- Interpretation
- Parametric
- Process
- Spacio-Temporal
- User
- Experimental



03	02	00	00	00	00	00	00	<b>Descriptive - Human Assigned</b>	Descriptors (Human Assigned) relating to analysis of the content			
03	02	01	00	00	00	00	00	<b>Categorisation</b>	Analytical categorisation of the content			
03	02	01	01	00	00	00	00	<b>Content Classification</b>	Content classification			
03	02	01	01	01	00	00	00	Content Coding System	The system of coding for programme classification eg Escort 2.4	ISO/IEC 646:1991 - ISO 7-Bit Coded Character Set	31 bytes max	
03	02	01	01	02	00	00	00	Programme Type	Type of programme (e.g., cartoon, film, ...)	ISO/IEC 646:1991 - ISO 7-Bit Coded Character Set	32 bytes max	
03	02	01	01	03	00	00	00	Genre	Programme genre (e.g., entertainment, current affairs magazine, Italo Western, ...)	ISO/IEC 646:1991 - ISO 7-Bit Coded Character Set	32 bytes max	
03	02	01	01	04	00	00	00	Target Audience	Target audience (e.g., children, 17 to 32, elderly, ...)	ISO/IEC 646:1991 - ISO 7-Bit Coded Character Set	32 bytes max	
03	02	01	02	00	00	00	00	<b>Cataloguing and Indexing</b>	Archival analysis of the essence metadata			
03	02	01	02	01	00	00	00	Status of Catalogue Data	The current status of the catalogue as a freeform text string	ISO/IEC 646:1991 - ISO 7-Bit Coded Character Set	127 bytes max	
03	02	01	02	02	00	00	00	Cataloguing, Indexing, Thesaurus or Gazetteer system used	The particular Cataloguing, Indexing or Thesaurus system used	ISO/IEC 646:1991 - ISO 7-Bit Coded Character Set	32 bytes max	
03	02	01	02	03	00	00	00	Theme	The category of the Theme of the content	ISO/IEC 646:1991 - ISO 7-Bit Coded Character Set	32 bytes max	
03	02	01	02	04	00	00	00	Subject Code	Subject Code.	ISO/IEC 646:1991 - ISO 7-Bit Coded Character Set	32 bytes max	
03	02	01	02	05	00	00	00	Key Words	Words or phrases summarizing an aspect of the data set.	ISO/IEC 646:1991 - ISO 7-Bit Coded Character Set	127 bytes max	
03	02	01	02	06	00	00	00	Key Frames	Freeform textual reference to a key frame of video in the data set	ISO/IEC 646:1991 - ISO 7-Bit Coded Character Set	127 bytes max	
03	02	01	02	07	00	00	00	Key Sounds	Freeform textual reference to a key sound in the data set	ISO/IEC 646:1991 - ISO 7-Bit Coded Character Set	127 bytes max	
03	02	01	02	08	00	00	00	Key data	Freeform textual reference to a key piece of data or program in the data set	ISO/IEC 646:1991 - ISO 7-Bit Coded Character Set	127 bytes max	
03	02	01	06	00	00	00	00	<b>Textual Description</b>	A textual characterization of the data set.			
03	02	01	06	01	00	00	00	Abstract	A brief narrative summary of the data set.	ISO/IEC 646:1991 - ISO 7-Bit Coded Character Set	1024 bytes max	
03	02	01	06	02	00	00	00	Purpose	A summary of the intentions with which the data set was developed.	ISO/IEC 646:1991 - ISO 7-Bit Coded Character Set	127 bytes max	
03	02	01	06	03	00	00	00	Description	A freeform textual description	ISO/IEC 646:1991 - ISO 7-Bit Coded Character Set	variable	
03	02	01	06	04	00	00	00	Colour descriptor	eg. Black and white, tinted etc	ISO/IEC 646:1991 - ISO 7-Bit Coded Character Set	32 bytes max	
03	02	01	06	05	00	00	00	Format descriptor	eg. Letterbox, Pillarbox etc	ISO/IEC 646:1991 - ISO 7-Bit Coded Character Set	32 bytes max	
03	02	01	07	00	00	00	00	<b>Stratum</b>	The descriptive stratum of the archival content analysis of the content			

# Metadata

## Dynamics and Persistence



- Metadata can (broadly) be static or dynamic
  - Static metadata applies unchangingly to an item of content
  - Dynamic metadata has values that change over time (e.g. timecode)
- Metadata can also be durable or transient
  - Durable metadata in a file will be unaffected by any exchange or further processing
    - And can be perpetually durable or finitely durable
  - Transient metadata reflects variable values such as used for business transactions
- There is also “Dark” Metadata
  - Dark metadata is that which is unknown to a decoder

# Descriptive Metadata Scheme 1

- Original objective was to do as much as we do now
  - Archives, tape labels, recording reports etc
  - Though it has grown since!
- Does *not* set out to Standardise the world
  - Does set out to accept organisational differences
  - Designed to extend easily as needs become apparent
  - Does *not* handle transient or in depth business metadata
  - Does *not* duplicate metadata in the structural metadata
    - Eg technical Standards, edit rates, durations etc
- Is *only* scheme 1
  - Nothing to stop writing another for a specific need
  - Document EG42 describes the generality

# Descriptive Metadata Scheme 1

- Split into Frameworks
  - To help with semantics
  - Example – location can be “real” or “pretend”
    - Think of Star Trek!!
- Sets can have many instances
  - Example – generating many shot descriptions to form a list.
- Sets have some “invisible” elements
  - For synchronising, identifying etc.
    - Instance numbers, UMIDs etc etc.
- The way Sets are presented to the User depend on the Application

# IRT Metadata Workshop

*February 2004*



Mike Cox  
*mike@miradortechniques.co.uk*