ONTOLOGY FOR MEDIA CREATION

PART 3C: AUDIO

Contents

This document is intended as a guide for companies developing or implementing products, solutions, or services for the future of media creation. No effort is made by Motion Picture Laboratories, Inc. to obligate any market participant to adhere to the recommendations in this document. Whether to adopt these recommendations in whole or in part is left to the discretion of individual market participants, using independent business judgment. Each MovieLabs member company shall decide independently the extent to which it will utilize, or require adherence to, these recommendations. All questions on member company adoption or implementation must be directed independently to each member company.

1 Introduction

The first moving pictures were silent. Muybridge's horse galloped noiselessly. Thomas Edison did some experiments combining his Kinetoscope with a phonograph, with little aesthetic or commercial success. This initial run of late 1920s talkies, including *Don Juan* (1926) and *The Jazz Singer* (1927), continued to use a combination of film and phonograph discs. By the very late 1920s, sound was being carried *on* the film itself, with systems such as Movietone.

After that, progress was rapid. *King Kong* (M. Cooper and E. Schoedsack, 1933) shows the complexity this allowed – combining dialog, music, and sound effects in the service of the story. The underlying technology changes constantly, and the transition from analog to digital has accelerated the rate of change. However, the underlying purpose – combining actors' voices, music, and narrative-driven sound effects – has remained the same, even though the underlying technology is completely different, as for example in *King Kong* (Peter Jackson, 2005)

In the modern world, there are thousands of bits and pieces of audio (the term OMC uses for recorded sound) that have to be tracked, managed, and combined to form a single final soundtrack. The OMC Audio Ontology provides mechanisms for doing that. It does not delve too far into the details of the audio essence (the underlying bits), relying instead on the work of standards groups such as SMPTE. As with similar areas (such as Images and Computer Graphics) OMC is about connecting elements within the production process and back to the narrative at an appropriate level of granularity.

The start and end points of getting sound into a finished Creative Work are simple enough: audio that goes with the visual portion is created (usually from many different captures), and the end result is a file or files that play alongside the video. It's the processing that happens in between that is complicated, and this part of OMC addresses the capture, the processing, and the finished product.

1.1 Generic Audio Workflow

Audio workflows can be very finely segmented. However, with modern tools the steps often blend into each other and collapse together.

There is some loose terminology in the workflow (as in most workflows), and one of OMC's goals is to provide precise and useful definitions of terms and concepts.

"Channel" and "track" are sometimes used interchangeably, which can cause confusion. Superficially, they are very similar: both are sequenced audio samples, but there are some differences:

- A track is a temporally continuous sequence of related audio samples, Tracks are an important part of the editing process, e.g. the music track or a voiceover track. "Track" can sometimes be thought of as a way of labelling pieces of audio. In OMC, tracks are most often used at the start of the audio pipeline.

- A channel is a distinct collection of sequenced audio samples intended for delivery to a single loudspeaker, loudspeaker array or other reproduction device, such as the left channel or the right channel in a stereo mix. In OMC, channels are most often used, with some exceptions, for the results of the audio pipeline.

A track might be divided across multiple channels, and a single channel can result from multiple tracks.

In OMC, playable sound data can be represented by a single-track audio file, a multi-track audio file, or an Asset Group containing multiple Audio Assets. For convenience, these things are sometimes collectively called “audio files.”

The sound editorial process deals with more than simple audio files. OMC defines various types of Audio Compositions (See ***Part 9: Utilities***), which are collections of Tracks and some instructions (called an Audio Session in OMC) for assembling them.

There is a potentially very complicated workflow between acquiring audio assets and generating final distributable audio. OMC defines Asset types and Composition types for a very fine-grained workflow. In OMC, the audio process starts with Assets captured on-set or elsewhere, uses a variety of Audio Compositions to process them, and ends with audio assets used as the sound for the distribution of a finished Creative Work.

1.2 Notational Conventions

In documents generally:

- The definition of a term included in the Dictionary is in bold, followed by the definition, e.g., **Creative Work:** A uniquely identified production.
- When a defined term is used in the text of a document, it is capitalized, for example in “The Production Scene is usually derived from a numbered scene in the Script,” Production Scene and Script are defined in the Ontology. (Note, a word that is part of defined term may sometimes be capitalized by itself as a shorthand, e.g., “Scene” may be used to indicate “Narrative or Production Scene.”)
- References to other Ontology Documents are in ***bold italic***, e.g., ***Part 3: Assets*** or ***Part 3A: Camera Metadata***

For Sample Attributes in the concept documents:

- If a data field or attribute is formally defined in this ontology or a connected ontology, it is italicized, e.g., *Setup* as an attribute refers to a defined concept.
- Attribute […] indicates an attribute can appear more than once, e.g., *Identifier* […]
- →Thing means that an attribute is expressed as a relationship to a Thing, e.g., the →*Script* attribute of Creative Work means there is a relationship Creative Work→*Script*
- A combination of the two indicates that the concept can have relationships to a set of things, e.g., →Components […]

- Many elements of the Ontology have a Context element. (See **Part 2: Context**.) Relationships declared in the Context are implied to have the item to which the Context is attached as their starting point, for example, Narrative Location→Context→Narrative Scene.

  Contextual relationships that are especially important to the concept being defined are given in the sample attributes tables as C→Thing or C→Thing […] as appropriate. These relationships can just as well be on the object that has the Context. For example, if Narrative Location has "C→Narrative Scene" as an attribute, it is ok to have the relationship directly on the Narrative Location or in its Context, e.g. Narrative Location→Narrative Scene or Narrative Location→Context→Narrative Scene.

  Some implementations (e.g. RDF) place these relationships directly on the class as well as allowing them in Context, and others (e.g. JSON) place all relationships in a Context.

2 Common Audio Terms

As mentioned, “track” and “channel” are often used imprecisely. Within the more formal OMC definitions there are many subtypes of Tracks and Channels. OMC does not formally subclass these subtypes, but distinguishes them with properties.

2.1 Audio Channel Name

Audio Channel Name: A formalization of the name of the loudspeaker the Audio Channel is intended to drive.

Notes:

The values for Channel should be taken from section 6 of SMPTE ST 428-12:2013 (e.g. “L” for Left and “R” for right), section 6 of SMPTE ST 2098-5:2018 (e.g. “Rts” for Right Top Surround), and section 5 of SMPTE 2067-8:2013. These shortened forms are used in many file-naming specs, interchange formats, and industry practices.

It is easy to convert the values in Channel to and from systems that use a single combined string, e.g. “l,r,c”.

2.2 Audio Content

Audio Content: Classification of the content type in a particular Audio Asset.

This is an extension of the concept of MCA Content and MCA Content Subtype, as described in SMPTE ST 377-41:2023. Besides Type and Subtype, OMC also allows an Audio Track to have a name, to allow distinguishing between different audio essence of the same type and subtype.

Attributes for Audio Content

Term	Value
MCA Content	String. This is a classification of the content in the audio essence. See SMPTE ST 377-4:2021, SMPTE ST 377-41:2023 and Appendix B.
MCA Content Subtype	String. Optional. This further differentiates the content in the audio essence. See SMPTE ST 377-4:2021, SMPTE ST 377-41:2023 and Appendix B.
Language	RFC 5646 compliant String indicating the language used, if the track is language-oriented.

Notes:

OMC takes an initial set of values for MCA Content and MCA Content subtype from SMPTE ST 377-41:2023. (See Appendix B.) These should be used whenever possible, although we expect implementations to extend it as needed. The fields should use the MCA Content and Content Subtypes symbols, rather than the value, e.g. "FX" rather than "Effects" or "ANN" rather than "Announcer" for MCA Content, and "CRD" rather than "Crowd" and "VI" rather than"Visually Impaired" for MCA Content Subtype.

MCA Content will often be one of Dialog, Music, and Sound Effect. It is used when audio relates to a particular aspect of the Creative Work's sound.

Many workflows treat separate components of various MCA Content types separately, for example by keeping character dialog and audio narration for the visually impaired separate, or processing a dedicated score and licensed music separately.

2.3 Soundfield

Digital audio is converted to sound during playback. This involves the codec, of course, but the sound has to come through one or more speakers. Loosely, the soundfield describes this arrangement.

Soundfield: The acoustical space created by simultaneously reproducing one or more Audio Channels.

Notes:

This definition is taken from SMPTE ST 428-12:2013

The string value in Soundfield should follow the guidelines for Soundfield Group naming in SMPTE ST 428-12:2013, SMPTE ST 2098-5:2018, and SMPTE ST 2067-8:2013, using the symbols rather than the names, e.g. "51" rather than "5.1."There are examples in Table 3 in the same document.

3 Audio Assets

There are two broad classes of Audio assets:

- For Audio files, playback of the essence within the file does not depend on other metadata. Audio files are a Structural Class.
- For Audio objects, the playback and spatialization of the Audio essence depend on properties, such as position, that vary over time. This information can be carried in metadata with the essence. Audio Objects are a Functional Class.

3.1 Digital Audio Structural Class

This describes an audio asset, independent of its use. It covers single files with a single track or channel, single files with multiple tracks or channels - such as a container file (e.g. MXF) or an interleaved file - and Asset Groups containing multiple files.

Sample Attributes for Digital Audio Structural Class

Attribute	**Description**
Identifier […]	One or more identifiers for the Digital Audio Structural Characteristics. At least one of these should be resolvable within the production environment.
Name	A human-readable name for the audio structural characteristics
Description	A description of this audio structural characteristics
File Details	See **Part 9: Utilities**
File Type	See list below
Codec	The Codec used for the Audio file. See ***Part 9: Utilities***
Audio Sample Rate	Number. See below
Audio Sample Size	Number. See below
Audio Bit Rate	Number. See below
Custom Data	Anything that is application or workflow dependent that can't be otherwise expressed in the Ontology or needs to be present in a particular format.
→ *Context* […]	Context related to this Asset Structural Characteristic
C->*Slate*	A capture is usually associated with a particular capture, and hence with a Slate.

Notes:

If the audio file is tightly coupled to a track or a channel, this should be expressed in its functional class.

Audio Sample Rate: The average number of samples per second taken from the source audio input in KHz,

Audio Sample Size: The number of bits per audio sample (also referred to as "Bit Depth)."

Audio Bit Rate: The number of bits in one second of sampled audio, expressed in Kbits/ per second.

3.2 Audio Functional Classes

The Audio Functional Classes may duplicate some of the information from the audio structural characteristics. This is unavoidable in current workflows and audio tools. Over time, we expect best practices to emerge that reduce this duplication and the accompanying data sync issues.

3.2.1 Audio Channel

Audio Channel: A distinct collection of sequenced audio samples that are intended for delivery to a single loudspeaker or other reproduction device.

This definition is taken from SMPTE ST 377-4:2021, SMPTE ST 428-12:2013, etc.

Sample Attributes for Audio Channel

Term	**Definition**
Identifier […]	One or more identifiers for the Audio Channel. At least one of these should be resolvable within the production environment.
Name	The name of the Audio Channel.
Description	A description of the Audio Channel.
→ *Audio Channel Name* […]	String. See Notes.
Custom Data	Anything that is application or workflow dependent that can't be otherwise expressed in the Ontology or needs to be present in a particular format.
→ *Context* […]	Any Context for the Asset. See ***Part 2: Context***

Notes:

Audio Channel Name has a single value if the audio structural characteristics to which this Audio Channel applies is a single file representing a single channel. If the structural element Is a single file with multiple channels or an Asset Group, Audio Channel Name has multiple values, one for each channel represented.

3.2.2 Audio Track

"Track" is a heavily overused term – a track was originally a physical thing – a recorded piece of media - and the name has stuck. "Track" can be used in many ways:

- Audio in a STEM is sometimes called a track.
- A file representing a 5.1 audio mix may be called a track and contains 6 audio tracks.
- An audio container file (e.g. MXF) contains multiple tracks (individual WAV files).
- People often refer to a video track or a subtitle track.
- An individual file can be called a track, e.g. the captured sound from one instrument.
- Two separate tracks can be combined into a single track (e.g. a vocal track and a piano track).
- A completed playable thing (e.g. track 2 on an album).
- And sometimes, "track" means "channel" – see above.

For this reason, OMC avoids unmodified "track" and uses "Audio Track" for some specific things.

Audio Track: A temporally continuous sequence of related samples.

Sample Attributes for Audio Track

Term	**Definition**
Identifier […]	One or more identifiers for the Audio Track. At least one of these should be resolvable within the production environment.
Name	The name of the Audio Track.
Description	A description of the Audio Track.
Audio Track Name	String. Further differentiation for the Audio Track, if needed.
Audio Content […]	See above.
Custom Data	Anything that is application or workflow dependent that can't be otherwise expressed in the Ontology, or needs to be present in a particular format.
→ *Context* […]	Any Context for the Asset. See ***Part 2: Context***

Notes:

Note that although the Audio Track is continuous, the samples may not be contiguous.

3.2.3 Audio Mix

At some point, a codec is used to turn all of the processed audio (often represented as an Audio Composition) into a playable form. The result of this is a Mix, which contains all of the necessary audio channels. Multiple Mixes can be produced from the same Audio Composition.

As with other audio functional classes, Mix can be used with files that contain a single channel, multi-channel files, and Asset groups. A Mix can contain Audio Objects.

Different kinds of Mix are used at various points in the production process.

Audio Mix: A combination of multiple Audio Assets into a particular format.

Sample Attributes for Audio Mix

Term	**Definition**
Identifier […]	One or more identifiers for the Mix. At least one of these should be resolvable within the production environment.
Name	The name of the Mix.
Description	A description of the Mix.
Audio Mix Type	String. See below.
Soundfield	String. See above.
Audio Channel Name[…]	String. See Notes.
Custom Data	Anything that is application or workflow dependent that can't be otherwise expressed in the Ontology or needs to be present in a particular format.
→ *Context* […]	Any Context for the Mix. See ***Part 2: Context***

Notes:

For an Audio Mix, Audio Channel Name should be a list of all the channels contained in the Mix, following the guidelines given in Audio Channel (above.)

An Asset Group that serves as the structural characteristics for a Mix can include Audio Objects and Audio Channels. See Audio Channel for how to represent single channel files and multichannel files.

3.2.3.1 Audio Mix Type

Mixes are used throughout the production process, and OMC defines a few standard Mix Types

On-set Mix: A Mix of the on-set captured Audio.

Temporary Mix: A non-final Mix used in the creative process, e.g. for review.

Notes:

Final mix represents the final stage of the creative process. It is generally treated as a verb, not a noun. See Printmaster and Audio Distribution Asset below.

For more Audio Mix types, see *Printmaster and Final Audio Deliverable*.

A Temporary Mix is sometimes called a Temp Mix, Work In Progress Mix or a Bounce Mix.

3.2.4 Audio Object

Audio Objects combine a set of audio samples and associated metadata that allows the samples to be played based on the metadata, which can include, for example, spatial position. Unlike simple audio files, the playback may not be associated with a single loudspeaker. Audio objects are rendered in a 3D sound field, the configuration of which is defined by vendor-specific formats.

Audio Objects and their metadata are not currently standardized, and are often vendor specific. For that reason, the OMC classes for them are basic shells, and important data that might be needed in the workflow should be kept in the Custom Data field.

The metadata and essence can be caried on a single file, or in multiple files.

Audio Object: A segment of audio essence with associated metadata describing positional and other properties which may vary with time.

Audio Objects can be used with a Digital Audio Structural Class or an Asset Group. The extra metadata can be carried in the Digital Audio Asset or in a separate file.

- If the Audio Object is a single file, it can be either a Track or Audio Object Metadata (see below.)
- If the Audio Object spans multiple files, the Asset Group should contain Assets for everything the Audio Object needs as either Tracks or Audio Object Metadata.

Sample Attributes for Audio Object

Attribute	**Description**
Identifier […]	One or more identifiers for the Audio Object. At least one of these should be resolvable within the production environment.
Name	A human-readable name for the Audio Object
Description	A description of this Audio Object.
Audio Content	See above. An Audio Object can have MCA Content information, e.g. “FX” for sound effects.
Custom Data	Anything that is application or workflow dependent that can’t be otherwise expressed in the Ontology or needs to be present in a particular format.
→ *Context* […]	Context related to this Audio Object.

Audio Object Metadata: An Asset containing metadata for an Audio Object.

Sample Attributes for Audio Object Metadata

Attribute	Description
Identifier […]	One or more identifiers for the Audio Object Metadata. At least one of these should be resolvable within the production environment.
Name	A human-readable name for the Audio Object Metadata
Description	A description of this Audio Object Metadata.
Custom Data	Anything that is application or workflow dependent that can't be otherwise expressed in the Ontology or needs to be present in a particular format.
→ *Context* […]	Context related to this Audio Object Metadata.

Notes:

Audio Object Metadata has very little information in the OMC class. It is assumed that applications using it will use Custom Data or be able to get relevant information from the file.

Audio Object Metadata can have Digital Data, Digital Document, and Digital Audio Asset as its structural class. These structural Assets contain X.Y.Z spatial data that changes over the time of the audio essence.

Sometimes this is inside the Audio Object and is not represented separately. However, if desired, an Audio Object Metadata element can be created as another Asset Functional Class to use with the Audio Object.

4 Examples of Audio Assets, Tracks, and Channels

Here are some common examples of using Audio Assets.

Single file from capture

- Asset Structural Class: Audio structural asset
- Asset Functional Class: An audio capture

Single file used as a Track

- Asset Structural Class: Audio structural asset
- Asset Functional Class: An Audio Track

Single file with left channel in it

- Asset Structural Class: Audio structural asset
- Asset Functional Class: Audio Channel, property ‘Left’

Single Interleaved stereo file

- Asset Structural Class: Audio structural asset
- Asset Functional Class: Mix, property ‘Stereo’

Two files for left and right stereo channels

- Asset Structural Class: Asset Group with two Assets, each of which has
    - Asset Structural Class: Audio structural asset
    - Asset Functional Class: Audio Channel with ‘left’ or ‘right’
- Asset Functional Class: Mix, property ‘Stereo’

5 Audio Compositions

Most of the editing and processing of audio in the production process is done with Compositions rather than on individual files.

As with other Compositions an Audio Composition consists of some data to work on and a set of instructions. As with other Compositions, the data can be individual Assets or other Compositions.

Audio editing software generally uses a set of audio files for the tracks and a separate file with instructions for adjusting and assembling the tracks. The assembly instructions are proprietary, and OMC abstracts this into an Audio Session, much as video EDLs, OTIO, and so on are abstracted into a Sequence Chronology Descriptor.

Audio Composition: A Composition consisting of Audio Assets, Audio Compositions, and an Audio Session.

Sample Attributes for Audio Composition

Term	**Definition**
Identifier […]	One or more identifiers for the Audio Composition. At least one of these should be resolvable within the production environment.
Name	The name of the Audio Composition
Description	A description of the Audio Composition.
Audio Content	See above. This is optional, but should be used when appropriate, e.g. for a dialog or music STEM or Sound Editorial Unit.
Soundfield	See above. Optional, but should be used if this Audio Composition is intended for a particular soundfield, e.g. a STEM.
->startHere	Audio Session. The starting point for loading and using this Audio Composition.
Asset ->[] *Asset Structural Class*->[] *Composition*->[…]	Inherited from Composition. NOTE: These should generally not be used.
->*Audio Track*[…]	Audio Tracks used directly by this Audio Composition. The relationship is includesAudioTrack/includedBy.
->*Audio Object*[…]	Audio Objects used directly by this Audio Composition. The relationship is includesAudioObject/includedBy
->*Audio Composition*[…]	Any Audio Compositions used directly by this Audio Composition. The relationship is includesAudioComposition/includedBy.
Custom Data	Anything that is application or workflow dependent that can't be otherwise expressed in the Ontology or needs to be present in a particular format.
→ *Context* […]	Any Context for the Audio Composition. See ***Part 2: Context***

Term	Definition
C→Production Scene	The Production Scene for this Audio Composition.
C→Editorial Sequence	The Editorial Sequence for this Audio Composition.
C→Partition	The Partition (Reel or Part) for this Audio Composition.

Notes:

An Audio Composition should have at least one of the indicated Context elements so it can be connected to the appropriate video.

If an Audio Composition uses a bare structural class for an Audio Object (i.e., one with no functional class) it should be included using the Asset Structural Class element inherited from a generic Composition.

A single Track as defined within the Audio Session can be composed of multiple files. There are two ways to handle this in OMC

- Include all the Tracks in the Composition and let the audio editing software determine what to do, based on information in the Audio Session.
- Bundle the separate files together into an Asset Group and include that Asset Group in the Audio Composition.

Audio Session: An Asset that represents the state of the audio mixing and editing process.

An Audio Session includes information about how to process the audio, but can also contain plugins or information about plugins.

Sample Attributes for Audio Session

Attribute	Description
Identifier […]	One or more identifiers for the Audio Session.
Name	A human-readable name for the Audio Session.
Description	A description of the Audio Session,
Audio Processing Action […]	Optional String. See Below
Session Type	Optional String. This can indicate things like “edit”, “mix”, and so on.
Custom Data	Anything that is application or workflow dependent that can’t be otherwise expressed in the Ontology or needs to be present in a particular format.
→*Context*	

Notes:

An Audio Session is an Asset Functional Class; its structural characteristics are usually Digital Data or Digital Document.

Common file types for an Audio Session are PTX and PTF (for ProTools sessions); AAF (Avid); OMF; and OTIO.

Audio Processing Action: Indication of what was done to the audio in an Audio Session.

This is a string, and provides some information about the processing that was done without having to open the Audio Session asset. It is not currently controlled vocabulary. Examples include “denoise” and “spatialization”.

An Audio Processing Action canbe very complex, combining many separate operations. The level of detail included in this field is workflow- and implementation-dependent.

5.1 Sound Editorial Unit

Some preprocessing may happen before everything is brought together. This is done with a Sound Editorial Unit.

Sound Editorial Unit is a named subclass of Audio Composition.

Sound Editorial Unit: A collection of sound samples that have been selected in reference to the Creative Work or a portion of it.

This is sometimes called “Unit” or “Sound Editorial” and is not included in all workflows.

5.2 Premix

Some work, such as spatial decisions and aligning ADR with other dialog, can be made before the final mix.

Premix is a named subclass of Audio Composition.

Premix: A collection of sound samples that have been processed for technical or creative review.

There are generally separate Premixes for Dialog, ADR, and various kinds of sound effects, but once again, this processing can be included in other parts of the workflow. These are sometimes called “virtual outputs”.

Note:

Premix is not an OMC Mix. Because it is a Composition, a Mix can be produced from it.

5.3 STEMs and Beds

5.3.1 STEM

A STEM is a completely mixed component (e.g., dialog, music, effects) of the soundtrack of a Creative work. There is generally a set of STEMs for each Soundfield, such as 5.1 or Atmos. (See Audio Classification.) Combining all the STEMs together at unity gain gives the final soundtrack for a particular Soundfield, such as 5.1 or Atmos (see Printmaster, below.) The process of producing the STEMs is called "final mix".

STEM is a named subclass of Audio Composition.

Audio Classification: One of "Dialog", "Effects", or "Music."

STEM: Final Audio Assets and Compositions combined based on their Audio Classification, used to produce the final sound for the Creative Work in a particular Soundfield.

Notes:

The STEM for the widest Soundfield[1] (loosely, the one with the most playback channels) is often produced first, with STEMs for narrower Soundfields derived from it. For example, a 5.1 STEM can be a derivation of an Atmos STEM.

In OMC, "Derivation" and "Variant" are both possible relationships between wider and narrower STEMS. *We expect that some best practices will emerge as implementation progresses.*

The Dialog, Effects, and Music STEMs can be combined into a single Audio Composition for convenience.

Audio Classification is a required property for STEMs.

Soundfield is a required property for STEMs.

In the case of immersive audio, the Audio Objects and their spatial metadata are caried forward from the STEMs into the Printmaster.

5.3.2 Bed

Bed: The Channel component of an immersive STEM or Printmaster.

The immersive stem beds are combined into a single bed in the immersive Printmaster. Any Audio Objects from the immersive STEMs are carried over into the immersive Printmaster.

Note:

This is a vocabulary term that can be used as a label, and has no associated data structure.

[1] The first Soundfield is often called the "native" Soundfield.

6 Printmaster and Audio Distribution Asset

Once the final STEMs are done, they must be converted into something playable for the final Creative Work. This can be thought of as a two-part process.

- The STEMs are used to produce a Printmaster, which has a producedBy relationship to the STEM.
- Audio Distribution Assets are derived from the Printmaster,

Processing the Printmaster produces audio files for particular formats. OMC calls these Final Audio Deliverables.

Printmaster: The combination of STEMs for a particular Soundfield at unity gain.

Audio Distribution Asset: Playable audio that can be packaged and distributed with a finished Creative Work.

Audio Distribution Asset is an OMC Audio Mix Type.

Different Soundfields have different OMC representations of the Printmaster. For Atmos, for example,

- The Printmaster is an Audio Composition, with the Bed, Audio Objects, and an Audio Session.
- The Final Audio Deliverables have a producedBy relationship to the Printmatser, and can be, for example, an ATMOS DCP or a Dolby Digital Plus (DD+) bitstream.

For a 5.1 Soundfield (e.g.),

- For a 5.1 Soundfield, the Printmaster is an Audio Mix Type.
- This Printmaster can be run through various codecs to produce, e.g., an MP3 or AC3 Digital Audio Asset. These Final Audio Deliverables have a representationOf relationship to the Printmaster.

7 Sample Audio Workflow

In many cases, stages of this workflow are collapsed together, and the OMC classes support this. Keeping in mind that these are logical tasks that can be combined into single tasks or stages of the workflow, a generic OMC audio workflow is as follows. Note that it starts with Audio Assets (Tracks), does most of its processing using Audio Compositions, and ends up with creating a Printmaster.

The diagram shows the flow of Assets and Compositions through the audio workflow, with some details of the various artifacts in the top row.

This example shows only dialog-related audio in detail. Other types (ADR, music, and sound effects) are similar. It does not include Audio Objects. It also shows how the steps before the final mix can be elided.

Capture

- Two Audio Tracks of production dialog are captured on set.
- Two Audio Tracks of ADR are captured; their structure is similar to that of the Dialog tracks, but their MCA Content would be “ADR”.

Sound Editorial

- The two dialog tracks are processed by sound editorial into a Sound Editorial Unit - an Audio Composition with an Audio Session and the two tracks.

Pre-mix

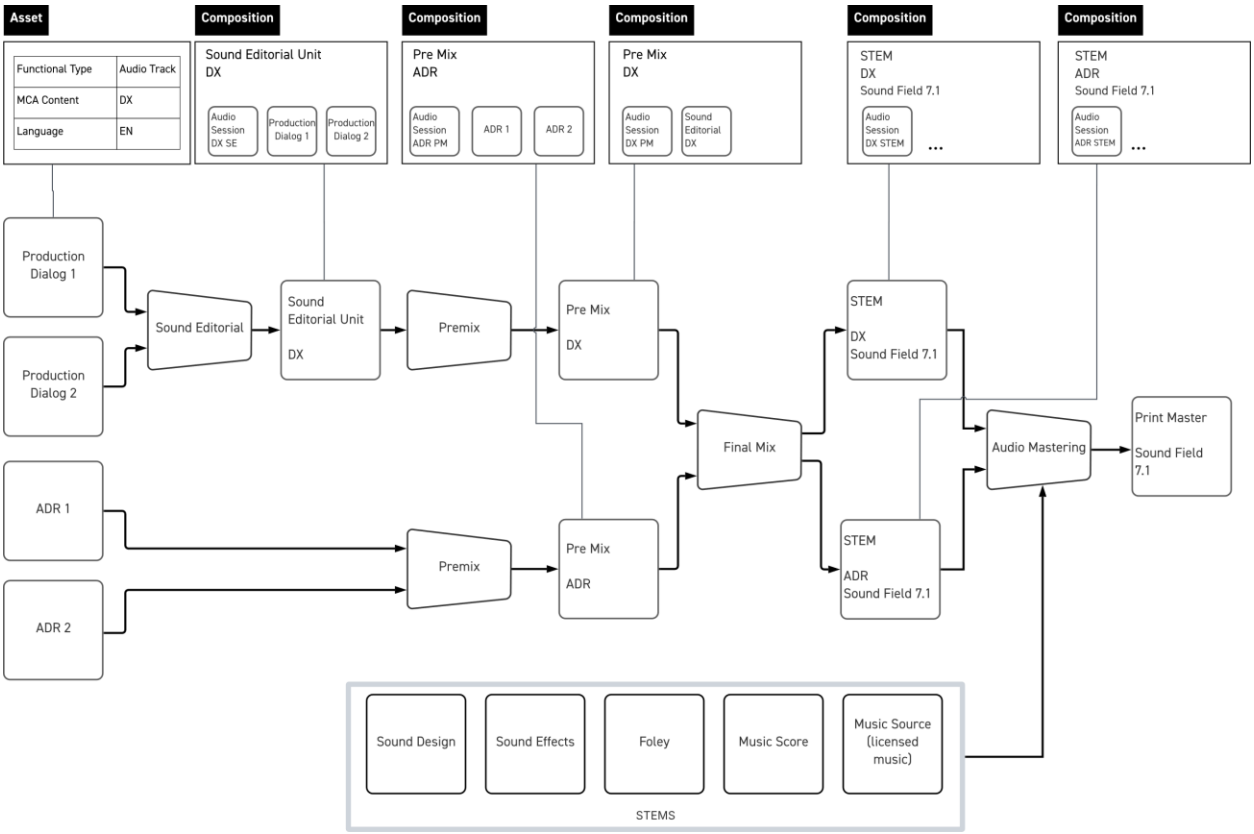
- The ADR premix is done directly from the captured Audio Tracks; this is expanded in the Pre Mix ADR Composition box (third box in the top row.)
- The Dialog premix is done from the dialog Sound Editorial Unit (shown in the fourth box in the top row.)

STEMs are derived from the pre-mixes, and are always associated with a particular Soundfield.

- The 7.1 Dialog STEM has an Audio Session and, in this case, the output of the dialog pre-mix, though it could have other audio tracks in it as well.
- This is also true for the 7.1 ADR STEM.

Final Mix

- The STEMs, including ones for other kinds of audio (some examples of which are given in the bottom row), are the input to the final mix process.
- For details on how the Print Master is processed into Audio Distribution Assets, see “Printmaster and Audio Distribution Assets” above.


Asset
Composition
Composition
Composition
Composition
Composition
Functional Type
Audio Track
MCA Content
DX
Language
EN
Sound Editorial Unit
DX
Audio Session DX SE
Production Dialog 1
Production Dialog 2
Pre Mix
ADR
Audio Session ADR PM
ADR 1
ADR 2
Pre Mix
DX
Audio Session DX PM
Sound Editorial DX
STEM
DX
Sound Field 7.1
Audio Session DX STEM
...
STEM
ADR
Sound Field 7.1
Audio Session ADR STEM
...
Production Dialog 1
Production Dialog 2
Sound Editorial
Sound Editorial Unit
DX
Premix
Pre Mix
DX
Final Mix
STEM
DX
Sound Field 7.1
Audio Mastering
Print Master
Sound Field 7.1
ADR 1
ADR 2
Premix
Pre Mix
ADR
STEM
ADR
Sound Field 7.1
Sound Design
Sound Effects
Foley
Music Score
Music Source (licensed music)
STEMS

Appendix A External references

A.1 OMC

These are terms defined elsewhere in the Production Ontology, included here for ease of reference.

Media Creation Context: Informs scope within the construction process of a Creative Work.

See **Part 2: Context**

Asset: A physical or digital object or collection of objects specific to the creation of the Creative Work.

See **Part 3: Assets**

Camera Metadata: Capture-specific details and information about the Camera itself.

See **Part 3A: Camera Metadata**

Participant: The entities (people, organizations, or services) that are responsible for the production of the Creative Work.

See **Part 4: Participants**

Task: A piece of work to be done and completed as a step in the production process.

See **Part 5: Tasks**

Creative Work: A uniquely identified production.

See **Part 6: Creative Works**

Relationship: Describes and defines the connections between elements of the Ontology, such as Assets, Tasks, Participants, and Contexts.

See **Part 7: Relationships**

Infrastructure: The underlying systems and framework required for the production of the Creative Work; it is generally not specific to a particular Creative Work.

See **Part 8: Infrastructure**

Utilities: Common data models and data structures used in multiple places and in multiple ways in a larger system.

See **Part 9: Utilities**

Identifier: An identifier uniquely identifies an entity within a particular scope.

See **Part 9: Utilities**

A.2 External Standards

The Audio Ontology makes use of the following IETF standards:

IETF, **RFC 5646**, “Tags for Identifying Languages”

The Audio Ontology makes use of the following SMPTE standards:

SMPTE, **ST 377-1:2019**, “Material Exchange Format (MXF) – File Format Specification”

SMPTE, **ST 377-4:2021**, “MXF Multichannel Audio Labeling Framework”

SMPTE, **ST 377-41:2023**, “MXF Multichannel Audio Controlled Vocabulary: Content Property Values”

SMPTE, **ST 428-12:2013**, “D-Cinema Distribution Master Common Audio Channels and Soundfield Groups”

SMPTE, **ST 2067-8:2013**, “Interoperable Master Format - Common Audio Labels”

SMPTE, **ST 2098-1:2018**, “Immersive Audio Metadata”

SMPTE, **ST 2098-5:2018**, “D-Cinema Immersive Audio Channels and Soundfield Groups”

Appendix B Sample Controlled Values

In general, OMC uses symbols rather than names.

B.1 MCA Content

MCA Content	Symbol	Notes
Primary	PRM	
Secondary Audio Program	SAP	
Hearing Impaired	HI	
Descriptive Video	DV	
Dialog	DX	
Music	MX	
Effects	FX	
Filled Effects	FFX	
Music and Effects	ME	
Optional Music and Effects	OP	
Music and Effects with Optional	MESP	
DME	DME	
NDME	NDME	
Program Narration	PNAR	
Optional Narration	ONAR	
Voice Over	VO	
Visually Impaired	VI	
Recorded Commentary	CM	
Live Commentary	LCM	
Silence	MOS	
Automated Dialog Replacement	ADR	
Group	GRP	
Walla	WLA	
Crowd	CRD	
Vocals	VOC	
Foley	FOL	
Backgrounds	BG	
Custom	x-<1-4 character tag>	Used for custom values

B.2 MCA Content Subtype

MCA Content Subtype	Symbol	
Director	DIR	
Technical	TECH	
Writer	WRT	
Cast	CAST	
Announcer	ANN	

MCA Content Subtype	Symbol	
Commentator	CTR	
Footsteps	FS	
Props	PRP	
Cloth	CL	
Other	OTHER	

B.3 Channel Names

This table lists standard Channel names from SMPTE ST 428-12:2013, SMPTE 2067-8:2013, and SMPTE ST 2098-5:2018. Please see those documents for formal definitions. This list can be extended either with new names and symbols or by extending existing ones

Channel Name	Channel Symbol	Notes
Left	L	SMPTE ST 428-12:2013
Right	R	SMPTE ST 428-12:2013
Center	C	SMPTE ST 428-12:2013
LFE	LFE	Low frequency effects SMPTE ST 428-12:2013
Left Surround	Ls	SMPTE ST 428-12:2013
Right Surround	Rs	SMPTE ST 428-12:2013
Left Side Surround	Lss	SMPTE ST 428-12:2013
Right Side Surround	Rss	SMPTE ST 428-12:2013
Left Rear Surround	Lrs	SMPTE ST 428-12:2013
Right Rear Surround	Rrs	SMPTE ST 428-12:2013
Left Center	Lc	SMPTE ST 428-12:2013
Right Center	Rc	SMPTE ST 428-12:2013
Center Surround	Cs	SMPTE ST 428-12:2013
Hearing Impaired	HI	SMPTE ST 428-12:2013
Visually Impaired	VIN	Video impaired narration. SMPTE ST 428-12:2013
Left Height	Lh	SMPTE ST 2098-5:2018
Right Height	Rh	SMPTE ST 2098-5:2018
Center Height	Ch	SMPTE ST 2098-5:2018
Left Surround Height	Lsh	SMPTE ST 2098-5:2018
Right Surround Height	Rsh	SMPTE ST 2098-5:2018
Left Side Surround Height	Lssh	SMPTE ST 2098-5:2018
Right Side Surround Height	Rssh	SMPTE ST 2098-5:2018
Left Rear Surround Height	Lrsh	SMPTE ST 2098-5:2018
Right Rear Surround Height	Rtsh	SMPTE ST 2098-5:2018
Left Top Surround	Lts	SMPTE ST 2098-5:2018
Right Top Surround	Rts	SMPTE ST 2098-5:2018
Top Surround	Ts	SMPTE ST 2098-5:2018

Channel Name	Channel Symbol	Notes
Mono One	M1	SMPTE 2067-8:2013
Mono Two	M2	SMPTE 2067-8:2013
Left Total	Lt	SMPTE 2067-8:2013
Right Total	Rt	SMPTE 2067-8:2013
Left Surround Top	Lst	SMPTE 2067-8:2013
Right Surround Top	Rst	SMPTE 2067-8:2013
Surround	S	SMPTE 2067-8:2013
Numbered Source Channel	NSC[Channel Number]	SMPTE 2067-8:2013

B.4 Soundfield Names

This table lists standard Soundfield names from SMPTE ST 428:12-2013 and SMPTE ST 2067-8:2013. See the Soundfield section of SMPTE ST 428:12-2013 for preferred practices for extending it.

Soundfield Name	Soundfield Symbol	Soundfield Channels	Source
5.1	51	L, C, R, LS, RS, LFE	SMPTE ST 428:12-2013
7.1DS	71	L, C, R, Lss, Rss, Lrs, Rrs, LFE	SMPTE ST 428:12-2013
7.1SDS	SDS	L, Lc, C, Rc, R, Ls, Rs, LFE	SMPTE ST 428:12-2013
6.1	61	L, R, C, Lss, Rss, Cs, LFE	SMPTE ST 428:12-2013
1.0 Monaural	M	C	SMPTE ST 428:12-2013
Standard Stereo	ST	L, R	SMPTE ST 2067-8:2013
Dual Mono	DM	M1, M2	SMPTE ST 2067-8:2013
Discrete Numbered Sources	DNS	NSC001, NSC002	SMPTE ST 2067-8:2013
3.0	30	L, C, R	SMPTE ST 2067-8:2013
4.0	40	L, C, R, S	SMPTE ST 2067-8:2013
5.0	50	L, C, R, Ls, Rs	SMPTE ST 2067-8:2013
6.0	60	L, C, R, Ls, Rs, Cs	SMPTE ST 2067-8:2013
7.0 DS	70	L, C, R, Lss, Rss, Rls, Rrs	SMPTE ST 2067-8:2013
Lt-Rt	LtRt	Lt, Rt	SMPTE ST 2067-8:2013
5.1 EX	51EX	L, C, R, Lst, Rst, LFE	SMPTE ST 2067-8:2013