

Common Extras Manifest Metadata

CONTENTS

1	Introduction	1
1.1	Overview.....	1
1.1.1	Scope	1
1.1.2	Extras Manifest.....	1
1.1.3	Relationship of Extras Manifest to Common Metadata.....	2
1.2	Document Organization	2
1.3	Document Notation and Conventions	3
1.3.1	XML Conventions	3
1.3.2	Use of Language	5
1.3.3	General Notes	5
1.4	Normative References	5
1.5	Informative References.....	6
2	Encoding.....	7
2.1	Identifiers	7
2.2	Asset References	7
2.3	General Types Encoding	7
2.3.1	Timecode Encoding.....	8
2.3.2	Clip References	8
3	Extras Manifest	10
3.1	Compatibility	10
4	Inventory	12
4.1	Inventory-type	12
4.2	Inventory Asset Types	12
4.2.1	InventoryAudio-type.....	12
4.2.2	InventoryVideo-type.....	13
4.2.3	InventorySubtitle-type.....	13
4.2.4	InventoryImage-type.....	13
4.2.5	InventoryInteractive-type	14
5	Track Groups and Chains	15
5.1	Track Group.....	15
5.1.1	TrackGroup-type	15
5.1.2	ContainerLanguagePair-type.....	17
5.1.3	Chapter Metadata.....	18
5.2	Chains	19
5.2.1	Chains Use Cases.....	20
5.2.2	Chapters and Chains.....	20
5.2.3	ChainList-type	20
6	Picture Groups and Galleries	23
6.1	Picture Group	23
6.2	Picture Group Type	23
6.2.1	Picture-Type	24
6.3	Gallery	25

6.3.1	Gallery-type	25
7	Title Catalog.....	27
7.1	Title Catalog List.....	27
7.2	Titles Catalog.....	27
7.2.1	Title.....	29
8	Annex A: Track Selection Process.....	30
8.1	Defined Preferences.....	31
8.2	Default Audio and Subtitle Track Selection.....	32
8.2.1	Default Audio Track Selection	33
8.2.2	Default Primary Subtitling Presentation Track Selection	34
8.3	Alternate Subtitling Presentation Track Selection.....	36
8.3.1	Select Alternate Subtitle Track	36

REVISION HISTORY

Version	Date	Description

1 INTRODUCTION

Electronic distribution of movies has traditionally just been audio, video and subtitles. This specification defines data to support additional functionality intended to meet or exceed a DVD and Blu-ray experience.

Functionality supported includes features such as

- Chapter selection
- Audio and subtitle track selection
- Supplemental audio, video and image materials

This specification is designed as a resource. Those using this specification may extend the definition with additional data element specific for their needs. They may replace elements with others perhaps more suitable to their needs; however, for interoperability all are highly encouraged to use the data elements exactly as defined.

Extras Manifest is part of the Common Metadata family of specifications.

1.1 Overview

1.1.1 Scope

This specification defines the underlying structure for information presented to a user in the form of 'extras' or 'interactivity' without defining the presentation itself. That is, it defines the algorithm for default track selection without defining the user interface for selecting tracks. This makes Extras Manifest useful across multiple user interface modalities and appearances.

The Extras Manifest can be used in either a B2B interface, such as between a studio and a streaming site, or in a B2C interface such as found on a tablet or home media player.

The Extras Manifest is designed to allow linkages to presentation components. For example, a separate menu specification can reference 'TitleID' or 'GalleryID' elements to link those menus to the definitions in this specification.

1.1.2 Extras Manifest

The Extras architecture has the following data objects

- Compatibility – What level of functionality is required to use this document
- Inventory – What media components comprise the Extras experience. This includes track metadata. Note that media components can be referenced as part of an Extras Package (container) or remotely on the Internet.

- Track Grouping – Which tracks can be played with each other; such as, related video, primary audio, commentary audio and subtitles. This includes information about track selection and chapters.
- Chains – Sequences of clips (Track Groups with entry and exit points) chained together. This is a simple composition playlist allowing video pieces to be strung together (e.g., distribution cards, followed by anti-piracy cards, followed by a main title).
- Image Grouping – Which images comprise a gallery
- Title Catalog – Indicates the main title, supplemental titles and galleries.

From these components a user experience can be created. A player starts with the Title Catalog and discovers a main title, alternate titles and picture galleries. From this the application can create the user experience including title selection, default track selection, gallery display, chapter selection, and so forth.

1.1.3 Relationship of Extras Manifest to Common Metadata

Extras Manifest is an extension to Common Metadata and may be used in conjunction with Common Metadata, or as its own entity.

Common Metadata includes elements that cover typical definitions of media, particularly movies and television. Common Metadata has two parts: Basic Metadata and Digital Asset Metadata. Basic Metadata includes descriptions such as title and artists. It describes information about the work independent of encoding. Digital Asset metadata describes information about individual encoded audio, video and subtitle streams, and other media included. Package and File Metadata describes one possible packaging scenario and ties in other metadata types. Ratings and Parental Control information is described.

Common Metadata is designed to provide definitions to be inserted into other metadata systems. A given metadata scheme, for example, the Entertainment Merchant's Association (EMA) may select element of the Common Metadata to be used within its definitions. EMA would then define additional metadata to cover areas not included in Common Metadata.

1.2 Document Organization

This document is organized as follows:

1. Introduction—Provides background, scope and conventions
2. Encoding
3. Extras Manifest
4. Inventory

-
5. Track Groups and Chains
 6. Picture Groups and Galleries
 7. Title Catalog
 8. Annex A: Track Selection Algorithm

1.3 Document Notation and Conventions

The key words “MUST”, “MUST NOT”, “REQUIRED”, “SHALL”, “SHALL NOT”, “SHOULD”, “SHOULD NOT”, “RECOMMENDED”, “MAY”, and “OPTIONAL” in this document are to be interpreted as described in [RFC2119]. That is:

- “MUST”, “REQUIRED” or “SHALL”, mean that the definition is an absolute requirement of the specification.
- “MUST NOT” or “SHALL NOT” means that the definition is an absolute prohibition of the specification.
- “SHOULD” or “RECOMMENDED” mean that there may be valid reasons to ignore a particular item, but the full implications must be understood and carefully weighed before choosing a different course.
- “SHOULD NOT” or “NOT RECOMMENDED” mean that there may be valid reasons when the particular behavior is acceptable, but the full implications should be understood and the case carefully weighed before implementing any behavior described with this label.
- “MAY” or “OPTIONAL” mean the item is truly optional, however a preferred implementation may be specified for OPTIONAL features to improve interoperability.

Terms defined to have a specific meaning within this specification will be capitalized, e.g. “Track”, and should be interpreted with their general meaning if not capitalized.

Normative key words are written in all caps, e.g. “SHALL”

1.3.1 XML Conventions

XML is used extensively in this document to describe data. It does not necessarily imply that actual data exchanged will be in XML. For example, JSON may be used equivalently.

This document uses tables to define XML structure. These tables may combine multiple elements and attributes in a single table. Although this does not align with schema structure, it is much more readable and hence easier to review and to implement.

Although the tables are less exact than XSD, the tables should not conflict with the schema. Such contradictions should be noted as errors and corrected.

1.3.1.1 Naming Conventions

This section describes naming conventions for Common Metadata XML attributes, element and other named entities. The conventions are as follows:

- Names use initial caps, as in InitialCaps.
- Elements begin with a capital letter, as in InitialCapitalElement.
- Attributes begin with a lowercase letter, as in initialLowercaseAttribute.
- XML structures are formatted as Courier New, such as md:rightstoken
- Names of both simple and complex types are followed with “-type”

1.3.1.2 Structure of Element Table

Each section begins with an information introduction. For example, “The Bin Element describes the unique case information assigned to the notice.”

This is followed by a table with the following structure.

The headings are

- Element—the name of the element.
- Attribute—the name of the attribute
- Definition—a descriptive definition. The definition may define conditions of usage or other constraints.
- Value—the format of the attribute or element. Value may be an XML type (e.g., “string”) or a reference to another element description (e.g., “See Bar Element”). Annotations for limits or enumerations may be included (e.g., “int [0..100]” to indicate an XML xs:int type with an accepted range from 1 to 100 inclusively)
- Card—cardinality of the element. If blank, then it is 1. Other typical values are 0..1 (optional), 1..n and 0..n.

The first row of the table after the header is the element being defined. This is immediately followed by attributes of this element, if any. Subsequent rows are child elements and their attributes. All child elements (i.e., those that are direct descendents) are included in the table. Simple child elements may be fully defined here (e.g., “Title”, “”, “Title of work”, “xs:string”), or described fully elsewhere (“POC”, “”, “Person to contact in case there is a problem”, “md:ContactInfo-type”). In this example, if POC was to be defined by a complex type defined as md:ContactInfo-type. Attributes immediately follow the containing element.

Accompanying the table is as much normative explanation as appropriate to fully define the element, and potentially examples for clarity. Examples and other informative descriptive text may follow. XML examples are included toward the end of the document and the referenced web sites.

1.3.2 Use of Language

This specification assumes that Devices have a parameter referred to here as System Language. The System Language is the current setting for the Device's interface language, perhaps set by the User. Users may also make independent language preference selections for audio language and for subtitle language.

Language preferences such as System Language are expressed as at least one language tag as per [RFC5646] and included in [IANA-LANG], possibly prioritized as a Language Priority List as per [RFC4647], Section 2.3. The assumed Priority List consists of at least the following language ranges:

- 1) The fully enumerated language tag including region, dialect or any other subtag element. For example, this would be a language tag from System Language, Audio User preference or Subtitle User preference.
- 2) The language tag from the first entry trimmed to the primary language tag, followed by a wildcard '*' subtag.

For example if the language is "en-GB", the Priority List will be "en-GB, en-*".

The best language match between a language preference (e.g., System Language) and one or more languages in a list (e.g., language tags in a list of audio tracks) is to be done in accordance with [RFC4647], Section 3.4 "Lookup".

1.3.3 General Notes

All required elements and attributes must be included.

When enumerations are provided in the form 'enumeration', the quotation marks (‘’) should not be included.

The term "Device" refers to an entity playing the interactive material specified here. It may be a standalone physical device, such as a Blu-ray player, or it might be an application running on a general purpose computer, a table, phone or as part of another device. The term 'User' refers to the person using the Device.

1.4 Normative References

[CM] Common Metadata, www.movielabs.com/md/md

[RFC4646] Philips, A, et al, *RFC 4646, Tags for Identifying Languages*, IETF, September, 2006.
<http://www.ietf.org/rfc/rfc4646.txt>

[ISO639] ISO 639-2 Registration Authority, Library of Congress.
<http://www.loc.gov/standards/iso639-2/>

[ISO3166-1] Codes for the representation of names of countries and their subdivisions -- Part 1: Country codes, 2007.

[ISO3166-2] ISO 3166-2:2007 Codes for the representation of names of countries and their subdivisions -- Part 2: Country subdivision code

[ISO4217] Currency shall be encoded using ISO 4217 Alphabetic Code.
http://www.iso.org/iso/currency_codes_list-1

[ISO8601] ISO 8601:2000 Second Edition, *Representation of dates and times, second edition*, 2000-12-15.

[TTML] Timed Text Markup Language (TTML) 1.0, W3C Proposed Recommendation 14 September 2010, <http://www.w3.org/TR/ttml1-dfxp/>

1.5 Informative References

2 ENCODING

2.1 Identifiers

Identifiers are as per Common Metadata [CM], Section 2.

The following identifiers are used in Extras:

- TrackGroupID – identifies a Track Group
- PictureGroupID – identifies a Picture Group
- ChainID – identifies a Chain.
- VideoTrackID, AudioTrackID, SubtitleTrackID – Identifies a video, audio or subtitle track from inventory (in its entirety)
- ImageID – Identifies an Image (just the image)
- PictureID – Identifies a Picture (Image plus description on its use)

2.2 Asset References

It is necessary to reference assets at the file level (e.g., A/V files, audio files, image files) and at the track level (e.g., audio, video, subtitle, etc.).

The exact reference depends on how the assets are packaged and therefore the reference encodings will be appropriate to that packaging.

2.2.1.1 Referencing Files

File level references are in the form of a URL.

Local files are of the form: [file://<filename>](#) where <filename> is the name of the physical file.

Internet files are of the form [<filelocation>](#) where <filelocation> is the fully enumerated URL for the file.

2.2.1.2 Referencing Tracks

Track references are of the form TrackReference as defined in Common Metadata [CM].

[CHS: Do we need to define this uniformly.]

2.3 General Types Encoding

General Types Encoding is as per Common Metadata [CM], Section 3.

2.3.1 Timecode Encoding

Timecode references a specific time in an audio, video or subtitle track.

Element	Attribute	Definition	Value	Card.
Timecode-type				
Timecode		Timecode for referenced point in an associated track.	extras:TimecodePattern-type	
	dropframe	Is timeframe dropframe used	xs:	

extras:TimecodePattern-type is xs:string with pattern '[0-9]+\.[0-9]+'

Timecode corresponds with a constrained form of the 'offset-time' syntax (without the metric field) of the media timebase defined in [TTML], Section 10.3.1, and corresponds with the referened video subtitle and/or audio tracks. The metric is in units of seconds.

In the case of a rounding error that doesn't result in an integer number of frames, the video and/or audio frame(s) Timecode refers to shall be the next decodable frame after the time in the media referenced by this value. For example, in a 30fps progressive video track, 0.1 = the 3rd frame. 0.101 = the 4th frame.

Encoding for dropframe is as follows:

- 'Drop' – Drop frame SMPTE timecode is used.
- 'Non-Drop' – Other timecode without drop frame
- 'EBU' – AES/EBU embedded timecome
- 'Other' – Other timecode

2.3.2 Clip References

A Clip is a subset of audio, video, subtitles or some combination. Currently, the only demand is for Audio (audio only) and Audiovisual (audio, video and subtitle). Typically, clips do not contain video and subtitles without audio.

The following defines and audio clip.

Element	Attribute	Definition	Value	Card.
AudioClipRef-type				
AudioTrackID		Identifier for an audio track from Inventory	extras:AudioTrackID-type	

EntryTimecode		Beginning timecode for clip. If absent, start is the beginning of the track.	extras:Timecode-type	0..1
ExitTimecode		Timecode for end of clip. If absent, end is the end of the track.	extras:Timecode-type	0..1

The following defines an audiovisual clip.

Element	Attribute	Definition	Value	Card.
AudiovisualClipRef-type				
TrackGroupID		Identifier for TrackGroup that may contain combinations of audio, video and subtitles.	extras:TrackGroupID-type	
EntryTimecode		Beginning timecode for clip. If absent, start is the beginning of the track.	extras:Timecode-type	0..1
ExitTimecode		Timecode for end of clip. If absent, end is the end of the track.	extras:Timecode-type	0..1

3 EXTRAS MANIFEST

The ExtrasManifest element is the top level definition of an Extras experience.

Element	Attribute	Definition	Value	Card.
ExtrasManifest-type				
	ExtraVersionReference	Version of the Extras (TBD)	xs:string	
Compatibility		Indicates which versions of Devices can fully use this instance of Extras.	extras:Capability-type	
Inventory		Inventory of audio, video, subtitle and image assets, regardless of where they are stored.	extras :Inventory-type	
TrackGroups		Groups of tracks that are intended to be played together. Also includes information about chapters and track selection.	extras :TrackGroupList-type	
Chains		Defines ordered sequences of TrackGroups that are intended to be played together.	extras :ChainList-type	
PictureGroups		Collections of related images.	extras :PictureGroupList-type	0..1
Behavior		Defines how Chains and PictureGroups are to be presented to a User.	extras :BehaviorList-type	1..n

3.1 Compatibility

The ExtrasManifest element defines the target Devices for this entry via the Compatibility element.

It consists of one or more instances of InteractiveCapabilityLevel.

The definition is TBD.

Element	Attribute	Definition	Value	Card.
Compatibility-type				

InteractiveCapabilityLevel		Indicates one versions of Devices can fully use this instance of Extras. Any Device that supports this level should play the Extras instance.	xs:string	1..n
----------------------------	--	---	-----------	------

DRAFT

4 INVENTORY

Inventory is a list of audio, video, subtitle and image elements that comprise the Extras experience.

4.1 Inventory-type

Element	Attribute	Definition	Value	Card.
Inventory-type				
Audio		Description of audio asset.	extras:TrackAudio-type	0..n
Video		Description of video asset.	extras:TrackVideo-type	0..n
Subtitle		Description of subtitle asset.	extras:TrackSubtitle-type	0..n
Image		Description of image asset.	extras:TrackImage-type	0..n
Interactive		Description of Interactive asset	extras:TrackAudio-type	0..n

4.2 Inventory Asset Types

The inventory consists of audio, video, subtitles and images.

Each entry consists of metadata based on Common Metadata types corresponding with the type (e.g., md:DigitalAssetAudioData-type for audio), along with a unique ID and information where to find the information.

Note that the inventory identifies complete audio, video and subtitle tracks. ‘Clips’ (subset) or Chains (sequenced unions) of tracks can be taken from these, but that is handled elsewhere.

4.2.1 InventoryAudio-type

Element	Attribute	Definition	Value	Card.
InventoryAudio-type			md:DigitalAssetAudioData-type (by extension)	
	AudioTrackID	Unique identifier for this asset.	extras:AudioTrackID-type	

ExternalTrackReference		Description of image asset.	Md:DigitalAssetExternalTrack Reference-type	0..1
------------------------	--	-----------------------------	--	------

4.2.2 InventoryVideo-type

Element	Attribute	Definition	Value	Card.
InventoryVideo-type			md:DigitalAssetVideoData- type (by extension)	
	VideoTrackID	Unique identifier for this asset.	extras:VideoTrackID-type	
ExternalTrackReference		Description of image asset.	md:DigitalAssetExternalTrack Reference-type	0..1

4.2.3 InventorySubtitle-type

Element	Attribute	Definition	Value	Card.
InventorySubtitle-type			md:DigitalAssetSubtitleData- type (by extension)	
	SubtitleTrackID	Unique identifier for this asset.	extras:SubtitleTrackID-type	
ExternalTrackReference		Description of image asset.	md:DigitalAssetExternalTrack Reference-type	0..1

4.2.4 InventoryImage-type

Element	Attribute	Definition	Value	Card.
InventoryImage-type			md:DigitalAssetImageData- type (by extension)	
	ImageTrackID	Unique identifier for this asset.	extras:ImageTrackID-type	
ExternalTrackReference		Description of image asset.	md:DigitalAssetExternalTrack Reference-type	0..1

VideoFrameTimecode		Timecode reference to a particular frame in a Video asset	extras:Timecode-type	0..1
	VideoTrackID	Video track from which frame is referenced.	Extras:VideoTrackID-type	

4.2.5 InventoryInteractive-type

Interactive elements may be included. These may define appearance and behavior based on the Extras Manifest.

Element	Attribute	Definition	Value	Card.
InventoryImage-type			md:DigitalAssetInteractiveData-type (by extension)	
ExternalTrackReference		Description of interactive asset.	md:DigitalAssetExternalTrackReference-type	0..1

5 TRACK GROUPS AND CHAINS

Track Groups and Chains together define what audiovisual material can be played for the User. Track Groups describe audio, video and subtitles that are played together simultaneously. Chains indicate which Track Groups are played together in sequence.

5.1 Track Group

The TrackGroup element provides information which tracks are intended to be played together. They are assumed to be of the same edit. Track Groups include tracks that are typically played together as part of a feature, such as video, audio and subtitle tracks. Track Groups also contain alternative material, such as commentary audio tracks.

The TrackGroup element also provides information to assist a Device and User selecting tracks in accordance with direction from the content creator. Included are track priority and which audio and subtitle language pair is preferred based on the language preferences. These data are used in conjunction with data in Inventory.

See *Section 7 Annex A: Track Selection Process* for information on expected interpretation of these data for the purpose of default track selection.

5.1.1 TrackGroup-type

The TrackGroupType defines which tracks are associated with each other. This allows a Device to determine which tracks should be played together. It also contains LanguagePairs that include information about which tracks language combinations the author recommends for a given a System Language.

Within an element of this type, any audio track is associated with any video track and any subtitle track; and any subtitle track is associated with any video track and any audio track.

For example, all video, audio and subtitle track relating to the main program, regardless of CODEC and language would be in the same element. However, commentary audio and subtitle tracks would be in a separate element. A TrackGroup would not include both a ‘primary’ audio track and a ‘commentary’ subtitles track that are not intended to be played together. A Device would know from this structure which subtitle track to play with a commentary audio track.

At least one instance of TrackGroup SHALL have TrackSelectionNumber='0'.

Each TrackGroup element SHALL have a unique value in TrackSelectionNumber.

Element	Attribute	Definition	Value	Card.
---------	-----------	------------	-------	-------

TrackGroup-type				
TrackSelectionNumber		A Track Selection Number assigned to the group of tracks that belong to the same type, such as normal or commentary tracks.	xs:nonNegativeInteger	
VideoTrackReference		Track Reference to a Video track in TrackMetadata.	xs:string	1..n
	priority	Relative priority of this track.	xs:positiveInteger	
AudioTrackReference		Track Reference to an Audio track in TrackMetadata.	xs:string	1..n
	priority	Relative priority of this track.	xs:positiveInteger	
SubtitleTrackReference		Track Reference to a Subtitle track in TrackMetadata.	xs:string	0..n
	priority	Relative priority of this track.	xs:positiveInteger	
LanguagePair		Defines which audio language and subtitle language are paired with a System Language. Each instance SHALL have a SystemLanguage element. With a unique language.	extras:ContainerLanguagePair-type	0..n
Chapters		Chapter stop definitions	extras:ChapterList-type	0..n

Within VideoTrackReference, AudioTrackReference and SubtitleTrackReference, the priority attribute is the relative priority of the track. A smaller number is a higher priority, with '1' being the highest priority.

Within a ContainerTrackGroup-type instance, each VideoTrackReference/priority child SHALL be unique.

Within a ContainerTrackGroup-type instance, each AudioTrackReference/priority child SHALL be unique.

Within a ContainerTrackGroup-type instance, each SubtitleTrackReference/priority child SHALL be unique.

Each TrackSelectionNumber represents a selection of tracks that belong to the same type. For example, primary audio tracks and normal subtitle tracks are associated with TrackSelectionNumber = '0', director's commentary audio tracks and subtitle tracks are associated with TrackSelectionNumber = '1', and so on.

Audio tracks of type ‘primary’ and subtitle tracks of Type ‘normal’ SHALL be associated with TrackSelectionNumber=‘0’.

VideoTrackReference, AudioTrackReference and SubtitleTrackReference elements, lists the track priority order for all video, audio and subtitle tracks associated with the TrackSelectionNumber. All tracks associated with a lower TrackSelectionNumber are higher priority than all tracks associated with a higher TrackSelectionNumber.

The priority attribute can be used to specify priority order amongst equivalent tracks. For example, given multiple AudioTrackReference instances that reference primary English tracks with different CODECs, the preferred order of these tracks would be indicated by the priority attributes, with the most preferred track having priority=‘1’. If there are multiple instances of SubtitleTrackReference elements for equivalent tracks with different Track/FormatTypes (Text or Image), authors can specify which FormatType has higher priority using the priority attribute. Within a TrackGroup, Priority is unique across all audio tracks and is unique across all subtitle tracks.

Note that CFF currently only allows one video track, so it is not meaningful to have more than one VideoTrackReference (i.e., a cardinality of 1). The schema allows multiple instances to support future growth.

5.1.2 ContainerLanguagePair-type

ContainerLanguagePair-type allows the author to specify audio and subtitle track pairs based on a User’s System Language.

A User preference for System Language does not always imply audio and subtitle tracks of the same language. For example, in some cases the best choice for a Japanese viewer would be Japanese language audio and no subtitle. In other cases, the best choice would be an English audio track and a Japanese subtitle.

TrackGroup/AudioReference and TrackGroup/SubtitleReference refer to a subset of tracks in Inventory/Audio and Inventory/Subtitle respectively. ContainerLanguagePair-type further constrains the track list by selecting tracks by language. That is, LanguagePair refers only to audio tracks where Inventory/Audio/Language equals AudioLanguage and to subtitle tracks where Inventory/Subtitle/Language equals SubtitleLanguage.

Element	Attribute	Definition	Value	Card.
ContainerLanguagePair-type				
SystemLanguage		The language scope for which the Language Pair applies. For example, if this element is ‘en-US’ then the Language Pair element applies to English	xs:language	

		spoken in the United States.		
AudioLanguage		Author recommended audio language for given SystemLanguage	xs:language	
SubtitleLanguage		Authore recommended subtitle language for given SystemLanguage	xs:language	

Within the set of LanguagePair elements, each LanguagePair element SHALL have a unique value in SystemLanguage.

5.1.3 Chapter Metadata

An A/V stream may be divided into chapters. The assumption is that a use may skip between chapters or jump to a chapter based on a list.

In an A/V stream, chapters are referenced to video and are timed or referenced to a specific video frame. Audio chapters are time referenced to the beginning of the audio.

It is best practice to encode audio and video to allow jumps to chapter starts.

Chapter start times are assumed against a/v stream of the same edit. A video with parts added or removed, such as a director's cut will have different chapter start times than the theatrical cut. This generally applies to Supplemental Material.

The challenge in defining chapters is referencing the correct frame. Depending on how the video is encoded, the time reference can be different.

Chapter metadata identifies the locations within a track where chapters begin. Each chapter has a numerical index and an entry point that defines where the chapter starts.

Note that Chapters are defined against a TrackGroup rather than a Chain. This allows the same chapter to be used across multiple Chains.

Element	Attribute	Definition	Value	Card.
ChapterList-type				
Chapter		Chapter entry point descriptor	extras:Chapter-type	

Elements in Chapter-type SHALL be in chapter order.

Element	Attribute	Definition	Value	Card.
Chapter-type				
	index	Chapter index.	xs:integer	
EntryTimecode		Entry point for chapter start.	xs:string, pattern [0-9]+\.[0-9]+	
DisplayLabel		Displayable text on a per-language basis for the chapter	xs:string	0..n
	language	Language of DisplayLabel. Must be included in all DisplayLabel elements if more than one DisplayLabel element is included. Matching is in accordance with Section 4.1.5.1 Use of Language	xs:language	0..1
ImageID		Reference to a chapter image.	extras:ImageID-type	0..n
	language	Language of image referenced by ImageID. Must be included in all ImageID elements if more than one ImageID element is included and there is language-specific text in the image (i.e., burned in text).	xs:language	0..1

The index attribute is a number starting with 0 and increasing monotonically for each subsequent chapter.

EntryTimecode corresponds with a constrained form of the ‘offset-time’ syntax (without the metric field) of the media timebase defined in [TTML], Section 10.3.1, and corresponds with the beginning of the chapter in the video and/or audio tracks for which the chapters are identified. The metric is in units of seconds.

In the case of a rounding error that doesn’t result in an integer number of frames, the video and/or audio frame(s) EntryTimecode refers to shall be the next decodable frame after the time in the media referenced by this value. For example, in a 30fps progressive video track, 0.1 = the 3rd frame. 0.101 = the 4th frame.

5.2 Chains

The Chain is the minimum unit of playback from the perspective of a User. A Chain references one or more Track Groups in sequence.

5.2.1 Chains Use Cases

There are two primary use cases for Chains, although others are allowed. The first is pre-roll or post-roll material unique to a region. In this case, the main title is a Track Group and each pre- and post-roll segment is its own Track Group. For each region, there is a Chain consistent of the pre-roll Track Group for that region and the main title Track Group.

The second primary use case is edits, such as a Theatrical Cut and a Director's Cut. In this case, each segment of unique video is a Track Group. The Chain is the sequence of Track Groups that results in the Theatrical Cut or the Director's Cut.

Seamless playback of Chains can be problematic. It requires careful authoring and interoperability testing within target environments. We advise caution with respect to the use of Chains, especially in the second use case.

5.2.2 Chapters and Chains

When Track Groups are sequenced using Chains, they must construct a complete chapter list correctly. Chapter times in each Track Group must remain at the same point in that Track Group. The full chapter list is constructed from the chapter list in each Track Group in sequence.

5.2.3 ChainList-type

Element	Attribute	Definition	Value	Card.
ChainList-type				
Chain		A sequenced segment of Chain. The order of Chain elements defines the order of playback.	extras:Chain-type	0..n

5.2.3.1 Chain-type

Element	Attribute	Definition	Value	Card.
Chain-type				
	ChainID	Identifier used to refer to the Chain. Must be unique within an Extras definition.	extras:ChainID-type	
Clip		Clip of audiovisual playback material. The order of Clip elements defines playback order.	extras:AudiovisualClip-type	1..n
BasicMetadata		Metadata describing the Chain. Note that the Chain is the end product and is therefore the appropriate unit for metadata description.	md:BasicMetadata-type	0..1
ReferenceID		Identifiers that identify the Chain. Additional information can be obtained using these Identifiers. We strongly suggest that at least one instance be an EIDR.	md:ContentIdentifier-type	0..n

[CHS: I'm a little concerned about track selection on a Chain since each segment has its own tracks. We'll likely need conventions to link tracks.]

5.2.3.2 ChainClip-type

Element	Attribute	Definition	Value	Card.
ChainClip-type				
TrackGroupID		Track Group to be played in this sequence.	extras:TrackGroupID-type	
EntryTimecode		Entry point for TrackGroup. If absent, the beginning of the Track Group assumed.	extras:Timecode-type	0..1
EntryTimecode		Exit point for TrackGroup. If absent, the end of the Track Group is assumed.	extras:Timecode-type	0..1

6 PICTURE GROUPS AND GALLERIES

Images may be provided with a main title, or as supplements to supplemental audiovisual material.

Images are grouped and sequenced. Basic models allow for a single sequence (slide show). More advanced models allow more complex navigation paths.

The Gallery, part of Behavior, defines how a Picture Group is displayed.

6.1 Picture Group

The top level definition for Picture Groups is PictureGroupList-type. It contains one or more Picture Groups.

Element	Attribute	Definition	Value	Card.
PictureGroupList-type				
PictureGroup		An unordered list of Picture Groups.	extras:PictureGroup-type	0..n

6.2 Picture Group Type

A Picture Group is an identified and sequenced collection of images with annotation.

Image annotation includes

- Localized captions to be displayed with images
- Languages of text in images (for localization)
- Intended sequence for playback in a gallery.

Element	Attribute	Definition	Value	Card.
PictureGroup-type				
	PictureGroupID	Identifier for the Picture Group. Must be unique within an Extras element.	extras:PictureGroupID-type	
Picture		An individual picture within the PictureGroup.	extras:Picture-type	1..n

6.2.1 Picture-Type

Picture-Type describes an individual picture, including how it relates to other pictures when sequenced within a Gallery. Note that a 'Picture' is more than an 'Image', so it has its own identity, PictureID.

Element	Attribute	Definition	Value	Card.
Picture-type				
PictureID		Identifier for this Picture.	extras:PictureID-type	
ImageID		Reference to the image for the Picture.	extras:ImageID-type	
LanguageInImage		If there is any text visible in the image, this element identifies this language. Anticipated use is to determine when alternate text is required.	xs:language	
AlternateText		Alternate text to be used for accessibility and Internationalization. This can be used to represent text on the screen.	xs:string	0..n
	Language	Language of AlternateText	xs:language	0..1
AlternateAudio		Audio corresponding with text in image. Anticipated use is accessibility.	extras:AudioClipRef-type	0..n
	Language	Language of AlternateAudio	xs:language	0..1
Caption		Caption for the image.	xs:string	0..n
	Language	Language of Caption	xs:language	0..1
AlternateCaptionAudio		Audio corresponding Caption. Anticipated use is accessibility.	extras:AudioClipRef-type	0..n
	Language	Language of AlternateCaptionAudio	xs:language	0..1

Sequence		Order of picture in picture group. [CHS: Should this be included, or is it too 'presentation?']	xs:nonNegativeInteger	0..1
----------	--	--	-----------------------	------

6.3 Gallery

The user interface for the presentation of images is called a Gallery. The Gallery contains enough information to provide a simple display of images.

A gallery contains

- Name – Used for gallery selection)
- Picture Group – Images associated with Gallery. The Gallery will include all images in the Picture Group.
- Background – Image or video background with optional audio.
- Auto-advance timing – If system is to display images automatically, how long to dwell on each slide.

6.3.1 Gallery-type

Element	Attribute	Definition	Value	Card.
Gallery-type				
Type				
PictureGroupID		Picture Group containing Pictures for gallery	extras:PictureGroupID	
GalleryName		Title of Gallery	xs:string	0..n
	language	Language of gallery	xs:language	0..1
GalleryNameAlternateAudio		Audio corresponding GalleryName. Anticipated use is accessibility.	extras:AudioClipRef-type	0..n
	language	Language of GalleryNameAlternateAudio	xs:language	0..1

GalleryID	Identifier that uniquely identifies this instance of Gallery-type. This is used as a reference for presentation implementations. For example, a menu intended to be associated with this gallery could refer to this GalleryID.	md:id-type	0..1
-----------	---	------------	------

DRAFT

7 TITLE CATALOG

The Title Catalog is the top-level element for determining what can be offered to a user. It defines which other elements comprise a user experience.

7.1 Title Catalog List

The TitleCatalogList element contains one or more instances of TitleCatalog. Each instance may be international, or it might be targeted towards some combination of region and language. Playback behavior is to select the best fit for the applicable region.

Element	Attribute	Definition	Value	Card.
TitleCatalogList-type				
TitlesCatalog		Each instance contains a collection of Title and Gallery elements, possibly specific to a combination of language, region and/or version.	extras:Behavior-type	0..n

TitleCatalog elements may have language attributes and Region elements that indicates which instance should be used. The first filter is language. That is, a player should first select Behavior elements with a suitably matching language. A Behavior element without a matching language is assumed to match all languages not present in other Behavior elements. Once language is matched, the Region should be matched.

The following is the order of preference for matches (highest first)

- Matching language and Region
- Matching Region element and no language attribute present
- Matching language attribute and no Region element present
- Any other match of language
- Any other match of Region
- No match (no rules on selection)

It is up to the implementation whether it processes the version attribute.

7.2 Titles Catalog

The TitleCatalog element defines which titles and galleries apply.

Element	Attribute	Definition	Value	Card.
TitlesRegional-type				
	version	Behavior version, used when player behavior accepts multiple versions. For this specification, version='1'.	xs:double	
	language	Language for which Behavior was authored	xs:language	0..1
Region		Region for which Behavior was authored	md:Region-type	0..1
MainTitle		The main title of the extras package. Typically, this is the feature. When there are multiple equivalent main titles (e.g., episode) they are all included. [CHS: Is this useful or should just rely on Organization?]	extras:Title-type	1..n
Title		Other titles, typically trailers, deleted scenes, and other value added material	extras:Title-type	0..n
Gallery		An image gallery	extras:Gallery-type	0..n
Organization		Detailed information about the organization of the titles.	md:CompObj-type	0..1

Organization uses the Common Metadata Completion Object (CompObj-type) to fully organize titles within the Title Catalog. This structure can include hierarchies such as a show comprised of seasons further comprised of episodes. In this usage, CompObjEntry-type is constrained as follows [CHS: these should probably be redefined.]:

- DisplayName shall not be included
- EntryNumber should be included
- EntryClass should be included
- Entry is included if appropriate
- ContentID is the only valid choice and it shall be a TitleID.

7.2.1 Title

A title is any collection of video, audio, subtitles and other behavior defined by a Chain. This can be the main feature, promotions such as trailers, or value added material such as making-of videos or deleted scenes.

Element	Attribute	Definition	Value	Card.
Title-type				
Type		Type of this title (see below)	xs:string	
SubType		Additional detail on type	xs:string	
TitleName		A consumer-facing name for this title.	xs:string	
ChainID		Identifier for the Chain associated with this title	extras:ChainID-type	
TitleID		Identifier that uniquely identifies this instance of Title-type. This is used as a reference for presentation implementations. For example, a menu intended to be associated with this Title could refer to this TitleID.	md:id-type	0..1

Type describes the top-level nature of the Title. Note that useful information in the Title can also be found in BasicMetadata or ReferenceID information in the referenced Chain. Type is enumerated as follows:

- ‘Main’ – Main title (typically the feature)
- ‘Promotion’ – Trailers, teasers, etc.
- ‘Bonus’ – Additional material related toward the Main Title, such as, deleted scenes, making-of, etc.
- ‘Other’ – Any other material included

SubType can add additional detail, especially to ‘Bonus’ Type. SubType may include terms such as ‘Deleted Scenes’ or ‘Making-of’. [CHS: This needs some work. I suspect player will need more information so it can present information in a more rational form.]

8 ANNEX A: TRACK SELECTION PROCESS

This section describes the intended use of Track Selection Data as described in Section 4.1.5. The following stages occur in track selection:

1. The Device assigns a default System Language
2. A User optionally changes System Language; and may select preferences such as audio and subtitle languages, and subtitle type
3. The Device selects default audio track and subtitle track (Primary Subtitling Presentation Track), if applicable
4. A User may optionally select specific audio track or subtitle track (Primary Subtitling Presentation Track)
5. The Device selects subtitle tracks for forced subtitles (Alternate Subtitling Presentation Track), if applicable
6. Playback can begin. User selections may require repeating some steps above. For example, changing tracks (Step 4) would require performing Step 5.

This Annex uses the following terminology:

- The following subtitle definitions are used to describe what is in a subtitle track
 - Forced Subtitle: A subtitle with only one instance of Inventory/Subtitle/Type where that instance equals 'forced'.
 - Other Subtitle: A subtitle with no instances of Inventory/Subtitle/Type equal to "forced"
 - Mixed Subtitle: A subtitle with at least one instance of Inventory/Subtitle/Type equal to 'forced'; and at least one instance of Inventory/Subtitle/Type not equal to "forced"
 - Within a Mixed Subtitle track, subtext and subpicture elements that are to be displayed as forced subtitles are referred to as 'forced elements' and elements that are not to be displayed as forced elements are referred to as 'non-forced elements'

-
- From a User’s perspective, subtitles are either “on” or “off”, however, in both cases subtitle elements may be displayed. The following definitions indicate what subtitle elements are presented when subtitles are off and on, what tracks contain those elements, and what audio track contains audio for playback
 - Primary Subtitling Presentation Mode: corresponds to subtitles are “on”. When in Primary Subtitling Presentation Mode, the Primary Subtitling Presentation Track will be presented.
 - Primary Subtitling Presentation Track: The subtitle track that is to be presented during Primary Subtitling Presentation. An Other Subtitle track or a Mixed Subtitle track will be decoded and presented during Primary Subtitling Presentation.
 - Alternate Subtitling Presentation Mode: corresponds to subtitles are “off”. When in Alternate Subtitling Presentation Mode, only forced elements within the Alternate Subtitling Presentation Track will be presented (if any). An Alternate Subtitle can be forced subtitle elements within a Mixed Subtitle track or a Forced Subtitle track.
 - Alternate Subtitling Presentation Track: The subtitle track that includes the forced subtitle elements to be presented during Alternate Subtitling Presentation. Forced subtitle elements within a Mixed Subtitle track or all elements in a Forced Subtitle track will be presented during Alternate Subtitle Presentation. Note that for a Mixed Track, the Selected Primary Subtitle Track and the Selected Alternate Subtitle Track might be the same track.
 - The following definition indicates what audio track contains audio for playback
 - Selected Audio Track: The audio track selected for play.

8.1 Defined Preferences

The following are Input Variables to default track selection and must be selected prior to default track selection.

- System Language (required)
- User Preferred Audio Type. The type of audio preferred by the user. Type enumeration is as per md:DigitalAssetAudioData-type/Type. By default this should be “primary”

-
- User Preferred Audio Language (optional) – User preference for audio language which applies to all DCCs
 - User Preferred Subtitle Language (optional) – User preference for subtitle language which applies to all DCCs
 - User Preferred Subtitle Type (optional) – The type of subtitle preferred by the User for the purposes of selecting default audio and subtitle tracks. Type enumeration is as per md:DigitalAssetSubtitleData-type/Type. By default this should be ‘normal’.

Devices are assumed to have the following capabilities

- Allow a User to override Input Variables
- Allow a User to select a specific audio track
- Allow a User to select a specific subtitle track for Primary Subtitling Presentation
- Allow a User to turn “on” and “off” subtitles
 - When “On”, decode and present the Primary Subtitling Presentation Track and display all forced and non-forced elements.
 - When “Off”: decode and present the Alternate Subtitling Presentation Track and only display forced elements

8.2 Default Audio and Subtitle Track Selection

This section defines algorithms for selecting default audio track and default subtitle track.

Default tracks are selected prior to initial playback and prior to User’s making specific tracks selections.

The following rules apply to the decision flow:

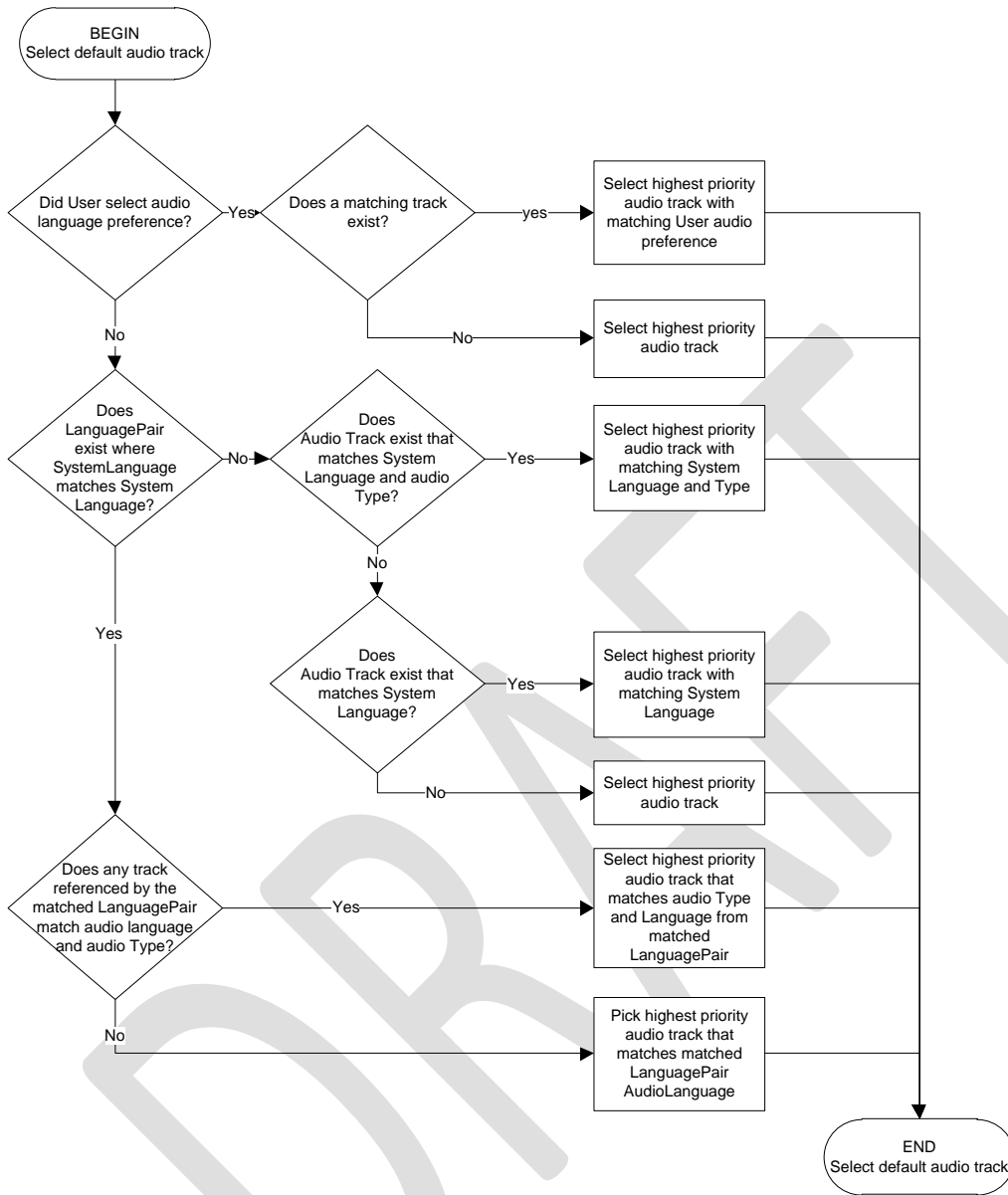
- When matching and selecting tracks, only tracks that are playable on the Device should be considered. Tracks that are not playable should be ignored. For example, a track with a CODEC not supported by the Device would never be selected.
- When multiple elements match equivalently
 - If there are additional User preference and at least one element matches this preference, filter elements based on the User preferences. For example, if the user

prefers original audio tracks, and an original audio track matches other criteria, select that track.

- Then, If elements are prioritized, return the element with the highest priority;
- Otherwise, return the element that appears first in the metadata. For example, if a language lookup matches two LanguagePairs equally well, the first LanguagePair to appear in the TrackGroup would be selected.
- If more than one TrackGroup element is present, the TrackGroup element with TrackSelectionNumber equal to 0 is referenced for automatic default track selection.
- In the diagrams, when an audio track is “selected” it is selected as the Selected Audio Track. When a subtitle track is selected, it is selected as a Selected Primary Subtitle Track, unless otherwise noted.
- In conditions referring to matching tracks of a given language, Inventory/Audio/Language is used for audio language matching and Inventory/Subtitle/Language is used for subtitle language matching.
- In conditions referring to matching tracks of a given type Inventory/Audio/Type is used for audio Type matching, and Inventory/Subtitle/Type is used for subtitle Type matching.
- When referring to Tracks referenced by LanguagePair this refers to all tracks referenced by TrackGroup/AudioTrackReference that match Inventory/Audio/Language in union with tracks referenced by TrackGroup/SubtitleTrackReference that match Inventory/Subtitle/Language.

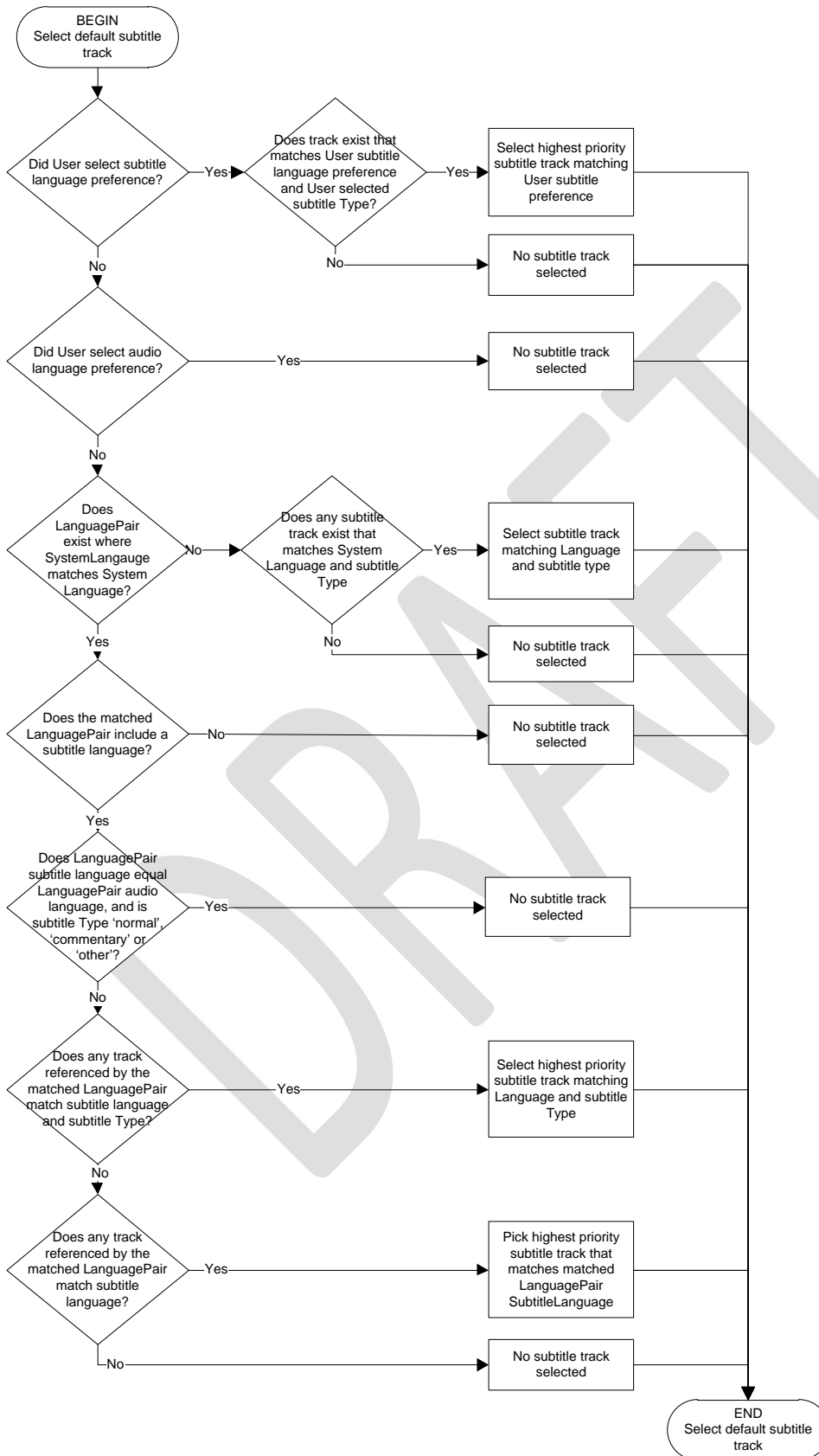
8.2.1 Default Audio Track Selection

This flow describes the assumed algorithm for selecting a Default Audio Track.



8.2.2 Default Primary Subtitling Presentation Track Selection

This flow describes the assumed algorithm for selecting a Default Subtitle Track.



8.3 Alternate Subtitling Presentation Track Selection

An Alternate Subtitle Track is used for Forced Subtitles.

Forced subtitles are displayed either in conjunction with other subtitles, or when subtitles are turned off. That is, if subtitle is off and a suitable forced subtitle track (i.e., either a Forced Subtitle track or a Mixed Subtitle Track) is present, it will be displayed.

A forced subtitle track is expected to match the language of a selected audio track.

If a subtitle track contains information that allows differentiation between elements that are forced and not forced, then the forced subtitle track should be interpreted as the mixed track with only forced elements presented.

8.3.1 Select Alternate Subtitle Track

This flow describes the assumed algorithm for selecting the Alternate Subtitle Track.

