Authoring IMSC documents

Best Practice

1. Scope

This document specifies best practices when authoring an IMSC[1] document against a video element.

2. Status of this Document

This Best Practice is published by the IMF User Group[2]. It may be updated, replaced or obsoleted by other documents at any time. Readers are encouraged to consult the following for a list of current issues, to which they are invited to contribute.

> https://github.com/imfug/003-imsc-authoring

3. Background

Historically, timed text has been synchronized with video using SMPTE timecode. As illustrated in Figure 1, the temporal extent of a timed text event would be defined by the (potentially discontinuous or drop frame) SMPTE timecodes of the starting and end frames.
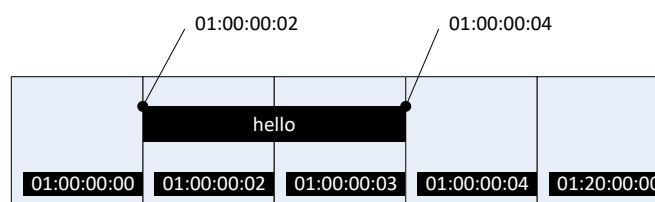

Figure 1. SMPTE Timecode Synchronization.

Such an approach is not possible with IMF or IMSC since:

- SMPTE timecode is not used for synchronization in IMF
- IMSC does not support SMPTE timecode, instead all time expressions, regardless of their syntax, correspond to offsets in seconds.

This document provides guidance when authoring IMSC documents. It can be applied equally to SMPTE-TT documents since IMSC is essentially a practical application of SMPTE-TT.

4. Recommendations

4.1   Time expression syntax

Time expressions in the IMSC document should use one of the following syntaxes:

- clock-time with fractions, e.g. "`00:11:12.234`"; or
- offset-time with frames, e.g. "`4564654f`", where `ttp:frameRate` and `ttp:frameRateMultiplier` are set to the video frame rate.

Time expression in the IMSC document should not use the clock-time with frames syntax, e.g. "`01:00:02:23`". This syntax, while it resembles SMPTE timecode, is not SMPTE timecode (see Annex I.2 at TTML2[4]).

Table 1. Time expression syntaxes.

✓	`begin="4564654f"` `end="4564754f"`	✓	`begin="00:11:12.234"` `end="00:11:14.124"`	✕	`begin="01:00:02:23"` `end="01:00:03:12"`

4.2   Time expressions are always relative to the first frame at 0s

The time expressions in the IMSC document should be relative to 0 seconds, which corresponds to the start of the first frame of the video element, including any slate, head format, etc.
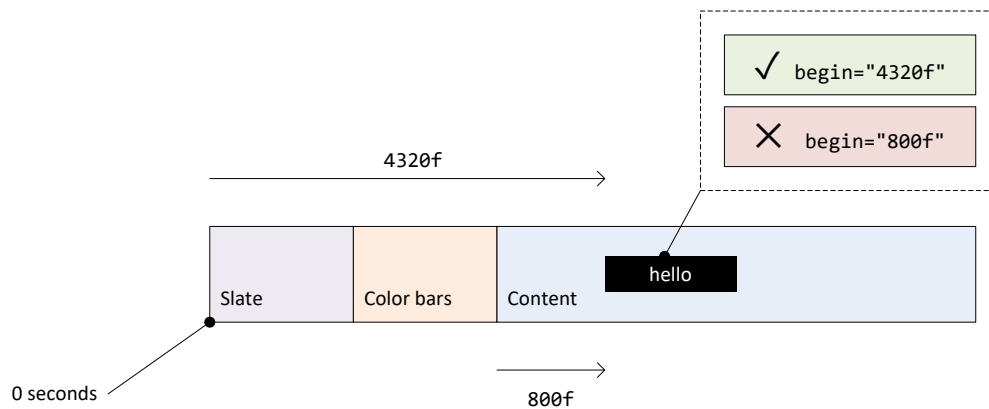

Figure 2. Time expression relative to first video frame.

4.3   Ignore video timecode

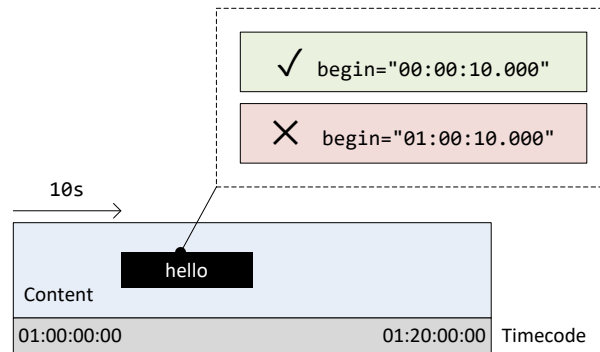The time expressions should ignore any timecode associated with the video.

Figure 3. Ignoring timecode.

4.4   Minimal empty IMSC document

IMF requires the presence of content at all times within a timed text virtual track. In order to fill a gap created by the addition of visual content for which no timed text is provided, e.g. when adding dub cards at the end of a Composition, a timed text virtual track typically references an IMSC document that contains no content.

The following is a minimal empty IMSC document that can be used to fill gaps of arbitrary duration in timed text virtual tracks:

```
<tt xml:lang="" xmlns="http://www.w3.org/ns/ttml"/>
```

NOTE 1: While the minimal empty IMSC document is not a valid EBU-TT-D document according to EBU Tech 3380[5], EBU Tech 3381[6] specifies that such a document is accepted by processors when the document is stored in an ISOBMFF file that conforms to ISO/IEC 14496-30.

NOTE 2: The XML prolog `<?xml … ?>` is not required by XML 1.0 since IMSC requires UTF-8 encoding.

5. Example #1

As illustrated in Figure 4, an IMSC document (en.ttml) is authored against a video proxy (video.mov). The video proxy consists of the same sequence of frames as a video master (video.mxf), which does not include any head format or slate. As specified in Section 4.2, each time expression in the IMSC document is relative to the start of the first frame of the video proxy, and thus to the start of the first frame of the video master. The IMSC document is then wrapped in MXF (en.tt.mxf) for use in a Composition.
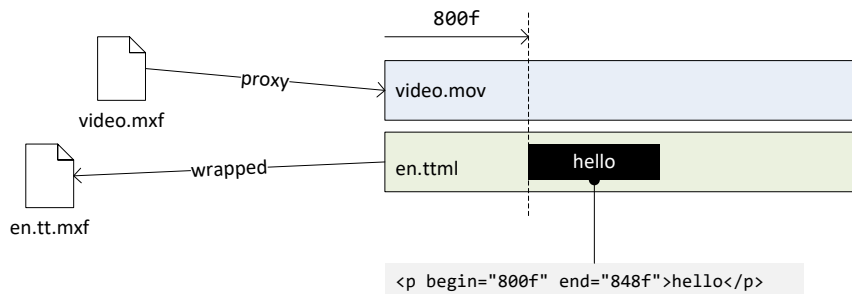
Figure 4. IMSC document authored against video that does not include a head format. The text "hello" is intended to appear at the beginning of the 801st frame of the video.mov file.

As illustrated in Figure 5, a Composition assembles the video master with a slate (slate.mxf). As a result, the first frame of the video master does not correspond to the first frame of the Composition. For the contents of the IMSC document authored above to remain synchronized with the video master, the empty IMSC document specified in 4.4 is added to the Composition. The duration of the Resource that references the empty IMSC document is equal to the duration of the slate.
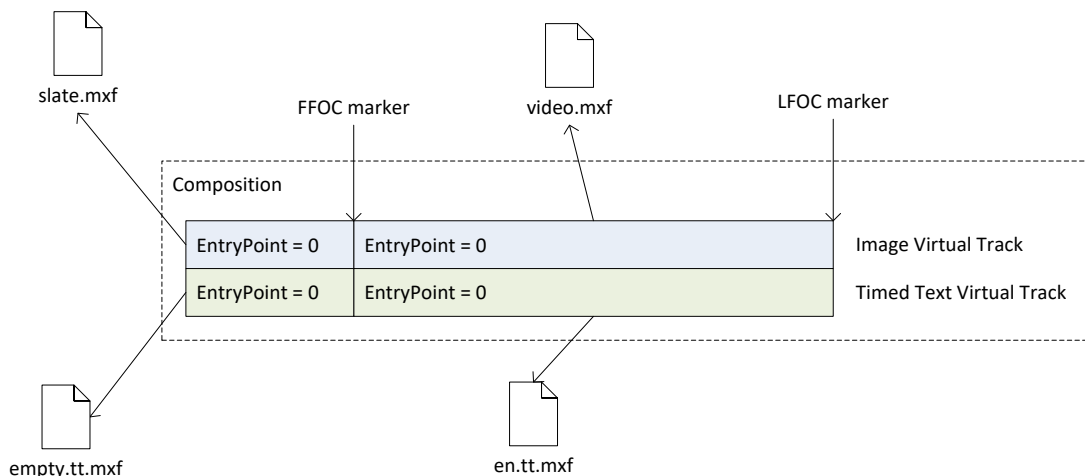

Figure 5. Synchronizing image and timed text essence in an IMF composition that includes a slate. Each of the image resources selects the entire contents of its referenced Track Files, i.e. its `EntryPoint` parameter is 0 and its `SourceDuration` parameter is equal to the duration of the referenced Track File. The `EntryPoint`, `SourceDuration` and `EditRate` parameters of the timed text resources match those of their corresponding image resources. The `FFOC` and `LFOC` markers signal the content for display, excluding the content of the slate.

6. Example #2

As illustrated in Figure 6, and similarly to Example #1, an IMSC document (en.ttml) is authored against a video proxy (video.mov). In contrast with Example #1, the video master (video.mxf) includes a head format and each time expression in the IMSC document is therefore relative to the first frame of the head format.
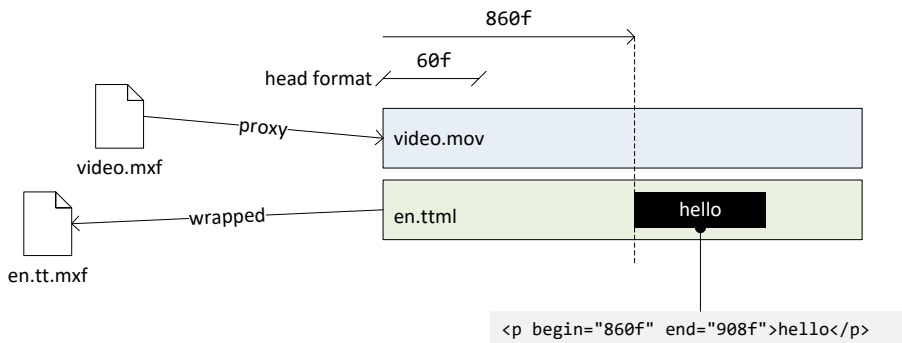
Figure 6. IMSC document authored against video that includes a head format. The text "hello" is intended to appear at the beginning of the 861st frame of the video.mov file. Frames 1-60 of the video.mov file comprises a head format.

As illustrated in Figure 7, a Composition assembles the video master, excluding the head format. This is achieved by setting the `EntryPoint` parameter of the Resource referencing the video master to the frame immediately following the last frame of the head format. For the contents of the IMSC document authored above to remain synchronized with the video master, the `EntryPoint` parameter of the Resource referencing the Timed Text Track File matches that of the Resource referencing the video master.
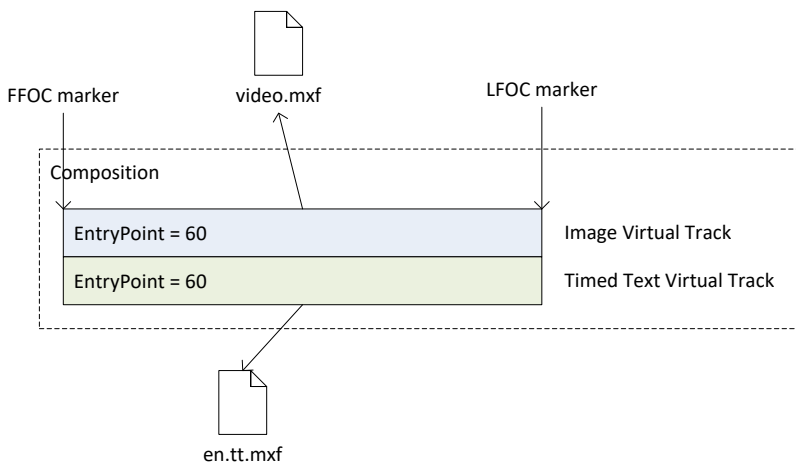

Figure 7. Synchronizing image and timed text essence in an IMF composition that removes a head format. The image resource that references the video master excludes the head format by setting the `EntryPoint` parameter just past the head format and the `SourceDuration` parameter to the duration of the video master minus the duration of the head format. The `EntryPoint`, `SourceDuration` and `EditRate` parameters of the timed text resources match those of the image resource. The `FFOC` and `LFOC` markers signal the content for display.

---

[1] https://www.w3.org/TR/ttml-imsc1.1/

[2] https://www.imfug.com/

[3] http://creativecommons.org/licenses/by-nd/4.0/

[4] https://www.w3.org/TR/ttml2/

[5] https://tech.ebu.ch/docs/tech/tech3380v1_0_1.pdf

[6] https://tech.ebu.ch/docs/tech/tech3381.pdf