ESUB-XF Specification Version 1.03

“European Subtitle Exchange Format”

XML Format for Exchange of Subtitles

Initial release: Version 0.8 / 2013-09-20 by F.A. Bernhardt GmbH, FAB / Miha Sokolov

Last modification: Version 1.03, 2020-06-19 by F.A. Bernhardt GmbH, FAB / Miha Sokolov

The latest version of this document can be obtained from:
http://www.fab-online.com/pdf/ESUB-XF.pdf

Table of Contents

1. Introduction

At the time of the first publication of this document in the year 2013 a number of different and incompatible subtitling and captioning formats were in use:

- Teletext subtitles defined in ETSI EN 300 706. The original specification was developed in the early 1970s and teletext subtitles have remained unchanged since then. Even today teletext subtitles are used as the main subtitle exchange format within EBU STL files and when subtitles have to be transferred as text encoded subtitles over TCP/IP, SDI or DVB links.
- Closed captions as defined in CEA-608 and CEA-708. The original specification was developed in the early 1970s. Even today CEA-608 is still used for captioning in Apple's recent HLS (http live streaming) and is also transported by CEA-708 in modern digital TV transmissions.
- DVB subtitling as defined in ETSI EN 300 743. The original specification was developed in 1997 and describes the transmission of bitmap subtitles in DVB.
- DCI Digital Cinema Specification that includes subtitling was developed in 2005 and later.
- DVD and Blu-ray specifications that include subtitling were developed in 1995 and later.
- EBU STL file format as defined in EBU Tech 3264 in 1991 and is currently still used as the only open file format for exchange of teletext subtitle files in Europe.
- WebVTT (Web Video Text Tracks) is a living standard and describes a plain-text subtitle format designed for use with HTML5 video.
- TTML, SMPTE-TT and EBU-TT file formats for subtitle files for internet video players.

Due to the fact that so many different, partly outdated (from 1970s) and incompatible subtitling and captioning formats were in use there was a need for a single subtitle exchange format that will include the functionality of all above standards and specifications in a modern and legacy free format, for example without obsolete character sets and obsolete transport encodings which are not necessary anymore but also without unnecessary complexity introduced in recent xml formats.

This document contains an XML based specification for storage, exchange, transport and display of subtitles, captions and text to speech audio description which:

- Initially was meant to replace teletext subtitles which date from the early 1970s.
- However now includes features offered by all of the above standards: multiple languages, text and bitmap encoded subtitles, transport over different and supports 3D, 4K, 8K formats and internet streaming.
- Is based on a modern and easy to understand human readable XML format which can be validated for correctness against the XSD (XML Schema Definition) file automatically.
- Supports Unicode and vertically written scripts and can therefore be used universally.
- Can be used as a replacement of all of the subtitling standards listed above.
- Allows conversion of subtitle and caption data to and from all of the above standards. A subtitle file in STL format can be converted from STL to ESUB-XF and back without any loss.
- Is strict and shall therefore result in compatible implementations.
- Is written in a compact form with examples which are easy to understand.
- Eliminates the need for multiple subtitling formats in the future by allowing inclusion of application specific information in the XML structure.
- Includes support for "Text-To-Speech Audio Description".

1.1 Purpose

This specification describes an XML based format for exchange of subtitles, captions and TTS (text to speech) audio description texts as a successor to the EBU STL Subtitle File Specification from 1991. The ESUB-XF specification defines:

- An XML based structure for storage of subtitles and texts for audio description.
- A file format for exchange of subtitle files based on the XML structure.
- A packet structure that is used for transport of subtitles over transport links based on the XML structure within multiple protocols.
- Protocols for transport of subtitles within the following transports: TCP/IP, Internet and Internet streaming. All of them are based on the same XML packet structure.

The main advantage of this specification is that it defines an XML structure for subtitles that remains unmodified throughout the complete subtitle preparation, storage, transport, transmission, decoding and display chain because:

- The same XML structure that contains subtitles is used throughout the complete chain.
- Subtitles are described using device independent information so that they can be imported, exported and displayed by any device or software that supports this specification correctly.
- Only the most basic subtitle properties are included in this specification. Subtitle properties which are specific to a broadcaster or manufacturer shall not be part of this specification.
- The XML structure may easily be extended by anybody to add application specific data which shall be ignored by parties who do not know how to interpret it.

1.2 Explication of Terms

This document contains both normative text and informative text. All text is normative except if marked as informative. Normative text describes indispensable or mandatory elements. It contains the conformance keywords ‘shall’ or ‘may’, defined as follows:

‘Shall’ and ‘shall not’ Indicate requirements to be followed strictly and from which no deviation is permitted in order to conform to the document.
‘May’ and ‘need not’ Indicate a course of action permissible within the limits of the document.

Other terms used in this document as defined as follows:

‘Captions’ describes the use of subtitling for the deaf or hard of hearing.
‘Subtitles’ describes the use of subtitles for translating purposes and for the deaf or hard of hearing.

‘Closed’ describes optionally displayed subtitles.
‘Open’ describes subtitles that are burnt-in the video material.

‘Text-To-Speech Audio Description’ describes text with timing information used by automated speech synthesis engines to reproduce human speech at the specified time.

CRLF describes Carriage Return (ASCII 13) and Line Feed (ASCII 10) characters.

1.3 General XML Usage Rules for this Specification

The XML format used within this specification is a standard XML format. However some limitations have to be respected which were introduced to avoid complexity and to improve human readability of the format.

The only encoding that shall be used is UTF-8. An example of an XML Schema for the XML format defined in this specification is available for free download here:

http://www.fab-online.com/schemas/esub-xf-1.0.xsd

When the XML Schema is not specified within the XML structure the above schema shall be used.

When XML nodes contain text, for example <text>, <line> and other nodes defined in this specification, the text within the node shall not contain spaces at the beginning and also not at the end of the line. The text shall also not contain any of the following characters: & (ampersand), < (less-than sign), > (greater-than sign). These characters shall be replaced with the sequence of characters defined in the following table:

Character	Replacement
&	&amp;
<	&lt;
>	&gt;

Example: The line text 'John & Mary' shall be stored as '<line>John & Mary</line>'.

The text within a node shall always contain only text for one line and shall not contain <br />. When using Unicode within a node the following page with information about 'Unicode in XML and other Markup Languages' shall be applied:

http://www.unicode.org/reports/tr20/

2. The ESUB-XF XML Structure

The structure of the XML shall be as following:

```
<?xml version="1.0" encoding="UTF-8"?>
<esub-xf xmlns="urn:esub-xf" framerate="25" timebase="smpte">
  <subtitlelist language="eng" langname="English" type="translation">
    <subtitle display="10:00:18:12" clear="10:00:21:03">
      <hregion>
        <line>First line of bottom justified text</line>
        <line>Second line of bottom justified text</line>
      </hregion>
    </subtitle>
    <subtitle display="10:00:25:01" clear="10:00:29:17">
      <hregion vposition="top">
        <line alignment="left">This is displayed in top left</line>
        <line alignment="left">corner of the screen</line>
      </hregion>
    </subtitle>
  </subtitlelist>
</esub-xf>
```

Every ESUB-XF XML structure shall contain the following XML nodes:

```
<?xml version="1.0" encoding="UTF-8"?>
<esub-xf />
```

The only encoding that shall be used is UTF-8. An example of an XML Schema is available here: http://www.fab-online.com/schemas/esub-xf-1.0.xsd When the XML Schema is not specified within the XML structure the above schema shall be used. The value of the attribute xml:space shall be assumed as default and the attribute shall not be specified. Other values shall not be used.

The <esub-xf> node contains subtitles for one or more languages. Subtitles for one language are stored within a <subtitlelist> node and multiple <subtitlelist> nodes may be present.

Every subtitle is stored within a node <subtitle>. Every <subtitle> node shall contain zero or one region which is stored within the <hregion> or <vregion> node. Every region is a rectangular area on the screen which contains the subtitle text.

Every region shall contain from zero up to twelve lines of subtitle text which are stored in <line> nodes. Lines of subtitle text shall be displayed continuously one below the other using the same line spacing.

Every node may contain attributes and nodes which are described in the following sections of this document. Attributes are not mandatory if not specified otherwise.

Additional nodes and attributes which are not defined in this document may be included in the XML structure however only when they follow the guidelines from the section '3. Application Specific ESUB-XF Extensions'. Other nodes and attributes are not permitted but may be included because they could be defined in a future version of this specification. When included they shall be ignored and shall not cause any malfunctioning of the device/software processing the XML structure.

2.1 The <esub-xf> Node

The <esub-xf> node contains subtitles for one or more languages for one program and may contain the following attributes:

`xmlns="urn:esub-xf"`
The namespace definition is necessary to allow validation of the XML with the schema .xsd file. This value shall always be present.

`framerate="25"`
This value defines the number of frames per second which were present in the original media for which the subtitles were prepared. This value shall always be present. The value shall be a whole positive integer number or a fraction specified as numerator/denominator. For NTSC 29.97 fps timecode the value shall therefore be "30000/1001". Other valid values include "24000/1001", "24", "25", "30", "50", "60000/1001", "60". Device/software processing the XML structure may decide to correctly support only some framerates values.

`dropframe="no"`
The value specifies whether timecode values are presented in dropframe mode or not. The value shall be one of the following: "yes" or "no". When this attribute is not present the value "no" shall be assumed.

`timebase="smpte"`
This attribute defines the format of timecode values used in the 'display' and 'clear' attributes as described later and shall always be present. The value shall be one of the following: "smpte" for timecode values in the format hh:mm:ss:ff or "msec" for timecode values specified as a whole positive integer number in milliseconds.

`start="hh:mm:ss:ff"` or `start="n"`
This timecode value defines the timecode of the first video frame of the video material used for preparation of subtitles. When not present the value 0 shall be assumed. The values of all 'display' and 'clear' attributes will be higher or equal than the value of 'start' except when timecode values for subtitles wrap around over midnight.

`generator="FAB Subtitler Version 9.66"`
This value defines the name and version of the writer of the ESUB-XF structure and it is optional.

The <esub-xf> node may contain exactly one <info> node which may contain comments, legal, copyright or other type of information about the subtitles in a simple text format. The text shall be included in one or multiple <text> nodes. All line breaks within the text in the <text> nodes shall be treated as spaces and not treated as text line breaks and <br /> shall not be used. Leading and trailing spaces of every line in the XML structure shall be ignored.

```
<?xml version="1.0" encoding="UTF-8"?>
<esub-xf xmlns="urn:esub-xf" framerate="30000/1001" dropframe="yes" timebase="smpte">
  <info>
    <text>The info block contains comments, legal, copyright</text>
    <text>or other type of information about the subtitles.</text>
    <text>special field 1=Content of the field 1</text>
    <text>second field name=Content of the second field</text>
  </info>
</esub-xf>
```

When data fields have to be included in the <text> nodes they shall be specified as fieldname=value.

2.2 The <subtitlelist> Node

The <subtitlelist> node contains all subtitles for one language and may contain the following attributes:

`language="eng"`
The language of subtitles contained in the <subtitlelist> node shall be specified using the ISO 639-2/T or ISO 639-2/B three letter code. Values between ‘qaa’ and ‘qtz’ may be used for user defined applications. This attribute is mandatory and shall always be present.

`langname="English"`
This value contains the descriptive name of the language. When not present the value is not defined. The value shall contain understandable and explicative texts which may be displayed to the viewer to select the type of subtitles to be displayed on the device i.e. "English – hard of hearing" or "French - translation".

`type="translation"`
This value specifies the type of text stream for this language and this attribute shall always be present. The following values shall be used: "translation" for subtitles which are a translation of the content or "hardofhearing" for subtitles which are intended for hearing impaired people or "ttsaudiodescription" for texts that shall be reproduced by display devices using an automated text to speech synthesis engine. Text to speech synthesis engines for all languages shall pronounce the text with at least 17 characters per second. Separation characters (.,!?) and spaces count as a normal character in this calculation.

```
<?xml version="1.0" encoding="UTF-8"?>
<esub-xf xmlns="urn:esub-xf" framerate="25" timebase="smpte">
  <subtitlelist language="eng" langname="English" type="translation">
  </subtitlelist>
</esub-xf>
```

2.3 The <subtitle> Node

The <subtitle> node contains information for one subtitle and may contain the following attributes:

`number="n"`
This attribute defines the subtitle number which can be used for easy identification of subtitles. The value shall start with one or more digits and may be followed by one or more letters from ‘a’ to ‘z’. When this attribute is not present then the subtitle number is not defined. The lowest possible subtitle number shall be 1.

`display="hh:mm:ss:ff"` or `display="n"`
This timecode value defines the timecode of the video frame on which the subtitle will first be displayed. When this attribute is not present then it shall be assumed that the subtitle does not have assigned ‘display’ timecode and can therefore not be used in operations that are based on time.

`clear="hh:mm:ss:ff"` or `clear="n"`
This timecode value defines the timecode of the video frame on which the subtitle will not be displayed any more. This value shall be higher than the value defined in ‘display’. When this attribute is not present then it shall be assumed that the subtitle does not have assigned ‘clear’ timecode and can therefore not be used in operations that are based on time.

Depending on the value of the ‘timebase’ attribute of the <esub-xf> node the timecode value can also be represented as a single whole number (0 or positive) which specifies the same value in milliseconds relative to the first video frame which has the number 0. When the timecode is specified as "hh:mm:ss:ff", the number of frames can consist of one, two or more digits.

The number of frames (:ff) in the timecode representation shall contain the correct number of frames without the correction defined in §12 SMPTE 12M-1. Therefore when using 720p60 the possible values for frames will be 0 to 59 and for 1080i60 the possible values will be 0 to 29.

When the ‘display’ value is lower or equal to the ‘clear’ value of the previous subtitle, this subtitle shall replace the previous subtitle on the screen at the time specified in ‘display’ without any visible flashing or flickering. Not even one frame without a visible subtitle shall be visible in this case.

`depth="0"`
This value defines the default Z-position of the subtitle text displayed in the subtitle for 3D video content. The value is a whole or decimal number with maximum 4 digits behind the decimal point from -100.0 to 1000.0. Comma shall not be used as the decimal point. The value 0 specifies that the text is displayed exactly on the screen surface. The value -100 specifies that the text is displayed in front of the screen exactly on the surface of the viewer’s eye. The value 100 specifies that the text is displayed behind the screen on the same distance from the screen as the distance between the viewer’s eye and the screen. When this attribute is not present this means that the subtitles were not prepared for 3D video content and the display device shall display them on the screen surface if not defined differently by default settings of the display device.

The following function converts the value ‘depth’ used for 3D video content to the number of pixels that the image for the left eye shall be shifted to the left and the image for the right eye shall be shifted to the right from its original position. When the result of the function is negative then the image for the left eye shall be shifted to the right and the image for the right eye shall be shifted to the left. The following parameters are used:

Depth_pct	The value ‘depth’ as defined in this specification
EyeDistance_cm	The distance between the left and right eye, normally 6.5 (cm)
WindowWidth_cm	The width of the screen which is used for display of the video in cm (i.e. 60.0)
VideoWidth_pixels	The horizontal resolution of the screen which is used for display (i.e. 1920)

```
function GetShiftInPixels (Depth_pct, EyeDistance_cm,
  WindowWidth_cm: Double; VideoWidth_pixels: Integer): Integer;
var
  ratio, shiftInCm: Double;
begin
  if Depth_pct = 0 then Result := 0
  else begin
    if Depth_pct < -90 then Depth_pct := -90;
    shiftInCm := (Depth_pct / 100 * EyeDistance_cm) / (2 * (1 + Depth_pct / 100));
    ratio := - shiftInCm / WindowWidth_cm;
    Result := Round (VideoWidth_pixels * ratio);
  end;
end;
```

appearance="nontransparentbox"
When this attribute is present and the value is ‘nontransparentbox’ then the subtitle text shall be displayed overlaid on a non-transparent box so that the picture below the box is not visible. When the value is ‘semitransparentbox’ then the subtitle text shall be displayed overlaid on a semi-transparent box. When the value is ‘border’ then the subtitle text shall be displayed as text with border. When the attribute is not present then the appearance shall be the standard appearance used for display of subtitles.

scrolllines="3"
When this attribute is present it shall contain a value between 1 and 12. In this case the display device shall process and display the subtitle in a special mode described under “2.9 Animated display mode with soft scroll”. Display devices may choose not to support the animated display. In this case subtitles with the defined attribute ‘scrolllines’ shall not be displayed at all.

The <subtitle> node may contain one or more <comment> nodes. Every <comment> node contains a comment text for the subtitle and shall not contain any CR or LF characters.

Examples of <subtitle> nodes

```
<?xml version="1.0" encoding="UTF-8"?>
<esub-xf xmlns="urn:esub-xf" framerate="25" timebase="smpte">
  <subtitlelist language="eng" langname="English" type="translation">
    <subtitle number="1" display="10:00:18:12" clear="10:00:21:03">
       <comment>This is a comment.</comment>
    </subtitle>
    <subtitle number="2" display="10:00:23:01" clear="10:00:25:19">
       <comment>This is comment line 1.</comment>
       <comment>This is comment line 2.</comment>
    </subtitle>
  </subtitlelist>
</esub-xf>
```

```
<?xml version="1.0" encoding="UTF-8"?>
<esub-xf xmlns="urn:esub-xf" framerate="25" timebase="msec">
  <subtitlelist language="eng" langname="English" type="translation">
    <subtitle display="5000" clear="7999">
       <comment>This is a comment.</comment>
    </subtitle>
  </subtitlelist>
</esub-xf>
```

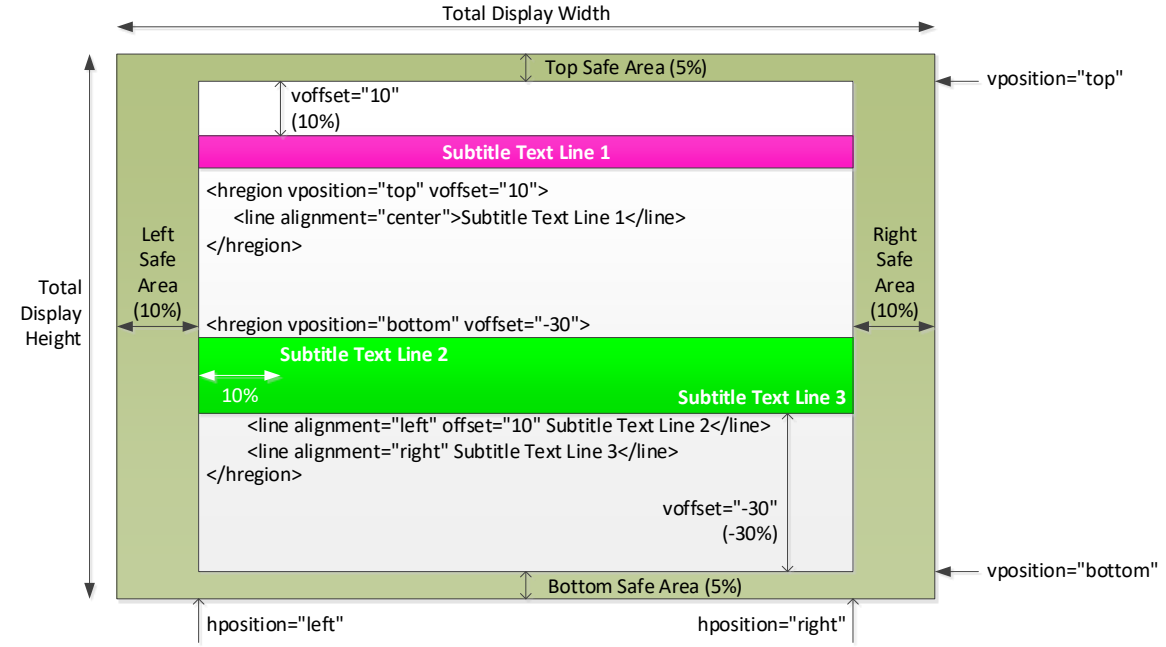
2.4 The <hregion> and <vregion> Nodes

Every <subtitle> node shall contain zero or one or two regions which are stored within <hregion> or <vregion> node. A subtitle without a region is an empty subtitle. The region is a rectangular area on the screen which contains subtitle text. Region contained in <hregion> node is horizontally oriented and outputs text in left-to-right or right-to-left mode. Region contained in <vregion> node is vertically oriented and outputs text in top-to-bottom or bottom-to-top mode. One <subtitle> node can contain only one horizontally or only one vertically oriented region.

The region shall contain from zero up to twelve lines of subtitle text which are stored in <line> nodes. The total number of lines in every subtitle shall not be higher than 12. Lines of subtitle text shall be displayed continuously one below the other for horizontal text and one next to the other for vertical text using the same line spacing.

Examples of a <hregion> node

```
<?xml version="1.0" encoding="UTF-8"?>
<esub-xf xmlns="urn:esub-xf" framerate="25" timebase="smpte">
  <subtitlelist language="eng" langname="English" type="translation">
    <subtitle display="10:00:18:12" clear="10:00:21:03">
      <hregion vposition="top" voffset="10">
        <line alignment="center">Subtitle text line 1</line>
      </hregion>
    </subtitle>
  </subtitlelist>
</esub-xf>
```




```
<?xml version="1.0" encoding="UTF-8"?>
<esub-xf xmlns="urn:esub-xf" framerate="25" timebase="smpte">
  <subtitlelist language="eng" langname="English" type="translation">
    <subtitle display="10:00:18:12" clear="10:00:21:03">
      <hregion vposition="bottom" voffset="-30">
        <line alignment="left" offset="10">Subtitle text line 2</line>
        <line alignment="right">Subtitle text line 3</line>
      </hregion>
    </subtitle>
  </subtitlelist>
</esub-xf>
```

A region is a rectangular area which has a well-defined non-overlapping position on the screen and a well-defined text output direction.

A horizontally oriented region is contained in <hregion> node:

- The region is horizontally oriented.
- The region has horizontal text output direction left-to-right or right-to-left.
- The region's vertical position is defined by the value of 'vposition' and 'voffset' attributes.
- The region starts horizontally on the left side at the right border of Left Safe Area and it ends on the right side at the left border of Right Safe Area.
- The text within <line> nodes will be displayed below the text from the previous <line> node.
- Constant line spacing shall be used between <line> nodes.
- The region's height is the sum of the height of all <line> nodes within the region.

A vertically oriented region is contained in <vregion> node:

- The region is vertically oriented.
- The region has vertical text output direction top-to-bottom or bottom-to-top.
- The region's horizontal position is defined by the value of 'hposition' and 'hoffset' attributes.
- The region starts vertically on the top at the bottom border of Top Safe Area and it ends at the bottom at the top border of Bottom Safe Area.
- The text within <line> nodes will be displayed either left of the text from the previous <line> node or right of the previous <line> node depending on the value of the attribute 'direction'.
- Constant line spacing shall be used between <line> nodes.
- The region's width is the sum of the width of all <line> nodes within the region.

One subtitle shall contain only horizontally or only vertically oriented regions.

The <hregion> node may contain the following attributes:

`vposition="bottom"`
The value defines the vertical position of the region on the screen and the value shall be one of the following: "bottom" when the region shall be at the bottom of the screen ending on the top border of Bottom Safe Area and "top" when the region shall be at the top of the screen starting at the bottom border of Top Safe Area. When this attribute is not present the value "bottom" shall be assumed.

`voffset="0"`
The value defines the vertical offset of the region on the screen in percentage of Total Display Height. The value is a whole or decimal number with maximum 4 digits behind the decimal point from -100.0 to 100.0. Comma shall not be used as the decimal point. A positive value moves the region down and a negative value moves the region up on the screen. When this attribute is not present the value 0 shall be assumed. In case that the subtitle text would have to be displayed outside of the screen after applying 'voffset', the position of the region shall be moved so that the complete subtitle text will be displayed inside of the screen. When converting teletext subtitles to ESUB-XF the 'voffset' value for one single height teletext row is 3.75 which is calculated from 24 teletext lines and 5% top and 5% bottom safe area. When converting open subtitles with 12 lines to ESUB-XF the 'voffset' value for one row is 7.5 which is calculated from 12 lines of text and 5% top and 5% bottom safe area.

The <vregion> node may contain the following attributes:

`hposition="left"`
The value defines the horizontal position of the region on the screen and the value shall be one of the following: "left" when the region shall be on the left side of the screen starting at the right border of Left Safe Area and "right" when the region shall be at the right side of the screen ending at the left border of Right Safe Area.

`hoffset="0"`
The value defines the horizontal offset of the region on the screen in percentage of Total Display Width. The value is a whole or decimal number with maximum 4 digits behind the decimal point from -100.0 to 100.0. Comma shall not be used as the decimal point. A positive value moves the region right and a negative value moves the region left on the screen. When this attribute is not present the value 0 shall be assumed. In case that the subtitle text would have to be displayed outside of the screen after applying 'hoffset', the position of the region shall be moved so that the complete subtitle text will be displayed inside of the screen.

`direction="ltr"`
This attribute is only used for vertically oriented regions and the value "ltr" specifies that lines will be displayed from left to right and "rtl" that lines will be displayed from right to left. When not present the value "ltr" shall be assumed.

The text within regions must be written in the reading order of the subtitle text so that systems that only extract the text from <line> nodes and do not process any positioning information will extract the text in the correct reading order.

Display devices shall use this algorithm when displaying a subtitle with a horizontally oriented region:

- Calculate the height and absolute vertical position on the screen for the region.
- Correct the vertical position for the region so that it is within the screen.
- Display the region until the last <line> which can be displayed within the screen.
- Do not display lines which do not fit into the screen when there are too many lines of text.

Please note that the subtitle shall generally be displayed inside a "Safe Area" which is defined horizontally as 10% of Total Display Width and vertically as 5% of Total Display Height. It is however possible to display the subtitle outside of the Safe Area by defining:

- A negative value for 'voffset' for horizontally oriented top positioned region.
- A positive value for 'voffset' for horizontally oriented bottom positioned region.
- A negative value for 'hoffset' for vertically oriented left positioned region.
- A positive value for 'hoffset' for vertically oriented right positioned region.

All display devices shall support horizontally oriented regions. Support for vertically oriented regions is optional and display devices may choose to support vertically oriented regions or not. When vertically oriented regions are not supported such subtitles shall be considered as empty subtitles.

2.5 The <line> Node

The region contains from zero to twelve <line> nodes. A region without <line> nodes is an empty region. Every <line> node contains information for one line of subtitle text within a <hregion> or <vregion> node and may contain the following attributes:

`alignment="center"`
This value defines the alignment of the subtitle text for the line within the rectangular region. The value shall be one of the following: "left", "center" or "right" for horizontally oriented regions. For vertically oriented regions it shall be one of the following: "top", "center" or "bottom". When this attribute is not present then the value "center" shall be assumed for horizontally oriented regions and "top" for vertically oriented regions.

`offset="0"`
This value defines the offset for which the text within the line shall be moved. For horizontally oriented regions the value is specified in percent of Total Display Width and for vertically oriented regions the value is specified in percent of Total Display Height. The value is a whole or decimal number with maximum 4 digits behind the decimal point from -100.0 to 100.0. Comma shall not be used as the decimal point. A positive value moves the text right/down and a negative value moves the text left/up on the screen. When this attribute is not present the value 0 shall be assumed. In case that the subtitle text would have to be displayed outside of the screen after applying 'offset', the position of the text shall be moved so that the complete subtitle text will be displayed inside the screen. Note that the 'offset' can be used to display the text outside of the Safe Area, for example in horizontally oriented regions when the value is negative for left aligned line or when the value is positive for right aligned line.

`direction="ltr"`
This value defines the default text output direction for the current line. The value for horizontally oriented regions shall be one of the following: "ltr" for 'left to right' or "rtl" for 'right to left'. The value for vertically oriented regions shall be one of the following: "ttb" for 'top to bottom' or "btt" for 'bottom to top'. When this attribute is not present the value "ltr" shall be assumed for horizontally oriented regions and "ttb" for vertically oriented regions.

`font="1"`
This value defines the font for display of subtitle text. The format of the value is deliberately not specified to allow any value which may or may not be interpreted by the display device.

The text contained within the <line> node is the subtitle text for one line. Leading and trailing spaces shall be ignored and multiple spaces shall be converted to a single space. The text shall also not contain any of the following characters: & (ampersand), < (less-than sign), > (greater-than sign). These characters shall be replaced with the sequence of characters defined in the following table:

Character	Replacement
&	&amp;
<	<
>	>

Example: The line text 'John & Mary' shall be stored as '<line>John & Mary</line>'.

All line breaks that may be present within the text of the <line> node shall produce a space in the subtitle text only when there is no preceding or succeeding space and shall not produce a line break within the subtitle.
 shall not be used. Leading and trailing spaces of every line in the XML structure shall be ignored.

The maximum number of characters per line is not defined by this specification because this depends on the font type and size which will be used for display of subtitles and the font definition is not part of this specification. As a general rule not more than 36 characters per line shall be used to allow for compatibility with legacy teletext subtitles. However depending on the final font size, screen size and screen aspect ratio the maximum number of characters per line may be lower or higher than 36. The number of lines of text per subtitle shall not exceed 12. Empty <line> nodes shall be permitted and may be used for vertical/horizontal positioning of the subtitle text in horizontally/vertically oriented regions.

Usage of within the <line> node

 shall be used when colors and text formatting are required for the subtitle text. When is present in the <line> node the subtitle text shall not be present outside of the and characters outside of shall be ignored. The following attributes are allowed:

`italic="on"`
This value defines the font style for the text within the . The value shall be one of the following: "off" or "on". When this attribute is not specified the value "off" shall be assumed.

`bold="on"`
This value defines the font style for the text within the . The value shall be one of the following: "off" or "on". When this attribute is not specified the value "off" shall be assumed.

`underline="on"`
This value defines the font style for the text within the . The value shall be one of the following: "off" or "on". When this attribute is not specified the value "off" shall be assumed.

`textcolor="white"`
This value defines the text color for the text within the . The value shall be one of the color names as defined below. When this attribute is not specified the value "white" shall be assumed.

`backcolor="black"`
This value defines the background color for the text within the . The value shall be one of the color names that are defined below. When this attribute is not specified the value "black" shall be assumed and the display device shall use the default presentation values, either black border surrounding the characters, black box or semi-transparent black box behind the characters. Display devices may decide to either use or ignore this value depending on whether they support the display of background colors. This value is mainly provided for compatibility with teletext subtitles.

The following color names (Color) shall be used in the ‘textcolor’ and ‘backcolor’ attribute. The RGB hex color values specified below are not permitted as values. They are recommended color values for display and may be redefined when displaying the subtitle. For HDR displays these values shall be redefined because white as FFFFFF will probably be too bright. Color values shall be kept similar to the color name because the color name might be used in the subtitle text (“Peter has the color

green" might appear in the text). When converting teletext subtitles from/to ESUB-XF the color violet shall be converted from/to teletext color black and purple from/to teletext color magenta.

Color	RGB hex
white	FFFFFF
red	FF2D34

Color	RGB hex
green	72FD59
blue	4545FF

Color	RGB hex
cyan	91FFFF
yellow	E8E858

Color	RGB hex
purple	F55FF5
violet	8505FD

`language="eng"`
This value defines the language of the text and is specified using the ISO 639-2/T/B three letter code. For subtitles of type "ttsaudiodescription" the text to speech engine shall pronounce the text in this language. When the language is not available, it shall pronounce the name of the language defined in the attribute 'langname' followed by the text, both spoken in the language defined for <subtitlelist>.

`langname="English"`
This value contains the descriptive name of the language. When not present the value is not defined.

`marker="AnyName"`
This value contains a name of a marker which is valid exactly at the position of the first character of the text within the . The name can be any text.

Limitations and instructions for usage of the within the text of the <line> node:

- Nested shall not be used. Therefore the following shall not be used:
  `<line><span italic="on">word1<span bold="on">word2</span></span></line>`
  Such formatting shall be written as a series of nodes. The above example could be written as:
  `<line><span italic="on">word1</span><span italic="on" bold="on">word2</span></line>`
- Every shall always contain complete words. When processing consecutive nodes a whitespace shall be inserted into the text. The above example shall produce the following text with a whitespace between word1 and word2: word1 word2
- Line breaks between nodes in the XML structure shall be ignored and shall not produce spaces in the text and <br /> shall not be used.
- Leading and trailing spaces within the text of the node shall be ignored and multiple consecutive spaces shall be converted to a single space. Empty nodes shall be used for multiple consecutive spaces.

Usage of <split /> within the <line> node

A special node <split /> may occur exactly in this form exactly once within the <line>. The presence of this node will instruct the display device to display the text in front of the <split /> starting on one side of the screen (left justified for ltr subtitles) and the text behind the <split /> on the other side of the screen (right justified for ltr subtitles). This enables displaying the text in multiple lines on two sides of the screen in case that there are two persons visible on the screen. Generally, a different text color will be used in front of the <split /> and behind the <split />.

Example of the and <split /> usage

```
<line alignment="left"><span textcolor="yellow" italic="on">Yellow person</span><split />
  <span textcolor="green" italic="on">Green person</span></line>
<line alignment="left"><span textcolor="yellow" italic="on">on the left.</span><split />
  <span textcolor="green" italic="on">on the right.</span></line>
```

2.6 The <image> Node

Every <subtitle> node may contain one or more <image> nodes. Every <image> node contains the image of the subtitle. The image has to be stored in the PNG format as a 32-bit bitmap with 8-bit alpha channel in the resolution of the whole screen. All PNG formats shall be supported, also those with palette colors. The image shall be stored in a separate file and the filename of the external file that contains the image must be specified. Multiple <image> nodes can be used if multiple bitmaps for different resolutions are necessary. When using ESUB-XF transport protocols image files shall be part of the same ESUB-XF packet that contains the ESUB-XF structure with the subtitle.

Following attributes can be present in the <image> node:

`format="png"`
This value defines the format of the image and it shall contain "png". The attribute is mandatory.

`width="1920"`
This value defines the width of the image stored in the <image> node. This attribute is mandatory.

`height="1080"`
This value defines the height of the image stored in the <image> node. This attribute is mandatory.

`filename="pngfile1.png"`
This value defines the name of the file that contains the image. When this attribute is present the image is stored in a separate file or in the <filelist> node within the same ESUB-XF XML structure.

`stereo="none"`
This value defines whether the image is for 3D video content and what type of 3D image it contains. Possible values are "none" if the image is not for 3D, "sidebyside" if the image is in 'Side by side' format, "topbottom" if the image is in 'Top Bottom' format, "left" if the image is for the left eye, "right" if the image is for the right eye (in this case two <image> nodes will be present for the same resolution) and "disparity" if the image is for the left eye and the additional attribute 'disparity' defines the disparity for the image for the right eye. The default value is "none".

`disparity="20"`
This value is a whole number which may be positive, 0 or negative. This attribute shall only be present if the value of the attribute 'stereo' is "disparity" and therefore the image is for the left eye. The value defines the number of pixels that the image shall be shifted to the right to obtain the image for the right eye.

Examples of <image> nodes

```
<subtitle display="10:00:18:12" clear="10:00:21:03">
  <hregion><line alignment="right">Subtitle text line 1</line></hregion>
  <image format="png" width="1920" height="1080" filename="s1.png" />
</subtitle>
<subtitle display="10:00:24:21" clear="10:00:27:11">
  <hregion><line alignment="right">Subtitle text line 1</line></hregion>
  <image format="png" width="1920" height="1080" filename="s2.png" />
</subtitle>
```

2.7 The <filelist> Node

The <filelist> node contains files encoded as BASE64 and allows using XML as transport for any kind of file. When present in the ESUB-XF file it must be included after the <subtitlelist> node. Files defined in the <image> node can be included in this node but also other types of files like any kind of original subtitle file in STL, PAC or any other format can be included.

Example of <filelist> node

```
<filelist>
  <file filename="s1.png">
iVBORw0KGgoAAAANSUhEUgAAAAoAAAAKCAIAAAACUFjqAAAAK3RFWHRDcmVhdGlvbiBUaW1lAE1p
IDEyIEZlYiAyMDIwIDE2OjMxOjEyICswMTAw8LnA5QAAAAd0SU1FB+QCDA8fIxV2Jz0AAAAJcEhZ
cwAACvAAAArwAUKsNJgAAAAEZ0FNQQAAsY8L/GEFAAAAFUlEQVR42mP8//8/A27AxIAXjFRpAKXj
AxFkRbxkAAAAAElFTkSuQmCC
  </file>
</filelist>
```

2.8 Example of a valid ESUB-XF XML Structure

The following is an example of a valid XML structure according to this specification:

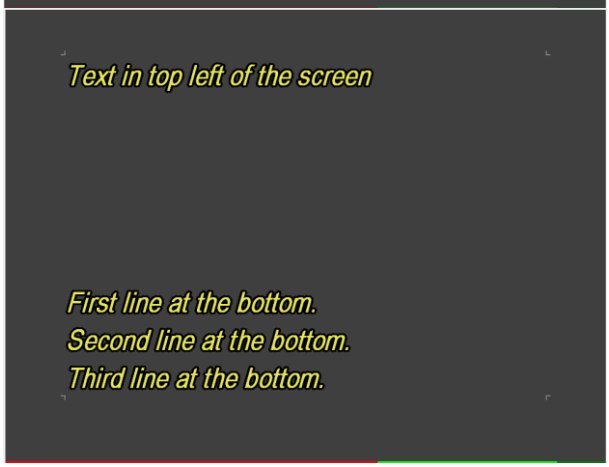
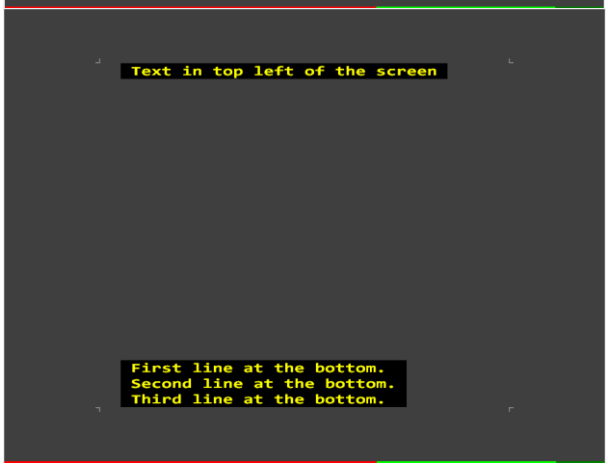
```
<?xml version="1.0" encoding="UTF-8"?>
<esub-xf xmlns="urn:esub-xf" framerate="25" timebase="smpte">
  <info>
    <text>The info block contains comments, legal, copyright</text>
    <text>or other type of information about the subtitles.</text>
    <text />
    <text>Empty lines can be used as well.</text>
  </info>
  <subtitlelist language="eng" langname="English" type="translation">
    <subtitle display="10:00:18:12" clear="10:00:21:03" depth="20">
      <comment>This is a comment</comment>
      <hregion vposition="bottom" voffset="-20">
        <line alignment="right"><span textcolor="yellow" italic="on">Yellow italic</span>
<span textcolor="cyan">and cyan non-italic.</span>
        </line>
        <line alignment="left"><span textcolor="red" bold="on">Red bold</span> <span
textcolor="green">and green non-bold.</span>
        </line>
      </hregion>
    </subtitle>
    <subtitle display="10:00:25:01" clear="10:00:29:17" depth="30">
      <comment>This is comment line 1</comment>
      <comment>This is comment line 2</comment>
      <hregion vposition="top">
        <line alignment="left"><span textcolor="yellow" italic="on">
          Text in top left corner</span>
        </line>
        <line alignment="left"><span textcolor="yellow" italic="on">of the screen</span>
        </line>
      </hregion>
    </subtitle>
  </subtitlelist>
</esub-xf>
```

2.8 Rendering of the Example XML Structure

The example XML structure above may be rendered in different ways, depending on capabilities of the rendering engine and the output device. The pictures below show two possible renderings.

In the left column, subtitles are rendered as a bitmap that is applied to the video. All formatting is preserved with the exception of the 'bold' attribute, which is not supported by the ESUB-XF rendering engine which was available at the time of publication of this specification.

In the right column, subtitles are rendered in the teletext format. Both 'italic' and 'bold' attribute were ignored as they are not supported by the teletext specification ETSI EN 300 706.

ESUB-XF subtitles	**Teletext subtitles**
	
	

2.9 Animated Display Mode with Soft Scroll

The display device shall process and display a subtitle in a special mode with soft scroll when the following conditions are met (when they are not fulfilled the subtitle shall be displayed normally):

- the attribute ‘scrolllines’ is present in the <subtitle> node and
- the display device is displaying text from <line> nodes and not <image> nodes and
- the previous subtitle is still displayed on the screen (defined by the attribute ‘clear’ of the previous subtitle) at the time when the current subtitle shall be displayed and
- the subtitle contains exactly one region and the value of the attributes ‘hposition’, ‘vposition’, ‘hoffset’, ‘voffset’ and ‘direction’ are equal in the current and previous subtitle

In such case the display device shall process and display the current subtitle as follows:

- If the previous subtitle, which is still displayed on the screen, contains more lines than specified in ‘scrolllines’ then the previous subtitle is displayed again immediately without flickering, however without first few lines so that ‘scrolllines’ number of lines is displayed.
- The text specified in <line> nodes of the current subtitle is added to the text of the previous subtitle so that first a space is added to the text from the last <line> of the previous subtitle and then the text from the <line> node is added. This way only one word can be added to the previous subtitle. The display device shall automatically apply a line break and add text to a new line if the text does not fit into one line. An empty <line> node creates a line break.
- If the number of lines of the resulting subtitle which shall be displayed is more than ‘scrolllines’ or if a new line of text appears in the new subtitle after adding the new text then first few lines shall be removed from the subtitle so that not more than ‘scrolllines’ of lines are present and then the “soft scroll effect” shall be applied to the previous subtitle before displaying the resulting new subtitle.
- The previous subtitle shall be replaced on the screen by the new subtitle. The viewer will generally only notice that a new word or line of text will appear because other lines of text will be displayed on exactly the same positions as in the previous subtitle after the movement was finished. There shall be no flashing or flickering of text visible on the screen.

The “soft scroll effect” shall appear similar as in CEA-608 and is defined as:

- The scroll direction for horizontally oriented subtitles is up. For vertically oriented subtitles the scroll direction depends on the value of the attribute ‘direction’. When ‘direction’ is "ltr" the scroll direction is left, otherwise it is right.
- The scroll effect shall start exactly at the display time of the subtitle, have a fixed duration of 400ms and shall appear soft to the viewer without irregular delays during the movement. Therefore the total display time of the subtitle will be 400ms less than defined by ‘display’ and ‘clear’ attributes.
- The number of pixels for the scroll movement is defined by the height of one line of text.
- If the number of lines which are currently displayed is ‘scrolllines’ then the first line of text of the displayed subtitle shall slowly disappear during the movement so that not more than the number of lines defined by ‘scrolllines’ shall be visible. Every time when the subtitle is moved for example 2 pixels up the two top most lines of pixels shall not be displayed any more. Analogous approach shall be used for other scroll directions.

There are three special cases that affect the display of the animated subtitle with soft scroll:

- When a subtitle is empty (the subtitle does not contain any region) then the complete subtitle which is currently on the screen shall be erased so that no subtitle text shall be displayed on the screen anymore.
- An empty <line> node shall create a line break.
- The <line> node may have an attribute ‘replacelastwords‘ with a value "n" between 0 and 99. When the attribute is not present the value 0 shall be assumed. When this attribute is present and it is not 0 the display device shall remove the last "n" words from the screen before displaying the text in the <line> node. Removing the last word from what is displayed on the screen means examining the text that is currently displayed on the screen, searching for the last normal whitespace character (dec. 32) or the last line break (whichever comes first) and removing all text after the last normal whitespace character or the last line break. All last "n" words shall be removed at the same time without any delay or pause and the text from the <line> node shall be displayed after that following the rules as described above. The attribute ‘replacelastwords‘ shall only be used in animated display mode with soft scroll (when the attribute ‘scrolllines’ is present). The attribute shall be ignored by the display device when the subtitle is not being displayed in animated display mode with soft scroll (when the attribute ‘scrolllines’ is not present).

3. Application Specific ESUB-XF Extensions

The ESUB-XF XML structure described in this document may be extended by anybody by including additional nodes and attributes which are not defined in this specification. For example, a broadcaster or a manufacturer may extend the XML structure to store additional information required by his system or workflow. The following rules shall be applied when using extensions within this specification:

- Additional nodes and attributes may be added within the xml structure as long as the xml structure remains valid.
- A separate namespace with a descriptive name of the extension shall be used for every node and attribute as shown in the following example.
- Attributes from a separate namespace may be used on any position within any node.
- Nodes from a separate namespace shall only be used after and not before or in between the nodes defined in this specification.
- Names of nodes and attributes without a namespace shall not be used if they are not part of the basic specification described in this document to avoid conflicts with future extensions of this specification.
- Systems that will read the xml structure shall ignore nodes and attributes from namespaces unknown to them or when they do not know how to interpret them.

3.1 Example of an Application Specific ESUB-XF XML Structure

The following is an example of a valid ESUB-XF XML structure with two custom extensions for ‘aaa’ and ‘bbb’ namespaces which in reality do not exist and are only used to explain how to use namespaces for extensions of this specification.

```
<?xml version="1.0" encoding="UTF-8"?>
<esub-xf xmlns="urn:esub-xf" xmlns:aaa="urn:esub-xf-aaa" xmlns:bbb="urn:esub-xf-bbb"
framerate="25" timebase="smpte">
  <info>
    <text>The info block contains comments, legal, copyright</text>
    <text>or other type of information about the subtitles.</text>
    <text />
    <text>Empty lines can be used as well.</text>
  </info>
  <aaa:metadata>
    <aaa:title>Title of the movie</aaa:title>
    <aaa:client>BR Munich</aaa:client>
  </aaa:metadata>
  <subtitlelist language="eng" langname="English" type="translation">
    <subtitle display="10:00:18:12" clear="10:00:21:03" depth="20" aaa:number="1">
      <comment>This is a comment</comment>
      <hregion vposition="bottom" voffset="-20">
        <line alignment="right"><span textcolor="yellow" italic="on" aaa:font="2">First
line in yellow italic</span> <span textcolor="cyan">and cyan non-italic.</span>
        </line>
        <line alignment="left"><span textcolor="red" bold="on" aaa:font="3">Second line in
red bold</span> <span textcolor="green">and green non-bold</span>
        </line>
      </hregion>
      <aaa:comment>This is a comment</aaa:comment>
    </subtitle>
    <subtitle display="10:00:25:01" clear="10:00:29:17" depth="30">
      <comment>This is comment line 1</comment>
      <comment>This is comment line 2</comment>
      <hregion vposition="top">
        <line alignment="left" bbb:speaker="Thomas"><span textcolor="yellow"
          italic="on">Text in top left corner</span>
        </line>
        <line alignment="left" bbb:speaker="Thomas"><span textcolor="yellow" italic="on">
          of the screen</span>
        </line>
      </hregion>
      <bbb:marker>This is BBB specific marker</bbb:marker>
    </subtitle>
  </subtitlelist>
</esub-xf>
```

The XML Schema for the ‘aaa’ and ‘bbb’ namespaces is available for download here:

http://www.fab-online.com/schemas/esub-xf-aaa-1.0.xsd
http://www.fab-online.com/schemas/esub-xf-bbb-1.0.xsd

3.2 Mapping of EBU STL Files to ESUB-XF XML Structure

The specification for EBU STL files (TECH. 3264-E from February 1991) defines content for subtitle files which is not included in the basic ESUB-XF XML Structure.

Therefore an extension of the ESUB-XF XML Structure using the 'ebu-stl' namespace shall be used as described below to store data from EBU STL files into the ESUB-XF XML Structure.

The fields of the General Subtitle Information (GSI) block from EBU STL files are stored within the <ebu-stl:metadata> node where every subnode contains the value of one field from the GSI. The <ebu-stl:metadata> node is stored within the <subtitlelist> node. The following table shows the GSI mnemonic names and their corresponding names of nodes within the <ebu-stl:metadata> node:

GSI mnemonic	**<ebu-stl:metadata> subnode name**	**Description**
CPN	<ebu-stl:cpn>	Code Page Number
DFC	<ebu-stl:dfc>	Disk Format Code
DSC	<ebu-stl:dsc>	Display Standard Code
CCT	<ebu-stl:cct>	Character Code Table Number
LC	<ebu-stl:lc>	Language Code
OPT	<ebu-stl:opt>	Original Programme Title
OET	<ebu-stl:oet>	Original Episode Title
TPT	<ebu-stl:tpt>	Translated Programme Title
TET	<ebu-stl:tet>	Translated Episode Title
TN	<ebu-stl:tn>	Translator´s Name
TCD	<ebu-stl:tcd>	Translator´s Contact Details
SLR	<ebu-stl:slr>	Subtitle List Reference Code
CD	<ebu-stl:cd>	Creation Date
RD	<ebu-stl:rd>	Revision Date
RN	<ebu-stl:rn>	Revision Number
TNB	<ebu-stl:tnb>	Total Number of Text and Timing Information (TTI) Blocks
TNS	<ebu-stl:tns>	Total Number of Subtitles
TNG	<ebu-stl:tng>	Total Number of Subtitle Groups
MNC	<ebu-stl:mnc>	Maximum Number of Displayable Characters in Any Text Row
MNR	<ebu-stl:mnr>	Maximum Number of Displayable Rows
TCS	<ebu-stl:tcs>	Time Code: Status
TCP	<ebu-stl:tcp>	Time Code: Start of Programme
TCF	<ebu-stl:tcf>	Time Code: First In-Cue
TND	<ebu-stl:tnd>	Total Number of Disks
DSN	<ebu-stl:dsn>	Disk Sequence Number
CO	<ebu-stl:co>	Country of Origin
PUB	<ebu-stl:pub>	Publisher
EN	<ebu-stl:en>	Editor's Name
ECD	<ebu-stl:ecd>	Editor's Contact Details
SB	<ebu-stl:sb>	Spare Bytes
UDA	<ebu-stl:uda>	User Defined Area

The data contained in the nodes is an exact copy of the field data from the EBU STL file converted from the STL codepage to UTF-8. Unicode format characters shall not be used in the node data. There shall be no spaces at the beginning and at the end of the node data. Characters with codes lower than 32 decimal shall be converted to spaces.

Example of a valid ESUB-XF XML Structure with EBU STL Mapping

The following is an example of a valid XML structure with EBU STL mapping:

```
<?xml version="1.0" encoding="utf-8"?>
<esub-xf xmlns="urn:esub-xf" xmlns:ebu-stl="urn:esub-xf-ebu-stl" framerate="25"
timebase="smpte">
  <ebu-stl:metadata>
    <ebu-stl:cpn>437</ebu-stl:cpn>
    <ebu-stl:dfc>STL25.01</ebu-stl:dfc>
    <ebu-stl:dsc>0</ebu-stl:dsc>
    <ebu-stl:cct>00</ebu-stl:cct>
    <ebu-stl:lc>0A</ebu-stl:lc>
    <ebu-stl:opt>Original Programme Title</ebu-stl:opt>
    <ebu-stl:oet>Original Episode Title</ebu-stl:oet>
    <ebu-stl:tpt>Translated Programme Title</ebu-stl:tpt>
    <ebu-stl:tet>Translated Episode Title</ebu-stl:tet>
    <ebu-stl:tn>Translator's Name</ebu-stl:tn>
    <ebu-stl:tcd>Translator's Contact Details</ebu-stl:tcd>
    <ebu-stl:slr>1234567890123456</ebu-stl:slr>
    <ebu-stl:cd>120922</ebu-stl:cd>
    <ebu-stl:rd>120923</ebu-stl:rd>
    <ebu-stl:rn>99</ebu-stl:rn>
    <ebu-stl:tnb>2</ebu-stl:tnb>
    <ebu-stl:tns>2</ebu-stl:tns>
    <ebu-stl:tng>1</ebu-stl:tng>
    <ebu-stl:mnc>37</ebu-stl:mnc>
    <ebu-stl:mnr>12</ebu-stl:mnr>
    <ebu-stl:tcs>1</ebu-stl:tcs>
    <ebu-stl:tcp>10000000</ebu-stl:tcp>
    <ebu-stl:tcf>10001812</ebu-stl:tcf>
    <ebu-stl:tnd>1</ebu-stl:tnd>
    <ebu-stl:dsn>1</ebu-stl:dsn>
    <ebu-stl:co>GBR</ebu-stl:co>
    <ebu-stl:pub>Publisher</ebu-stl:pub>
    <ebu-stl:en>Editor's Name</ebu-stl:en>
    <ebu-stl:ecd>Editor's Contact Details</ebu-stl:ecd>
    <ebu-stl:sb>Spare bytes</ebu-stl:sb>
    <ebu-stl:uda>User defined area</ebu-stl:uda>
  </ebu-stl:metadata>
  <subtitlelist language="eng" langname="English" type="translation">
    <subtitle display="10:00:18:12" clear="10:00:21:03">
      <hregion>
        <line>First line of bottom justified text</line>
        <line>Second line of bottom justified text</line>
      </hregion>
    </subtitle>
    <subtitle display="10:00:25:01" clear="10:00:29:17">
      <hregion vposition="top">
        <line alignment="left">This is displayed in</line>
        <line alignment="left">top left corner</line>
      </hregion>
    </subtitle>
  </subtitlelist>
</esub-xf>
```

The XML Schema for the ‘ebu-stl’ namespace is available for download here:

http://www.fab-online.com/schemas/esub-xf-ebu-stl-1.0.xsd

4. The ESUB-XF Storage Formats

The ESUB-XF XML structure can be stored in plain XML text files as described in the ESUB-XF File Format and also in MXF files.

4.1 The ESUB-XF File Format

The ESUB-XF XML structure can be stored directly into a file using these rules:

- The file shall not have a BOM (Byte Ordering Mark) at the beginning.
- The file shall always be stored using UTF-8 encoding.
- CRLF ("\r\n") shall be used for line breaks and XML shall be stored in a formatted human readable form.
- The file extension for ESUB-XF files shall always be ".esub".
- The file shall contain only the ESUB-XF XML structure and not the complete ESUB-XF Packet

Valid filenames include, but are not limited to: House_Episode_29.esub, SherlockHolmes_2009_English.esub, Two-And-A-Half-Men-Epi37.esub

It is recommended that ESUB-XF files are used for one language only and that multiple files are used for multiple languages. Multiple languages (multiple <subtitlelist> nodes) in one ESUB-XF file can however be used in scenarios where the reader of the file expects all languages in a single file.

4.2 Embedding ESUB-XF into MXF Files

SMPTE 410 shall be used for embedding of the complete ESUB-XF XML Structure with all subtitles for all languages as a clip-wrapped Generic Container.

The ESUB-XF XML Structure containing all subtitles for all languages can be embedded into an MXF file as SMPTE 410 Essence Data. The content of the SMPTE 410 Generic Stream Data Element Value shall always be the ESUB-XF XML Structure in the same format as when stored into a file and shall never contain the ESUB-XF Packet Header. The “ESUB-XF Data” Key of the KLV is the same as for Timed Text XML files.

The following rules shall apply for storing the ESUB-XF XML Structure in a Generic Stream Partition:

- The value of the Data Element KLV shall contain one ESUB-XF XML Structure.
- The Key of the Data Element KLV shall be the “ESUB-XF Data” Key.
- CRLF ("\r\n") shall be used for line breaks within the ESUB-XF XML Structure.
- There shall not be a BOM (Byte Ordering Mark) before the ESUB-XF XML Structure.
- There shall be one Generic Stream Partition for each ESUB-XF XML Structure.
- A Generic Stream partition shall contain only one ESUB-XF XML Structure.
- Content shall be Clip-Wrapped and stored as “Type C1” (Figure A.5 in S410M).

5. The ESUB-XF Transport Protocols

Often there is a need for transport of subtitles from one device to another like:

- During live transmission of subtitles where a system with speech recognition sends subtitles one after the other to the subtitle transmission system
- During transmission of subtitles to a subtitle decoder/display device

In above scenarios subtitles are sent one by one instead of sending the complete list of subtitles. The ESUB-XF transport protocols define how subtitles can be transported within different signals. Implementation of these protocols is not mandatory.

5.1 The ESUB-XF Packet

The ESUB-XF Packet is used for encapsulation of the ESUB-XF XML Structure in files, metadata containers or transport protocols and consists of:

- ESUB-XF Packet Header followed by
- ESUB-XF Payload Data consisting of the ESUB-XF XML Structure

CRLF ("\r\n") shall be used for line breaks within the ESUB-XF XML Structure. There shall not be a BOM (Byte Ordering Mark) before the ESUB-XF XML Structure and there may be a CRLF between the ESUB-XF Packet Header and the ESUB-XF XML Structure.

The ESUB-XF Packet Header is UTF-8 encoded text in the following form:

```
<esub-xf,size=xxx,type=yyy>
```

The text between <esub-xf, and > is separated by commas into multiple keyword=value pairs. Values shall never contain the characters “,”, “<” and “>”. These characters shall be replaced by “&”, “<” and “>”.

The following keywords shall always be present in the ESUB-XF packet header. Other keywords may be present but this depends on whether the ESUB-XF packet is used for storage or transport.

size= defines the size of the ESUB-XF Payload data following the ESUB-XF Packet Header in bytes including all CRLF characters. The value shall be a whole number greater than or equal to zero.

type= defines the type of data following the packet header, the following values are defined:

0 The data following the ESUB-XF Packet Header is ESUB-XF Payload Data.

1 The data following the ESUB-XF Packet Header is compressed binary 8-bit ESUB-XF Payload Data. The compression is done using the standard GZip Deflate algorithm.

Other “type=” values are reserved for other types of packets which may be transported in the future and the receiver shall ignore the content of all packets of unknown type and return the ESUB-XF Reply packet with the error text.

Example of a valid ESUB-XF Packet

```
<esub-xf,size=359,type=0>
<?xml version="1.0" encoding="UTF-8"?>
<esub-xf xmlns="urn:esub-xf" framerate="25" timebase="msec">
  <subtitlelist language="eng" langname="English" type="translation">
    <subtitle display="1000" clear="3500">
      <hregion>
        <line>First line of bottom justified text</line>
        <line>Second line of bottom justified text</line>
      </hregion>
    </subtitle>
  </subtitlelist>
</esub-xf>
```

5.2 Usage of the ESUB-XF Packet in ESUB-XF Transport Protocols

All ESUB-XF transport protocols transfer subtitles encapsulated in ESUB-XF packets as described earlier in this document. When the ESUB-XF packet is used in a transport protocol it is essential that:

- The ESUB-XF XML structure shall contain zero or more <subtitlelist> nodes but not more than one <subtitle> node shall be present in every <subtitlelist> node because only one subtitle for one language shall be transferred within one ESUB-XF packet.

- The ESUB-XF XML structure shall always contain all <subtitlelist> nodes for languages that are currently active and present in the subtitle stream. The receiver shall use the presence of <subtitlelist> nodes for signalization of presence of a subtitle language in the subtitle stream. If a <subtitlelist> node will be missing then the receiver shall consider that the language is not present and may react appropriately, for example by updating its internal list of present subtitle languages which is presented in the subtitle language selection menu.

- After receiving the ESUB-XF packet the receiver shall evaluate it and display the subtitle on the screen by considering the delay specified in the ‘dly’ keyword of the ESUB-XF packet header and the attribute ‘display’ shall not be used. The content of the ESUB-XF XML Structure shall be evaluated to see which <subtitlelist> nodes are present and thereby see which languages are present. Only languages that contain a <subtitle> node within the <subtitlelist> node shall be updated on the screen. If a <subtitle> node is not present in the <subtitlelist> node this means that the subtitle shall not be updated with new content on the screen as in this case the <subtitlelist> node is only used for signalization of presence of the language (KEEPALIVE).

- The receiver shall evaluate the attributes ‘display’ and ‘clear’ from the ESUB-XF XML structure, use the values to calculate the duration of the subtitle (clear-display) and delete the subtitle from the screen automatically after the duration has passed. When the ‘display’ or ‘clear’ parameters are missing, invalid or the duration is invalid or more than one minute, the duration of one minute shall be assumed.

- Subtitles with empty <subtitle> nodes shall be assumed as empty subtitles and can be used to delete a previously displayed subtitle before it is deleted automatically after the specified duration. If a new subtitle is received and shall be displayed before the duration of the

previously displayed subtitle is reached then the previously displayed subtitle on the screen shall be replaced at the display time of the new subtitle.

- To allow the receiver to detect the presence of subtitle data and to obtain the list of currently transmitted languages in short time at least one ESUB-XF packet can be transmitted. KEEPALIVE packets contain <subtitlelist> nodes for all languages but do not contain any <subtitle> nodes.

The ESUB-XF Packet Header used in ESUB-XF transport protocols uses additional keywords:

```
<esub-xf,size=xxx,type=yyy,cmd=c,dly=d,sid=s,ctr=c>
```

The keywords ‘cmd’, ‘sid’, ‘dly’ and ‘ctr’ shall only be evaluated when the type of the packet is 0, 1 or 2. Other types of packets may use different keywords.

The keyword ‘cmd’ defines the command and shall always be present. The only defined values are 0 which means DISPLAY SUBTITLE according to the rules as defined above and 1 which means END OF SUBTITLES (ESUB-XF Payload Data shall be ignored for value 1) and indicates that the sender with the ID defined in the keyword ‘sid’ will not transfer more subtitles. When ‘cmd’ is not present, the value 0 shall be assumed. The receiver shall ignore the complete ESUB-XF packet when the ‘cmd’ keyword is present and the value of the ‘cmd’ keyword is not 0 or 1 to allow future extensions of this protocol. The value of cmd may be alphanumeric and not just a simple number.

The keyword ‘dly’ is optional and defines the delay in milliseconds (a whole positive number or 0) that shall be applied after receiving the complete ESUB-XF packet and before displaying it on the screen.

No delay shall be applied and the subtitle shall be displayed immediately when ‘dly’ is not present. Generally all ESUB-XF packets shall be sent approximately 1 second before the subtitle shall be displayed on the screen to provide enough time to the receiver to render the subtitle. Only in live subtitling situations the delay may be 0.

sid= defines the ID of the source. The value shall be a string that shall not include spaces, ‘,’, ‘<’ and ‘>’. This keyword is optional and when not present the value is empty. The decoder shall evaluate the value on every received packet. When the value is different than the value from the previously received packet for the same channel the decoder shall delete all subtitles that are currently displayed on the screen for all languages before evaluating the packet. This value is used to inform the receiver that the video source has been changed (for example from movie to commercial) and therefore any subtitle that is on the screen shall be deleted.

ctr= contains a sequence counter for the source defined in ‘sid’. It is a value from 1 to 9999 and 0 is a special value. Every packet shall contain the value from the previous packet incremented by 1. 9999 shall be followed by 1. The value shall be ignored by display devices. It shall only be evaluated by monitoring equipment which shall display a warning when the received sequence counter value is not the one that was expected. The value 0 is used as the initial value after the start of the protocol and therefore a warning shall never be displayed when the value is 0. Consider that the value is valid only for packets of type 0, 1 and 2 because other types of packets may be transferred within the same stream. Only one counter is used for all types of packets (type 0, 1 and 2). If a type 0 packet is

transmitted with ctr value “n” and it is followed by a type 1 packet then the type 1 packet must have the ctr value “n+1” or “0”. The next packet must have the value “n+2” or “0”.

did= defines the ID of the destination. The value shall be a string that shall not include spaces, ‘,’, ‘<’ and ‘>’. This keyword is optional and may define the name of the channel, i.e. ZDF, ORF1. This keyword shall be ignored by the receiver. It may be used for filtering in special applications.

Example of a valid ESUB-XF Packet used in ESUB-XF transport protocol

```
<esub-xf,size=611,type=0,cmd=0,sid=1234,dly=1000,ctr=0>
<?xml version="1.0" encoding="UTF-8"?>
<esub-xf xmlns="urn:esub-xf" framerate="25" timebase="msec">
  <subtitlelist language="eng" langname="English" type="translation">
    <subtitle display="1000" clear="3500">
      <hregion>
        <line>First line of bottom justified text</line>
        <line>Second line of bottom justified text</line>
      </hregion>
    </subtitle>
  </subtitlelist>
  <subtitlelist language="fra" langname="French" type="translation">
    <subtitle display="1200" clear="3100">
      <hregion><line>Premiere ligne</line></hregion>
    </subtitle>
  </subtitlelist>
  <subtitlelist language="pol" langname="Polish" type="translation">
  </subtitlelist>
</esub-xf>
```

5.3 The ESUB-XF Generic Protocol

The ESUB-XF Generic Protocol can be used for transmission of subtitles on any kind of a serial connection. It consists of a series of ESUB-XF packets sent over the serial line without any response from the receiver.

The serial connection may be of any kind, for example a serial RS232/RS422 connection, a TCP socket connection, a virtual data stream within another data stream but is not limited to the mentioned connection types.

5.4 The ESUB-XF TCP Protocol

The ESUB-XF TCP Protocol can be used for transmission of subtitles using the TCP protocol.

The standard TCP port number that shall be used is 8120 but other port numbers can be used if required.

The system that sends the subtitles can establish a TCP connection to the system that will receive the subtitles or it can wait to receive a TCP connection from the receiving system. When the connection is closed the receiving system shall automatically delete all subtitles that are still displayed.

The data sent over the TCP connection is the ESUB-XF Packet for every subtitle.

The receiving system shall send a reply packet for every received ESUB-XF Packet latest within 100 milliseconds after reception. The transmitting system shall not send the next packet before receiving the reply packet or earliest after a timeout of 150 milliseconds.

The reply packet is also an ESUB-XF Packet with the same value of keywords as in the received ESUB-XF Packet and which contains all <subtitlelist> nodes from the received packet including all attributes and their values but it shall not contain any <subtitle> nodes. Every <subtitlelist> node shall contain a node <message> with the attribute:

`code="ok"`
This value defines the result of the reception of the subtitle for the selected language. The value shall be one of the following: "ok" or "error" but also any other value is permitted, i.e. "warning". When this attribute is not specified or when the <message> node is not present the value "ok" shall be assumed.

The text for the node <message> shall contain descriptive text for the "ok" or "error" from the 'code' attribute. It can be any text without any formatting information. All line breaks within the text in the <message> nodes shall be treated as spaces and not treated as text line breaks and <br /> shall not be used. Leading and trailing spaces of every line in the XML structure shall be ignored.

The same XML Schema as for the ESUB-XF Request shall be used:

http://www.fab-online.com/schemas/esub-xf-1.0.xsd

Example of a valid ESUB-XF Reply Packet

An example of a valid TCP reply packet is:

```
<esub-xf,size=434,type=0,cmd=0,sid=1234,dly=1000,ctr=0>
<?xml version="1.0" encoding="UTF-8"?>
<esub-xf xmlns="urn:esub-xf" framerate="25" timebase="smpte">
  <subtitlelist language="eng" langname="English" type="translation">
    <message code="ok" />
  </subtitlelist>
  <subtitlelist language="fra" langname="French" type="translation">
    <message code="error">Language fra is not supported</response>
  </subtitlelist>
  <subtitlelist language="pol" langname="Polish" type="translation">
    <message code="ok" />
  </subtitlelist>
</esub-xf>
```

In case that the receiver wishes to report to the sender that the ESUB-XF packet could not be processed, because it is of type which is not supported by the receiver, only the ESUB-XF header must be sent as reply where the reply must be "error" and replytext may be any text without spaces:

```
<esub-xf,size=0,reply=error,replytext=NotImplemented>
```

User Authentication

The system receiving subtitles over the ESUB-XF TCP protocol may require user authentication. In such case the sender must provide a username and password before ESUB-XF packets will be accepted by the receiver.

When the sender sends a request and the TCP session has not been authenticated yet, the server shall respond with a special header containing the keyword 'auth' (authentication challenge). The authentication challenge shall be a random string consisting of letters and numbers in UTF-8 representation, but it shall not include the characters , < and >. This header shall have 'size' set to zero and shall not contain the ESUB-XF XML structure.

```
<esub-xf,size=0,type=0,cmd=0,sid=1234,dly=1000,ctr=0,auth=0FA0D134E15A6EF8>
```

The client shall resend the request with incremented ctr value and with two additional keywords:

user= Contains the username as a string of letters and numbers in UTF-8 representation, but it shall not include the characters , < and >.

pass= Contains the password hash calculated by the following algorithm:

- The password is converted to the UTF8 representation (without the byte order mark).
- The challenge is appended to the UTF8 representation of the password.
- The resulting string is packed into 32-bit words, byte by byte (see the example below). The remaining space is filled with the character 0x00.
- The resulting memory block is passed to the SHA-256 message digest algorithm.
- The result is encoded as a hexadecimal string as the value of the 'pass' keyword. Both capital and lower case letters shall be permitted in the hexadecimal string.

Example: If the password is 'passwd-12%' and challenge is as shown in the example above, their combination shall be stored in the memory as shown by the table below.

	Bit 31			Bit 0
Word 0	's'	's'	'a'	'p'
Word 1	'1'	'-'	'd'	'w'
Word 2	'F'	'0'	'%'	'2'
Word 3	'1'	'D'	'0'	'A'
Word 4	'1'	'E'	'4'	'3'
Word 5	'E'	'6'	'A'	'5'
Word 6	0x00	0x00	'8'	'F'

The client would repeat the same request as before but with added 'user' and 'pass' keywords:

```
<esub-xf,size=403,type=0,cmd=0,sid=1234,ctr=1,user=testusername,pass=08ED5A4983BA143CA1BF5C
7A5495D33BBB6D09EAADC359FF6DFB38639F2A3C7E><?xml…>
```

The response in this example can be recalculated by the following JavaScript fragment:

```
<script src="https://cdnjs.cloudflare.com/ajax/libs/crypto-js/3.1.2/components/sha256.js">
</script>
<script>
    var hash = CryptoJS.SHA256(unescape(encodeURIComponent("passwd-12%")) +
      "0FA0D134E15A6EF8").toString();
    alert(hash);
</script>
```

The functions unescape and encodeURIComponent are used to convert the password to UTF-8.

5.5 The ESUB-XF REST Protocol

The ESUB-XF REST protocol can be used for the transfer of a complete subtitle file or for the transfer of a single subtitle for live subtitling. The following rules shall be followed when using the ESUB-XF REST Protocol:

- Two URLs are defined by the system which accepts HTTP Post requests with subtitle data. One is for the DISPLAY SUBTITLE (cmd=0) and another one is for END OF SUBTITLES (cmd=1)
- The ESUB-XF Packet Header Keywords (without size and type) shall be included in the URL
- The same ESUB-XF Payload Data is included in the Body of the HTTP Request and Reply as defined in the ESUB-XF TCP Protocol
- The ESUB-XF XML or JSON Data is included in the body of the HTTP POST request without the ESUB-XF Packet Header
- When transferring JSON Data instead of XML in the body of the HTTP request, then names of XML attributes shall be preceded by @. Names of terminal XML #text item types like the <file> node within the <filelist> node shall be preceded by #text.
- To delete a subtitle send an ESUB-XF structure without a region (empty subtitle).

Example of a valid ESUB-XF XML HTTP POST request

URL Examples:

To display/transmit a subtitle:

http://ipaddress/livesubtitling/esub-xf/displaysubtitle?ctr=1&sid=PCNAME

To signalize that the client will not send any more subtitles:

http://ipaddress/livesubtitling/esub-xf/endofsubtitles?ctr=1&sid=PCNAME

The keyword 'ctr' shall be omitted or it shall include an incrementing counter as described earlier. The keyword 'cmd' shall not be used because different URLs are used for different cmd values. The keyword 'sid' shall contain a unique name of the sending client, i.e. the computer name.

Body example when using Content-Type: application/xml; charset=UTF-8 in the HTTP Header:

```
<?xml version="1.0" encoding="UTF-8"?>
<esub-xf xmlns="urn:esub-xf" framerate="25" timebase="msec">
  <subtitlelist language="eng" langname="English" type="translation">
    <subtitle display="1000" clear="3500">
      <hregion>
        <line alignment="center">First line of bottom justified text</line>
        <line alignment="center">Second line of bottom justified text</line>
      </hregion>
      <image format="png" width="1920" height="1080" filename="s1.png" />
    </subtitle>
  </subtitlelist>
  <filelist>
    <file filename="s1.png">
iVBORw0KGgoAAAANSUhEUgAAAAoAAAAKCAIAAAACUFjqAAAAK3RFWHRDcmVhdGlvbiBUaW1lAE1p
IDEyIEZlYiAyMDIwIDE2OjMxOjEyICswMTAw8LnA5QAAAAd0SU1FB+QCDA8fIxV2Jz0AAAAJcEhZ
cwAACvAAAArwAUKsNJgAAAAEZ0FNQQAAsY8L/GEFAAAAFUlEQVR42mP8//8/A27AxIAXjFRpAKXj
AxFkRbxkAAAAAElFTkSuQmCC
    </file>
  </filelist>
</esub-xf>
```

Body example when using Content-Type: application/json; charset=UTF-8 in the HTTP Header:


```
{
   "@framerate": "25",
   "@timebase": "msec",
   "subtitlelist": {
      "@language": "eng",
      "@langname": "English",
      "subtitle": {
         "@display": "1000",
         "@clear": "3500",
         "hregion": {
            "line": [
               {
                  "@alignment": "center",
                  "#text": "First line of bottom justified text"
               },
               {
                  "@alignment": "center",
                  "#text": "Second line of bottom justified text"
               }
            ]
         },
         "image": {
            "@format": "png",
            "@width": "1920",
            "@height": "1080",
            "@filename": "s1.png"
         }
      }
   },
   "filelist": {
      "file": {
         "@filename": "s1.png",
         "#text":
"iVBORw0KGgoAAAANSUhEUgAAAAoAAAAKCAIAAAACUFjqAAAAK3RFWHRDcmVhdGlvbiBUaW1lAE1p\nIDEyIEZlYiAy
MDIwIDE2OjMxOjEyICswMTAw8LnA5QAAAAd0SU1FB+QCDA8fIxV2Jz0AAAAJcEhZ\ncwAACvAAAArwAUKsNJgAAAAEZ
0FNQQAAsY8L/GEFAAAAFUlEQVR42mP8//8/A27AxIAXjFRpAKXj\nAxFkRbxkAAAAAElFTkSuQmCC"
      }
   }
}
```


Note that the “image” and “file” is included in above examples only to explain the possibility to transfer an image and a file. In normal usage scenarios the “image” and the “file” will not be included. Also note that the content of the “line” is converted in a special way.

Every client that will stop sending subtitles must send a HTTP POST to the URL /endofsubtitles to signalize that no more subtitles are expected for a certain time (for example at the end of the program):

http://ipaddress/api/livesubtitling/esub-xf/endofsubtitles?ctr=1&sid=PCNAME

5.6 The ESUB-XF Streaming Protocol

In subtitling applications like video streaming over the internet subtitles will be transported to the display device which is reproducing the video and must be displayed synchronized to the video stream. The transport protocol used is irrelevant because live video streaming may use many different transport protocols like HLS, RTMP and others and subtitles may be transported within different protocols as well (HTTP download of ESUB-XF .esub files, ESUB-XF Generic Protocol, …). It is however important to define how subtitle timecodes from the ESUB-XF structure correspond to Presentation Time Stamps (PTS=Presentation Time Stamp) within the video stream.

The ESUB-XF Streaming Protocol defines an extension to the ESUB-XF structure so that the display device can easily read the time stamp values for display and deletion of the subtitle on the screen. The following shall be respected:

- A special attribute ‘streaming:basepts’ in the <esub-xf> node specifies the PTS value of the video stream which corresponds to the ‘display’ and ‘clear’ timecode value "0" from the ESUB-XF structure.

The display device shall evaluate the value of the attribute streaming:basepts and add the ‘display’ timecode value to calculate the PTS of the video frame on which the subtitle shall be displayed. The same calculation shall be applied to the ‘clear’ attribute to calculate the PTS of the video frame on which the subtitle shall be cleared from the screen. The subtitle shall only be displayed if the calculated ‘display’ PTS value corresponds to the PTS value of the video stream. In systems where PTS values are limited to a certain interval range the calculation shall wrap around the value if necessary to make sure that the resulting value is within the permitted interval.

Example of a valid ESUB-XF Structure used in ESUB-XF Streaming Protocol

```
<?xml version="1.0" encoding="UTF-8"?>
<esub-xf xmlns="urn:esub-xf" xmlns:streaming="urn:esub-xf-streaming" framerate="25"
timebase="msec" streaming:basepts="12345678">
  <subtitlelist language="eng" langname="English" type="translation">
    <subtitle display="1000" clear="3500">
      <hregion>
        <line>First line of bottom justified text</line>
        <line>Second line of bottom justified text</line>
      </hregion>
    </subtitle>
  </subtitlelist>
  <subtitlelist language="fra" langname="French" type="translation">
    <subtitle display="1200" clear="3100">
      <hregion><line>Première ligne</line></hregion>
    </subtitle>
  </subtitlelist>
  <subtitlelist language="pol" langname="Polish" type="translation">
  </subtitlelist>
</esub-xf>
```

The XML Schema for the ‘streaming’ namespace is available for download here:

http://www.fab-online.com/schemas/esub-xf-streaming-1.0.xsd

5.7 The ESUB-XF Timed Text Compatibility Protocol

Both ESUB-XF and Timed Text files are XML files that contain subtitles. Any protocol which is defined for transport of Timed Text files can be used also for transport of ESUB-XF files. Instead of embedding Timed Text XML content simply embed ESUB-XF XML content. The receiver will notice that the content is not Timed Text and will ignore it as long as it does not support ESUB-XF XML content. Receivers that support ESUB-XF content will automatically detect whether they have to deal with timed text or with ESUB-XF content.

5.8 Examples of ESUB-XF Structures for Live Subtitling

The following examples contain ESUB-XF XML Structures that can be used for different functionality in live subtitling when using any ESUB-XF Transport Protocol. In Live Subtitling applications only one subtitle shall be transferred within the ESUB-XF XML Structure and display and clear attributes of the <subtitle> node shall not be present. It is also important that the keyword 'sid' contains a unique name of the sender (i.e. a combination of the computer name and session id).

To transmit a live subtitle send a request with cmd=0 (DISPLAY SUBTITLE):

```
<?xml version="1.0" encoding="UTF-8"?>
<esub-xf xmlns="urn:esub-xf" framerate="25" timebase="msec">
  <subtitlelist language="eng" langname="English" type="translation">
    <subtitle>
      <hregion>
        <line>First line of bottom justified text</line>
        <line>Second line of bottom justified text</line>
      </hregion>
    </subtitle>
  </subtitlelist>
</esub-xf>
```

To delete a subtitle from the screen (transmit an empty subtitle) send a request with cmd=0:

```
<?xml version="1.0" encoding="UTF-8"?>
<esub-xf xmlns="urn:esub-xf" framerate="25" timebase="msec">
  <subtitlelist language="eng" langname="English" type="translation">
    <subtitle>
    </subtitle>
  </subtitlelist>
</esub-xf>
```

To signalize that the sender will not send more subtitles send a request with cmd=1 (END OF SUBTITLES) and do not include any ESUB-XF Payload Data because it will be ignored.

To transmit a live subtitle in word by word mode by only sending additional words send a request with cmd=0:

```
<?xml version="1.0" encoding="UTF-8"?>
<esub-xf xmlns="urn:esub-xf" framerate="25" timebase="msec">
  <subtitlelist language="eng" langname="English" type="translation">
    <subtitle scrolllines="3">
      <hregion>
        <line>some more words.</line>
      </hregion>
    </subtitle>
  </subtitlelist>
</esub-xf>
```

To delete last 3 words from the subtitle which is currently on the screen before displaying new words add the attribute replacelastwords="3" to the <subtitle> node.

6. Checklist for Implementation

Systems claiming to support ESUB-XF shall respect these rules:

- The system shall only implement the ESUB-XF transport protocols which it requires. Support for all transport protocols is not a mandatory requirement.

- The ESUB-XF specification does not define the font type or size for display of subtitles. The display device shall automatically choose a font size that fits the screen size so that 36 characters can always be displayed in every line and that 12 lines of text can be displayed. The display device may give the user the possibility to select a different font size. When the font is so large that not all characters of one line can be displayed on the screen it shall automatically reduce the font size and keep using the smaller font for following subtitles.

- The display device may allow the selection of font face and size to the viewer so that the viewer can choose a larger or smaller font size.

- The default font shall not be a “Serif” font. A font similar to “Arial” shall be used by default.

- When a new subtitle shall be displayed on the screen before the previous subtitle was cleared the new subtitle shall directly replace the previously displayed subtitle without any flashing. This mode of transmission may be used for subtitling of live events where every subtitle may contain a new word or line of text. The viewer will generally only notice that a new word or line of text will appear because other lines of text will be displayed on exactly the same positions as in the previous subtitle. There shall be no flashing or flickering of text visible because the previous subtitle shall be replaced directly by the new subtitle on the screen without any pause in between.

- Support for vertically oriented regions is not mandatory. Subtitles with vertically oriented regions shall be displayed as empty subtitles by devices that do not support them.

- Support for the ‘scrolllines’ attribute in the <subtitle> node is not mandatory. Subtitles with this attribute shall only be displayed by display devices that support this functionality.

- Support for text to speech audio description is not mandatory. When implemented, the text shall be obtained from the <line> nodes and speech reproduction shall start exactly at the time specified by ‘display’ and terminate before the time specified by the ‘clear’ attribute.

- Support for all other functionality for rendering of the ESUB-XF XML structure to display the subtitle including the rendering of the <image> node is mandatory for all display devices.

- The display device shall allow the user to select the default subtitling and audio description:
    - The subtitles that shall be displayed: none, translation, hard of hearing.
    - The language of subtitles that shall be displayed.
    - The language of text to speech audio description that shall be reproduced.

7. Bibliography

ETSI EN 300 706 V 1.2.1 (2003-04)	Enhanced Teletext Specification
ETSI EN 300 743 V 1.3.1 (2006-11)	Digital Video Broadcasting (DVB): Subtitling Systems
ETSI EN 300 775	Digital Video Broadcasting (DVB): Specification of the carriage of Vertical Blanking Information (VBI) data in DVB bitstreams
CEA 608-E	Line 21 Data Services
CEA-708-D	Digital Television (DTV) Closed Captioning
EBU Tech 3264	Specification of the EBU subtitling data exchange format
EBU Tech 3350	EBU-TT Part 1 Subtitling format definition
SMPTE ST 2052-1:2010	Timed Text Format (SMPTE-TT)
WebVTT	The Web Video Text Tracks Format – W3C Community Group 13 May 2013
SMPTE ST 291-2011	Television – Ancillary Data Packet and Space Formatting
SMPTE RP 214-2002	Packing KLV Encoded Metadata and Data Essence into SMPTE 291M Ancillary Data Packets
SMPTE ST 376-2003	Material Exchange Format (MXF) – File Format Specification
SMPTE ST 2075	Mapping EBU TECH 3264 (STL) into the MXF Generic Stream Container
EBU R 133	EBU Recommendation 133 – Transport of subtitles using MXF in an IT-Based Television Production Environment
DC System Specification	Digital Cinema System Specification Version 1.2 30 August 2012
W3C PNG Specification	W3C Portable Network Graphics (PNG) Specification (Second Edition) - Information technology — Computer graphics and image processing — Portable Network Graphics (PNG): Functional specification. ISO/IEC 15948:2003 (E) - W3C Recommendation 10 November 2003

8. Revision History

Version 0.8, 2013-09-20 by F.A. Bernhardt GmbH, FAB / Miha Sokolov
- Initial Release

Version 0.81, 2014-01-24 by F.A. Bernhardt GmbH, FAB / Miha Sokolov
- 2.5: The attribute ‘backcolor’ is defined for the in the <subtitle> node for compatibility with teletext subtitles

Version 0.82, 2014-04-09 by F.A. Bernhardt GmbH, FAB / Miha Sokolov
- 5.1: Compressed data in the ESUB-XF Packet may be transported within multiple packets

Version 0.83, 2014-06-22 by F.A. Bernhardt GmbH, FAB / Miha Sokolov
- 2.3 Clarification of display time of subtitle when ‘display’ is lower or equal to previous ‘clear’
- 2.5: Number of characters in one line is limited to 36 for compatibility with teletext subtitles

Version 0.84, 2014-11-24 by F.A. Bernhardt GmbH, FAB / Miha Sokolov
- 2.3 The attribute ‘appearance’ with the value “nontransparentbox” is defined

Version 0.85, 2015-03-13 by F.A. Bernhardt GmbH, FAB / Miha Sokolov
- 2.5 The node <split /> has been added

Version 0.86, 2015-04-23 by F.A. Bernhardt GmbH, FAB / Miha Sokolov
- 2.4 A maximum of two regions can be present in a subtitle

Version 0.87, 2015-07-01 by F.A. Bernhardt GmbH, FAB / Miha Sokolov
- 5.1 Packets of type=2 are used for encrypted data transfer within the ESUB-XF TCP Protocol

Version 0.88, 2015-07-16 by F.A. Bernhardt GmbH, FAB / Miha Sokolov
- 2.5 The attributes ‘language’ and ‘langname’ are introduced for ttsaudiodescription

Version 0.89, 2015-08-05 by F.A. Bernhardt GmbH, FAB / Miha Sokolov
- 5.2 The keyword ‘did’ is introduced for the destination channel name

Version 0.90, 2016-07-13 by F.A. Bernhardt GmbH, FAB / Miha Sokolov
- 2.3 & 6 Support for the attribute ‘scrolllines’ and vertical regions is not mandatory
- 5.6 Timestamps in transport streams are redefined

Version 0.91, 2016-09-05 by F.A. Bernhardt GmbH, FAB / Miha Sokolov
- 2.2 Attribute values "simplified" and "forced" removed in the attribute type.

Version 0.92, 2017-02-09 by F.A. Bernhardt GmbH, FAB / Miha Sokolov
- 2.1 Added the attribute ‘start’
- 2.5 Usage of within <line> nodes is redefined

Version 0.93, 2017-08-21 by F.A. Bernhardt GmbH, FAB / Miha Sokolov
- 5.4 ESUB-XF SDI Protocol was renamed to ESUB-XF VANC Protocol

Version 0.94, 2017-11-06 by F.A. Bernhardt GmbH, FAB / Miha Sokolov
- 2.4 Second region is not allowed anymore for better compatibility with other formats

Version 1.0, 2018-08-29 by F.A. Bernhardt GmbH, FAB / Miha Sokolov

- Added hint that hex color values shall be redefined for HDR because FFFFFF will be too bright

Version 1.01, 2018-12-03 by F.A. Bernhardt GmbH, FAB / Miha Sokolov

- 2.9 The attribute “replacelastwords” can be used for animated display with soft scroll

Version 1.02, 2020-03-03 by F.A. Bernhardt GmbH, FAB / Miha Sokolov

- 2.1 The attribute “generator” can be used within the <esub-xf> node
- 2.3 The attribute “appearance” can have multiple values
- 2.5 The attributes “font” and “marker” can be used within the <line and nodes
- 2.6 Transport of BASE64 encoded bitmaps within the <image> moved to <filelist>
- 5.1 Transport of files is added in 2.7 <filelist> node
- 5.4 The standard reply header is defined for error reporting of the receiver to the sender and the content of the reply is redefined
- 5.5 and 5.6 VANC and DVB protocol descriptions are not included anymore
- 5.5 The ESUB-XF REST Protocol is included
- 5.7 The ESUB-XF Timed Text Compatibility Protocol is included
- 5.8 Examples of ESUB-XF Structures for Live Subtitling are included

Version 1.03, 2020-06-19 by F.A. Bernhardt GmbH, FAB / Miha Sokolov

- 2.2 The attribute language allows both ISO 639-2/T and ISO 639-2/B
- 5.5 The ESUB-XF REST Protocol uses different URLs for different values of “cmd”