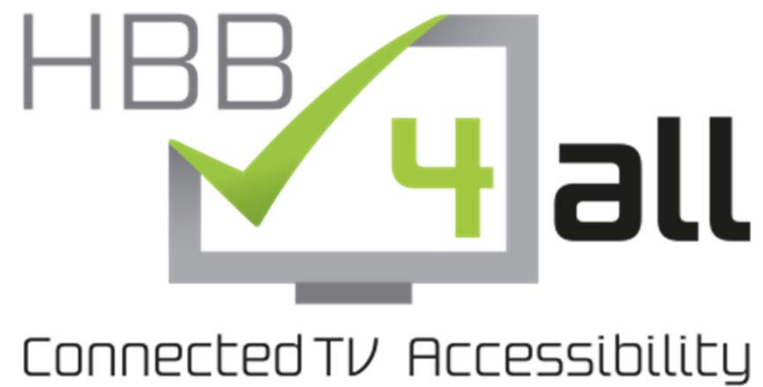


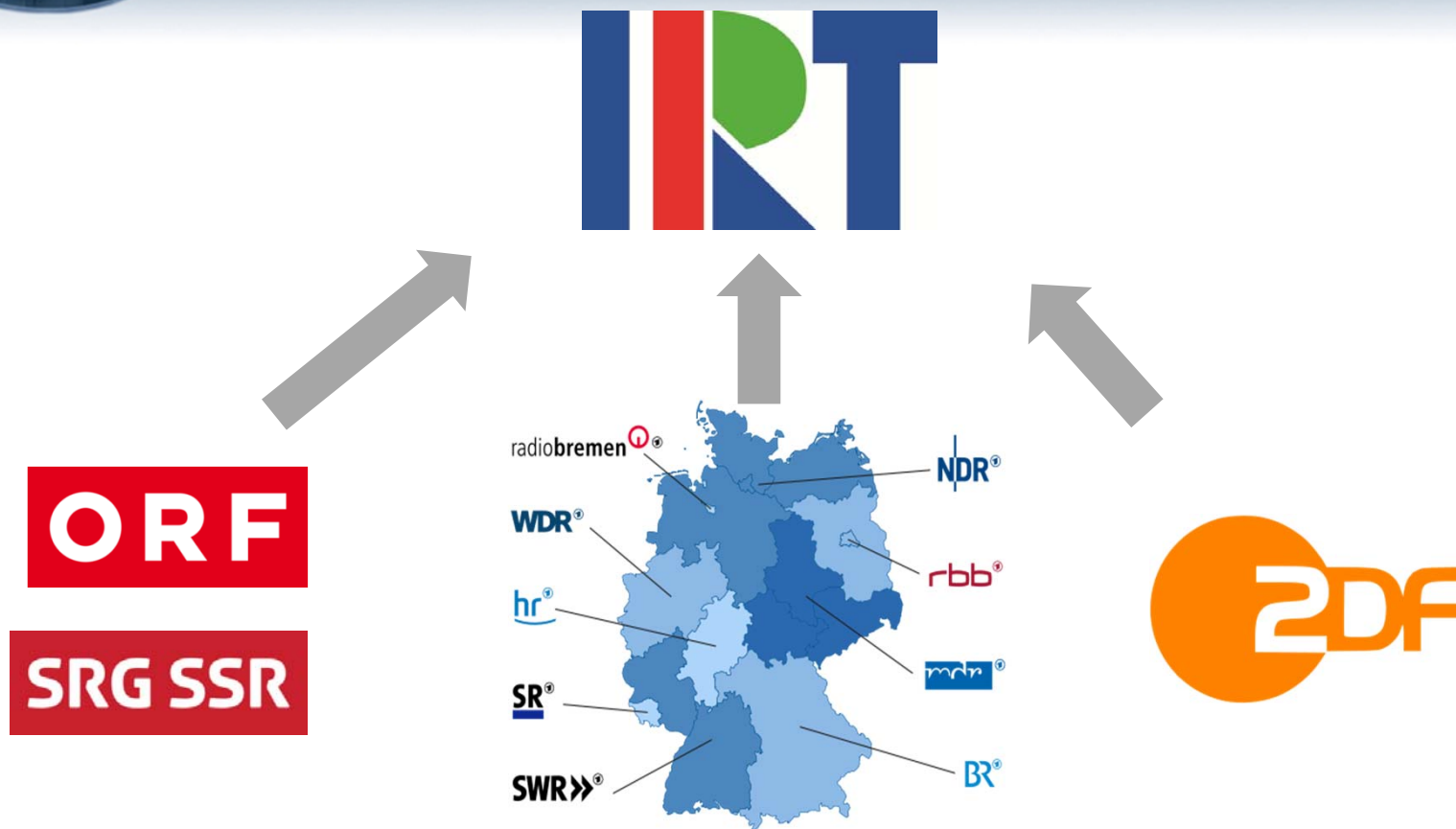
Open source tools for new subtitle standards

Fosdem¹⁵

Andreas Tai
Production Systems Television



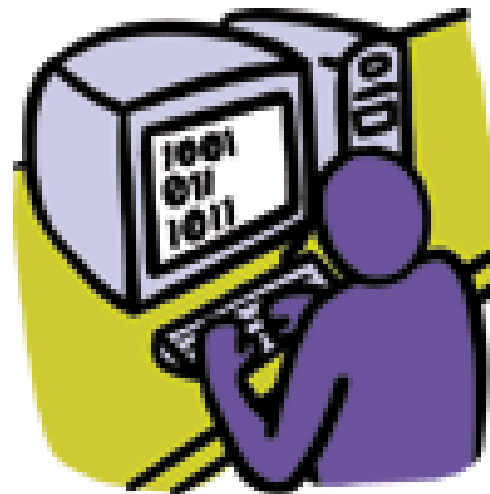
2 Academic Institutions	 Universitat Autònoma de Barcelona  
4 Public Broadcasters	 RUNDFUNK BERLIN-BRANDENBURG  RADIOTELEVISÃO PORTUGUESA  Corporació Catalana de Mitjans Audiovisuals, SA  
2 Research Institutes	 Institut für Rundfunktechnik  IK4 Research Alliance
4 SMEs	  PEOPLE'S PLAYGROUND   Holken Consultants & Partners



Standards

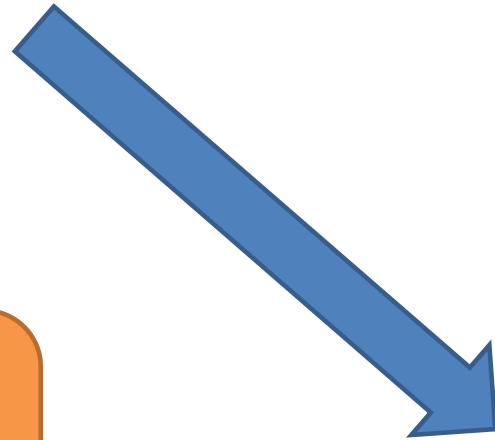


(Open Source) Code

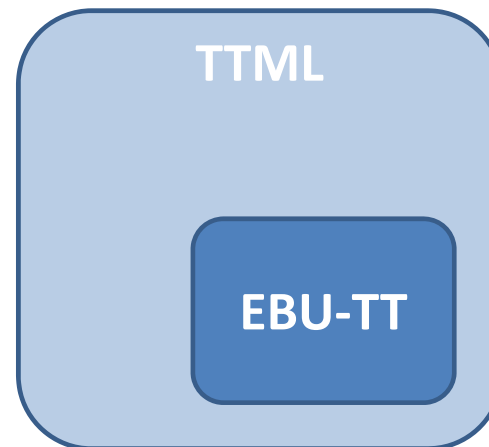


Different Subtitle Standards

EBU



EBU-STL



TTML

EBU-TT



WebVTT



Use case format conversion

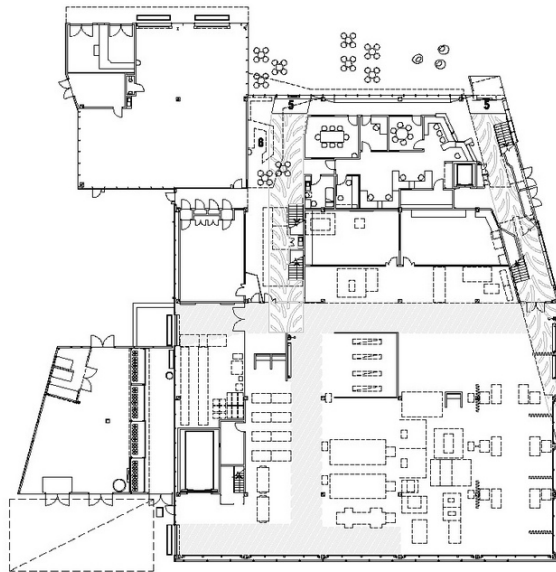
EBU-STL, binary based

[illegible]

EBU-TT XML Based

```
<tt:div style="defaultStyle" xml:id="sgn1">
  <tt:p xml:id="sub1a" region="DoubleHeightRow20" begin="00:00:00:00" end="00:00:08:00">
    <tt:span style="style1">These people are amongst the <tt:br/>greatest quiz players in Britain.</tt:span>
  </tt:p>
</tt:div>
```

Why reference implementation are so important....



Lyons Architects

**A standard is only an architectural plan
... not the house!**

Why open source?

Commercial Software



Customer requests

Open source software



Attribution: lewing@isc.tamu.edu

Standard conformance
higher priority ... possible

Subtitle Conversion Framework - SCF

IRT-Open-Source / **scf** Unwatch 8 Unstar 5 Fork 0

Subtitling Conversion Framework — Edit

28 commits 1 branch 2 releases 1 contributor

branch: master **scf** / +

UPDATE DOCUMENTATION + DELETE TEST FILES ...

tairt authored 3 days ago latest commit f8b5fd6736

documentation	ADD AND CHANGE DOCUMENTATION	a month ago
modules	UPDATE DOCUMENTATION + DELETE TEST FILES	3 days ago
CHANGELOG.txt	ADD AND CHANGE DOCUMENTATION	a month ago
LICENSE	Initial commit	5 months ago
NOTICE	ADD AND CHANGE DOCUMENTATION	a month ago
README.md	UPDATE DOCUMENTATION + DELETE TEST FILES	3 days ago

README.md

Code

Issues 0

Pull Requests 0

Pulse

Graphs

Settings

HTTPS clone URL

<https://github.com/IRT-Open-Source/scf>

You can clone with [HTTPS](#), [SSH](#), or [Subversion](#).

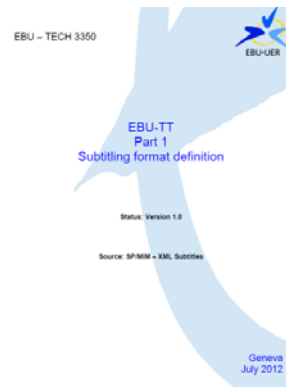
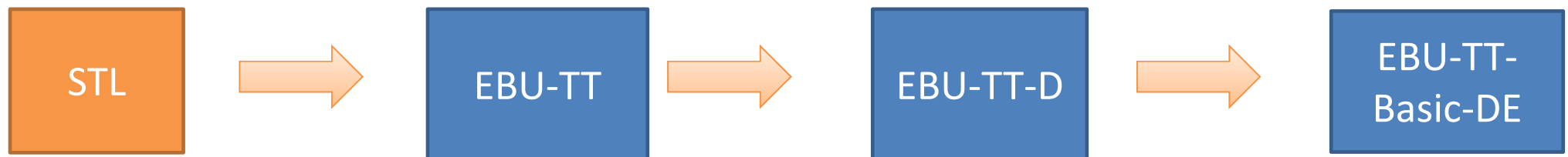
Clone in Desktop

<https://github.com/IRT-Open-Source/scf>

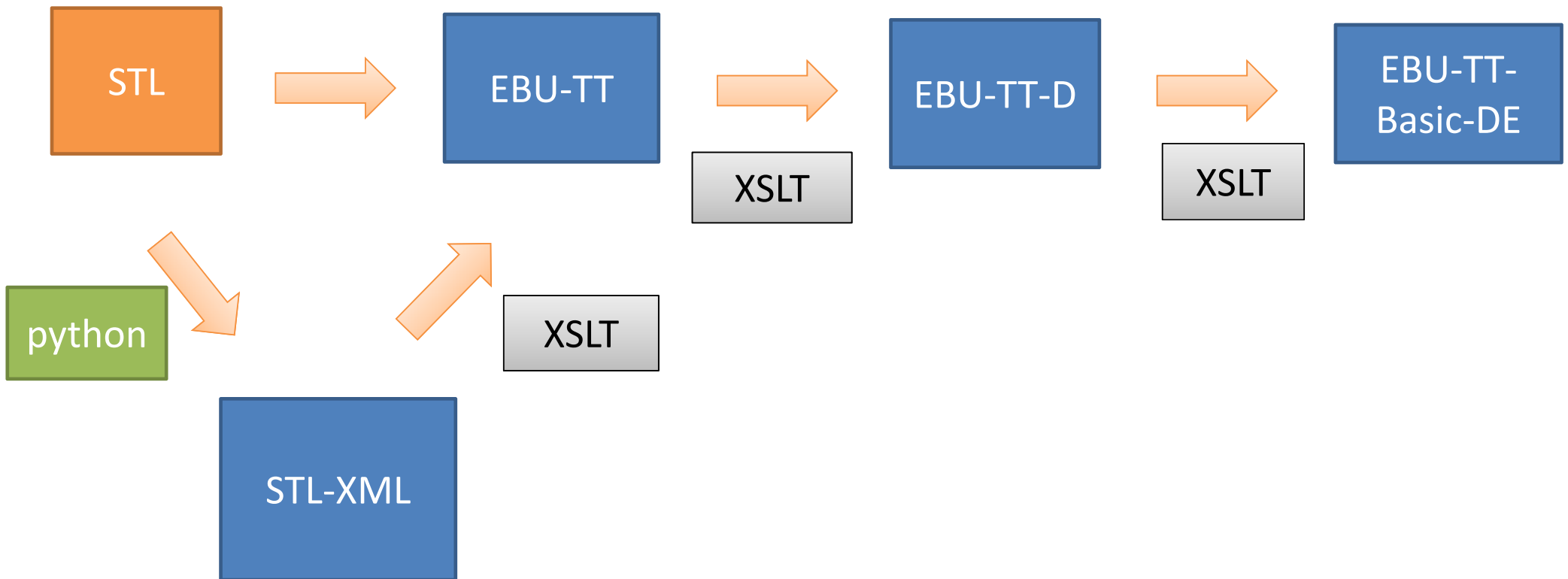
SCF - „Non Functional Requirements“

- Integration
 - different platforms
 - different programming languages
- Transparency
 - requirements
 - source code

SCF – Functional Requirements



SCF – Implementation



Simple Pipeline

```
python stl2stlxml.py %1 > STLXML.xml
```

```
java -jar saxon9he.jar -s:STLXML.xml -xsl:STLXML2EBU-TT.xslt -o:EBU-TT.xml
```

```
java -jar saxon9he.jar -s:EBU-TT.xml -xsl:EBU-TT2EBU-TT-D.xslt offsetInSeconds=36000 o:EBU-TT-D.xml
```

```
java -jar saxon9he.jar -s:EBU-TT-D.xml -xsl:EBU-TT-D2EBU-TT-D-Basic-DE.xslt o:EBU-TT-D-Basic-DE.xml
```

- Python 2.7 (!)
- XSLT 1 Processor (+ EXSLT nodeset func)
 - Saxon from 6.5.5
 - libxslt
 - Xalan
 - Web Browser



XProc: An XML Pipeline Language

W3C Recommendation 11 May 2010

This Version:

<http://www.w3.org/TR/2010/REC-xproc-20100511/>

Latest Version:

<http://www.w3.org/TR/xproc/>

Previous versions:

<http://www.w3.org/TR/2010/PR-xproc-20100309/>

<http://www.w3.org/TR/2010/WD-xproc-20100105/>

<http://www.w3.org/TR/2009/CR-xproc-20090528/>

Editors:

Norman Walsh, MarkLogic Corporation <norman.walsh@marklogic.com>

Alex Milowski, Invited expert <alex@milowski.org>

Henry S. Thompson, University of Edinburgh <ht@inf.ed.ac.uk>

Please refer to the [errata](#) for this document, which may include some

See also [translations](#).

This document is also available in these non-normative formats: [XML](#)

Copyright © 2010 W3C® (MIT, ERCIM, Keio), All Rights Reserved. W3C [liability](#), [trade](#)


[Home](#)
[Documentation](#)
[Extensions](#)
[Download](#)
[Overview](#)
[Contact us](#)

Welcome to Calabash

XML Calabash is an implementation of [XProc: An XML Pipeline Language](#).

See the XML Calabash [project status page](#) for more details.

You can [download](#) Calabash and/or read the (very little bit of) [documentation](#). Calabash also ships with the [<ox](#)

There is a [reference guide](#) to *XML Calabash*.

Announcing...

XML Calabash version 1.0.23 was released on 16 December 2014!

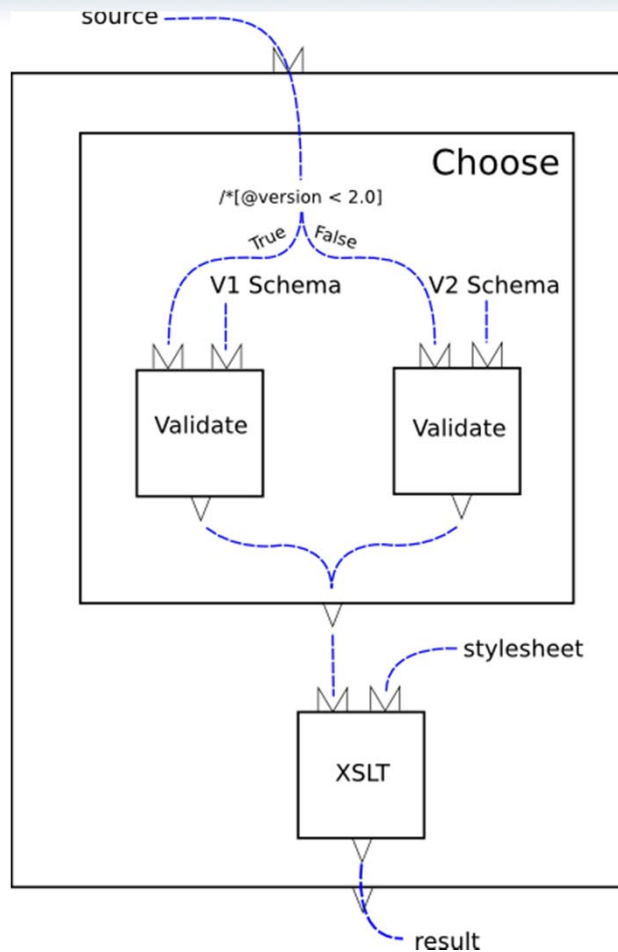
Building up pipeline using XPROC

```
<input port=„source">  
  <document href=„StlXml.xml"/>  
</input>
```

```
<xslt>  
  <input port="stylesheet">  
    <document href="STLXML2EBU-TT.xslt"/>  
  </input>  
</xslt>
```

```
<xslt>  
  <input port="stylesheet">  
    <document href="EBU-TT-D2EBU-TT-D-Basic-DE.xslt"/>  
  </input>  
</xslt>
```

XPROC pipeline



From XProc: An XML Pipeline Language
Copyright © 11.10.2010
World Wide Web Consortium

Documentation SCF requirements

s

requirement-339: ttp:timeBase attribute mapping

Description

The EBU-TT-D file's tt:tt root element has exactly one ttp:timeBase attribute with a value of type xs:string set to "media" according to EBU Tech 3380 (p.13).

Area

EBU-TT2EBU-TT-D

Requirement Review Status

accepted

Status Implementation

reviewed

Priority according to MoSCoW

m

SCF – Requirements covered by tests

IRT-Open-Source / **scf** Unwa

branch: **master** **scf** / modules / EBU-TT2EBU-TT-D / tests / **test_files** / +

ADD TESTS STLXML2EBU-TT AND EBU-TT2EBU-TT-D ...

fichte authored on 28 Oct 2014
→ tairt committed on 14 Nov 2014

..	
requirement-0339-001.xml	Initial commit
requirement-0340-001.xml	Initial commit
requirement-0340-002.xml	Initial commit
requirement-0341-001.xml	Initial commit
requirement-0350-001.xml	Initial commit
requirement-0351-001.xml	Initial commit
requirement-0351-002.xml	Initial commit
requirement-0351-003.xml	Initial commit
requirement-0351-004.xml	Initial commit
requirement-0351-005.xml	Initial commit

SCF Licence

- Apache Licence, Version 2.0
- Allows: integration in commercial software
- Goal:
 - standard compliance in open source **and** commercial software

Big thanks to Yann Coupin

yanncoupin / stl2srt

Watch 5

A python command-line utility to convert EBU STL and TT (W3C, SMPTE and EBU variants) subtitles files into the simpler SRT format

21 commits 1 branch 12 releases 1 contributor

branch: master stl2srt / +

Switch to Apache License, Version 2.0

yanncoupin authored on 22 Sep 2014 latest commit 9ce2584242

LICENSE	Switch to Apache License, Version 2.0	4 months ago
README.md	Switch to Apache License, Version 2.0	4 months ago
to_srt.py	Switch to Apache License, Version 2.0	4 months ago

README.md

to_srt.py

A command line utility written in Python to convert some variants of subtitles in to the much simpler SRT file format with optional formatting. SRT also has the benefit of being trivial to convert to WebVTT, the W3C sanctioned subtitle format. Supported input file formats are:

Stl2stlxml module is

derived from

stl2srt from Yann Coupin!



More.....

Subtitles in ISOBMFF



The screenshot shows the GPAC Multimedia Open Source Project website. The header features the GPAC logo and the tagline 'Multimedia Open Source Project'. Below the header is a navigation bar with links: 'What is GPAC?', 'Downloads', 'MP4Box', 'Osmo4', 'DashCast', 'Authoring', and 'Jobs'. A search icon is also present. Below the navigation bar, there is a section titled 'Subtitling with GPAC' with a date of '4 September 2014', tags 'GPAC, MP4, Subtitle', and the author 'cconcolato'. The text of the post begins: 'This post is the first post in a series I'm writing to help you discover the many different ways to handle tracks in MP4 files using MP4Box and other GPAC tools, with a particular focus on three types of tracks: subtitles, metadata and graphics tracks. Let me start in this post with subtitle tracks.'

- Mux ISOBMFF
- DASH Segmentation
- Different Subtitle formats
 - WebVTT, TTML/EBU-TT, 3GPP Timed Text
- Standard Compliance
- ISO/IEC 14496-30

WebVTT parser and validator (Anne van Kesteren)

Live [WebVTT](#) Validator

WEBVTT

```
00:11.000 --> 00:13.000 vertical:rl
<v Roger Bingham>We are in New York City

00:13.000 --> 00:16.000
<v Roger Bingham>We're actually at the Lucern Hotel, just down the street

00:16.000 --> 00:18.000
<v Roger Bingham>from the American Museum of Natural History

00:18.000 --> 00:20.000
<v Roger Bingham>And with me is Neil deGrasse Tyson

00:20.000 --> 00:22.000
<v Roger Bingham>Astrophysicist, Director of the Hayden Planetarium
```

Kind:

This is boring, your WebVTT is valid! (8ms)

[Source code](#) available under [CC0](#).

TTML Validator

GitHub

This repository Search

Explore Features Enterprise Blog



skynav / ttv

Timed Text Markup Language (TTML) Validation Tools

103 commits

3 branches

3 releases

1 contributor



branch: master

ttv / +



Complete initial ISD transformation support.



skynavga authored on 31 Aug 2014

latest commit 34d7ed0a90



.eclipse/ttv

Preliminary ISD transformation support.

5 months ago



dst

Release 1.0.0a2

2 years ago



ext/lib

Progress servlet support (WIP). Initial overhaul of Reporter et al.

11 months ago



Image: Jürg Vollmer / maiak.info

Connected TV Accessibility

Advanced Subtitle features in VLC



Subtitle track synchronization:

-10,000 s (Hastened)

Summary

- Open Source Software is essential for subtitle standards
 - market adoption
 - standard compliance
 - transparent reference
- Industry and Broadcaster can help through:
 - Test content, Test Content, Test Content



Thank you!

Andreas Tai
Production Systems Television

Institut für Rundfunktechnik
Floriansmühlstraße 60
80939 München

Tel. +49-(0)89-32399-0
Fax +49-(0)89-32399-351
E-Mail: tai@irt.de