

Adobe Position Paper

Submitted for The Third W3C Web and TV Workshop

19-20 September 2011

Executive Summary:

- W3C is faced with a significant technical and standardization problem – the possible creation of two competing specifications for video captioning, depending upon whether the video is TV or browser based.
- The issue is divisive and can potentially fragment the market.
- Adobe believes that the situation does admit of a technical solution – but one which will require people on both sides of the argument to willingly work together.
- Adobe believes that the W3C should act to correct this situation based upon our suggestion for a possible way forward.

Overview:

The purpose of any standardization effort is to create a specification that is accepted and implemented by all of the elements of society that create “the market”. A specification must recognize the needs of all stakeholders in order to avoid either being ignored or fracturing the market. For HTML5 and its ancillary technologies to succeed as the media platform of choice for unifying the World Wide Web and television, the specification will need to be an inclusive, and not exclusive, force. It is this issue, particularly with respect to captioning, that Adobe wishes to address with this position paper.

Background:

Beginning in the 1970s, work on captioning has seen a tremendous amount of interplay of academic research, industry implementation in both software and hardware, regulatory influence, and the creation of an enormous amount of very valuable legacy content. Recognizing this effort, and also recognizing the amount of legacy content, the W3C undertook to standardize captioning on the Web. This resulted in the creation of the W3C Timed Text Working Group in 2003. The group was made up of representatives from throughout the industry, and the format was undertaken with an expressed goal of “... transcoding or exchanging timed text information among legacy distribution content formats presently in use

for subtitling and captioning functions.”¹ This resulted in the Timed Text Markup Language (TTML) format widely implemented by content producers and tool vendors. Captioning houses generate it. Existing platforms for video such as Adobe Flash Player and Microsoft Silverlight support it. And the Society of Motion Picture and Television Engineers (SMPTE) has adopted a profile called SMPTE-TT as a core format for broadband captioning while bridging existing CEA-608 captions for the Web: “SMPTE-TT can be used in an end-to-end broadband infrastructure as a native format – including authoring and rendering.”²

The depth of knowledge and breadth of participation found in the Timed Text Working Group (TTWG) reflects a broad cross industry view, with elements of most impacted stakeholders. The group’s goal was well-defined, included input from all the relevant parties, and it produced a concrete specification meeting its mandate of compatible captions for the Web.

However, the changes in the browser market and the growing capabilities of HTML5 in being able to present video resulted in new market requirements from some of the stakeholders - specifically browser makers. This resulted in the creation of a specification called Web Video Text Tracks (WebVTT), which has been forwarded to the W3C for possible approval to start a W3C Working Group with the intent of a W3C Recommendation as the ultimate goal.

On October 8, 2010, the 21st Century Communications and Video Accessibility Act (CVAA) in the United States was signed into law. As enabling legislation begins, we expect to see the emergence of captioning as a legal requirement in some applications, including broadcast content also made available online. We anticipate that other nations, especially in the European Union, will follow suit explicitly with captioning legislation, or implicitly by adopting the W3C/WAI Web Content Accessibility Guidelines (WCAG) 2.0 as policy. Absent a single specification, this raises the specter of an absolutely fragmented market, based on national borders and legislative fiat. This would lead to the failure of not only the captioning standard(s), but may also have a chilling effect on technology adoption inside of HTML5, a situation that Adobe wishes to avoid.

Proposed solution:

We propose that the W3C and the industry engage in the creation of a single working group to standardize on Web captioning. The current distinction between broadcast TV and Web video is rapidly vanishing. To distinguish captioning standards based on a distinction that is going away

¹ <http://www.w3.org/TR/ttaf1-dfxp/>

² <https://smppte.org/sites/default/files/st2052-0-2010.pdf>

is not a viable approach. We need one captioning standard that can take a form most appropriate for a given situation.

Adobe believes that the W3C should empower a single working group for video captioning. The charter for this group should be to:

- Document the requirements for both broadcast video captioning (sympathetic to TTML) and for web video captioning (sympathetic to WebVTT) and it should point out the differences of those requirements, if any;
- Document a base semantic model that would encompass the requirements expressed by both constituencies;
- Define representations for that model that work in the contexts where TTML is used in a Web browser;
- Define a concrete syntax for that model that recreates TTML and a concrete syntax for that model informed by WebVTT.

The members of this video captioning group **must** include the experienced experts that have been in the TTML working group and **must** include experts from the HTML5 proponents of WebVTT. Introducing a captioning standard for HTML video should be as compatible with the current practice as possible.

Conclusion:

We think that the idea of defining one common abstract model can unify the existing concepts and will provide a solid base for further concrete realizations, for well-defined conversions and for future extensions. We believe that defining a core semantic model for video captioning, which meets the requirements exposed while developing TTML and exposed while developing WebVTT is possible. Once that common semantic model has been defined, then two concrete representations for expressing it can be developed, one being TTML and the other being heavily informed by WebVTT and meeting the more dynamic requirements for Web video captioning. This will make conversion between these two formats very well defined and will provide a common base for future work.

Adobe believes that if we cannot create a common standard to bridge this gap in an area as elemental as captioning, the potential for achieving commonality across a broader range of more complex technical and market issues is limited. There exists a great body of legacy content, a great deal of technical expertise, and a goodly amount of both opportunity and need for a common solution. Adobe will commit to working with the W3C to encourage this merger of these two similar technical efforts and encourage all stakeholders to join us in that effort.