

Multimedia Standards: Building Blocks of the Web

Lloyd Rutledge
CWI Amsterdam

You've probably heard this joke before: "The nice thing about standards is that there are so many to choose from." It used to mean that for any one task, there were many standards from which to choose. However, these days in multimedia, the joke takes on a whole new meaning—there are many standards for composing multimedia on the Web, but now rather than choosing one, you use all of them together.

Proliferations of standards, by no means unique to multimedia, always bring about analogies with the Tower of Babel. With the old version of the joke, developers wrote multimedia presentations in different formats for different browsers, so they weren't portable. This degenerated into a world of presentations and systems that couldn't communicate with each other, and the idea of one language for all was lost.

In the new age of cooperative formats, we're still surrounded by many different languages, but we must understand them all at once. As we struggle to build our ziggurat to the heaven of an all-reaching and all-knowing multimedia-based Web, we should stop and ponder if the effort will crumble under its own mass and complexity or if the goal is even reachable. On the other hand, we can compare this effort to the space shuttle—a system that successfully combines the many different and complex interwoven components needed to build a stairway to the heavens.

So far, the media standardization effort has been working, with the emerging formats for the most part having successful implementation and adoption. In my opinion, here are the steps in the proverbial multimedia stairway so far.

The foundation

The Extensible Markup Language (XML) provides the bricks and mortar for building the Web. By first glancing at XML, you would think the Web was held together by angle brackets—and

indeed it is. These characters form a syntax that defines elements, attributes of these elements, and contents of these elements made up of either text or other elements. Sharing XML as their common foundation, other languages can be built upon it.

The Hypertext Markup Language (which needs no introduction) was originally defined in Standard Generalized Markup Language, XML's predecessor. The release of XHTML, an XML version of HTML, brings it fully into the XML family of languages. All browser wishing to keep up-to-date with the last HTML version will need to process XHTML correctly. All this has XHTML emerging as the XML-defined structure bearing presentable text to the world over. XHTML is one set of XML's elements and attribute types. Its contribution to multimedia is its support for in-line images and links to media of any format.

Although XHTML defines the text's structure, it leaves out the details of its final presentation. These details, such as background color and font type, are the document's presentation style, as opposed to its content and structure. Because one style can apply to many documents and one document can have many different styles, it's best to define the presentation style in a separate file. On the Web, developers define such styles in cascading style sheets (CSS), which define how XHTML- and XML-defined content and structure will appear. Thus, when you see an XHTML page in a Web browser, there are two pieces of code at work: XHTML and CSS.

SMIL

At the center of the new standards for multimedia is the Synchronized Multimedia Integration Language, the World Wide Web Consortium's recommendation for multimedia on the Web. The W3C released SMIL 1.0 in 1998 as the basic foundation for distributed multimedia. SMIL 2.0, in the later stages of development, defines state-of-the-art Web-based multimedia with many

new features such as event-based timing, animation, and transitions. SMIL 1.0 has been widely adopted, with millions of RealPlayers browsers distributed that play it. SMIL 2.0 already has a prerelease implementation in the beta version of Oratrix's Grins player(see <http://www.oratrix.com/GRiNS>).

SMIL's primary contribution to the new babble of formats is that as multimedia, it's an integration format; it doesn't define presentation but refers to other files in other formats and integrates their presentation. Thus, SMIL presentation writers must know SMIL and the formats of the various media to be included.

One of SMIL's main gifts to the Web is adaptivity. SMIL presentations can adapt gracefully to varying bandwidth and system hiccups. They also adapt to different users, enabling multilingual presentation and accessibility to users with different perceptual abilities. SMIL's powerful timing mechanisms also provide dynamic and smooth response to user interaction with mouse events, transitions, and animations.

A further linguistic dependency comes from SMIL's relationship with CSS. SMIL layout constructs have the same name and behavior as corresponding constructs in CSS. Furthermore, CSS can define an alternative layout for a SMIL presentation. However, CSS don't provide all the function needed for multimedia layout. Thus, SMIL can't use CSS exclusively, and CSS-defined alternative layouts might have limited functionality. Future versions of CSS might include style and layout for multimedia as well as for hypertext, but what "multimedia style" is remains a complex issue that's not likely to be resolved soon at the standards level.

Modules in SMIL

SMIL 2.0 adds a new layer of linguistic complexity by introducing profiles and modularization to SMIL. These features make SMIL, like XML, a metalanguage from which we can define conforming languages. Sublanguages of SMIL 2.0 are SMIL profiles, and they're made up of combinations of the modules into which SMIL 2.0 is split.

The primary profile is the SMIL 2.0 Language Profile, which represents state-of-the-art multimedia, using almost all SMIL 2.0 constructs. There's also the profile SMIL Basic, which is for displaying multimedia on limited-resource devices such as mobile phones and personal data assistants. SMIL Basic only has the modules of SMIL 2.0 that can be readily processed on small machines. Finally, there's the profile XHTML+SMIL, which adds SMIL's adap-

tivity and dynamic timing to XHTML's text content and text-based layout. Microsoft has already implemented and widely distributed XHTML+SMIL's prototype, HTML+TIME, in Internet Explorer 5.5. Microsoft is active in XHTML+SMIL's development as well. The SMIL 2.0 specification also states how we can make more SMIL profiles.

SMIL 2.0's modularization technique is basically the same as Modularization of XHTML, which recently became an official W3C recommendation. The XHTML modularization lets us develop XHTML sublanguages tailored, for example, for mobile devices. The SMIL and XHTML modularizations also let us combine modules of each format with each other and with other XML-based metalanguages to form new specially tailored sublanguages. XHTML+SMIL is one example of the potential of such combinations between formats.

SVG

Among the media formats that SMIL integrates, Scalable Vector Graphic (SVG) has been getting the most attention lately. SVG is an XML-based graphics format that provides rich adaptivity and user interaction. It can handle complex diagrams gracefully. SVG's abstract level of graphics representation enables much flexibility, such as zooming in without producing aliasing effects. There are also many implementations of it available, including a plug-in player from Adobe, SVG's primary industrial supporter.

Typical of currently emerging W3C formats, SVG has intricate tie-ins with other formats. SVG can incorporate images files in other visual formats, as does XHTML and SMIL, along with other media. Also like XHTML, SVG uses CSS to define the presentation style and, to some degree, the layout of the graphics it represents. Finally, SVG incorporates constructs from SMIL's timing modules into its format to provide animation of its graphic components.

The Semantic Web

The Web's founder, Tim Berners-Lee, is pushing the vision of the Semantic Web. Currently, the meaning within the content of Web documents is perceivable only to a human reader. Putting semantics into the Web infrastructure lets machines process a document's meaning as well. Such machine processing would greatly enhance searching on the Web and help provide other functions such as automatically drawing conclusions for the user and generating user-tailored presentations.

The W3C is pursuing this vision with several

Web Resources

To access specifications for W3C's completed and developing recommendations, visit <http://www.w3.org>. Here are the direct links to the standards I've mentioned:

CSS: <http://www.w3.org/TR/REC-CSS2/>
 DAML: <http://DAML.SemanticWeb.org>
 MPEG-4: <http://www.cselt.it/mpeg/standards/mpeg-4/mpeg-4.htm>
 MPEG-7: <http://www.cselt.it/mpeg/standards/mpeg-7/mpeg-7.htm>
 MPEG-21: <http://www.cselt.it/mpeg/standards/mpeg-21/mpeg-21.htm>
 OIL: <http://www.ontoknowledge.org/oil/>
 PNG: <http://www.w3.org/TR/REC-png>
 RDF: <http://www.w3.org/RDF/>
 RDF Schema: <http://www.w3.org/TR/2000/CR-rdf-schema-20000327/>
 Semantic Web: <http://www.semanticweb.org/>
 SMIL: <http://www.w3.org/TR/smil20/>
 SVG: <http://www.w3.org/Graphics/SVG/Overview.htm#>
 XHTML: <http://www.w3.org/TR/xhtml1/>
 XML: <http://www.w3.org/TR/REC-xml>
 XML Schema: <http://www.w3.org/TR/xmlschema-0/>, <http://www.w3.org/TR/xmlschema-1/>, and <http://www.w3.org/TR/xmlschema-2/>

layered standards. The foundational format for semantics on the Web is the Resource Description Format. RDF provides a simple data model for expressing statements using subject, predicate, and object triples and an associated syntax in XML. A set of RDF statements uses a particular vocabulary that defines the properties and data types that are meaningful for the application at hand. We can define such an RDF vocabulary with another W3C format, RDF Schema. Finally, DAML+OIL is a language being developed by a joint US and European Union committee seeking to put RDF and RDF Schema on top of additional features from the artificial intelligence community for more efficient inference processing and more formal semantics. By applying these formats to Web multimedia presentations, we can say how they should appear and what they mean.

Other standards

Of course, the list goes on. A portable network graphic (PNG) is a recent W3C image format that provides partial transparency, enhancing the potential for visually blending multiple images. The International Organization for Standardization is providing MPEG-7 for annotating media, and it may end up with an XML encoding. The ISO is also developing MPEG-4 for the representation and unified transmission of multimedia presentations, providing a means of packaging collections of media in these formats for delivery.

In the final phases of development are the W3C's XLink for defining relations between document components and XPointer for locating portions of XML documents.

Legal babble

Providing an economic and legal environment for open, interoperable multimedia is proving to be as complex as developing the technical environment.

Recently, the W3C working draft for XPointer started linking to a patent claim by Sun Microsystems on portions of the standard. The claim is, by all appearances, purely defensive. Sun doesn't want to charge anyone for using XPointer; they just don't want anyone else to charge for it either. But similar defensive patents are appearing on other emerging standards as well. This is creating an atmosphere of distrust and fear similar to TV's Wild West, where everyone needs a gun because everyone else has one and you never know when you might be shot at. There's widespread concern that this proliferation of patents will hamper industrial development of tools for standards and go against what standards are supposed to be about. A generally applicable solution that fully protects both standards and the companies that help development them remains unseen.

Get involved

Standards creation is, for the most part, an open process, and it depends on involvement from a breadth of communities. Increasingly, the research community is having a say and enjoying rapid expression of their research ideas as standards and industrialization. All this means, of course, that volunteer-based standardization needs your ideas and expertise. Take a look at the emerging formats, choose your favorite, and help guide its development. If you can't find a standard that suits your fancy, then you can always start a new one. **MM**

Readers may contact Rutledge at CWI, Kruislaan 413, PO Box 94079, 1090 GB Amsterdam, The Netherlands, email Lloyd.Rutledge@cwi.nl.

Contact Media Impact editor Frank Nack at CWI, Kruislaan 413, PO Box 94079, 1090 GB Amsterdam, The Netherlands, email Frank.Nack@cwi.nl.