

Overview of the W3C

Ray Denenberg
Library of Congress

March 2000



*Library of Congress
Network Development & MARC Standards Office*

Background

The W3C (Web Consortium) was established in 1994 to develop common protocols for the web, promote its evolution, and improve interoperability. Its stated goal is to "lead the Web to its full potential" and ensure that the web remains a robust, scalable, and adaptive infrastructure; make the web accessible to as many users as possible; and promote technologies that bridge differences in culture, education, ability, material resources, and physical limitations.

W3C is an international industry consortium, jointly hosted by MIT (in the US), INRIA (in Europe); and the Keio University (in Japan). Currently (as of late February 2000) W3C has 400 members worldwide. These are mostly commercial companies, but also include academic institutions, consortiums, and government agencies. Commercial companies include hardware, software, and telecommunications companies; content providers; and corporate users.

Members pay an annual fee of \$50,000 or \$5000, for full or affiliate membership respectively. Affiliate membership is available only to government agencies, non-profit companies, and companies with limited revenue; however, there are no differences in privileges between full and affiliate membership.

Services and Products

The primary service provided by the W3C to its members is a vendor-neutral forum to address web-related issues. In addition, W3C provides information to its members about its proceedings that the W3C considers proprietary and which is not available to non-members, and which is considered by some members to be crucial for strategic planning. (This is one of the primary reasons that some commercial companies join; on the other hand, for some companies this is not a consideration at all.) Additional services include reference code implementations (for web standards) and various prototype and sample applications to demonstrate use of new technology.

The primary product of the W3C (and its members) are technical specifications, notably, the W3C "Recommendation", which is the W3C version of "standard". XML, for example, is a W3C Recommendation. Recommendations must be formally approved by the Membership, after proceeding through review stages of Working Draft and Proposed Recommendation. There are approximately 20 W3C Recommendations.

Goals and Principles

As noted above W3C's stated goal is to "lead the Web to its full potential", by developing common protocols, promoting the web's evolution, and improving interoperability. To do so, W3C articulates and claims commitment to a number of principles: Universal access, trust, decentralization, evolvability, improved multimedia, interoperability, and semantic interoperability.

The latter two differ as follows: "interoperability" (unqualified) refers to hardware and software portability and protocol operation, as well as the open process of developing protocols; *semantic* interoperability refers to the ability to express semantics in a machine-understandable, processible form. RDF (as well as the more general metadata activity) is aimed at semantic interoperability.

In describing "universal access", W3C articulates a view of the web as the universe of net-worked information available to anyone

with access to the network. Several W3C activities are aimed at universal access: Mobile Access, TVWeb, Voice Browser, and Web Accessibility Initiative.

Activities such as digital signatures, annotation mechanisms, group authoring, and versioning are aimed at the principle of "trust": ensuring confidentiality, instilling confidence, and allowing people to take responsibility (or be accountable) for what they publish. The principle of improved multimedia is represented by such activities as Scalable Vector Graphics (SVG) language and the Synchronized Multimedia Integration Language (SMIL). The remaining two principles, evolvability and decentralization, are fairly general and self-explanatory; they are not articulated in any detail nor represented by any specific activities.

The W3C Process

The W3C organizational components are the members, staff (approximately 60, referred to as the "team"), and host institutions. Members are entitled to assorted privileges, including a seat on the Advisory Committee, the opportunity to participate in W3C Activities and to provide strategic direction, access to information not available to non-members including newsletters, announcements, and information on events, technologies, software releases, discussion forums, and mailing lists. The team provide direction (keeping up-to-date on new technology, market fluctuations, activities of related organizations, etc.), keeps members (as well as the public) informed of W3C Activities, gathers ideas and input from outside sources, markets W3C results to gain acceptance in the web community, and attracts new members. The team consists of a Chairman, Director, and staff. The Chairman manages the general operation, chairs Advisory Committee and Advisory Board meetings, oversees the development of the W3C international structure, coordinates liaisons with standards bodies, and addresses legal and policy issues. The current Chairman is Jean-François Abramatic (Director of Development and Industrial Relations at INRIA - Rocquencourt, and formerly Associate Director of the MIT Laboratory for Computer Science). The Director is the lead architect for the technologies developed, and also approves Recommendations, Activity proposals, and charters, and designates Group Chairs. The current director is Tim Berners-Lee (web founder).

The W3C Advisory Committee meets twice a year and is composed of a single designated representative from each member organization, who is the official link between the Member organization and the team.

The W3C Advisory Board ensures that the W3C remains responsive to the needs of the members as well as entities outside of W3C, notably, standards bodies. It provides rapid feedback to the team on issues that are vital to W3C operation and which cannot wait until the next AC meeting for resolution. The Advisory Board consists of individuals belonging to member organizations, but not necessarily Advisory Committee representatives. AB members are elected.

Development of technical specifications is carried out largely by Working Groups, composed primarily of representatives from member organizations, and administered by team members. There are 30 or so working groups, with several hundred participants total. The level of effort per working group member per working group varies, but can be as much as 25% of a members time. The aggregate effort per year might therefore be estimated (very roughly) to be 100 man-years.

W3C articulates a consensus policy whereby deliberations must address all participants' views and objections, attempt to resolve them, and determine consensus, which is considered to be established when "substantial agreement" has been reached. "Substantial agreement" means more than a simple majority, but not necessarily unanimity, for example consensus may be declared when the minority no longer wishes to articulate its objections.

W3C Activities and Domains

W3C has various *activities*, organized into four *domains*: User Interface, Technology and Society, Architecture, and the Web Accessibility Initiative.

User Interface Domain

The premise of the User Interface domain is that as web information continues to grow, both in volume and in variety, it will be used by a much greater diversity of people, and simplicity and interoperability will take on even greater importance. The UI domain focuses on formats and languages that will present information to users with more accuracy and a higher level of control.

The most visible activities within the User Interface Domain are HTML, style sheets, and graphics. Also included are internationalization, document object model, mobile access, synchronized multimedia, voice browsers, Web television, and math.

HTML

The HTML activity itself is closely related to other activities, including style sheets and internationalization. Version 4.01 is the current W3C recommendation for HTML (as of December 1999) and fixes bugs in the HTML 4.0 specification (earlier versions were 3.2 and 2.0). HTML 4.0 (and 4.01) supports style sheets -- to control color, font and layout - and (for internationalization purposed) text written right to left, as, for example, in Hebrew, as well accessibility features - speech synthesizers or braille readers. HTML, is now being re-cast in XML, as "XHTML", to be used in combination with XML applications. XHTML1.0 became a W3C Recommendation in January.

Style Sheets

Style sheets control the presentation of web pages, allowing designers to specify visual effects as well as voice, pitch and other aspects of how text will sound when rendered into speech. Within the style sheet activity, there is the Cascading Style Sheet (CSS) specification. Level 1 (CSS1) became a W3C recommendation in 1996. CSS1 is a simple style sheet mechanism that allows authors and readers to attach style (e.g. fonts, colors and spacing) to HTML documents. "cascading" means that HTML authors can attach a preferred style sheet, while the reader may have a personal style sheet to adjust for human or technological handicaps.

W3C is also developing XSL, "Extensible Style Sheets Language" which builds upon CSS, and is intended for complex publishing tasks such as automatically producing a table of contents, and for converting documents written in XML into HTML for publication.

Graphics

Graphics is one of the most visible and popular features of the web. Successful use of graphics depends on interoperability across platforms, output resolutions, color spaces, and software products.

PNG, Portable Network Graphics, which became a W3C recommendation in 1996, is a format for bitmapped images and now (according to W3C) has widespread support on the latest generation of browsers. The initial motivation for developing PNG graphics was to improve on GIF, and PNG adds many new features. It is a lossless format (when the image is decompressed, the exact pixel values are preserved) supporting true color (in addition to indexed color, whereas only the latter is supported by GIF, so GIF is lossy for truecolor images). In true color, values for each pixel are specified directly, while indexed color represents each pixel as an entry in a palette. At relatively little extra cost, true color gives greatly enhanced graphical quality and dispenses with all the bother of dealing with palettes.

Graphics traditionally have consisted of images effectively represented as a sequence of colored pixels. GIF, JPEG and PNG are examples of bit-mapped graphic formats based on this principle. However, rather than transferring pixel values, an alternative approach lies in vector graphics, where instructions for drawing lines or curves (vectors) and filling shapes are transmitted. (Graphics are often created this way but then converted to raster images, and the scalable vector information is lost.)

There are some clear advantages of vector-graphics over bit-mapped graphics; the latter may look good when first created, but deteriorate when viewed on different kinds of screen at different sizes. When a bit-mapped image is scaled, lines become jagged and the image blurs; often, after compression, vector graphics images are considerably smaller than the bit-mapped equivalent. Vector graphics are easier to manipulate and revise, search engines can index them meaningfully, and they are accessible to people with visual disabilities.

There are some proprietary vector graphic formats, generally implemented by plug-ins. No single format is widely and well supported by tools for creating pages, and in general there is no cross-platform support.

There are two different markets for vector graphics. One is "technical documentation for industry", covering aerospace, defense, automotive and electronics industries. They have no requirement for restylable graphics or for graduated fills, but need precise specification of line and hatch styles. These industries have standardized on CGM, an ISO standard for vector graphics, but they lack a vendor-neutral, interoperable standard. W3C has developed the WebCGM profile developed in collaboration with CGM Open. The WebCGM profile adds additional constraints to improve interoperability, and supports hyperlinking and HTML.

The other market, graphic design for advertising, clip art, business presentations and general web use, needs complex fills, restyling, image clipping and manipulation, and re-usable components. For this market, use of CGM is less important than good integration with XML. W3C is therefore developing a standard vector graphics format, Scalable

Vector Graphics , SVG, written in XML, designed to work well across platforms, output resolutions, color spaces, and a range of bandwidths.

Internationalization

Issues addressed by the internationalization activity include promotion of Unicode, urls (specifically, the language of human readable identifiers), non-western character sets, languages and writing systems, and representation of time. The internationalization effort attempts to facilitate authoring and publishing in Arabic, Chinese, French, Japanese, Hebrew, Thai and other languages.

Document Object Model

The Document Object Model (DOM) is an abstract general representation of a document's structure and how a document is accessed and manipulated, intended to facilitate programmer access to document components to enable addition, editing, and deletion of content, attributes and style. DOM allows programmers to write applications which work across different browsers, servers, and platforms.

Vendors support the DOM as an interface to their proprietary data structures and APIs, and content authors can write to the standard DOM interfaces rather than product- specific APIs.

DOM originated to allow Java portability among browsers. Though originally thought of largely in terms of browsers, when the DOM Working Group was formed it was by vendors in other domains, including HTML/XML editors and document repositories and vendors who had worked with SGML before XML was developed. Some of these vendors had developed their own object models for documents to provide an API for SGML/XML editors or document repositories, and these object models have also influenced the DOM.

Level 1 of DOM addresses HTML and XML and is already approved as a W3C Recommendation. Level 2 includes a style sheet object model, and defines functionality for manipulating the style information attached to a document. It also enables filters on the document, defines an event model and provides support for XML namespaces. Level 2 is a candidate recommendation. Level 3 will address document loading and saving, as well as DTDs and schemas and document validation. A working draft of Level 3 is expected in April.

Mobile Access

There is interest in mobile web access among hardware manufacturers, software providers, communication service providers, content providers, and end user organizations. W3C is working towards making web information accessible to mobile devices, characterized by small screens, limited keyboard, low bandwidth, small memory, etc.

Synchronized Multimedia

W3C has developed the Synchronized Multimedia Integration Language (SMIL), a language for scheduling multimedia presentations combining audio, video, text and graphics.

It defines a simple XML-based language allowing authors to write interactive multimedia presentations, describe its temporal behavior, associate hyperlinks, and describe the screen-layout.

Voice Browser

"Voice Browsers" will allow people to interact with web sites via spoken commands, and listen to prerecorded or synthetic speech and music. This will allow a telephone to access web services, and help people with visual impairments and those needing web access while their hands or eyes are otherwise occupied.

Math

W3C has developed a specification for embedding mathematical expressions in HTML and XML documents. Mathematical Markup Language (MathML) provides a way of encoding both mathematical content and visual presentation for mathematics at all levels, from elementary school to scientific research.

Web and TV

There is much interest in integration of web and television technologies to allow web browsing using a television set, or to use web content formats to annotate a television broadcast to make it more interactive. There are efforts to standardize TV/Web integration in several standardization bodies, and the W3C activity hopes to ward off regional, incompatible solutions (typical of TV standards). See also the WebTV activity in the architecture domain.

Technology and Society Domain

The technology and Society Domain looks at ethical and legal issues including electronic commerce, metadata, and privacy. It concentrates on issues that arise from applications of web technology including international public policy issues. It attempts to educate the public as well as policy makers about web capabilities and how these affect and are affected by proposed policies, and to establish trust in the web. The objective is to support a diverse range of public policy options without cultural fragmentation or domination.

W3C describes this domain as "trying to understanding the social impact of the web and reach out to affected communities, to look at ethical and legal issues from an international perspective, and seek to understand these issues in light of new technology, both by changing the technology and by educating users about the technology's benefits, costs, and limits."

Much of the focus of the Technology and Society domain is to establish and maintain trust on the web: The ability to make statements that have agreed upon meanings is addressed by the Metadata Activity; the ability to know who made the statement and to be assured that the statement is really theirs is addressed by the Digital Signature Initiative. There are also various privacy and security-related initiatives within this domain.

Metadata Activity

The Metadata Activity is concerned with modeling and encoding metadata. It focuses on two primary areas: the Resource Description Framework, RDF, and Platform for Internet Content Selection, PICS. PICS is now complete; work on RDF continues.

PICS enables distribution of metadata about the content of digital material in the form of "labels", containing machine-processible information about content, allowing computers to filter out undesirable material or direct users to sites that may be of interest.

Though PICS is said to be "complete", its practical status remains unclear; it may be that PICS is superseded by RDF (their relationship has not been clarified), which provides a more general treatment of metadata.

RDF is a declarative language for using XML to represent metadata in the form of properties and relationships among resources. RDF provides a framework for independent communities to develop vocabularies that suit specific needs while sharing vocabularies with other communities. To share vocabularies, terms must be described in detail. Descriptions of these vocabularies are called RDF Schemas. A schema defines the meaning, characteristics, and relationships of a set of properties, perhaps including constraints on potential values and inheritance of properties from other schemas. RDF allows each document containing metadata to clarify which vocabulary is being used, by assigning each vocabulary a unique namespace identifier.

The W3C thinks that there are many practical uses of RDF, including thesauri and library classification schemes, web sitemaps, description of the contents of web pages, rating systems, and digital signatures.

Digital Signature Initiative

While metadata provides a means to create machine-processible statements about information, digital signatures provides the ability to know who made the statement and to be assured that the statement is really theirs. The Digital Signature Initiative attempts to provide a mechanism for signing metadata in order to establish who is making the statement.

Privacy Activity

The Privacy Activity enables web sites to express privacy policies, indicating what data is collected from users and what the site will do with that data. The Privacy Activity will allow users to control the flow of data to Web sites, based on their privacy preferences and on the Web sites' privacy policies.

The Privacy Activity was originally chartered in 1997 and was originally scheduled to produce a Recommendation called P3P (Platform for Privacy Preferences Project) by the end of 1998. The activity has been delayed and P3P is still under

development, because of the complexity of designing a vocabulary that accommodates all of the various privacy policy regimes worldwide.

Intellectual Property Rights

A number of members have expressed interest in having the W3C address issues of protecting intellectual property on the Internet, investigating policy issues, and clarifying corporate positions in this area. The W3C is currently studying this as a possible activity, and considering organizations to collaborate with.

Architecture Domain

The Architecture Domain covers HTTP, Web television, XML, naming and addressing and web characterization.

HTTP

HTTP is the backbone protocol of the web, but it has become strained both in terms of modularity and performance. There are potential web applications that cannot be implemented because they are too complex to deploy in the current infrastructure given the architectural constraints of HTTP.

A now-dormant activity, HTTP-NG (new generation) attempted to address these concerns, but got only as far as to outline potential directions for the future development of HTTP. At this point, W3C does not plan any follow-up activities and the future development of HTTP is unclear.

Web Television

The use of the television set as display terminal for web access has increased over the past few years. A trend triggered by this development is the integration of web and TV content, for example transmitting a web page showing player statistics during a sports game. The development of television/web integration is advanced by the use of digital TV, which more closely aligns transmission technologies for television and web content, and increases the intelligence available in a television set, and will lead to the necessity of hardware upgrades of TV sets that in many cases will include web functionality. Furthermore, many of the services associated with digital interactive TV are available today on the web. W3C has thus initiated an activity involving the television community and the Internet community as focusing on "Real-Time Multimedia and the Web".

See also the "Web and TV" activity in the User Interface Domain.

XML

XML was originally intended and designed to facilitate the use of SGML -- define document types, author and manage SGML-defined documents, and transmit and share them -- but it has taken on a life of its own.

XML defines a simple dialect of SGML whose goal is to allow generic SGML to be transmitted, received, and processed in the way that is now possible with HTML.

The current version of XML, version 1.0, reached W3C Recommendation status, February 1998. The companion specification, Namespaces in XML, became a recommendation January 1999.

Naming and Addressing

This activity pertains to URIs, including the development of URI schemes. This activity is currently dormant.

Web Characterization

The Web Characterization activity attempts to provide information and test scenarios about the web and how it is used now and will be used in the near future, in order to allow web evolution and ensure its long term interoperability and robustness.

Web Accessibility Initiative

The Web Accessibility Initiative (WAI) promotes web-usability for people with disabilities.

Web Content Accessibility Guidelines, version 1.0, became a W3C Recommendation May 1999. These guidelines explain how to make web content accessible to people with disabilities and are intended for web content developers and developers of authoring tools. The primary goal is to promote accessibility with respect to a variety of constraints (e.g., noisy surroundings, under- or over-illuminated rooms, hands-free environments, etc.) and disabilities: inability to see, hear, move, process certain types of information; difficulty reading or comprehending text; inability to use a keyboard or mouse; a text-only screen, a small screen, or a slow Internet connection; lack of fluency of the language in which a document is written; situations where eyes, ears, or hands are busy (e.g., driving to work, working in a loud environment, etc.); early or different version of a browser, a voice browser, or a different operating system.

The guidelines establish principles for accessible design, such as the need to provide equivalent alternatives for auditory and visual information. Each guideline has checkpoints explaining how these accessibility principles apply to specific features of sites. For example, providing alternative text for images ensures that information is available to a person who cannot see images. Providing captions for audio files makes information available to someone who cannot hear audio. The guidelines are designed to be compatible with evolving technologies, yet operate with legacy browsers.

In January 2000, "User Agent Accessibility Guidelines" (1.0) became a W3C Candidate Recommendation. The guidelines explain how to design user agents that are accessible to people with disabilities -- graphical desktop browsers, multimedia players, text browsers, voice browsers, plug-ins, and other assistive technologies that give full access to web content.

In February, "Authoring Tool Accessibility Guidelines" (1.0) became a W3C Recommendation. It provides guidelines to assist developers in designing authoring tools that produce accessible web content and creating accessible authoring interfaces

W3C Open Source Releases

As a compliment to their specifications, W3C also provides several Open Source Releases. These include:

- *Jigsaw*, W3C's Web server platform, providing a sample HTTP 1.1 implementation and several oother features on top of an advanced architecture implemented in Java.
- *Amaya*, intended to be a comprehensive client environment for testing and evaluating new proposals for web standards and formats. A large part of the intended features of Amaya are implemented in the most recent release, but some of them are not complete yet.
- *Winie*, a Java network utility to 'put' and 'get' files on the Web using HTTP/1.1.
- *Charlint* , a Character Normalization Tool. It implements Normalization Form C of Unicode TR 15.
- HTML Validation Service: Source Code.
- CSS Validator.
- SiRPAC - Simple RDF Parser & Compilers.

Other W3C Functions

Aside from official Activities carried out by Working Groups who develop Recommendations, there are other, less-formal W3C functions. W3C holds occasional "events": workshops, symposia, and conferences. It sponsors various "interest groups and produces various technical reports. For more information see [W3C's public home page](#).



*Library of Congress
Network Development & MARC Standards Office
Updated: March 9, 2000
[comments](#)*