

## XML in 2000—The Impact on Technical Communicators

Leonor Ciarlone, CAP Ventures, Inc.  
Senior Consultant

Robert Boeri, Forefield Inc.  
Information Architect

© 2000 and 2001



Updated October 13, 2001

## Handout: HTML, SGML, XML Glossary of Terms

### Anchor

Target for a hypertext link. Can be a section of text, a graphic or position on a graphic, or any other defined element. In SGML (without HyTime capabilities) and HTML, that means an element with an ID attribute. In XML (via XLL), that means any element, regardless of its attribute.

### Attribute

Additional markup information added to an element definition. Attributes are to elements as adjectives are to nouns. Attributes continue the goal of structuring documents to add value and intelligence.

### CSS (Cascading Style Sheet)

Defines the language in which style sheets for HTML documents are written. Provides a set of properties that designers use to describe how documents should be presented. Developed as early as 1994. Current released version is CSS2 (May, 1998), although CSS3 was a working draft Feb, 2000.

### Deprecated

A term used by the World Wide Web Consortium (W3C) to describe obsolete HTML elements (tags). A deprecated element (or attribute) is one that has been outdated by newer constructs and may become obsolete in future versions of HTML. In general, the W3C has either “deprecated” or is planning on “deprecating” HTML tags that purely presentational and have style sheet mechanism alternatives (via CSS).

### DHTML (Dynamic HTML)

A term being applied by Microsoft and Netscape to describe a collection of technologies developed for making HTML more dynamic and interactive. Each vendor “technology” is unique and proprietary. Not a standard that has been enacted, proposed or even developed by any Internet standards organization.

### Document Analysis

A stage in the development of an XML Document Type Definition (DTD). Document analysis defines the goals for the logical document model and produces the specification used to develop the DTD.

## **Document Model**

The document's logical organization. Similar to a database schema that describes information types and field relationships. The foundation of and framework for an SGML environment.

## **DOM (Document Object Model)**

A platform- and language-neutral interface that will allow programs and scripts to dynamically access and update the content, structure and style of XML documents.

## **DSSSL (Document Style Semantics and Specification Language)**

An International Standards Organization (ISO) 1996 SGML styling standard not implemented by any commercial software vendor.

## **DTD (Document Type Definition)**

Defines the structure for an SGML, HTML, or XML document. Lets processors automatically identify where every element comes from and how elements relate to each other. Allows stylesheets, navigators, browsers, search engines, databases, printing routines, and other applications to easily integrate and operate with the document. Also sometimes called a schema, although --strictly speaking-- schemas are a new, richer XML-based standard for expressing document models.

## **Element**

Logical units of information, surrounded by tags. Elements are an integral part of XML and consistent and vendor-neutral way of adding markup.

## **Entity**

Any data that can be treated as a unit. Often used to point to non-XML resources within an XML document or represent text or characters that are either too difficult or long to type repeatedly.

## **Frameset HTML**

An HTML 4.01 "flavor." For developers who understand how to and can support (through compliant browsers) splitting a browser window into two or more frames.

## **FTP (File Transfer Protocol)**

A set of communications protocols used to transfer files between computer systems.

## **Generalized Markup Language**

A language that is user defined, platform-independent, and vendor-neutral. A discipline that promotes modularity, portability, and flexibility. Designed for organization rather than presentation. Separates structure from format.

## **HTML – Hypertext Markup Language**

An SGML application. Development began as early as 1992. Officially replaced by an XML rendition called XHTML on January 26, 2000. Expresses all HTML 4.01's capabilities using XML syntax.

## **HTTP (Hypertext Transport Protocol)**

The protocol for moving hypertext files across the Internet. It requires an HTTP client program at one end and an HTTP server program at the other end.

## **HyTime (Hypermedia/Time-Based Structuring Language)**

An SGML standard providing the inspiration for XML's Synchronous Multimedia Integration Language (SMIL), on which RealNetwork's RealPlayer is based.

## **ICE (Information and Content Exchange)**

Protocol based on XML that automates content syndication on the Web. Describes how content is used, how often it is refreshed and what attributes it contains. Enables a developer to manage and exchange content without manual packaging or having knowledge of remote Web site structures.

## **IETM (Interactive Electronic Technical Manual)**

Electronic technical manuals that use conditional, branching logic to design training or support systems. Marked by the ability to adjust responses according to user input. Often designed based on SGML information repositories. A term defined by the U.S. Department of Defense.

## **Instance**

Also known as document or web page. The actual information (content) to be conveyed.

## **ISO (International Standards Organization)**

A worldwide federation of national standards bodies, whose work is carried out through technical committees. International organizations, governmental and non-governmental, in liaison with ISO, take part in the work defined within each committee. See and contrast with "World Wide Web Consortium."

## **Markup Language**

Any additional information added to the text of a document.

## **Metadata**

Data about data. Information about information. Classifies and categorizes pieces of information typically stored separately from a document.

## **Parser**

A software program that reads, interprets, and processes SGML or XML documents and Document Type Definitions (DTDs). A validating parser locates and reports errors in DTDs and document instances using a DTD.

## **SGML – Standard Generalized Markup Language**

A flexible, platform-independent, vendor neutral, meta-language. "Mother of all markup languages," including HTML and XML. Distantly derived from GML (Generalized Markup Language, Charles Goldfarb, 1960). Maintained by the International Standards Organization (ISO).

## **Strict HTML**

An HTML "flavor." For developers who want "clean markup" (markup without presentation instructions) that incorporates references to CSS (Cascading Style Sheets) components.

## **Stylesheets**

Provides the formatting information to the DTD and the SGML document. Describe how each element should look. Provides a separate, consistent method of applying format for each tag.

## **Tag**

A term defined by the SGML standard. The markup strings that name SGML, HTML, or XML elements.

## **Transitional HTML**

Another HTML 4.01 “flavor.” For developers who don’t want to or cannot support the separation of presentation from content. An example would be development for audiences whose browsers do not support CSS.

## **Valid XML Documents**

XML documents that have a Document Type Definition (DTD) and that conform to it. All valid XML documents are also well-formed XML documents, although the reverse is not true.

## **Well-formed XML Documents**

Well-formed XML documents follow standard XML rules including “balanced” tags (each must have a start tag and an end tag), mandatory root elements, and proper tag nesting that results in a data hierarchy. Unlike valid XML documents, well-formed XML documents don’t require an explicit Document Type Definition (DTD).

## **World Wide Web Consortium (W3C)**

The World Wide Web Consortium, with over 400 member organizations from around the world, was created in October 1994 to lead the World Wide Web to its full potential by developing common protocols that promote its evolution and ensure its interoperability. By promoting interoperability and encouraging an open forum for discussion, W3C commits to leading the technical evolution of the Web.

## **XHTML**

HTML, expressed in XML, and the preferred expression for web content. Like HTML, there are three sets of tags, actually three DTDs: strict, transitional, and frameset.

## **XLink**

Part of the XML linking solution. See XLL (Extensible Linking Language), XPointer and XPath standards on <http://www.w3.org> or summaries on <http://world.std.com/~bboeri/TRC-XML.htm/>. Specifies the syntax required to create uni-directional links (like HTML’s) as well as more sophisticated bi-directional and multi-choice links.

## **XML – Extensible Markup Language**

A subset rather than an application of SGML Like SGML, XML is also a flexible, platform-independent, vendor neutral, meta-language --but XML is designed to work at Web-speed by being less complex than SGML, but about 80% as powerful. Work on XML began in October 1996, and XML became a W3C Recommendation on February 2, 1998.

## **XPointer**

Part of the XML linking solution. See XLL (Extensible Linking Language)  
Defines a language that supports addressing into the internal structure, or model, of an XML document. It provides for specific reference to elements, character strings, and other parts of XML documents, whether or not they bear an explicit ID attribute."

## **XSL (Extensible Style Language)**

The XML stylesheet solution. Based on DSSSL and compatible with CSS. Current status: Working draft at the W3C, March 27, 2000 as of this writing.