



# Brief History of HTML, XML and XHTML

Release: 2008-05-22

Table Of Contents

- [Markup](#)
- [What are elements, attributes and entities in a markup language?](#)
- [Continuing Markup](#)
- [HTML goes standard and rakes in accessibility](#)
- [eXtensible Markup Language](#)
- [eXtensible HTML](#)
- [Field Testing XHTML](#)
- [HTML 5](#)



## **Markup**

In the dawn of styled information, when typing meant using typewriters, the content also had written markings to describe parts of a document such as emphasis, layout and comments.

Today written markings or 'markup' are used practically everywhere. You may not even be aware that you are using it! For example today's word processing applications such as OpenOffice.org, Microsoft Word, Microsoft Excel, Corel Wordperfect, all use markup languages to describe your stylizations and layouts with keywords such as bold, italic, tables and lists.

An International Standard of describing Markup Languages is SGML - Standard Generalized Markup Language. This is a robust technical language that is used to create technical project documentation and other information documents.

### ***What are elements, attributes and entities in a Markup Language?***

[\(If you already know, you can skip this section\)](#)

Markup information are the keywords encapsulated or enclosed within a set of pointy brackets '<' and '>'.

An example could be that a sentence "The dog ran across the road, *barking*". Could be enhanced by markup:

"The <animal>dog</animal> ran across the road, <em>barking</em>."

Here, dog has been described with the keyword of 'animal' - it is an animal. Also the word barking has been enhanced by the markup of 'em' stating that it is emphasized. Usually markup have a start 'tag' (<em>) and an end 'tag' (</em>) surrounding the text in question. This together is an Element. Markup is used for document structure. Some elements such as line breaks don't contain any main text so could only require a start tag or 'Empty Element' or 'Void Element'. For instance:

"For the first time this letter had been  
styled to provide increased readability."

Would be marked up as:

"For the first time this letter had been<br>styled to provide increased readability."

Comments could be marked up by <comment>This is a comment</comment>. But SGML provides a standard comment syntax:

## Brief History of HTML, XML and XHTML - Legend Scrolls

-- This is an SGML comment --

These comments cannot have two dashes within the comment because the processor (or parser) would think you are ending the comment before you intended. These comments are used within elements but are typically within its own special Void Element:

<!-- This is a comment -->

Attributes provide extra information to the element or void element by having `name="value"` pairs within the start tag or void element after the keyword.

For example:

```
<object data="images/display.png" type="image/png" width="22" height="22">
  A Portable Network Graphic
</object>
```

An entity reference, or entity, refers to a string of characters like a variable in a programming language. So you could have a commonly used line of text and put it into an entity then just use the entity throughout the document. Usually a set of entities can be declared for a document and each entity refers to a character in the current character set to make sure that a document displays the right character properly. Entities can take the form of a numbered entity: `&#109;` or a hex entity: `&#x00A9;` or even a named entity: `&copy;`. Each of those three entity references would produce a copyright symbol: ©.

### ***Continuing Markup***

In applications such as OpenOffice.org, Microsoft Word, Corel Wordperfect and Microsoft Wordpad markup is used behind the scenes: selecting a word and then activating the bold button or menu item makes the word appear bold. When you activated the button or menu item, the application wrapped the word in a markup element for applying bold styles which is visually interpreted as changing the word to appear bold.

SGML has been used to develop various markup languages such as Rich Text, Cold Fusion and also one of the most popular Internet information documents: HTML (HyperText Markup Language).

HTML was invented by Tim Burners-Lee (now in charge of the World Wide Web Consortium (W3C)) and then a new set of formal specifications developed by the Internet Engineering Task Force (IETF). Later passed the baton to the World Wide Web Consortium (W3C), gained more and more presentational elements and attributes. Also web browser vendors started adding non-standard, browser-specific markup to try and compete with other browser organizations which began a race of buggy, insecure, inaccessible, things rendering differently, poor excuse for web browsers. Finally the global community of Web Designers, Web Developers and general web users in discussion with browser vendors and the W3C began an initiative for Web

Standards.

### ***HTML goes standard and rakes in accessibility***

From version 4, HTML embraces Web Standards including accessibility. Providing a strict flavour that cuts out support for most presentational and other deprecated markup and forces proper structure rules. Strict flavour is to be used with Cascade Stylesheets which provide a more realistic way of providing presentation and layout to HTML documents. A Transitional and Frameset flavours allow the use of presentational and deprecated markup. But all three have features for alternative content for non-text material and support for internationalization.

### ***eXtensible Markup Language***

Browser vendors were screaming to have an easier way of adding more Markup that doesn't break Web Standards and generally in the Information Technology domain wanting a standard flexible document exchange format. The Web community with W3C developed a specification that is a strict subset of SGML: XML (eXtensible Markup Language). Even though it is smaller, the spec is strict to continue the power to create flexible Markup Languages. The strict structure rules include normal elements must have a start and end tag, empty elements are defined with a slash before the ending pointy bracket as `<linebreak/>` (empty elements can be a normal element too as long as there is absolutely nothing for the element content), attribute values must always be quoted and comments must only be used within its own special empty element: `<!-- An XML comment -->`. Also as most SGML documents like HTML are case-insensitive: don't care if the element or attribute names are uppercase, lowercase or mixed, XML does care about the casing: is case sensitive. Each element cannot have more than one attribute with the same name and an entity reference must be any of the three forms Named Entity, Hex Entity or Numbered Entity such as `&apos;` (if defined), `&#x0027;`, `&#39;`. This allows a flexible construct to create markup languages where the XML Author can define their own XML Document elements and attributes. This is the foundation of XML.

### ***eXtensible HTML***

As more and more Markup Languages based on XML were popping up from W3C and other companies; a consensus was brought to update and refine HTML into the XML domain. XHTML, eXtensible HyperText Markup Language, does exactly that. The first version is basically the last HTML version (4.01) exposed as an XML language obeying the strict rules of XML but also basking in the benefits of a more solid structure (less error-prone), standard language

## Brief History of HTML, XML and XHTML - Legend Scrolls

and space handling, native Unicode support and bound to a Namespace allowing other XML structures to be used in the same document.

The XHTML Namespace URI is: <http://www.w3.org/1999/xhtml> and is usually a Default Namespace.

Coding in XHTML 1.0 and following the HTML Compatibility Guidelines will make your webpages not only backwards compatible - browsers that don't support native XHTML, or even XML will process it as HTML; but it is also forwards compatible - already well-formed and valid XML which will allow you to use any XML editing tool or any XML/native XHTML -only browsers currently and in the future plus being ready for newer XHTML versions.

Further XHTML versions include XHTML Basic which is the consensus choice for current and future mobile Internet and low-process Internet. XHTML Basic is made up with the minimal XHTML Modules from the Modular XHTML Collection - breaking up the language into more appropriate modules allows document authors to subset or extend XHTML with other modules and other XML-based languages. XHTML 1.1 is a typical XHTML Family specification using the standard set of modules and is a refined XHTML 1.0 Strict. All Modular XHTML like Basic and version 1.1 are native XHTML - processed by most XML supported browsers. Currently the W3C are developing XHTML 2 which is the ultimate XHTML specification with true accessibilities including a better way of handling non-text features and use of other XML languages such as XForms and XML Events.

### ***Field Testing XHTML***

Not all browsers support native XHTML let alone adding support for XHTML2 as in order to implement HTML into programs like web browsers, HTML had to be altered from the specifications into a Custom HTML form so that it was possible to implement it. On top of that, XHTML2 was not backwards compatible enough and imposes a massive learning curve for both developers of web browsers and webpages. Still there were many issues to address from HTML and XHTML in regards to streamlining the basics of webpages, the differences between SGML and XML processing some text content models and the confusion of MIME Type serving with XHTML. Many webpages seem to have an identity crisis with using a mixture of HTML and XML syntaxes under either a HTML or XHTML Doctype compounded by some XHTML 1.1 webpages using the text/html MIME Type which is invalid.

With the strictness of XHTML came intolerance for errors even with a single subtle misspelling which may be fine in a controlled environment but not in the wild wild web. Many features of validating and extending XHTML were not handled due to some browsers inability to support such features. Plus the emergence of Web Applications requiring persistent storage, dynamic pixel imagery, improved audio and video processing and even basic sectioning of information needed to be addressed.

## Brief History of HTML, XML and XHTML - Legend Scrolls

It is now clear that XML based languages are not best suited for general webpages, websites and web application user interfaces. HTML handles errors more gracefully and features can be incrementally supported.

### **HTML 5**

HTML has updated to handle all these issues and has changed how it is described and what it is modelled on with more focus on reality. As web browsers internally represent most web technologies as an object model and specifically markup languages as a Document Object Model (DOM) plus the fact that web browsers provide the practical implementation of webpages, HTML 5 is described in terms of the DOM rather than a Document Type Definition (DTD). With backwards compatibility with web browsers' implementation of HTML, the basics of HTML 5 or the HTML 5 Subset is already supported in web browsers. This Subset as the base and receiving feedback from web browser vendors, other application vendors, web authors amongst many users of the web, HTML 5 provides an optimum structure and related Application Programming Interfaces (APIs) to represent information on the World Wide Web.

As so many have attempted XHTML structured webpages, and HTML5 is an abstract language, the Custom HTML form of HTML 5 (which is the default) has, in addition to the HTML syntax, support of XML-like syntax to easily allow an upgrade path to this practical webpage language. Plus there is a native XHTML 5 form of HTML 5 that strictly obeys the rules of the latest XML 1.0 and 1.1 for those actually using pure native XHTML.

If you are using native XHTML in a closed controlled environment then you can upgrade to the XHTML 5 Subset by the [From XHTML To XHTML 5 Subset](#) article.

The majority of you will be using old HTML or HTML-Compatible XHTML and so you can upgrade to the HTML 5 Subset by the [From HTML To HTML 5 Subset](#) article.

In either case, or you are new to webpage development and have read the basics on markup from this article, can continue on to [Structure Your Webpages With HTML 5](#) article.

## Brief History of HTML, XML and XHTML - Legend Scrolls

This article and others are available online and in other document formats at:

<http://www.legendscrolls.co.uk/webstandards/>

Author: Peter Davison from [Legend Scrolls](#).

Copyright ©2005-2008 Legend Scrolls and Peter Davison.

The Globe icon from Crystal Project Icons: LGPL, Copyright © [Everaldo](#).

All rights reserved.