

# A BRIEF HISTORY OF MARKUP

HTML IS THE UNIFYING LANGUAGE of the World Wide Web. Using just the simple tags it contains, the human race has created an astoundingly diverse network of hyperlinked documents, from Amazon, eBay, and Wikipedia, to personal blogs and websites dedicated to cats that look like Hitler.

HTML5 is the latest iteration of this lingua franca. While it is the most ambitious change to our common tongue, this isn't the first time that HTML has been updated. The language has been evolving from the start.

As with the web itself, the HyperText Markup Language was the brainchild of Sir Tim Berners-Lee. In 1991 he wrote a document called "HTML Tags" in which he proposed fewer than two dozen elements that could be used for writing web pages.

Sir Tim didn't come up with the idea of using tags consisting of words between angle brackets; those kinds of tags already existed in the SGML (Standard Generalized Markup Language)

format. Rather than inventing a new standard, Sir Tim saw the benefit of building on top of what already existed—a trend that can still be seen in the development of HTML5.

## FROM IETF TO W3C: THE ROAD TO HTML 4

There was never any such thing as HTML 1. The first official specification was HTML 2.0, published by the IETF, the Internet Engineering Task Force. Many of the features in this specification were driven by existing implementations. For example, the market-leading Mosaic web browser of 1994 already provided a way for authors to embed images in their documents using an `<img>` tag. The `img` element later appeared in the HTML 2.0 specification.

The role of the IETF was superseded by the W3C, the World Wide Web Consortium, where subsequent iterations of the HTML standard have been published at <http://www.w3.org>. The latter half of the nineties saw a flurry of revisions to the specification until HTML 4.01 was published in 1999.

At that time, HTML faced its first major turning point.

## XHTML 1: HTML AS XML

After HTML 4.01, the next revision to the language was called XHTML 1.0. The X stood for “eXtreme” and web developers were required to cross their arms in an X shape when speaking the letter.

No, not really. The X stood for “eXtensible” and arm crossing was entirely optional.

The content of the XHTML 1.0 specification was identical to that of HTML 4.01. No new elements or attributes were added. The only difference was in the syntax of the language. Whereas HTML allowed authors plenty of freedom in how

they wrote their elements and attributes, XHTML required authors to follow the rules of XML, a stricter markup language upon which the W3C was basing most of their technologies.

Having stricter rules wasn't such a bad thing. It encouraged authors to use a single writing style. Whereas previously tags and attributes could be written in uppercase, lowercase, or any combination thereof, a valid XHTML 1.0 document required all tags and attributes to be lowercase.

The publication of XHTML 1.0 coincided with the rise of browser support for CSS. As web designers embraced the emergence of web standards, led by The Web Standards Project, the stricter syntax of XHTML was viewed as a “best practice” way of writing markup.

Then the W3C published XHTML 1.1.

While XHTML 1.0 was simply HTML reformulated as XML, XHTML 1.1 was real, honest-to-goodness XML. That meant it couldn't be served with a mime-type of `text/html`. But if authors published a document with an XML mime-type, then the most popular web browser in the world at the time—Internet Explorer—couldn't render the document.

It seemed as if the W3C were losing touch with the day-to-day reality of publishing on the web.

## XHTML 2: OH, WE'RE NOT GONNA TAKE IT!

If Dustin Hoffman's character in *The Graduate* had been a web designer, the W3C would have said one word to him, just one word: XML.

As far as the W3C was concerned, HTML was finished as of version 4. They began working on XHTML 2, designed to lead the web to a bright new XML-based future.

Although the name XHTML 2 sounded very similar to XHTML 1, they couldn't have been more different. Unlike XHTML 1, XHTML 2 wasn't going to be backwards compatible with existing web content or even previous versions of HTML. Instead, it was going to be a pure language, unburdened by the sloppy history of previous specifications.

It was a disaster.

## THE SCHISM: WHATWG TF?

A rebellion formed within the W3C. The consortium seemed to be formulating theoretically pure standards unrelated to the needs of web designers. Representatives from Opera, Apple, and Mozilla were unhappy with this direction. They wanted to see more emphasis placed on formats that allowed the creation of web applications.

Things came to a head in a workshop meeting in 2004. Ian Hickson, who was working for Opera Software at the time, proposed the idea of extending HTML to allow the creation of web applications. The proposal was rejected.

The disaffected rebels formed their own group: the Web Hypertext Application Technology Working Group, or WHATWG for short.

## FROM WEB APPS 1.0 TO HTML5

From the start, the WHATWG operated quite differently than the W3C. The W3C uses a consensus-based approach: issues are raised, discussed, and voted on. At the WHATWG, issues are also raised and discussed, but the final decision on what goes into a specification rests with the editor. The editor is Ian Hickson.

On the face of it, the W3C process sounds more democratic and fair. In practice, politics and internal bickering can bog down progress. At the WHATWG, where anyone is free to contribute but the editor has the last word, things move at a faster pace. But the editor doesn't quite have absolute power: an invitation-only steering committee can impeach him in the unlikely event of a Strangelove scenario.

Initially, the bulk of the work at the WHATWG was split into two specifications: Web Forms 2.0 and Web Apps 1.0. Both specifications were intended to extend HTML. Over time, they were merged into a single specification called simply HTML5.

## REUNIFICATION

While HTML5 was being developed at the WHATWG, the W3C continued working on XHTML 2. It would be inaccurate to say that it was going nowhere fast. It was going nowhere very, very slowly.

In October 2006, Sir Tim Berners-Lee wrote a blog post in which he admitted that the attempt to move the web from HTML to XML just wasn't working. A few months later, the W3C issued a new charter for an HTML Working Group. Rather than start from scratch, they wisely decided that the work of the WHATWG should be used as the basis for any future version of HTML.

All of this stopping and starting led to a somewhat confusing situation. The W3C was simultaneously working on two different, incompatible types of markup: XHTML 2 and HTML 5 (note the space before the number five). Meanwhile a separate organization, the WHATWG, was working on a specification called HTML5 (with no space) that would be used as a basis for one of the W3C specifications!

Any web designers trying to make sense of this situation would have had an easier time deciphering a movie marathon of *Memento*, *Primer*, and the complete works of David Lynch.

## XHTML IS DEAD: LONG LIVE XHTML SYNTAX

The fog of confusion began to clear in 2009. The W3C announced that the charter for XHTML 2 would not be renewed. The format had been as good as dead for several years; this announcement was little more than a death certificate.

Strangely, rather than passing unnoticed, the death of XHTML 2 was greeted with some mean-spirited gloating. XML naysayers used the announcement as an opportunity to deride anyone who had ever used XHTML 1—despite the fact that XHTML 1 and XHTML 2 have almost nothing in common.

Meanwhile, authors who had been writing XHTML 1 in order to enforce a stricter writing style became worried that HTML5 would herald a return to sloppy markup.

As you'll soon see, that's not necessarily the case. HTML5 is as sloppy or as strict as you want to make it.

## THE TIMELINE OF HTML5

The current state of HTML5 isn't as confusing as it once was, but it still isn't straightforward.

There are two groups working on HTML5. The WHATWG is creating an HTML5 specification using its process of “commit then review.” The W3C HTML Working Group is taking that specification and putting it through its process of “review then commit.” As you can imagine, it's an uneasy alliance. Still, there seems to finally be some consensus about that pesky

“space or no space?” question (it’s HTML5 with no space, just in case you were interested).

Perhaps the most confusing issue for web designers dipping their toes into the waters of HTML5 is getting an answer to the question, “when will it be ready?”

In an interview, Ian Hickson mentioned 2022 as the year he expected HTML5 to become a proposed recommendation. What followed was a wave of public outrage from some web designers. They didn’t understand what “proposed recommendation” meant, but they knew they didn’t have enough fingers to count off the years until 2022.

The outrage was unwarranted. In this case, reaching a status of “proposed recommendation” requires two complete implementations of HTML5. Considering the scope of the specification, this date is incredibly ambitious. After all, browsers don’t have the best track record of implementing existing standards. It took Internet Explorer over a decade just to add support for the `abbr` element.

The date that really matters for HTML5 is 2012. That’s when the specification is due to become a “candidate recommendation.” That’s standards-speak for “done and dusted.”

But even that date isn’t particularly relevant to web designers. What really matters is when browsers start supporting features. We began using parts of CSS 2.1 as soon as browsers started shipping with support for those parts. If we had waited for every browser to completely support CSS 2.1 before we started using any of it, we would still be waiting.

It’s no different with HTML5. There won’t be a single point in time at which we can declare that the language is ready to use. Instead, we can start using parts of the specification as web browsers support those features.

Remember, HTML5 isn't a completely new language created from scratch. It's an evolutionary rather than revolutionary change in the ongoing story of markup. If you are currently creating websites with any version of HTML, you're already using HTML5.