Why HTML5?
Why not XHTML2?

Learning from history how to drive the future of the Web

Michael(tm) Smith
mike@w3.org
http://people.w3.org/mike
sideshowbarker on Twitter, GitHub, &c

W3C Interaction domain

- HTML Working Group
- Web Applications Working Group
- CSS Working Group
- SVG Working Group
- more...

From 1997 through the end of 2006, work on HTML within the W3C focused exclusively on the XHTML dialect.

A government in exile...

From June 2004 to March 2007, work on the (non-XHTML) HTML language took place outside of the W3C.

About HTML5 (and HTML forms)...

HTML5 in the words of the W3C HTML WG...

HTML design principles

http://w3.org/TR/html-design-principles/

HTML design principles

- Support existing content
- Ensure interoperability
- Precisely define UA behavior
- Handle errors (non-draconically)
- Evolution not revolution

“Draconically”=
“Draconian”=
“catch fire and fail”

About XHTML2 (and XForms)...

XHTML2 in the words of the W3C XHTML WG...

XHTML2 Design Aims

http://w3.org/TR/xhtml2/introduction.html#aims

XHTML2 Design Aims

- Use existing XML facilities rather than duplicating them (implies namespace support)
- Less scripting (vs declarative approach)
- Integration with Semantic Web

What does
“declarative” mean?

Declarative programming
success story: SVG
(XSLT also? XForms?)

HTML5 and XHTML2
in contrast...

Things HTML5 doesn't do

- Does not favor XML facilities
- Does not avoid scripting
- Does not consider integration with the SemWeb a priority
- No arbitrary namespaces

Things XHTML2 did not do

- Does not support existing content in the same way that HTML5 does
- Does not precisely define UA behavior
- Does not handle errors non-draconically (uses “catch fire and fail” error handling)

Important point:
XHTML2 was a
different language than
XHTML1

…“different language” in that XHTML2 does not fully support existing XHTML1 content (not backward compatible)

A representative statement about the difference in philosophy:

"HTML is the assembly language of the Web."

Important point: In some cases HTML5 offers a choice of *both* declarative and scripting approaches.

About error handling...

Which of these are errors?

- Well-formed XML:
  ```
  <input disabled="disabled">
  ```
- Empty attribute: `<input disabled>`
- Without quotes: `<input value=yes>`
- Single quotes: `<input type='checkbox'>`
- Double quotes: `<input name="be evil">`

This is a real error

misnested tags

HTML5 parsers can handle real errors interoperably and gracefully.

Why is it important to handle errors?

More than 93% of Alexa Top 500 sites contain HTML conformance errors.

A little history...

(About draconian error handling in XML)

I think users and application builders should have a choice with what they do with invalid data... I therefore plan to continue to provide it even if the spec says that this is non-conforming.

April 1997

After careful consideration, the HTML Working Group has decided that the goals for the next generation of forms are incompatible with preserving backwards compatibility with browsers designed for earlier versions of HTML.

August 1999

W3C has no intention to extend HTML 4 as such. Instead, further work is focusing on a reformulation of HTML in XML

November 1999

...while the ancestry of XHTML 2 comes from HTML 4, XHTML 1.0, and XHTML 1.1, it is not intended to be backward compatible with its earlier versions

August 2002

XHTML 2.0 seems to me the live proof that something is going wrong at W3C... I strongly suggest dropping all XHTML 2.0 efforts in favor of a new “xHTML 5.0” language. Clearly a successor to HTML 4, feature-oriented, made for the *web.*

December 2002

The W3C had so far failed to address a need in the Web community: There is no language for Web applications... I intend to do something about this (hopefully within a W3C context, although that will depend on the politics of the situation).

January 2004

The dream of a new web, based on XHTML+SVG+SMIL+XForms, is just that — a dream... The best way to help the Web is to incrementally improve the existing web standards... so that web content authors can actually deploy new formats interoperably.

June 2004

We need to specify error handling behavior to ensure interoperability "even in the face of documents that do not comply to the letter of the specifications".

Authors will write invalid content regardless of what we spec. So the spec states “what authors must not do, and then tells implementors what they must do when an author does it anyway”.

It is necessary to evolve HTML incrementally. The attempt to get the world to switch to XML, including quotes around attribute values and slashes in empty tags and namespaces all at once didn't work...

October 2006

more HTML history

http://w3.org/html/wg/wiki/History

HTML5 has a major focus on facilitating use of a browser as a Web application platform (or Web application runtime environment).

XHTML2 had a major focus on providing a general-purpose document language with declarative mechanisms to enable interactive features.

HTML5 support

- specific native browser support being implemented by all major browser vendors
- spec remains in active development

XHTML2 support

- no specific client-side native browser support from any major browser vendor
- … but was intended to be possible to "bolt on" some level of support using CSS+JS
- last WD: 2006

The bottom line…

HTML5 is the only HTML dialect that will be natively supported in browsers on the client side.

Some HTML5 differences...

HTML5 defines HTML as an abstract language with two standard syntaxes supported by browsers:

- a text/html syntax, with parsing rules defined by the HTML5 spec
- an XML syntax, with parsing rules defined by the XML spec

Similarly, applications can potentially represent HTML in memory in any number of ways.

http://software.hixie.ch/utilities/js/live-dom-viewer/

However, there's only one standard in-memory representation supported by browsers: The W3C DOM.

The HTML5 spec precisely defines the DOM representation that browsers must use to represent HTML content in memory.

<!DOCTYPE html PUBLIC
"-//W3C//DTD XHTML 1.0
Transitional//EN"
"http://www.w3.org/TR/xhtml1/
DTD/xhtml1-transitional.dtd">

<!DOCTYPE html>

```
<meta http-equiv="Content-Type"
 content="text/html; charset=utf-8">
```

```
<meta charset="utf-8">
```

Important point:

HTML5 includes XHTML

(but *not* XHTML**2**...)

Frequently Asked Questions (FAQ) about the future of XHTML

http://www.w3.org/2009/06/xhtml-faq.html

Michael(tm) Smith
mike@w3.org
http://people.w3.org/mike
sideshowbarker on Twitter, GitHub, &c