



HTML+XML: The W3C HTML/XML Task Force

**Norman Walsh
MarkLogic Corporation**

Agenda

- **History**
- **Use Cases**
- **Conclusions**
- **Ask for your help**

Disclaimer

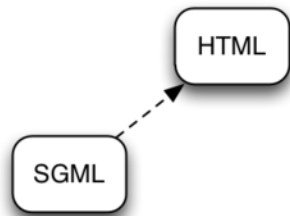
Facts are facts. But any opinions expressed are the opinions only of myself and may or may not reflect the opinions of anybody else with whom I may or may not have discussed the issues at hand.

In particular, I do not claim that what follows is necessarily the consensus opinion of the HTML/XML Task Force.

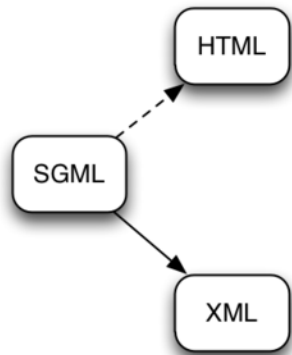
History (In the beginning...)

SGML

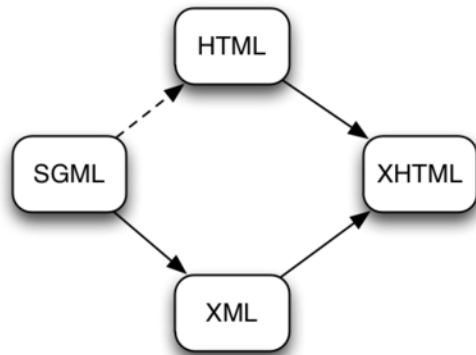
History (HTML)



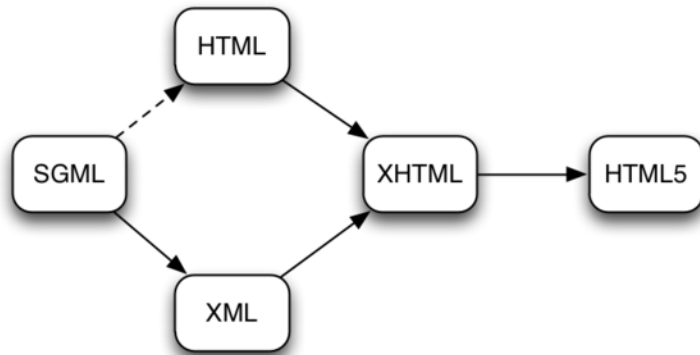
History (XML)



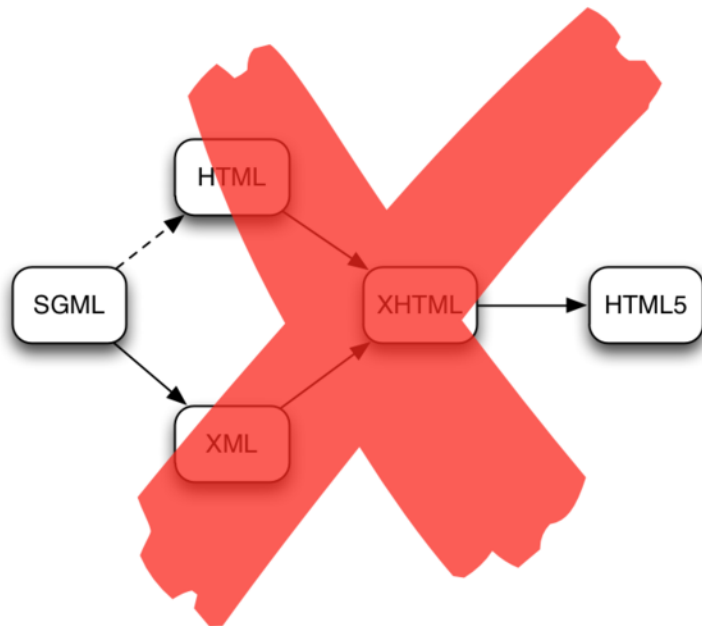
History (XHTML)



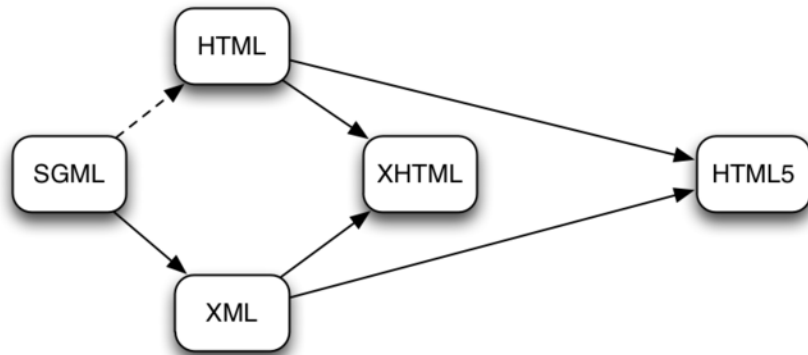
History (Alternative history 1)



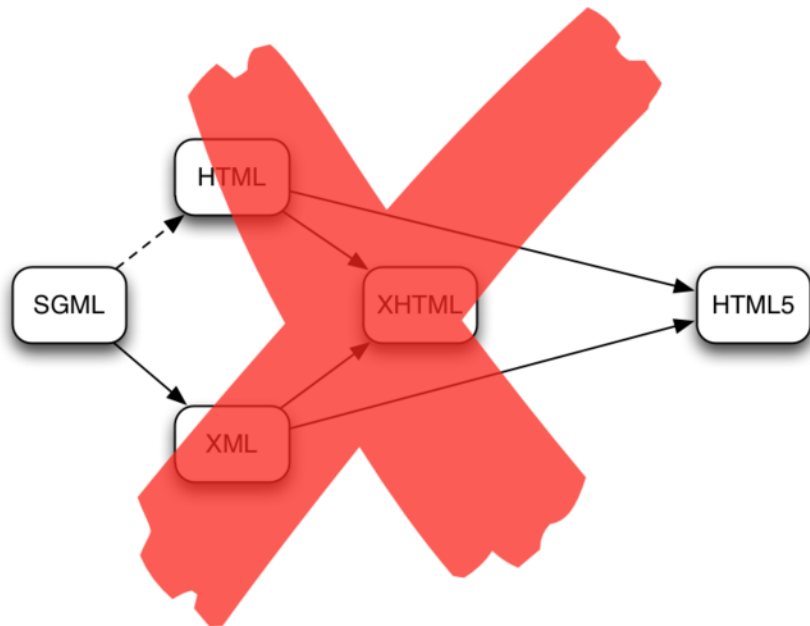
History (Bzzzt! No! ...)



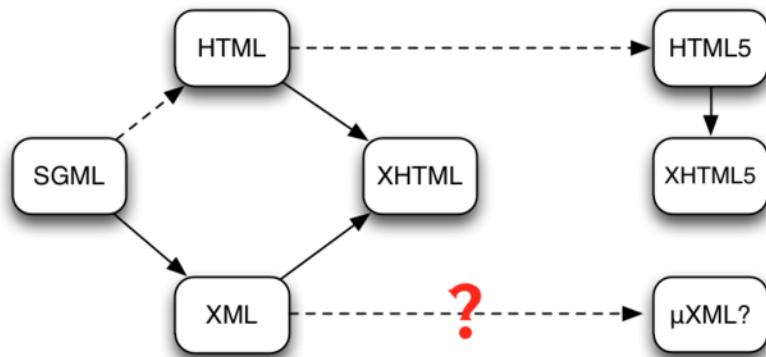
History (Alternative history 2)



History (Bzzzt! No! ...)



Present day



What's wrong?

Perhaps the biggest challenge that faces the W3C's technical work on the Web is the growing chasm between HTML and XML

—T. V. Raman

- **TAG Issue-67: HTML and XML Divergence**
 - > **Tag soup**
 - > **Namespaces**
 - > **Syntactic differences (quoted attribute values)**
 - > **DOM differences (`tbody` insertion)**
 - > **Distributed extensibility**

DOM differences (markup)

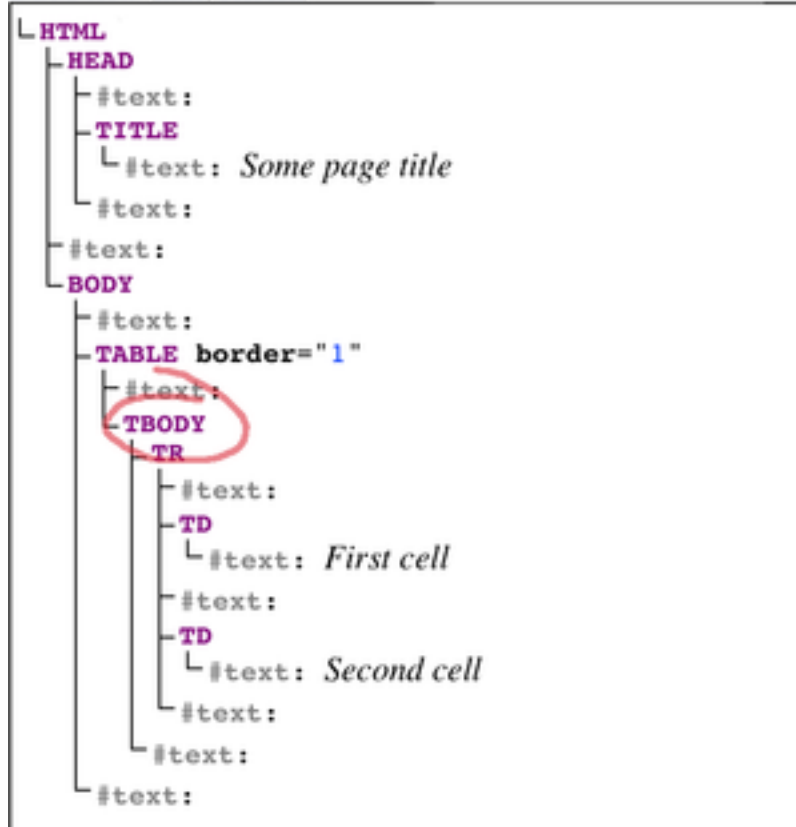
Markup to test ([permalink](#), [hide](#)):

```
<!DOCTYPE html>
<html xmlns="http://www.w3.org/1999/xhtml">
  <head>
    <title>Some page title</title>
  </head>
  <body>
    <table border="1">
      <tr>
        <td>First cell</td>
        <td>Second cell</td>
      </tr>
    </table>
  </body>
</html>
```

(Using the HTML5 Live DOM Viewer)

DOM differences (DOM)

DOM view ([hide](#), [refresh](#)):



Distributed extensibility

- **In practice**
 - > **SVG in W3C and HTML WG**
 - > **RDFa in W3C not in HTML WG**
 - > **FBML not in W3C**
- **In practice**
 - > **Namespaces**
- **In principle**

W3C TAG HTML/XML Task Force

- Robin Berjon
- Michael Champion
- James Clark
- John Cowan
- Michael Kay
- Anne van Kesteren
- Yves Lafon (staff contact)
- Noah Mendelsohn
- Henri Sivonen
- Norman Walsh (chair)

Use Case: Consume HTML

How can an XML toolchain be used to consume HTML?

- Author HTML5 with polyglot markup.
- Add an HTML5 parser to the front end of your toolchain.
 - > Doesn't solve the pernicious “`document.write`” problem or other script-related problems.
 - > But short of running a JavaScript engine on the content, nothing is likely to solve those problems.

Use Case: Consume XML

How can an HTML5 toolchain be used to consume XML?

- **HTML5 tools won't be designed to deal with arbitrary element names in arbitrary namespaces.**
- **Transforming to HTML5 is probably the best route.**
 - > **Even a partial transformation to remove namespaces, PIs, etc. might prove valuable.**
- **It's probably best not to encourage users to imagine this will be broadly successful.**

Use Case: Embed HTML

How can islands of HTML be embedded in XML?

- Use the XML serialization of HTML5.
- Escape the markup.
- Rely on more sophisticated multipart-message handling systems.

Regardless, some care may be necessary. How are the HTML islands going to be processed? By “clipping” them out and processing them with an HTML5 tool, or by passing the whole DOM to the tool?

Use Case: Embed XML

How can islands of XML be embedded in HTML?

- **The HTML5 parser interprets unfamiliar markup as an error and corrects for it.**
- **Correction can include changing the order and nesting of elements.**
- **Practically speaking: you can't embed a “naked” island of XML in HTML5.**

Use Case: Embed XML

Putting clothes on your XML

```
<script type="application/xml">
  <data>
    <title>Your XML</title>
    <gpx xmlns="http://www.topografix.com/GPX/1/1">
      <wpt lat="50.077484" lon="14.443800">
        <ele>200.08</ele>
        <time>2007-01-06T17:33:04Z</time>
        <name>001</name>
        <sym>Restaurant</sym>
      </wpt>
    </gpx>
  </data>
</script>
```

Use Case: Embed XML

Belt and suspenders

```
<script type="application/xml">
  &lt;data&gt;
    &lt;title&gt;Your XML&lt;/title&gt;
    &lt;gpx xmlns="http://www.topografix.com/GPX/1/1"&gt;
      &lt;wpt lat="50.077484" lon="14.443800"&gt;
        &lt;ele&gt;200.08&lt;/ele&gt;
        &lt;time&gt;2007-01-06T17:33:04Z&lt;/time&gt;
        &lt;name&gt;001&lt;/name&gt;
        &lt;sym&gt;Restaurant&lt;/sym&gt;
      &lt;/wpt&gt;
    &lt;/gpx&gt;
  &lt;/data&gt;
</script>
```

Use Case: Embed XML

Use the HTML5 extensibility mechanisms

```
<div class="data">
  <h1 class="title">Your XML</h1>
  <div class="gpx">
    <div class="wpt"
      data-lat="50.077484" data-lon="14.443800">
      <span class="ele">200.08</span>
      <span class="time">2007-01-06T17:33:04Z</span>
      <span class="name">001</span>
      <span class="sym">Restaurant</span>
    </div>
  </div>
</div>
```


Use Case: a more forgiving XML

How can XML be made easier to use?

- **Guaranteeing that you'll get well-formed XML out of naive attempts to generate it with “print” statements is tricky.**
- **Rules could be devised for providing some degree of markup minimization/error correction in XML.**
- **It's possible to consider other simplifications as well, for namespaces, for example.**

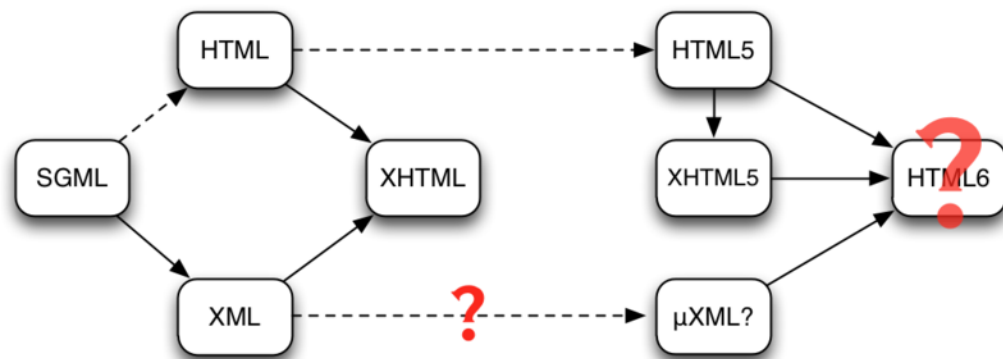
Use Case: XForms?

- **Is XForms a use case or a specific solution to the use case of better form controls?**
- **Is XForms different in some substantial way than the general “embedding XML in HTML” use case?**

What you can do

- **Review the Task Force Report**
- **Talk to the communities you know about the use cases *they* have.**
- **Report use cases that you think are not met.**

Future history?



Thank you!

- **Join the mailing list**
 - > <http://lists.w3.org/Archives/Public/public-html-xml/>
- **Read the draft report**
 - > <http://www.w3.org/2010/html-xml/snapshot/report.html>