Introducing MicroXML

Uche Ogbuji, XML Prague, 10 February 2013

uche@ogbuji.net

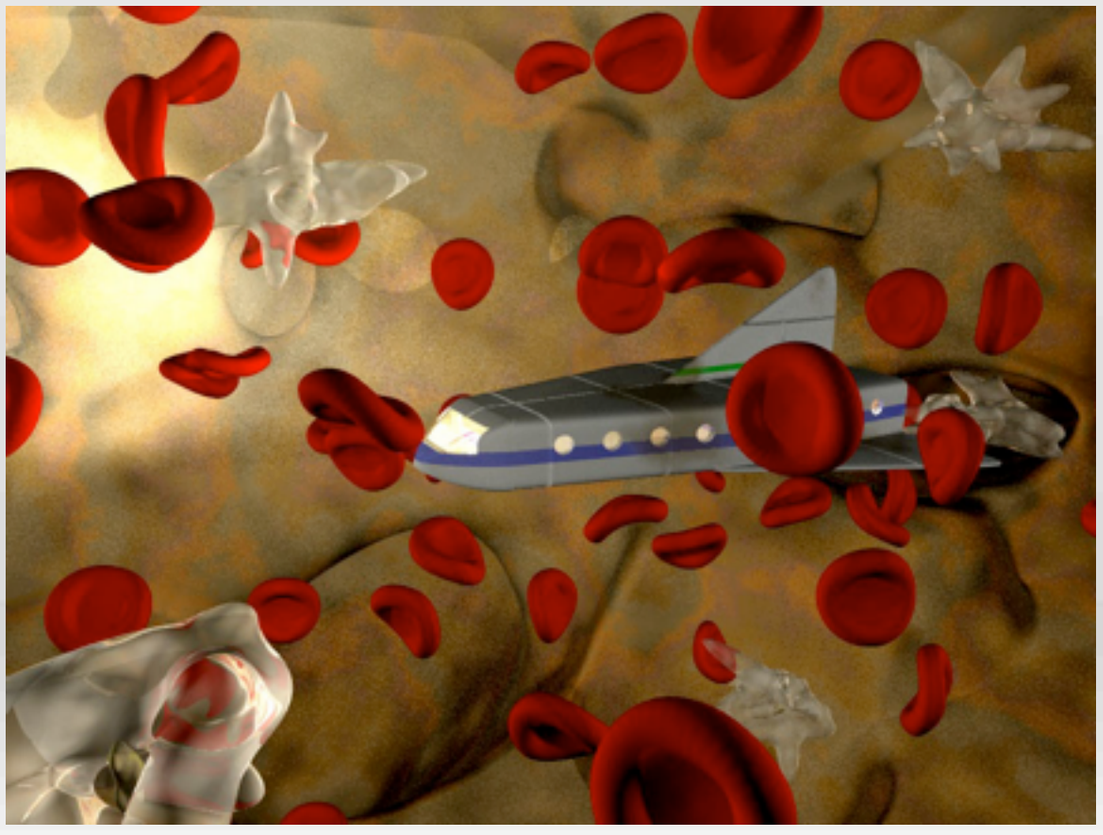
Twitter: @uogbuji

Presenters - Uche

- Pioneer/expert in XML, RDF/semantic web & SOA/Web services
  - Data architecture best practices
- Prolific writer & presenter
- Partner at **Zepheira**
  - Leads data architecture practice
- Lead developer of influential open specifications & open source software
  - EXSLT
  - MicroXML
  - 4Suite / Akara (+Amara)
  - Python/XML

Fantastic Voyage

- Know your history (because those who don't…)
- A micro tour
- OK now what?



XML: 15 Years Old Today!

Extensible Markup Language (XML) 1.0

www.w3.org/TR/1998/REC-xml-19980210

Reader

Apple Yahoo! Google Maps YouTube Wikipedia News Popular

W3C

REC-xml-19980210

Extensible Markup Language (XML) 1.0

W3C Recommendation 10-February-1998

This version:

http://www.w3.org/TR/1998/REC-xml-19980210
http://www.w3.org/TR/1998/REC-xml-19980210.xml
http://www.w3.org/TR/1998/REC-xml-19980210.html
http://www.w3.org/TR/1998/REC-xml-19980210.pdf
http://www.w3.org/TR/1998/REC-xml-19980210.ps

Latest version:

http://www.w3.org/TR/REC-xml

Previous version:

http://www.w3.org/TR/PR-xml-971208

Editors:

Tim Bray (Textuality and Netscape) <tbray@textuality.com>
Jean Paoli (Microsoft) <jeanpa@microsoft.com>
C. M. Sperberg-McQueen (University of Illinois at Chicago) <cmsmcq@uic.edu>

Abstract

The Extensible Markup Language (XML) is a subset of SGML that is completely described in this document. Its goal is to enable generic SGML to be served, received, and processed on the Web in the way that is now possible with HTML. XML has been designed for ease of implementation and for interoperability with both SGML and HTML.

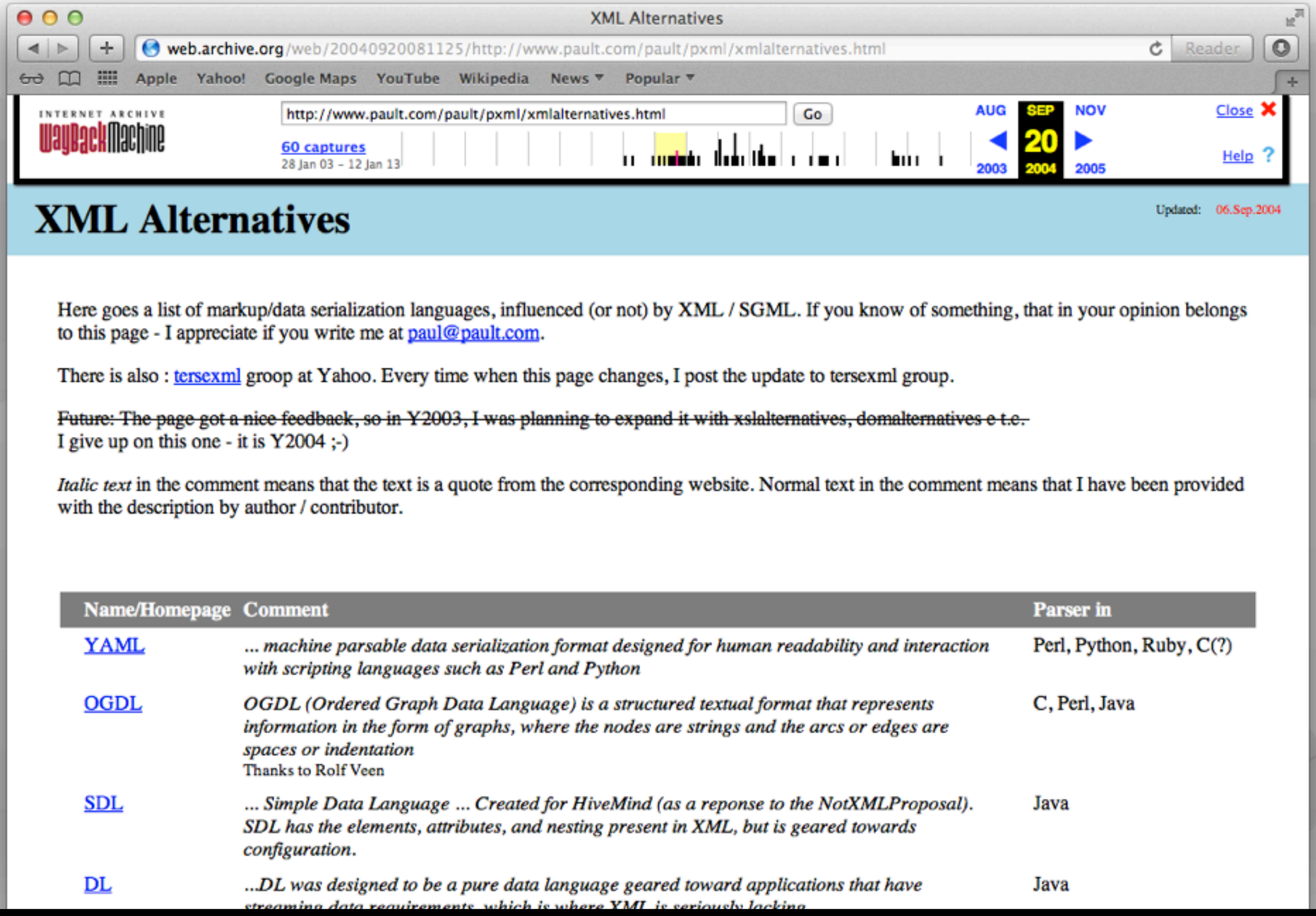
Status of this document

This document has been reviewed by W3C Members and other interested parties and has been endorsed by the Director as a W3C Recommendation. It is a stable document and may be used as reference material or cited as a normative reference from another document. W3C's

Didn't take long before the assassination attempts

Scene: Ides of March, 1998

PaulT's XML Alternatives



http://web.archive.org/web/20040920081125/http://www.pault.com/pault/pxml/xmlalternatives.html

"XML Alternatives"

Just the first section, 2004

- YAML
- OGDL
- SDL
- DL
- Boulder
- ONX
- JSON
- SMEL
- Property Lists
- ATerms
- LMNL
- JITTs
- ConsiceXML
- SML [Common XML]
- TexMecs
- A Specification Language
- Waterken Doc
- UBF
- Xqueeze

SML/Common XML

1999 on XML-DEV: Don Park's tentative definition of SML, a subset of Canonical XML, but having:

- No Attributes
- No Processing Instructions
- No Document Type Declaration
- No non-character entity-references
- No CDATA marked sections
- Support for only UTF-8 character encoding
- No optional features

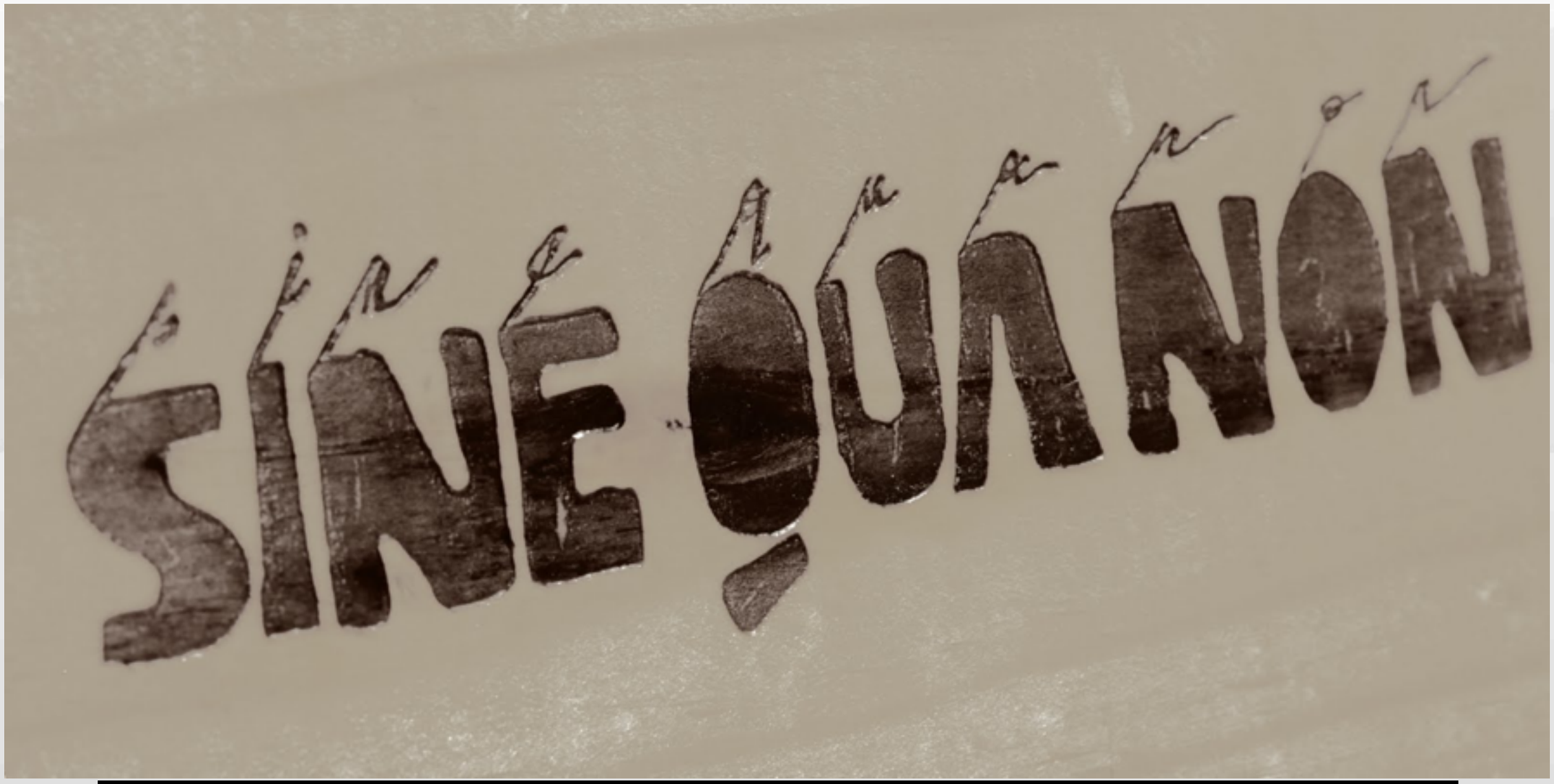
Yay! Mixed content survived!

Syntactic sugar?

Entities?

PIs?

Attributes?



SINE QUA NON XML?

DTDs?

Unicode++?

Comments?

Colons?

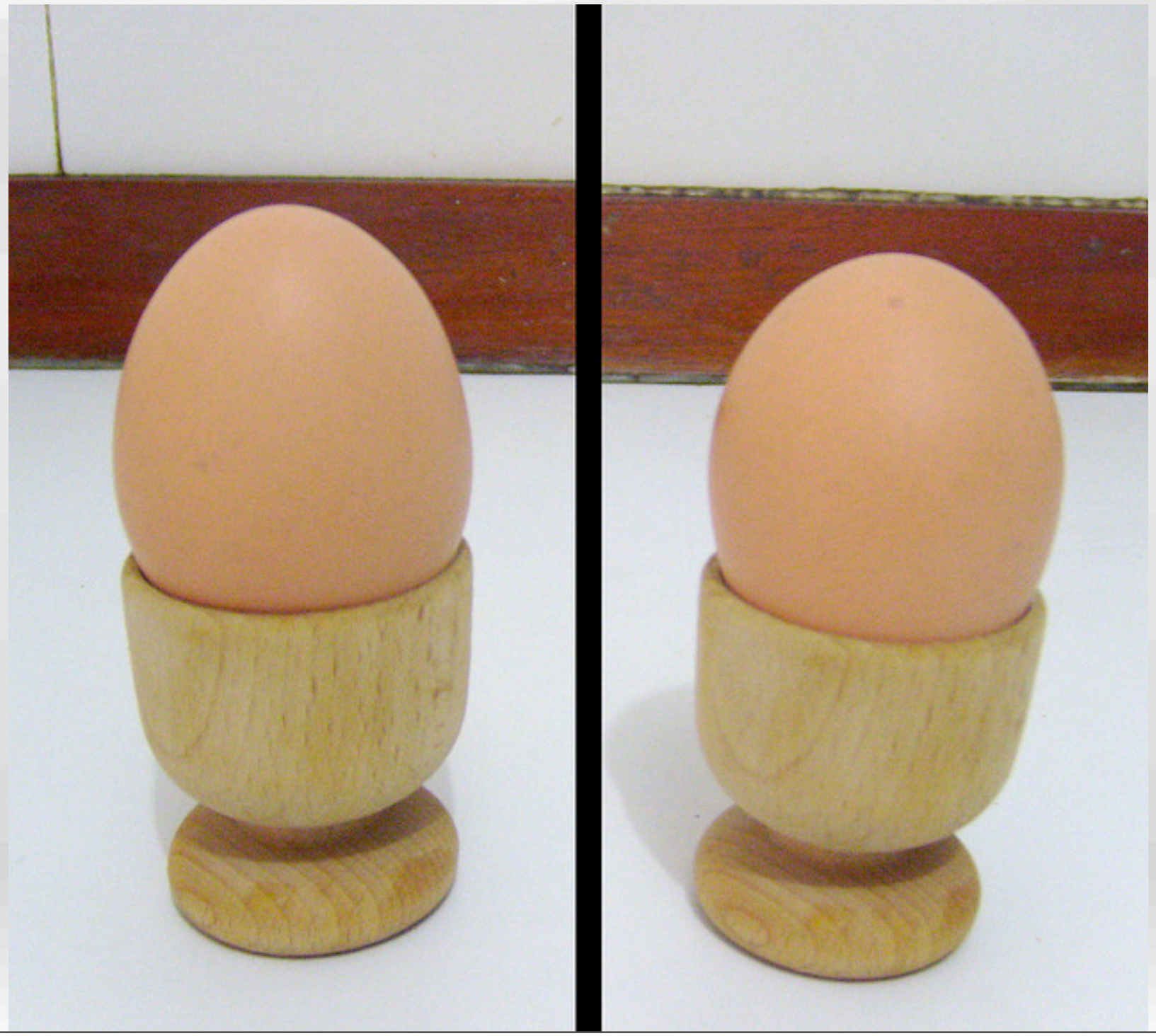
Mixed content?

IMNSHO

SINE QUA NON XML?

- Attributes
  - Metadata, anchors for processing
- Mixed content
  - Computer people like neat, "square" data, but so much real-world information is in ragged, semi-structured documents
- Unicode
  - Because I've wrassled e.g. ASCII/EBCDIC (*nightmares*!) & prefer to defer to a standardized worse-is-better

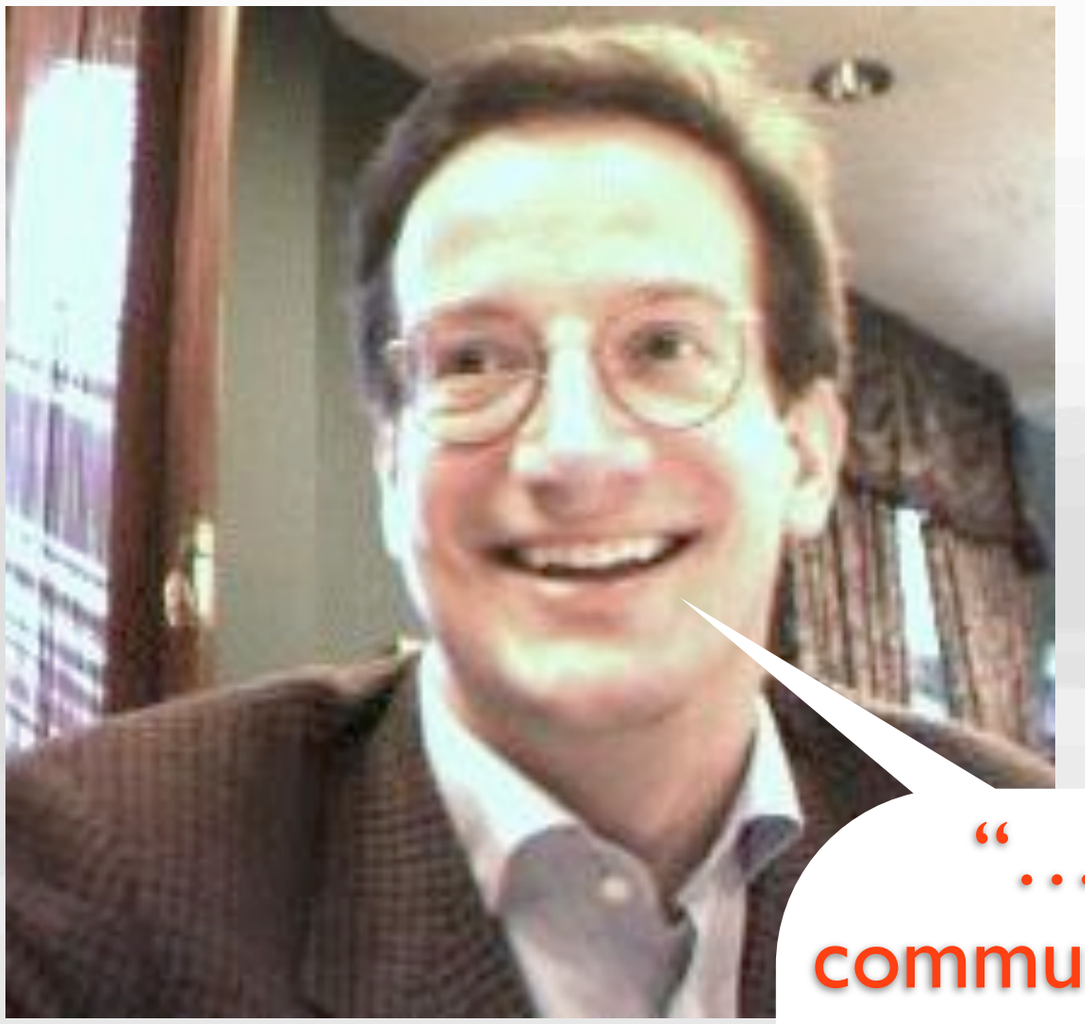
But for the longest time such debate was all…



Then one day in 2010…

Twitter & Foursquare dropped XML from APIs. Went to JSON only

http://norman.walsh.name/2010/11/17/deprecatingXML



"…there's a gulf between the XML community and the broader Web developer community; all the stuff that's been piled on top of XML, together with the huge advances in the Web world in HTML5, JSON and JavaScript, have combined to make XML be perceived as an overly complex, enterprisey technology, which doesn't bring any value to the average Web developer."

http://blog.jclark.com/2010/11/xml-vs-web_24.html

December, 2010

- Much discussion/debate on future of XML, XML 2.0, etc.
- James Clark came along and crystallized the matter

MicroXML is born:

http://blog.jclark.com/2010/12/microxml.html

...

http://blog.jclark.com/2010/12/more-on-microxml.html

2011-

- John Cowan drafts a spec on his blog, with his own variations on James's proposed grammar
- John Cowan implements his MicroXML flavor as MicroLark

MicroXML is first implemented:

http://recycledknowledge.blogspot.com/2011/01/microlark-parser.html

2012-

- Media catches up to MicroXML (my articles in developerWorks lead to SD Times coverage)

MicroXML gets noticed:

http://www.ibm.com/developerworks/library/x-microxml1/

http://sdt.bz/36778

- A W3C community group is launched, led by James Clark, John Cowan & Uche Ogbuji

MicroXML goes semi-official:

http://www.w3.org/community/microxml/

A look at MicroXML

This won't take long.

That's the idea!

MicroXML, as agreed

Open membership, 36 participants

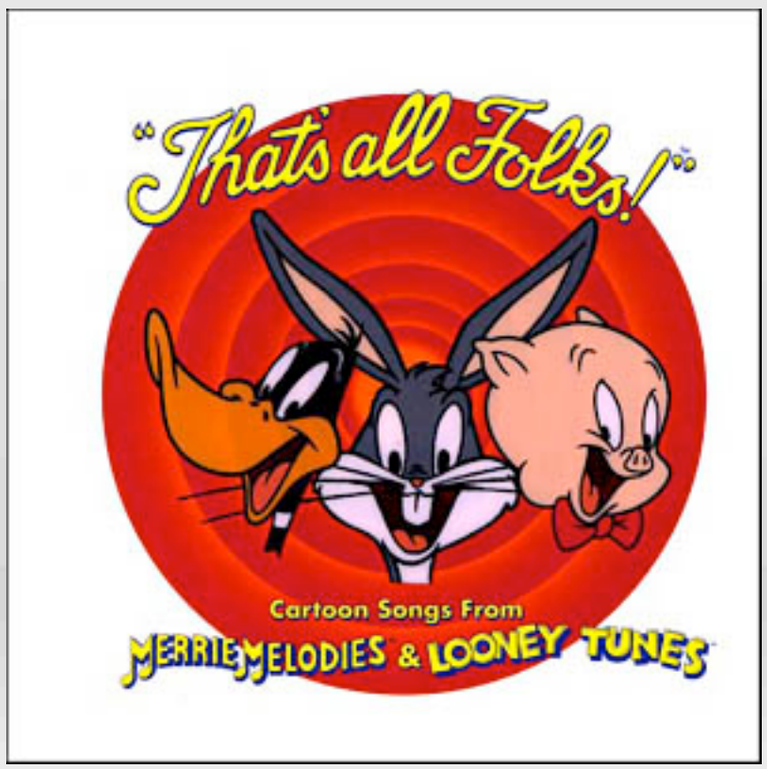
- Draconian error handling not required
- No DTD features or XML declaration
- No colons in element & attribute names, no “xmlns”
- Only hex character entity references
- No PIs or CDATA Sections
- Only UTF-8 character encoding
- Tweaks to character syntax/model

#3 means no XML Namespaces. Yay!

So basically it's

- Elements
- Attributes
- Content

- Structure
- Metadata
- Content



Ill formed? Now what?!

```
<para>Hello, I claim to be
<strong>MicroXML</para>
```

No namespaces? Now what?!

- First of all, are you sure you need 'em, or have you just always been told that?
- No one understands them, or gets them right
- They're just tripling everyone's code
- Where do you stand on the Treaty of Wulai?

Can we coax that poisoned candy from your hands?

- Hey, consider this different perspective on layering?
  - Ever heard the XLink vs HLink story?
- Use out of band annotations (e.g. in schemata) to interpret vocabulary "islands" in the XML
  - OK ever heard of Architectural Forms?
  - No? OK just use processing pipelines, XSLT…

OK screw that. Bad cop's here

- Bottom line: Namespaces can **always** be replaced by well-designed transformation within the processing chain
  - That might sound tricky, but it's much less tricky than XMLNS
  - They're gone from MicroXML. Consider it an intervention from your loved ones. You'll thank us later.

Recent activity

- First CG draft October 2012
- Several implementations since then
- Many follow-up discussions
  - Error recovery
  - Microschemata
  - Microtransforms
  - Micro arch forms

MicroLark Parser

- First implementation
- Written in Java, by John Cowan
- Based on Lark, by Tim Bray
- Push, pull and tree APIs
- About a third as much code as Lark

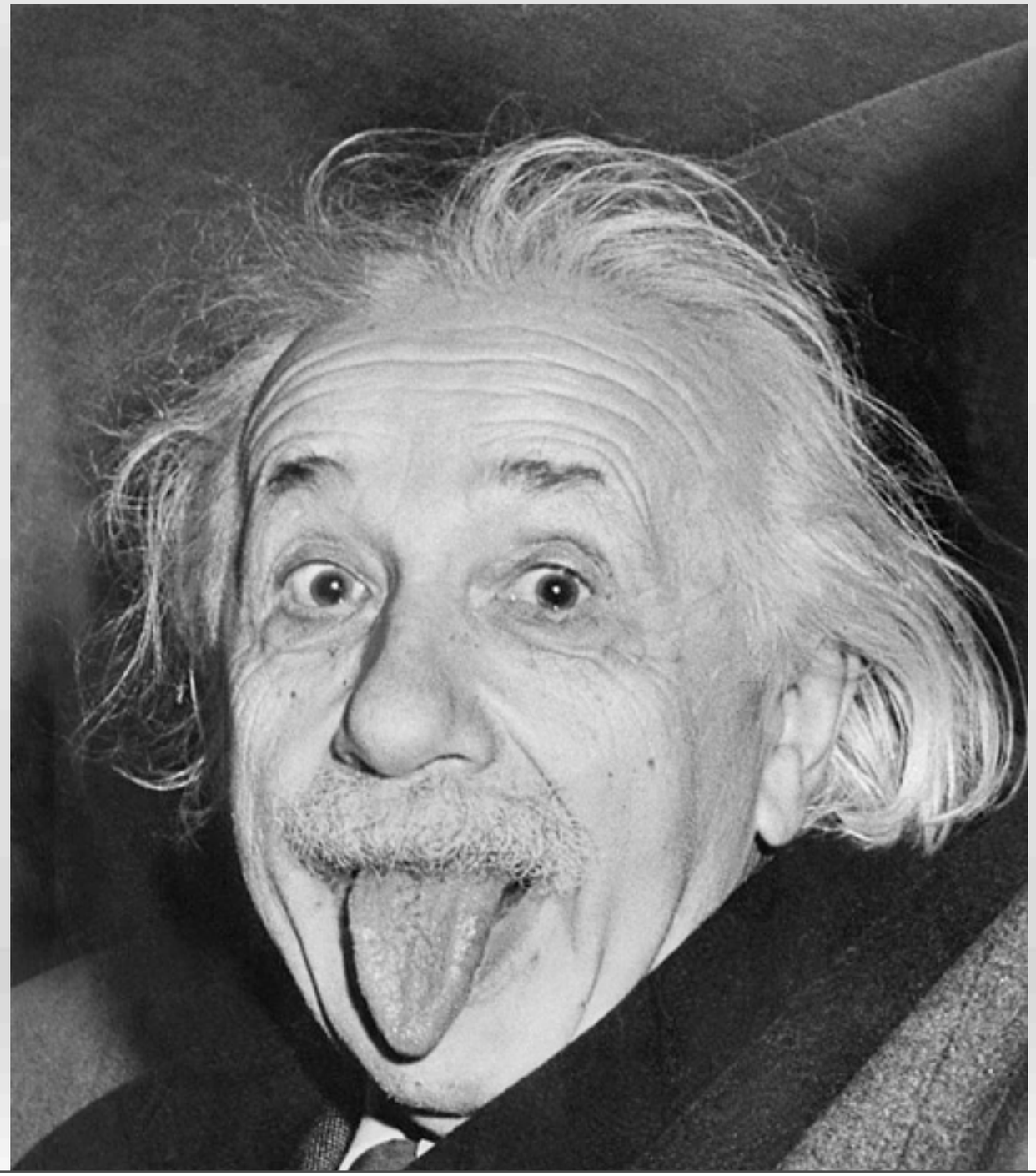
Other implementations

- MicroXML-js, a little MicroXML parser in JavaScript, by James Clark
  - Also seems to be a Java version in incubation
- microxml-parsers, by James Fuller a MicroXML parser back end in REx
  - Generated parsers in a number of languages: microxml.hpp (C++), microxml.xquery (XQuery), microxml.js (Javascript), microxml.java (Java)
- Amara 3 (by Uche Ogbuji)

So who is it for?

- People who like JSON?
  - Probably not
- People who like XML?
  - Maybe

It's for those who just don't give a damn anymore!

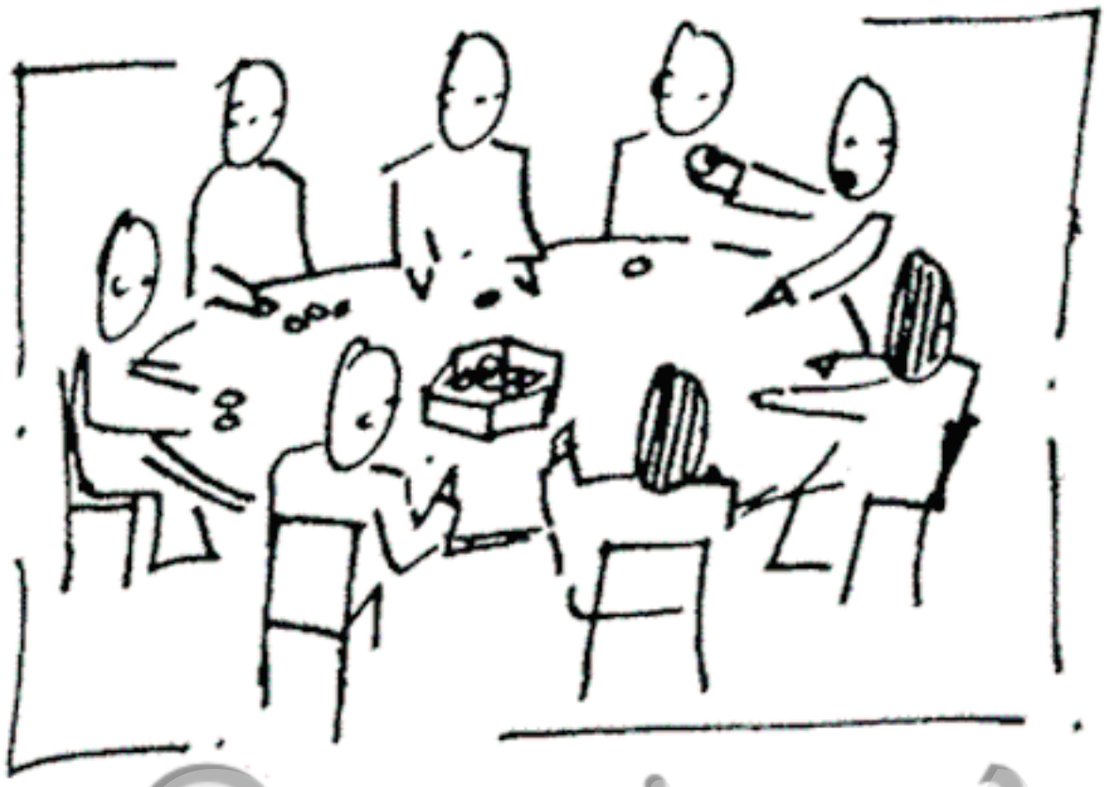


You're there, or almost there. Admit it!

- Try it!
  - Won't take very long
- Implement it!
  - Won't take very long

Join the rebellion!

You'll thank us later!



Questions?

Discussion!

uche@ogbuji.net
@uogbuji

http://www.w3.org/community/microxml/

http://zepheira.com