UPFRONT XHTML:
For Workflow, Not Just the Web

Bill Kasdorf

Vice President, Apex Content Solutions

General Editor, *The Columbia Guide to Digital Publishing*

Chair, BISG Content Structure Working Group

If HTML is so ubiquitous and essential . . .

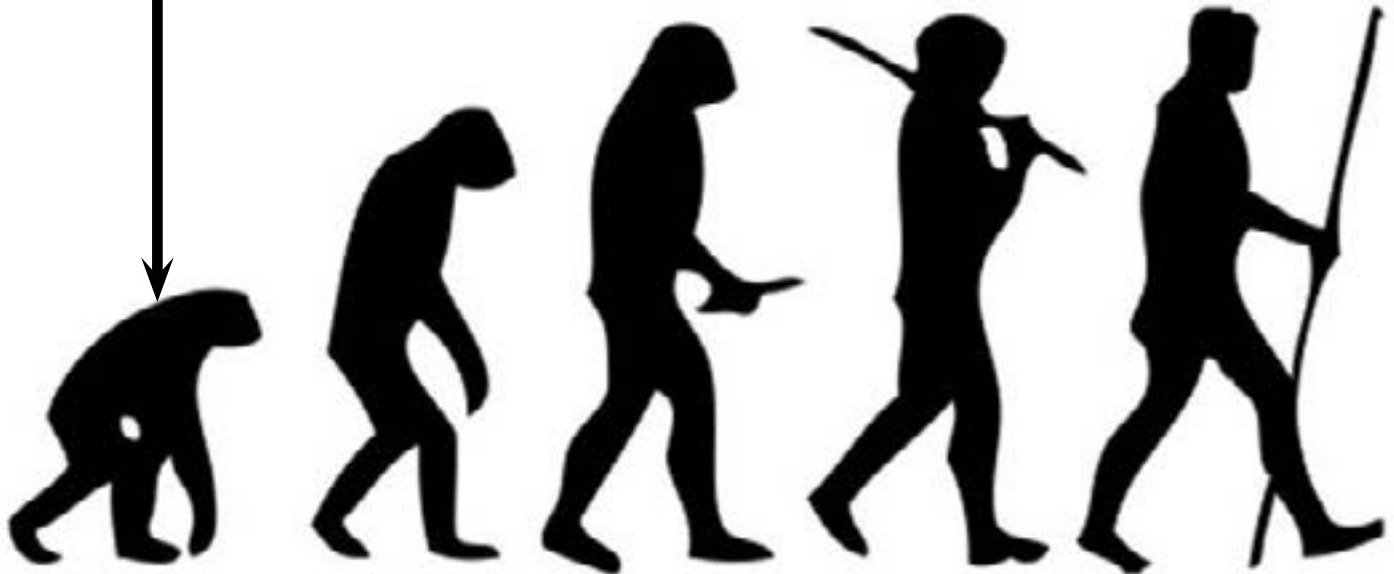How come it don't get no respect?

The Fundamental Reasons:

The tangled, troubled history of XHTML.

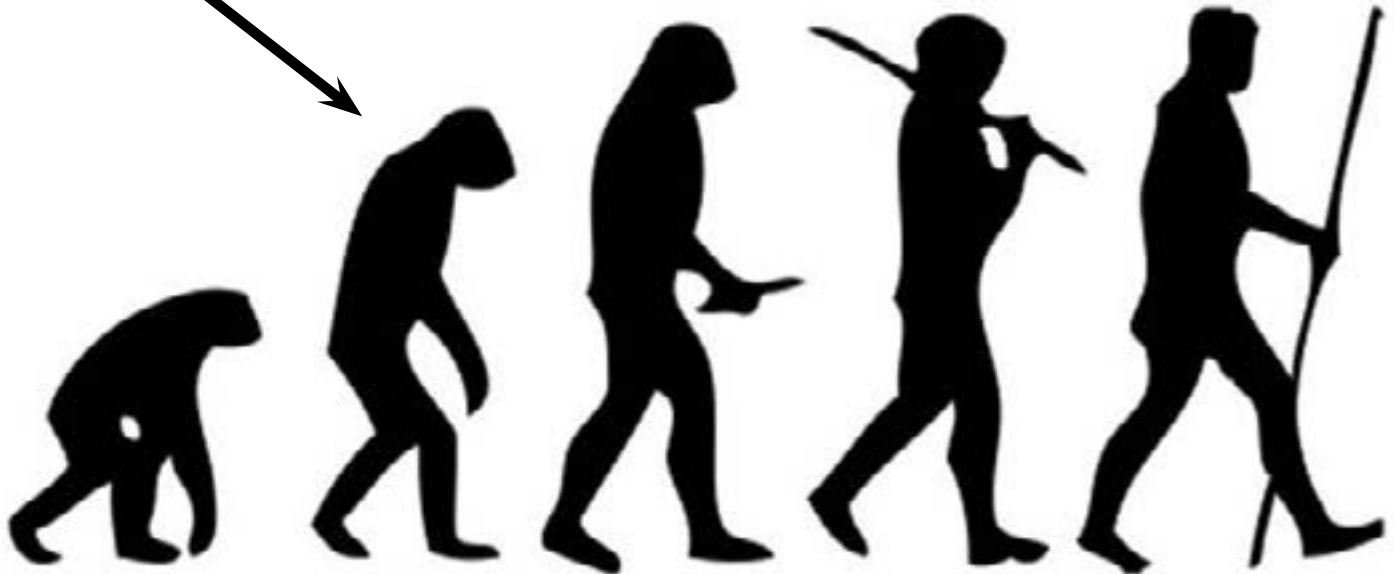The big sea change on presentation.

THE EVOLUTION OF HTML

HTML1:
"Look, Ma,
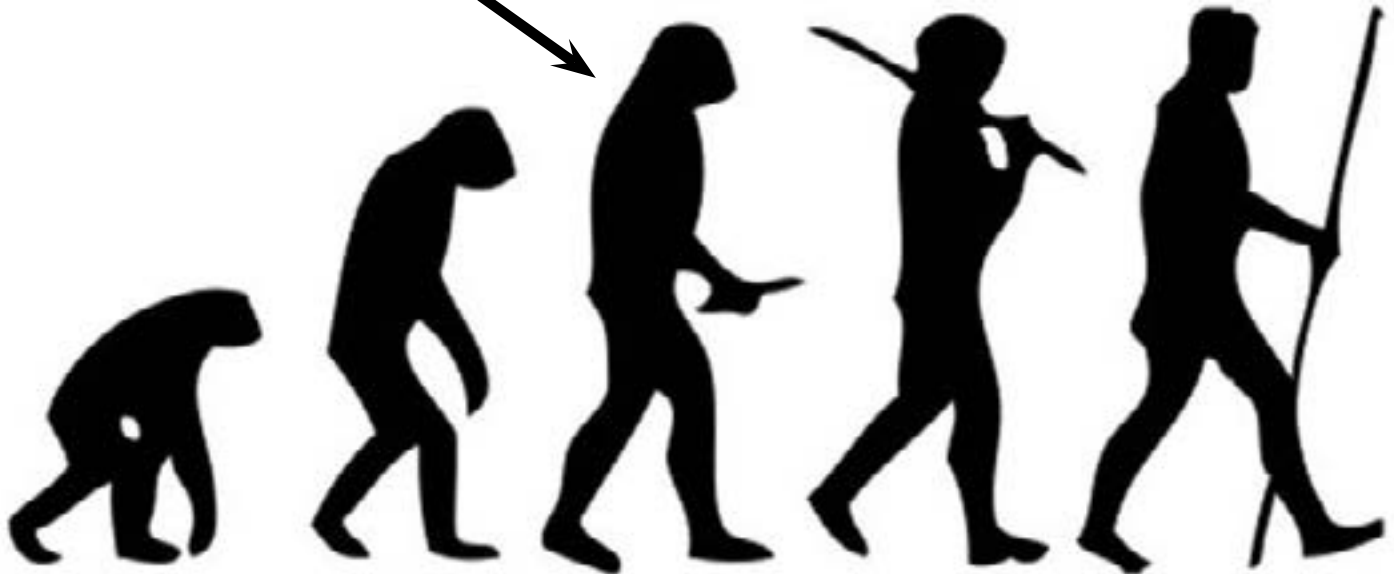**I made a web page!"**

THE EVOLUTION OF HTML

HTML2:
"Somebody actually PAID me to make a web page!"

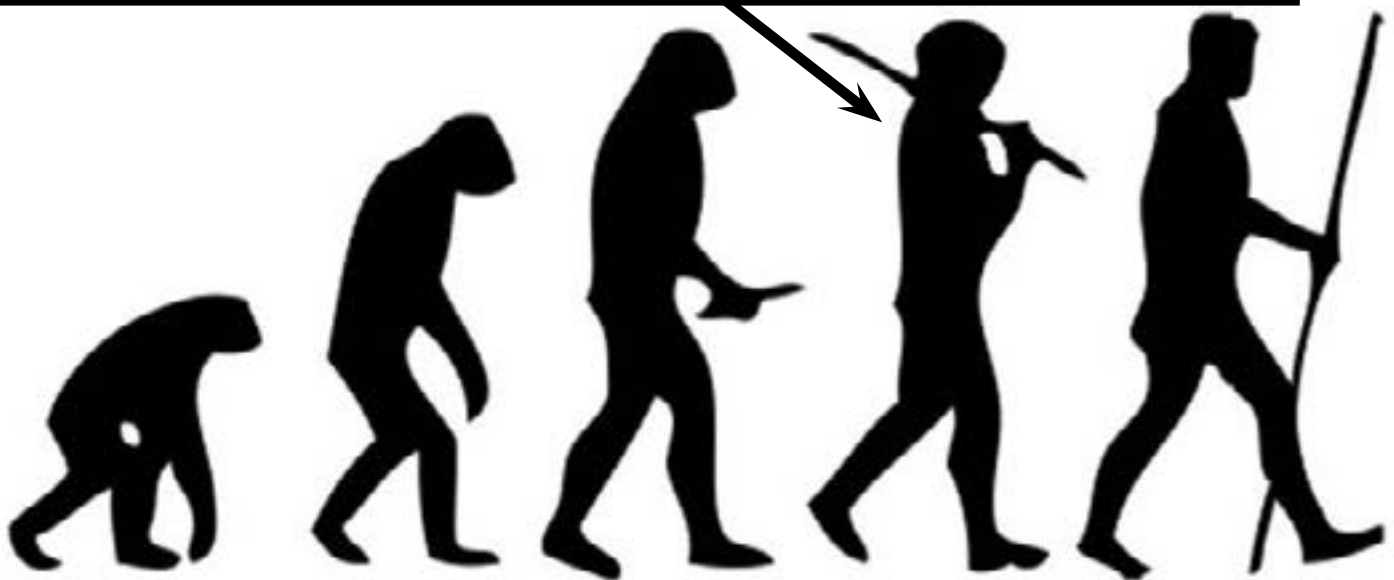# THE EVOLUTION OF HTML

HTML3:
Early signs of
primitive tool use.

THE EVOLUTION OF HTML

HTML4:
Don't use presentational markup.
Unless you want to. [It's "deprecated."]

THE EVOLUTION OF HTML

HTML5:
No presentational markup. Use CSS. Period.
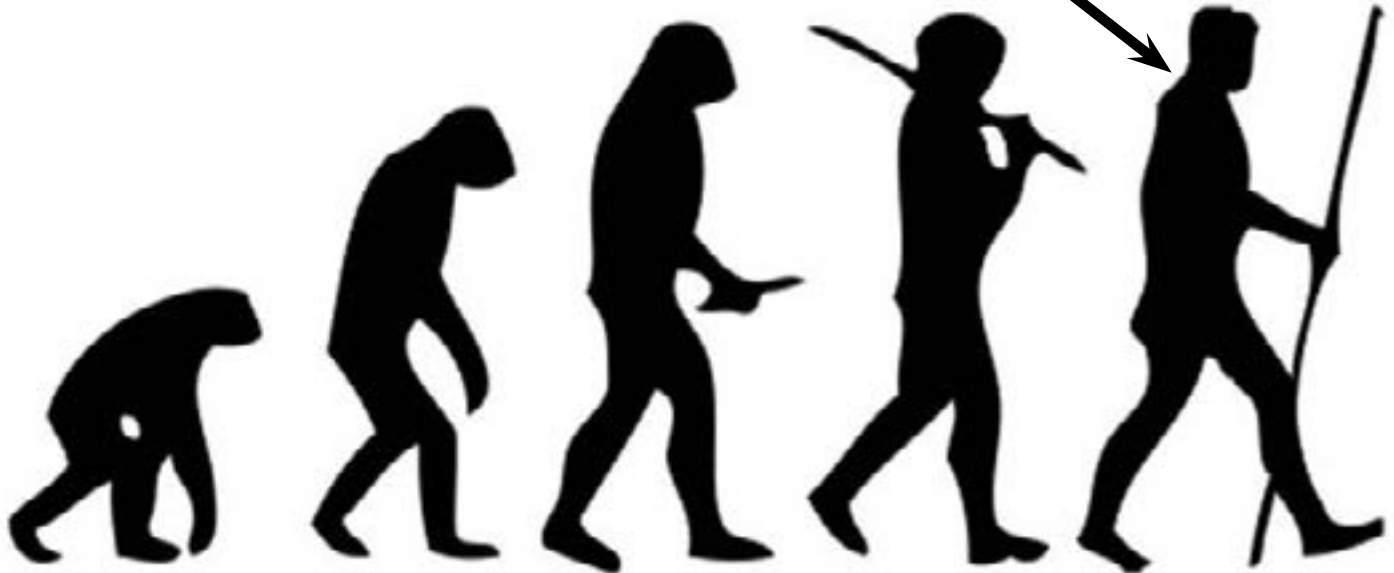
And it's XML . . .
except when it isn't.

THE HOLY TRINITY

DocBook
NLM
TEI

Originally created for specific communities.
Extended to be "general purpose" [generic].
Almost always customized unless they're a direct fit.
(If they are a direct fit, use them!)

THE HOLY TRINITY

DocBook
NLM
TEI
Okay, DITA too…

Originally created for specific communities.

Extended to be "general purpose" [generic].

Almost always ~~customized~~ **specialized** unless they're a direct fit.

(If they are a direct fit, use them!)

But you still always need HTML.

Always.

But you still always need HTML.

Always.

Online? HTML.

But you still always need HTML.

Always.

Online? HTML.
EPUB? HTML.

But you still always need HTML.

Always.

Online? **HTML.**
EPUB? **HTML.**
Apps? **HTML.**

XHTML IS XML.

XHTML is HTML that follows XML rules.

XHTML IS XML.

XHTML is HTML that follows XML rules.

XHTML IS XML.

**It can be simple or complex.
It can be rigorously structured.
It can be richly semantic.
It can contain tons of metadata.**

The Upfront XHTML Strategy

Start with the vocabulary.

- What are the *components* of your content?
- What do you *call* them?
- What do you need to *do* with them?
- What do you need to *distinguish for rendering*?

Everybody understands the vocabulary. Only the geeks understand the XML.

The Upfront XHTML Strategy

APEX

Use these terms consistently for:

- MS Word style names for copyediting.
- InDesign style names for typesetting.
- Values of @class or @role attributes for CSS.
- Values of @epub:type attribute for EPUB 3;
  can align with Structural Semantics Vocabulary
- Values of @class attribute for CSS for PDF. . . .

XHTML 1.1 or HTML5?

Known, formal schema, used by systems & tools

No schema, evolving, more semantic

XHTML is the FRAMEWORK.

Within that framework, incorporate best-of-breed markup that does what you need to do.

Building blocks are <section>/<div>, <p>, <span>, <h*n*>, etc.

Use your vocabulary in attributes.

Use Schematron to QC it.

Big benefit: proof and edit via browser-based tools.

Need math?

Incorporate MathML.

Need CALS tables?

Incorporate the CALS/OASIS table model.

Need granular citation markup?

Incorporate the NLM/JATS citation markup.

(CrossRef!)

Need rich metadata?

Leave <meta> for what it's meant for. Devote the first <div> to metadata that can = JATS header.

(Or whatever metadata is most useful to you.)

Need to use HTML1.1-based tools but want HTML5 semantics?

Capture the HTML5 semantics in attributes.

E.g., <div class="aside"> as a "container" element to wrap sidebars, pull quotes, etc.

Need JATS?

Align your XHTML-based markup with JATS.

UPFRONT.

Get the benefit of XHTML for online/web-based technologies and JATS for hosting/interchange.

The Upfront XHTML Strategy

APEX

Example 1: **SAGE Publications**

- SAGE titles all use TEI
- CQPress titles all use DocBook
- Needed common master XML model for
  - —Delivery by composition vendors
  - —Repository
  - —Transform to existing TEI & DocBook models
- **Solution:** *Base model on XHTML, incorporate vocabularies needed for TEI & DocBook uses*

The Upfront XHTML Strategy

APEX

Example 2: **University of Toronto Press**

- Publish monographs, textbooks, and journals
- Want to provide full services to distribution clients
- Needed well integrated workflow
  - —Editing in Word, using eXtyles
  - —Comp by various vendors using InDesign
  - —XML to generate EPUB, NLM, DocBook, TEI
- **Solution:** *Base model on XHTML, develop UTP vocabulary that spans Word, InDesign, CSS*

The Upfront XHTML Strategy

APEX

Example 3: The World Bank

- Big range of publications, some highly designed
- 50% of books planned to be only digital, no print
- Outsource all production to various vendors
- Had NLM model, not widely used
- Committed to open access, online, accessibility
- **Solution:** *Base model on XHTML, develop rich vocabulary, schema, plain-English specifications; works with InDesign or CSS-driven PDF+EPUB*

The Upfront XHTML Strategy

Example 4: **Harvard Business Publishing**

- Publish thousands of cases authored by faculty
- Word template for authoring, creating final PDFs
- Need XML repository to generate online, EPUB, and accessible versions, + future repurposing
- Need highly automated workflow
- **Solution:** *Base model on XHTML, use upCast to generate XHTML1.1, modified by XSLT to HBP Case XML + EPUB, QC with Schematron*

Thanks!

APEX

Bill Kasdorf

bkasdorf@apexcovantage.com

+1 734 904 6252

@BillKasdorf