Why is Markup Still Important after 30 years and Will it Still be Important in Another 30 ?

Sharon Adler
IBM Research
W3C XSLT WG Chair

In the Beginning – there was SGML

SGML – A document processing standard

- Make certain document content would outlive the application that created it or the processor that formatted it for rendering and print. The importance of clear text.
- Early SGML ANSI Committee - 6 people – DCF users – one vendor of sophisticated typesetting technology at small printing company
- Focus on separating content and its processing – primarily formatting
- Original "reuse" concept focused more on single content being able to be printed using many different devices, laser printers, line printers, photocomposition systems and presentation formats possibly.
- Asynchronous entities – big deal that did not carry over to XML but we still live with a cumbersome mechanism that lots of people hate.
- Processing instructions – page ejects/line breaks, etc.
- Whitespace –whitespace particularly in mixed content
- Marked sections – so you could publish tech doc for SGML-based systems

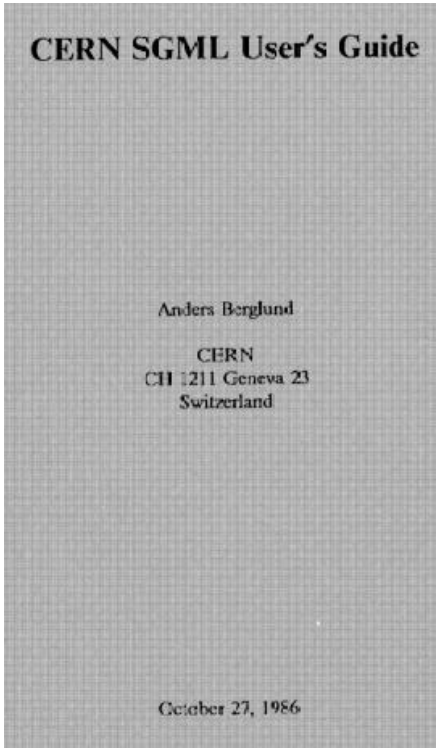
Early Tools

- Early tools had all the focus on “how content was authored/created” not how that content was to be used in an enterprise.
- Lots of limitations on the tools as they co-existed with simple word processors and large scale technical publishing systems - got caught in the middle.
- In the standards world the big competition was ODA – Office Document Architecture (later changed to “Open” Document Architecture) – needed a “reference” implementation
- Big companies put marketing effort into ODA and technical departments used SGML
- Began to change with the advent of applications such as SoftQuad Panorama and EBT’s DynaText. SGML browsers, indexers, converters, transformers, programming languages, search tools were born in late 80s early 90s.
- First SGML conference in Europe was in 1982 in Oxford England sponsored by the GCA.
- Community grew – conference attendance grew slowly from about 89

Early applications

- SGML “the methodology” as opposed to SGML the syntax. Conventional wisdom, folk lore grew but the standard allowed a good deal of flexibility
- Document analysis was the early technique for building markup vocabularies
- No notion of universality – vocabularies were developed at the department or enterprise level - not as a single application or vocabulary for entire national government
- More focus on “structure markup” rather than “content markup”
- First instance of “if you think you can’t beat them – join them” philosophy. ODIF and SGML encoding
- Syntax could be used for anything – formatting standardization became the problem
- OmniMark and Balise – SGML transformation procedural programming languages

XML was born



- Advent of the Web and HTML
- HTML was derived from CERN SGML publishing system
  http://cdsweb.cern.ch/record/997909/files/cer-002659963.pdf
- Need for greater participation by bigger tools vendors
- Bring "SGML" to the Web
- Early initiative
  - Syntax - XML
  - Linking - XLink
  - Style – XSL
- Emulating subset of ISO WG8 structure – SGML, HyTime, DSSSL
- XML became a phenomenon
  - Morphed into many things
  - Each group thought the other had it "wrong" (documents, DataBase, messages, academics, video and other graphic formats)

Flexibility and Longevity

- "XML is part of the DNA of computing" – Jim Melton, Oracle Corp.
- "Document" versus "data" wars
  - Regular, consistent flat data
  - Irregular, recursive
- The XML "good news" / "bad news" story
- Lost some focus but is more inclusive and broader market appeal
- New applications – primarily an interchange format for messages
- Work together, share and reuse – join together rather than squabble
- Family of standards
- How to stay relevant for the next 30 years

So what's left

- **XML has become the de facto standard for representing information and business objects in enterprise computing, and serves as the foundation for Web Services.**
- **XML is one of the most important technologies for business integration both inside and between enterprises because it makes it much easier to share data and business objects between incompatible systems.**

Is it Enough?

- **Semantics – as much as necessary to get the job done**
- **Have been many attempts – either not gone far enough or not enough similar attempts**
- **Better techniques for vocabulary development**
- **Data integration, universal access, are a given/required – what else is needed**
- **It's up to YOU**

Future

- **Wanted to ensure that our marked up content was able to survive long time (although we only talked about decades – now that is not sufficient) — see the Balisage conference**
- **Future is now**