

III. XML Evolution

University of California Extension
Sunnyvale, June 10, 1999

Jon Bosak
Sun Microsystems

The XML Standards Infrastructure.....	A-1
Continuing Work: Schemas and Linking	B-1
Continuing Work: Infoset, Fragments, and Profiles.....	C-1
Continuing Work: Layout and Formatting	D-1

The XML Standards Infrastructure

Key XML standards organizations	A-2
Other significant XML organizations	A-3
The W3C	A-4
W3C XML Working Groups	A-5
Closely related W3C working groups	A-6
Other related W3C working groups.....	A-7
XML activity coordination	A-8

Key XML standards organizations

- ISO/IEC JTC1/SC34
 - Maintains SGML, from which XML is derived
 - Also DSSSL, HyTime, Topic Maps
- World Wide Web Consortium (W3C)
 - Hosts the "XML Activity"
 - Develops other standards related to XML
 - Develops other web technologies not directly related to XML
- OASIS
 - Organization for the Advancement of Structured Information Standards
 - Formerly SGML Open
 - Dedicated to XML interoperability and conformance
 - Rapidly expanding role in XML implementation

Other significant XML organizations

- Graphic Communications Association (GCA)
 - Old-time printing standards association
 - Historic home of SGML events
 - Now hosts XML events such as XML Europe
- eCo Commerce Framework
 - CommerceNet initiative
 - Developing methods for the electronic negotiation of XML commercial protocols

The W3C

- Close to 300 members
- Working group decisions are only advisory; all final W3C decisions are made by the Director
- To protect against premature disclosure of product plans, discussions are mostly visible only to members
- XML activity has opened this up significantly through the publication of Requirements Documents
 - Show technical goals that each working group is trying to accomplish
 - Publicly visible (<http://www.w3.org/TR/>)
 - Open to public comment
 - Process includes periodic "checkpointing" and revision in light of comments
 - Process is currently experimental -- XML activity only

W3C XML Working Groups

The original XML work has been divided up among six W3C working groups.

- In the W3C Architecture Domain:
 - **XML Linking WG:** Next-generation hypertext
 - **XML Schema WG:** Next-generation DTDs
 - **XML Fragment WG:** Partial XML documents
 - **XML Infoset WG:** XML object model *per se*
 - **XML Syntax WG:** XML profiles, stylesheet linking...
- In the W3C User Interface Domain:
 - **XSL (Extensible Stylesheet Language) WG:** Industrial-strength layout and formatting language for XML uses on and off the Web

Closely related W3C working groups

- Document Object Model (DOM) WG
 - Not actually an object model (though one is implied)
 - Provides a generalized tree-oriented API for HTML and XML document structures
 - (There is also a widely used stream-oriented XML API called SAX that was produced outside of W3C.)
 - DOM Level 1 done, Level 2 in progress
- I18N (Internationalization) WG
 - Coordinates I18N issues among W3C WGs and between W3C and outside standards organizations (in particular, the Unicode Consortium)
 - Responsible for the W3C character model
 - Example: is "être" ê + tre or e + ^ + tre?

Other related W3C working groups

- HTML WG
 - Creating the XML version of HTML: XHTML
 - Maximum compatibility with old web browsers (complete compatibility probably not possible)
 - Backward compatibility will require users to follow specified guidelines
- Cascading Style Sheets and Formatting Properties (CSS&FP) WG
 - Stylesheets for HTML and simple XML documents
 - Lacks ability to transform document structure
 - Insufficient for current print formatting or future online display formatting
 - Will probably be superseded by XSL for print and industrial-strength web applications

XML activity coordination

- XML Coordination Group
 - Schedules the work of the XML WGs
 - Coordinates dependencies between XML specifications
 - Interacts with standards efforts outside of W3C
- XML Plenary
 - Union set of participants in the XML working groups
 - Decides major XML policy issues

Continuing Work: Schemas and Linking

XML schemas	B-2
XML linking (XLink and XPointer).....	B-3
The importance of XML linking.....	B-4

XML schemas

1. Provide existing DTD functionality in XML syntax
2. Enhance existing DTD functionality
 - Data typing
 - Inheritance
 - Namespace validation
- The need to improve DTDs has been known for years
- Two submissions to start from: DCD and SOX
- Also relates to XML query languages
- Massive involvement from major software companies, especially those with database or electronic commerce concerns

Doing a good job with schemas will take a while!

XML linking (XLink and XPointer)

XML linking builds on 20 years of hypertext research.

- Fully extensible -- any element can be a link
- Links to collections
- Automatic traversal
- Transclusion
- Addressing by structure
- Links outside of documents
 - Links can be managed separately from objects
 - Links can be applied to read-only objects
 - Links can be made long after the things linked have been released

The importance of XML linking

- Doesn't get much publicity but has deep implications
- Allows new ways of associating information
- Promotes the creation of advanced information structures and site management
- Makes possible an industry devoted to knowledge management
- Keep your eye on XLink and XPointer:
<http://www.w3.org/TR>

Continuing Work: Infoset, Fragments, and Profiles

Infoset WG.....	C-2
Fragment WG.....	C-3
Profiling (Syntax WG).....	C-4

Infoset WG

Basic problem: What's in a document after it's parsed

- Developing a data model for XML
- Defining the objects that an XML application can expect to find in a document
 - Elements and attributes? *Of course.*
 - Comments? *Maybe.*
 - Individual characters? *Maybe not.*
- Other key groups are heavily dependent on the data model
 - DOM
 - XSL
 - Linking
 - Schemas

Fragment WG

Basic problem: How to work with part of a document in isolation

- Classic example: autonumbering
- Document editing
- Partial document transmission
- W3C work builds on an existing solution developed over several years by OASIS (SGML Open) for SGML documents

Profiling (Syntax WG)

- Originally, XML subsets (simpler versions)
- Turning into an attempt to define classes of XML processors
- Requirements are hard to agree on
- Growing ubiquity of full XML 1.0 processors has taken a lot of the urgency out of this work item
- But the discussion has exposed weaknesses in the way the the specification defines XML processors
- My guess is that we will get cleaner definitions of the categories implied in the 1.0 Recommendation

Continuing Work: Layout and Formatting

Completing global publishing.....	D-2
XSL transformational requirements	D-3
XSL formatting and layout requirements	D-4
Misconceptions about XSL.....	D-5
XSL scheduling.....	D-6

Completing global publishing

Large-scale cross-platform XML publishing demands that XML deliver on the *display-oriented* promises:

- User-configurable views
- More powerful display-centric client-side applications
- Asian-language rendering support
- Media-independent publishing
- In particular, printed and online deliverables (now and in the future) from the same source

XSL provides the foundation for professional publishing in the 21st century.

- A **single technology** for online and print publishing
- A **common infrastructure** (training, tools, and techniques) for online and print publishing

XSL transformational requirements

- Generated structures (e.g., TOCs)
- Reordering (e.g., sort ordering)
- Logical objects to formatting objects (FOs)
 - In: **Book**- Chapter - Section - **Para**
 - Out: **Book**- Page sequence - Column set - **Para**

The logical tree and the formatting tree look similar at first glance, but in reality, they are profoundly different.

- Logical objects to HTML (to use existing web browsers)

XSL formatting and layout requirements

- Character alignment (top, middle, bottom)
- Writing direction (left, right, vertical)
- Mixed-width column sets
- Multiple flows
- Footnote zones
- Callouts
- Autonumbering
- Automatic copyfitting
- Content-driven headers and footers
- Formatting changes at non-tag boundaries
- Hyphenation controls
- etc., etc., etc.: <http://www.w3.org/TR/WD-XSLReq>

Misconceptions about XSL

1. **Misconception:** XSL is competition for CSS
 - **Fact:** XSL is a parallel formatting language designed to solve a set of problems unique to the page-oriented display of structured documents
2. **Misconception:** XSL is something new
 - **Fact:** A formatting language separate from CSS and specially designed for the needs of XML was part of the XML charter from the beginning (1996)
3. **Misconception:** XSL is just a transformation language (XML to HTML)
 - **Fact:** The transformation component of XSL exists to satisfy the formatting requirements; the formatting requirements of XSL are primary
 - Rendering through HTML generation is a stopgap to get us past the transition to better XML formatting; it is not and never has been the end goal of XSL

XSL scheduling

- New working draft, April 1999 (check it out!)
- Support for the transformational component is catching on
- Support for the formatting objects is much slower in coming
 - Formatting is **hard**
 - Little incentive for magazine-quality design in current web browser displays
 - No instant gratification for low-end users
 - No immediate reason for vendors of high-end publishing systems to standardize on a completely interoperable page design language
 - In sum: not many people get the point yet (like XML 3-4 years ago)