What is XML
and
Why Should You Care

B. Tommie Usdin
Mulberry Technologies Inc.
17 West Jefferson St.
Suite 207
Rockville MD 20850
Phone: 301/315-9631
Fax: 301/315-8285
btusdin@mulberrytech.com
http://www.mulberrytech.com

July 2000
Version 1.0

Mulberry
Technologies, Inc.

What is XML and Why Should You Care

What is XML?

Parts of an XML Application

Where XML can be Used in Publishing

Future of XML

Where to Get More Information

What is XML?

slide 1

The Word "XML" is Used to Mean:

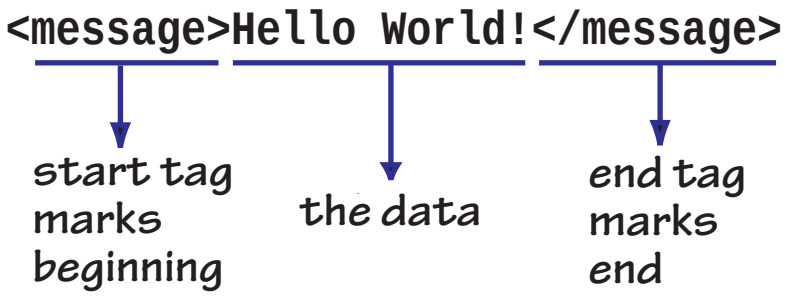
- An open standard (well... a W3C recommendation) that provides:
  - A data format
  - A data modeling language
- The use of XML-formatted data in an application (like a browser)
- A metalanguage for creating markup languages
- A set of associated recommendations and specifications

  (link, style, query, language specification, transformation, APIs, etc.)

XML Works Through Tags

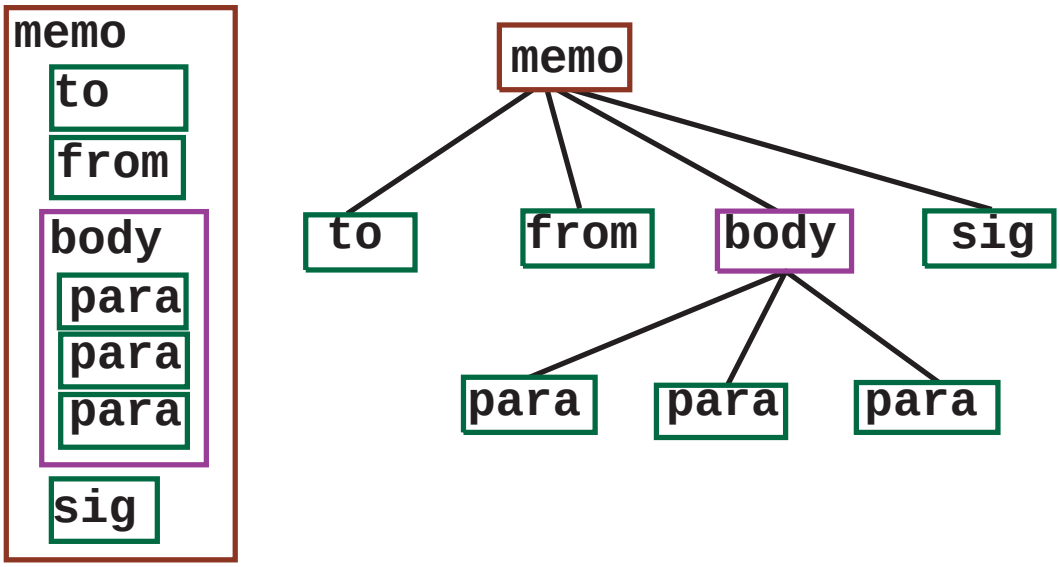
Paired tags:

- Enclose data
- Identify/name the data
- Named component called an “element”



slide 3

Elements Can Nest



slide 4

XML Isn't:

- A programming language
  Does not replace C++, Java, perl, Python, ...
- A user interface
- A presentation format
- A formatting or processing system
- A standard set of tags
- A recommended set of tags

XML Is a Data Format

"Employee Record" Example

In XML we can separate data content from behavior/presentation.

```
<employee-record type="dog" empno="9">
<name><first>Sarsparilla</first><last>Usdin</last></name>
<affiliation>
<title>Deputy in Charge of Chewables</title>
<company>Mulberry Technologies</company>
<location>Rockville, MD 20850</location>
<email-name>sassy</email-name>
</affiliation>
<height unit="in">36</height><weight unit="lb">70</weight>
</employee-record>
```

slide 6

View This in a Browser

Convert into HTML (today). Or in an XML browser (tomorrow):

slide 7

A Familiar Print Application

slide 8

Same Data, Different Application

New Employee Announcements

Sasparilla Usdin
has recently joined Mulberry Technologies, Inc.'s Rockville staff as Deputy in Charge of Chewables.

Welcome to the team, Sassy!

- XML elements rolled into "form letter"
- Something (perhaps employee-id) linked to photo

slide 9

Same Source: Load a Database

```
Key: 00095AUS
EMPNO: 009
001:USDIN
002:Sasparilla
008:36
014:70
020:Deputy in Charge of Chewables
```

slide 10

Ultimate Purpose of XML: Encode Data Once

- Enable semantically complex search
- Produce many products from that markup
- Reuse data (in whole or part) many times
- Interchange data freely
- Enable machine–machine communication
- Let whole communities agree on data content
- Let data live a long time

Parts of an XML Application

slide 11

Logical Components of an XML Application

- XML document (tags and text)
- DTD or Schema (the model)
- Output Specifications (how looks/behaves)
- Transformations (from here to there)

slide 12

Component: XML Document

The tags (markup) and the text (content)

- Two types
  - Well-formed
  - Valid (has a model)
- Usually created
  - using an XML editor (authoring)
  - by a program from
    - a database
    - another type of XML file by transform
    - conversion from another format (like Quark or Word)

slide 13

Component: DTD (Document Type Definition)

The modeling mechanism specified by the XML specification

- Models one type/class of information (a "document")
  (reference book, bank transfer, journal article, memo, help-topic)
- Is a set of rules describing how documents of that type can be marked up
- Is written in the formal syntax of XML

slide 14

DTDs Express Rules

for example:

- **article** = *metadata* followed by *article body*, followed by optional *back matter*
- **paragraph** = data characters and may include any of the following: *Person Name*s *URL*s, and/or *Geographic Region*s
- **Reference book** =
  Front Matter followed by *Body* followed by *Back Matter*
- **Purchase Order** =
  Order Header followed by *List of Order Detail*, followed by optional *Order Summary*

slide 15

Why Use a DTD?

- DTD is a contract between producers and consumers
  (Both can validate to see if they got/sent what they expected)
- Formal specification of information *types* allows consistent downstream processing
- Support for interoperable families of documents and document types
  - To ensure information conforms to model (validation mechanism)
  - Parties don't have to share software or applications

To share information, share the DTD

slide 16

DTDs today, Schemas tomorrow

- Provide data typing
- Model multi-element dependencies, attribute dependencies, real inheritance, or other complex data dependencies
- Perform real data modeling
- Enforce rules about the data *content* (as opposed to the tag structure)

Component: Output Specification/Stylesheet

slide 17

XML Separates Content from Format/Behavior

How it looks (16 pt Helvetica Bold) or what it does (start a javascript)

- Is based on the tagging
- Is the same for every tag *in the same context*
  - NOT one tag per one format
  - Table title may differ from Figure title from Chapter title

slide 18

Use an Output Specification to get There

(Frequently called “Stylesheet”)

- Says what XML data will look like or how to behave
  - On screen or paper
  - Or in other media (for example in audible output)
- Defines an appearance or rendition or behavior
  - For each element
  - In each of its contexts within a document

slide 19

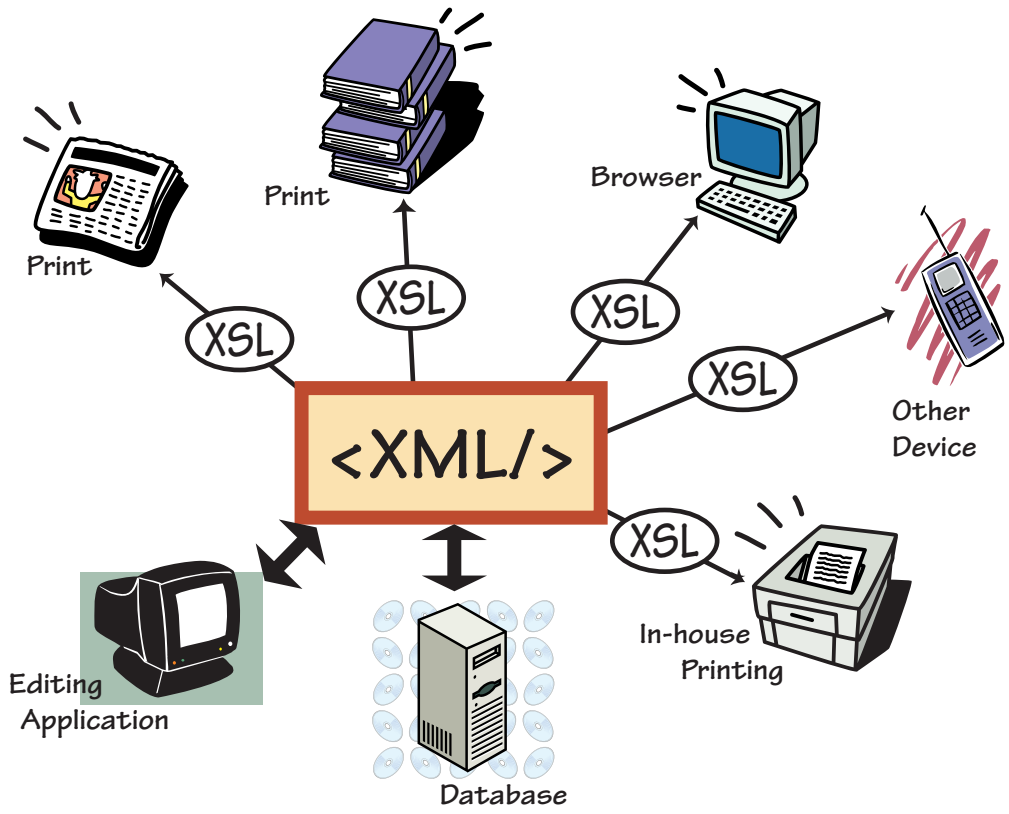
Documents and Stylesheets

- One stylesheet, many documents
  - Maintain consistency of format, “look and feel” across documents
  - House style is easy to develop, maintain, apply
- One document, many stylesheets
  - Different media types: print, on-line, etc.
  - Different derivative documents: selections, summaries, indexes, catalogs....

Component: XML Transforms

slide 20

XML and XSL



slide 21

XSL For XML Transformation (XSLT)

Transforms from one set of tags directly into another
Transform XML into

- HTML for browsers
- Other (XML) tag sets for further processing
- Plain text formats (e.g., loader files for databases)
- Non-XML tag sets

Where Does XML Fit in the Publishing Process?

slide 22

XML Files

- Are “plain text” underneath
  - Use any text editor or any word processor that can handle plain text
  - Built on Unicode (represents all major scripts of the world)
- XML could be
  - Native file format for a software package
  - Something you “import” and “export”

slide 23

XML for Print and Web Publishing

- Many different outputs, one manageable source
  - Many media/device types
    (Web, CD-ROM, handheld PDAs, voice-synthesis)
  - Many styles of print/display
- Different hardware, software, OS for input, manipulation, display
- Publish on demand/Customized output

slide 24

XML Integrates Services on a Web Base

- Web-based “data-centric” integration
  - Allows loose coupling of applications: robust and open
- Web becomes friendly to other media
  - Information syndication, on-demand publication
  - Web and print services complementary
  - Information into and out of databases
- From existing data systems directly to the web

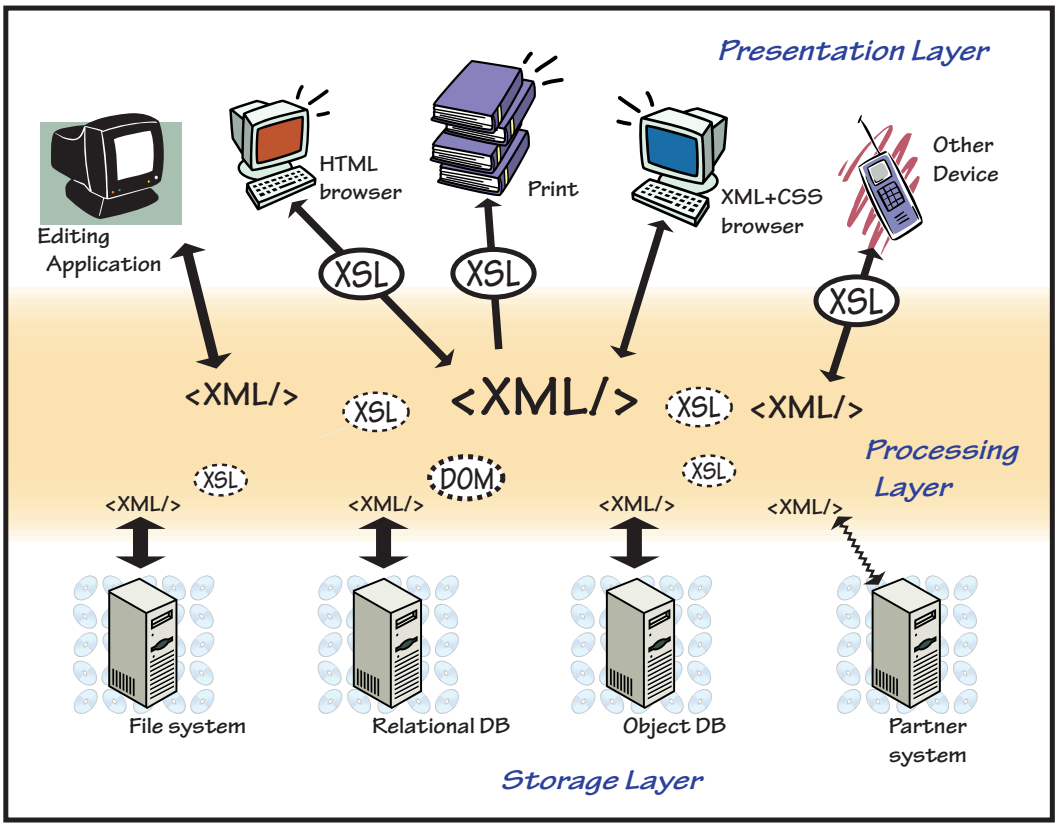
slide 25

XML Between Application Layers

- In the "Three-tier" system model:
  - Presentation/User interface layer
  - Processing or "business logic" layer
  - Storage or repository layer
- XML used in any of the three tiers, *especially in the middle*
- XSL is used for any processing
  - Within the middle tier, and
  - *Between* tiers

slide 26

A Typical "3-tiered" Model



slide 27

XML Tagged Data for Information Interchange

- By data aggregators (scientific and journal websites, semiconductor industry, aircraft)
- Through the life-cycle of a product (among divisions)
- Direct machine-to-machine transfer
- Between rival proprietary formats
- Inter-process communication (IPC)
- Among business partners (E-Business transactions B2B, B2C, EDI replacing EDIFACT or proprietary formats)

slide 28

XML Repositories and Databases

- XML is used inside:
  - OO databases
  - Relational databases
  - Object-relational and hierarchical databases
- XML is used to communication among
  - Databases and applications
  - Databases and databases
- XML Repositories
  - Store content for use and reuse
  - Enable repurposing

slide 29

XML for Search and Retrieval

- Enhance documents with metadata
- Use grammar to increase relevance
  limit to or exclude in named element(s)
- Use vocabulary to increase recall

Future of XML

XML Versus HTML

slide 30

XML Will Replace HTML for:

- Users running into limitations of HTML
  - Retrieval
  - Formatting
  - Richness of encoding
- Users with complex data requirements (semi-conductors, airlines)
- Users with complex security requirements
- Users with SGML data

slide 31

XML Won't Replace HTML for:

- Huge amounts of legacy HTML
- Simple display-only pages
- Unstructured data
- Write-once web pages

XML and PDF

slide 32

NOT XML Versus PDF

- We will continue to need both
- XML source used to produce:
  - PDF
  - HTML for older browsers
  - Search service proprietary format (or specific-DTD tagging)
  - eBook format(s)
  - CD-ROM internal format

slide 33

XML *and* PDF

- PDF will still be a popular pre-press/archival form
- XML systems will produce PDF as an output product (as they do now)
- Searchable PDF will make in-roads against XML but be dropped as lesser functionality is recognized.
- It will take pretty pages and snazzy behavior to wean people from PDF, which is
  - easy
  - expensive

XML and Proprietary Formats

slide 34

Proprietary Formats

- Optimized for one use (and one set of tools)
- Faster, easier for intended use
- Controlled by vendors or other non-public process
- Often change with version of software or application

slide 35

XML *feeds* Proprietary Formats

- XML will be transformed into proprietary formats for:
  - Print and display
  - Database loads and device drivers
  - Existing processes and procedures
- Proprietary formats will be transformed into XML
  - (Generally harder than XML to proprietary)
  - Allows people to use known tools
  - Captures existing content

slide 36

XML Tools

(XML got its start from SGML tools:

- XML is SGML minus some features
- XML tools have been available since XML announced)

New XML Tools announced every week

- Authoring applications (word processors, spread sheets, etc.)
- Document management and workflow applications
- Databases with XML smarts (or XML-based databases)
- XML-based search and retrieval
- XML formatting and display

slide 37

XML Ubiquitous Under the Covers

- All major software companies using XML internally
- XML the format of choice for process-to-process interchange
- XML will dominate e-commerce and e-business in a year or two

slide 38

XML Tag Sets will Proliferate

- Discipline based
  - CML (Chemistry Markup Language)
  - MathML (Math Markup Language)
- Subject or industry domain (financial services, health care, auto and aircraft maintenance, ...)

  Groups of users creating shared XML applications
  literally hundreds of efforts announced and more daily
- SVG (Scaleable Vector Graphics)

Where to Get More Information

slide 39

The Source for XML and Related Information

- Robin Cover's SGML/XML Web Page:
  `http://www.oasis-open.org/cover/sgml-xml.html`

slide 40

General XML Information

- W3C's XML page: **`http://www.w3.org/XML/`**
- XML FAQ (Peter Flynn): **`http://www.ucc.ie/xml/`**
- XML.com: **`http://www.xml.com`** (industry coverage and tools)
- XMLinfo.org **`http://www.xmlinfo.org`** (covers tools and development)
- XSLinfo.org: **`http://www.xslinfo.org`** (covers XSL development and implementation issues)

XML E-Business and EDI Information

- The OASIS and UN/CEFACT entry **`http://www.ebxml.org`**
- CommerceOne's entry**`http://www.commerceone.com`** (Common Business Library etc.)
- The Microsoft entry **`http://www.BizTalk.org`**
- RosettaNet consortium **`http://www.rosettanet.org`** (vocabulary and process issues, industry coverage)
- XML.ORG **`http://www.xml.org`**
- Commerce XML (cXML) **`http://www.cxml.org`**
- Oracle's XML material **`http://www.oracle.com/xml/content.html`**
- IBM is heavily into this, too. **`http://www.developer.ibm.com`** [for Business Rules Markup Language (BRML) and Trading Partner Agreement Markup Language (tpaML)

And others too numerous to mention: CommerceNet's eCoFramework, SAIC's Universal Commerce Language and Protocol (ULCP), XEDI.org, etc.

Printed Books on Concepts

- **SGML: the Billion-Dollar Secret**, by Chet Ensign (Prentice-Hall PTR, 1997)
  - Manager level. Written about SGML (XML's parent standard), but almost entirely applicable: excellent on issues of scalable system development.
- **ABCD… SGML**, by Liora Alschuler (Thompson Computer Press, 1995)
  - Written about SGML (XML's parent standard), but change the word "SGML" to "XML" as you read it and it still applies. Talks about work process changes an XML system can bring.
- **XML: A Manager's Guide**, by Kevin Dick (Addison-Wesley Information Technology Series, 2000)
  - Manager level. Solid view, but stays at 10,000 feet up.
- **The XML Companion (2nd Edition)**, by Neil Bradley (Addison-Wesley, 2000)
  - Very good basic technical introduction.
- **Professional XML**, by Richard Anderson, Mark Birbeck and ten more authors. (Wrox Press Ltd.)
  - Light technical level. Each author wrote an introduction and then examples/case study for one technical topic. Introduces the problems of XML and databases, the XML APIs DOM and SAX, server to server XML (XML-RPC, SOAP, etc.) and more.

slide 43

Other Information Sources

- **Markup Languages: Theory and Practice** (a quarterly journal): **`http://mitpress.mit.edu/MLANG`**
- OASIS Home Page (vendor consortium): **`http://www.oasis-open.org`**
  - XML.ORG: **`http://xml.org`** (document model repository and support materials)
  - OASIS XML Conformance Subcommittee: **`http://www.oasis-open.org/committees/xmlconf-pub.html`**
- Graphic Communications Association: **`http://www.gca.org`** (sponsors conferences including Extreme Markup Languages, August 2000 and XML 2000, December 2000)
- XML.COM: **`http://www.xml.com`**

slide 44

Still More Information Sources

- Basic newsgroup: `comp.text.xml` (also some on `comp.text.sgml`)
- Useful Lists
  - XML-L: **`http:/listserv.heaanet.ie/xml-l.html`** (for newcomers)
  - XML-Developer's List: **`http://www.lists.ic.ac.uk/hypermail/xml-dev`** (heavy technical discussion)
  - XSL-List: **`http://www.mulberrytech.com`**