

What is XML and Why Should You Care?

Deborah Aleyne Lapeyre
Mulberry Technologies, Inc.
17 West Jefferson St.
Suite 207
Rockville, MD 20850
Phone: 301/315-9631
Fax: 301/315-8285
info@mulberrytech.com
<http://www.mulberrytech.com>

Version 1.0 (January 2006)
©2006 Mulberry Technologies, Inc.



What is XML and Why Should You Care?

Administrivia	1
What is XML?	
What XML Means.....	2
XML Works Through Tags.....	2
XML Documents.....	3
How XML works	
An XML Document is a Sequence of Elements.....	3
XML Elements.....	5
Elements Contain Other Elements.....	5
Elements Identify Many Kinds of Content.....	6
Attributes Add Further Description.....	6
Technical Note: Editing XML Files.....	7
What XML Isn't.....	7
XML Is a Data Format	
XML Unites/Divides Two Very Different Data Visions.....	8
One XML Document Produces Many Results	
“Employee Record” Example	13
Why Use XML?.....	16
Real World Examples of XML	
XML is a “Metalanguage”.....	16
New XML Markup Languages.....	17
XML Initiatives (Very Partial List).....	17
Parts of an XML Application	
Logical Components of an XML Application.....	27
Component 1: XML Document	28
Component 2: The Document Model	28
DTDs / Schemas Express Rules	29
Why Use a DTD or Schema?.....	29
To Share Information, Share the DTD / Schema	30
Current XML Modeling/Constraint Languages.....	30

What is XML and Why Should You Care?

Component 3: Formatting (and Behavior)	31
Remember What XML Looks Like Without Formatting.....	31
What We Would Like to See (Print or Screen).....	33
XML Design Feature.....	33
Display Specifications Give Instructions.....	34
Component 4: XML Transformation	
XSLT for XML.....	36
Why This Is Exciting.....	36
Component 5: XML Repository	37
Content Management Using an XML Repository.....	37
XML in Non-XML Databases.....	38
Lots of XML is Managed in File Systems.....	38
Conclusion	
The Big News: XML DOES NOT DO ANYTHING!.....	39
<i>You</i> (and Your Software) Can Do a Lot with XML.....	39
The Good News: You Can Do XML and Benefit.....	40
Where to Get More Information	
<i>The Source</i> for XML and Related Information.....	40
General XML Information on the Web.....	41
Books on XML Concepts (Not So Technical).....	41
XML Books We Recommend (More Technical).....	41
Other Information Sources.....	42
Colophon.....	43

What is XML and Why Should You Care?

slide 1

Administrivia

- Questions are always in order
 - Why this talk
 - Anything else?
-

slide 2

Where We Are *Not* Going in This Talk

- Specific XML vocabularies/languages (UbXML, IFX, DocBook, etc.)
 - XML for Ecommerce, eBusiness, B2B, B2C
 - JDF (Job Definition Format)
 - PPML (Personalized Print Markup Language)
 - Content syndication (PRISM, et al.)
 - Interchange and packaging of XML (SOAP, XML-RPC, etc.)
 - How to solve your particular business problems
 - Specific XML tools
-

slide 3

Where We Are Going Today: XML as Content

XML as the text and the data (content)

- What is XML
- How XML works
- Why XML is important
- Typical components of an XML application

What is XML?

slide 4

The Word “XML” is Used to Mean:

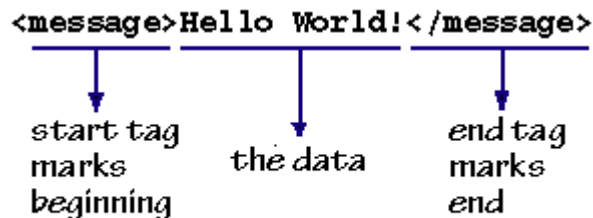
- An open standard (well ... a W3C Recommendation) that provides
 - a data format
 - a data modeling language
- The use of XML-formatted data in an application (like a browser)
- A metalanguage for creating markup languages
- A set of associated recommendations and specifications (style, transformation, query, link, APIs, etc.)

slide 5

XML Works Through Tags

A tag is a word surrounded by “pointy brackets”
<phone-number>411</phone-number>

- Start tag marks start of some data
- End tag marks end of some data



XML Documents

- In XML jargon, your data (no matter what form) is called a “document”
- A document is a coherent, ordered collection of information
 - journal article
 - invoice
 - reference book
 - chapter in a reference book
 - sales catalog
 - drug monograph

(Sometimes called “document instance” or just “instance”)

How XML works

An XML Document is a Sequence of Elements

Documents

- are made up of *Elements*
- consisting of *Markup* (“tags”)
- ... and *Element Content*

```
<book-title>Economics for Everyone</book-title>
```

```
<publisherLoc>New York</publisherLoc>
```

```
<abstract>
```

```
<p>This chapter focuses on the role of optical fibers as a communication channel in lightwave systems. We use geometrical optics to explain the guiding mechanism and introduce the related basic concepts. </p>
```

```
</abstract>
```

Sample XML Document: A Chapter

```
<chapter>
<title>Mysterious Doings</title>
<para>It was a dark and stormy night ...</para>
</chapter>
```

Sample XML Document: A Recipe

```
<recipe>
<ingredients>
  <item>
    <qty>3</qty>
    <ingredient>eggs</ingredient>
  </item>
  <item>
    <qty>1/4</qty>
    <measure>cup</measure>
    <ingredient>milk</ingredient>
  </item>
</ingredients>

<directions>
<step><para>Break the eggs into a bowl.</para>
<para>Add the milk and mix with a fork.</para></step>

<step><para>Pour the mixture into a frying pan
over medium-low heat. Tilt the pan to cover
the bottom.</para></step>

<step><para>When the eggs have set to your liking,
turn onto a plate.</para></step>

<step><para>Enjoy!</para></step>
</directions>
</recipe>
```

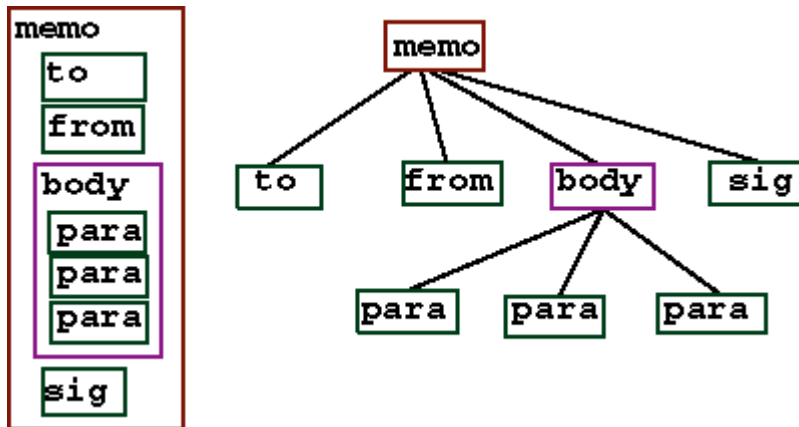

XML Elements

An *element* is an identifiable, named component of a document (paragraph, author's name, article title, unit price, bulleted list)

- Can have content (data, other elements)
- Can be a pointer to information (hypertext link, table reference)
- Must be contiguous (one start and one end; no holes in the middle)

Elements Contain Other Elements

(this nesting makes a "Tree")



Elements Identify Many Kinds of Content

(No limit to the number of *possible* elements)

Structure	What part of the document? (article, title, paragraph, list, footnote)
Metadata	<i>About</i> the document (issue number, first page, article title, DOI, journal abbreviation)
Named content	What is this text/data? (genus-species, surname, glossary, gene, question and answer)
Navigation/Links	Value-added for searching/linking (bibliographic citations; links to other articles, index terms, related material; figure references)
Presentation	How text should <i>look</i> (typographic emphasis, superscript, forced line breaks)

Attributes Add Further Description

- Live inside start tags
- Say something *about* the data
- Add information to our knowledge of the element
- Are made up of
 - a “name”
 - an equal sign
 - a quoted “value”

```
<productTitle language="EN">Hoboken Highlights</productTitle>
```

```
<printDate year="2002" month="07" day="00"/>
```

```
<phoneNumber type="unlisted" rate="premiumplus"  
assigned="1996-04-01">301/315-9631</phoneNumber>
```

Technical Note: Editing XML Files

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)
- Are “human readable”
- Are machine processable

XML Isn't Any of the Following

- 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

slide 16

XML Unites/Divides Two Very Different Data Visions

- Data Folks
 - All the world is a database
 - I know how big everything is and its data type
 - Things come in this order/arrangement
 - That's the whole point of information!
- Text Folks
 - Databases are limited
 - I have free-flowing content
 - If you have to ask how long a paragraph is or how many times it may repeat, you don't understand
 - I need named things that float *somewhere* inside other things (like a Part Number or a Person's Name somewhere inside a paragraph)
 - That's the whole point of information!

One XML Document Produces Many Results

slide 17

“Text Book” Example

```

<section id="F8493842" lastupdate="2001-05-22">
<title>Compounds</title>
<para>
A <keyterm>compound</keyterm> is a
substance containing at least two elements combined
chemically in definite proportions by mass. A compound
can be chemically broken up into its constituent elements
or simpler compounds. There are two types of compounds,
<term>ionic</term> and <term>molecular</term>.
<question-and-answer>
Testbank <testgroup>GDW</testgroup>
<question-group>
<question>6</question><question>7</question>
<question>9</question><question>54</question>
</question-group>
</question-and-answer>
</para>

<para>An <keyterm>ion</keyterm>
(<pronunc>eye-on</pronunc>) is an atom or group of
atoms that is positively or negatively charged. A
negatively charged ion is an <keyterm>anion</keyterm>
(pronounced <pronunc>an-eye-on</pronunc>) while a
positively charged ion is a <keyterm>cation</keyterm>
(pronounced <pronunc>cat-eye-on</pronunc>). An
<keyterm>ionic compound</keyterm> is a compound that
is held together by the attractive forces between
positively and negatively charged ions.
<question-and-answer>
Testbank <testgroup>GDW</testgroup>
<question-group><question>6</question>
<question>7</question>ionic compounds</question-group>
<question-group><question>9</question>cations<question-group>
<question-group><question>25</question>
<question>26</question>anions<question-group>
</question-and-answer>
</para>
...</section>

```

We Still Print Textbooks

Chapter 6: Classification	Page 55
6.9 Compounds	
compound	A compound is a substance containing at least two elements combined chemically in definite proportions by mass. A compound can be chemically broken up into its constituent elements or simpler compounds. There are two types of compounds, <i>ionic</i> and <i>molecular</i> .
ion	An ion (pronounced <i>eye-on</i>) is an atom or group of atoms that is positively or negatively charged. A negatively charged ion is an anion (pronounced <i>an-eye-on</i>) while a positively charged ion is a cation (pronounced <i>cat-eye-on</i>). An ionic compound is a compound that is held together by the attractive forces between positively and negatively charged ions.

Textbooks May Have Instructor's Manuals

Jeremy's Chemistry
Chapter 6: Classification

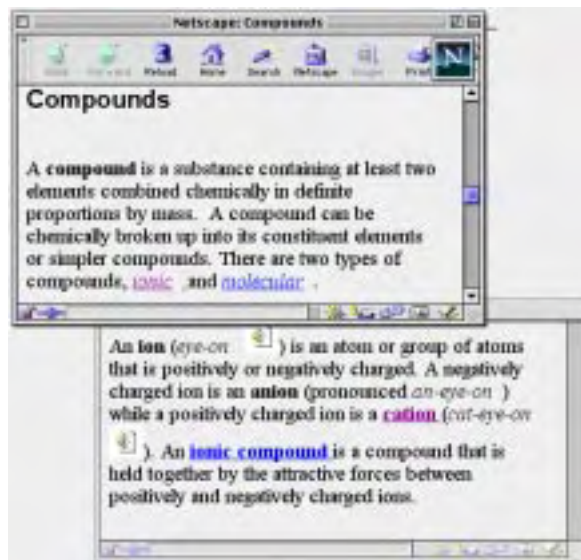
INSTRUCTOR GUIDE
Page 55

6.9 Compounds

compound	A compound is a substance containing at least two elements combined chemically in definite proportions by mass. A compound can be chemically broken up into its constituent elements or simpler compounds. There are two types of compounds, <i>ionic</i> and <i>molecular</i> .	Testbank GDW 6, 7, 9, 54
ion	An ion (pronounced <i>eye-on</i>) is an atom or group of atoms that is positively or negatively charged. A negatively charged ion is an anion (pronounced <i>an-eye-on</i>) while a positively charged ion is a cation (pronounced <i>cat-eye-on</i>). An ionic compound is a compound that is held together by the attractive forces between positively and negatively charged ions.	Testbank GDW ionic compounds GDW 6, 7 cations GDW 9 anions GDW 25,26

View This in a Web Browser/eBook

Convert into HTML or HTML-like format



Automatically Generated Section of Same Textbook

Chapter 6: Classification Page 48

6.0 Key Concepts To Define and Review

- ◆ anion
- ◆ cation
- ◆ compound
- ◆ ion
- ◆ ionic compound
- ◆ molecular compound

Same Source, Different Results

- On the Web, eBook, and CD versions
 - tie the pronunciations to audio files
 - link keywords to definitions in the dictionary
- Make large print, voice synthesis, and Braille
- Collect statistics on which test questions are used, how often, and where

“Employee Record” Example

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

View This in a Browser

Convert into HTML (today), or in an XML browser (tomorrow)



A Familiar Print Application



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

Same Source: Load a Database

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

Why Use XML?

The ultimate purpose

- Encode (mark up) data only once
- Produce many products from that markup
- Enable semantically complex searching
- Reuse data (in whole or part) many times
- Interchange data freely
- Enable machine-to-machine communication
- Let whole communities agree on data content
- Let data live a long time

Real World Examples of XML

XML is a “Metalanguage”

Used to define custom tag sets

- Tag sets get called “languages”
- Languages can be built for
 - problem domains (journal publishing, textbooks, computer manuals)
 - applications (like eBusiness, content-management)
 - vertical markets (airplanes, tourist industry, financial)
 - information collections (reference works, laws and statutes, biographies, dictionaries)
- Different markup languages for different *information types*

New XML Markup Languages

- Not really “languages” but a set of agreements
- Agreements may include:
 - sets of tags
 - problem and process models
 - document or message models (DTDs and schemas)
 - vocabularies and dictionaries
 - business rules
- Discipline-oriented, like CML (chemistry) and MathML (mathematics)
- Process-oriented, like SVG (Scalable Vector Graphics)
- Industry oriented, like Airlines/aircraft and Semiconductors

XML Initiatives (Very Partial List)

(with thanks to OASIS and Robin Cover)

- Text Encoding Initiative (TEI)
- Markup Language for Complex Documents (Bergen MLCD Project)
- Manuscript Access through Standards for Electronic Records (MASTER)
- XCES: Corpus Encoding Standard for XML
- Global Document Annotation Initiative (GDA)
- Electronic Metadata for Endangered Languages Data (EMELD)
- Electronic Text Corpus of Sumerian Literature (ETCSL)
- XML System for Textual and Archaeological Research (XSTAR)
- Metadata Encoding and Transmission Standard (METS)
- Image Metadata Aggregation for Enhanced Searching (IMAGES)
- Encoded Archival Description (EAD)
- Encoded Archival Context Initiative (EAC)
- Linking and Exploring Authority Files (LEAF)
- Channel Definition Format, CDF (Based on XML)
- RDF Rich Site Summary (RSS)
- Open Content Syndication (OCS)
- Web Modeling Language (WebML)
- Portable Site Information (PSI)
- XHTML and 'XML-Based' HTML Modules

What is XML and Why Should You Care?

- Meta Content Framework Using XML (MCF)
- Resource Directory Description Language (RDDL)
- Resource Description Framework (RDF)
- Ontology Interchange Language (OIL)
- Meaning Definition Language (MDL)
- The Australia New Zealand Land Information Council (ANZLIC) - Metadata
- NISO Circulation Interchange Protocol (NCIP)
- Alexandria Digital Library Project
- European Visual Archive Project (EVA)
- ATLA Serials Project (ATLAS)
- BiblioML - XML for UNIMARC Bibliographic Records
- bibteXML: XML for BibTeX
- e-Government Interoperability Framework (e-GIF)
- Governmental Markup Language (GovML)
- US Federal CIO Council XML Working Group
- United States Congress: XML for Legislative Documents
- US Patent and Trademark Office Electronic Filing System
- WIPO XML DTDs for the Electronic Patent Cooperation Treaty Application
- Election Markup Language (EML)
- Environmental Protection Agency (EPA) Central Data Exchange (CDX)
- PEO Interchange XML Initiative (PIXIT)
- Tukwila Data Integration System (University of Washington)
- UML to XML Design Rules Project
- XML Metadata Interchange Format (XMI) - Object Management Group (OMG)
- OMG Common Warehouse Metadata Interchange (CWMI) Specification
- OMG Model Driven Architecture (MDA)
- Object Management Group XML/Value RFP
- MDC Open Information Model (OIM)
- Dublin Core Metadata Initiative (DCMI)
- RSLP Collection Description Project
- Vocabulary Markup Language (VocML)
- Open Archives Metadata Set (OAMS)
- Xyleme Project: Dynamic Data Warehouse for the XML Data of the Web
- Publishing Requirements for Industry Standard Metadata (PRISM)
- Platform for Internet Content Selection (PICS)
- Extensible Graph Markup and Modeling Language (XGMML)
- Structured Graph Format (SGF)
- Graph Exchange Language (GXL)
- Petri Net Markup Language (PNML)
- Outline Processor Markup Language (OPML)
- ParlML: A Common Vocabulary for Parliamentary Language
- Tax Information Group for ECommerce Requirements Standardization (TIGERS)
- COSCA/NACM JTC XML Court Filing Project
- New Mexico District Court XML Interface (XCI)
- Georgia State University Electronic Court Filing Project
- Web Standards Project (WSP)
- BIC Workgroup for XML-based eBusiness Standard Convergence

What is XML and Why Should You Care?

- XML Mail Transport Protocol (XMTP) for XML SMTP and MIME Representation
- HTML Threading - Use of HTML in Email
- Open Software Description Format (OSD)
- Log Markup Language (LOGML)
- XLF (Extensible Log Format) Initiative
- ALURe (Aggregation and Logging of User Requests) XML Specification
- Apache XML Project
- Relational Markup Language (RML)
- WAP Wireless Markup Language Specification
- Common Profile for Instant Messaging (CPIM)
- XML Encoding for SMS (Short Message Service) Messages
- MessageML
- Multi-Channel Access XML (MAXML)
- The SyncML Initiative
- XML Document Navigation Language (XDNL)
- HTTP Distribution and Replication Protocol (DRP)
- Materials Property Data Markup Language (MatML)
- Measurement Units Markup Language
- XML-Based 'Chem eStandard' for the Chemical Industry
- Chemical Markup Language (CML)
- Molecular Dynamics [Markup] Language (MoDL)
- NCBI Molecular Biology Data Model
- StarDOM - Transforming Scientific Data into XML
- Bioinformatic Sequence Markup Language (BSML)
- BIOPolymer Markup Language (BIOML)
- CellML
- Gene Expression Markup Language (GEML)
- GeneX Gene Expression Markup Language (GeneXML)
- Genome Annotation Markup Elements (GAME)
- MicroArray and Gene Expression Markup Language (MAGE-ML)
- Microarray Markup Language (MAML)
- XML for Multiple Sequence Alignments (MSAML)
- Systems Biology Markup Language (SBML)
- OMG Gene Expression RFP
- Protein Extensible Markup Language (PROXIML)
- Taxonomic Markup Language
- XDELTA: XML Format for Taxonomic Information
- The Species Analyst Project
- Virtual Hyperglossary (VHG)
- Weather Observation Definition Format (OMF)
- Open Philanthropy Exchange (OPX)
- Open Financial Exchange (OFX/OFE)
- Interactive Financial Exchange (IFX)
- FinXML - 'The Digital Language for Capital Markets'
- Investment Research Markup Language (IRML)
- Extensible Financial Reporting Markup Language (XFRML)
- Extensible Business Reporting Language (XBRL)
- XMLPay Specification
- Financial Products Markup Language (FpML)
- Treasury Workstation Integration Standards Team (TWIST)
- Market Data Definition Language (MDDL)

What is XML and Why Should You Care?

- Market Data Markup Language (MDML)
- Weather Markup Language (WeatherML)
- MarketsML Initiative
- Research Information Exchange Markup Language (RIXML)
- Data Link for Intermediaries Markup Language (daliML)
- SwiftML for Business Messages
- Straight Through Processing Markup Language (STPML)
- FAML DTD for Financial Research Documents
- XML-MP: XML Mortgage Partners Framework
- Mortgage Bankers Association of America MISMO Standard
- EcoKnowMICS ML
- Trading Partner Agreement Markup Language (tpaML)
- Internet Open Trading Protocol (IOTP)
- XML Voucher: Generic Voucher Language
- XML Messaging (IETF)
- XML Common Biometric Format (XCBF)
- Security Assertion Markup Language (SAML)
- Web Services Security Specification (WS-Security)
- Security Services Markup Language (S2ML)
- XML Access Control Language (XACL)
- AuthXML Standard for Web Security
- Intrusion Detection Message Exchange Format
- Digital Signatures for Internet Open Trading Protocol (IOTP)
- XML Encoding of SPKI Certificates
- Digital Receipt Infrastructure Initiative
- Digest Values for DOM (DOMHASH)
- Signed Document Markup Language (SDML)
- Financial Services Markup Language (FSML)
- FIXML - A Markup Language for the FIX Application Message Layer
- Electronic Commerce Modeling Language (ECML)
- Bank Internet Payment System (BIPS)
- smartX ['SmartCard'] Markup Language (SML)
- Real Estate Transaction Markup Language (RETMML)
- OpenMLS and RELML (Real Estate Listing Markup Language)
- Data Consortium (Real Estate Standards)
- Comprehensive Real Estate Transaction Markup Language (CRTML)
- ACORD: XML for the Insurance Industry
- iLingo XML Schemas for Insurance
- Customer Profile Exchange (CPEX) Working Group
- Customer Support Consortium
- Standards for Technology in Automotive Retail (STAR)
- XML for the Automotive Industry - SAE J2008
- Spacecraft Markup Language (SML)
- Electronic Business XML Initiative (ebXML)
- Universal Business Language (UBL)
- UN/CEFACT XML Business Document Library Project (XBDL)
- DISA Registry Initiative (DRIVE)
- BASDA eBIS-XML
- Portal Markup Language (PML)
- EDGARspace Portal
- Global Commerce Initiative (GCI)
- DII Common Operating Environment (COE) XML Registry

What is XML and Why Should You Care?

- StarOffice XML File Format
- Open eBook Initiative (OEB)
- ONIX International XML DTD
- NISO Digital Talking Books (DTB)
- Mathematical Markup Language (MathML)
- OpenMath Standard
- OMDoc: A Standard for Mathematical Documents
- Re-Useable Data Language (RDL)
- Metadata - PICS
- MIX - Mediation of Information Using XML
- CDIF XML-Based Transfer Format
- Synchronized Multimedia Integration Language (SMIL)
- Multimodal Presentation Markup Language (MPML)
- Moving Picture Experts Group: MPEG-7 Standard
- MPEG-21 Part 2: Digital Item Declaration Language (DIDL)
- DIG35: Metadata Standard for Digital Images
- Extensible Metadata Platform (XMP)
- W3C Scalable Vector Graphics (SVG)
- WebCGM
- Precision Graphics Markup Language (PGML)
- Vector Markup Language (VML)
- Image Markup Language (IML)
- VRML (Virtual Reality Modeling Language) and X3D
- Covad xLink API (XML-Based DSL Provisioning)
- WebBroker: Distributed Object Communication on the Web
- Web Interface Definition Language (WIDL)
- XML/EDI - Electronic Data Interchange
- Global Engineering Networking Initiative (GEN)
- XML/EDI Repository Working Group
- TranXML
- Value Chain Markup Language (VCML)
- TransportationXML (tXML)
- Tracker XML (TXML)
- Uniform Code Council (UCC) XML Program
- Physical Markup Language (PML)
- VICS CPFR XML Messaging Standard
- Global Uniform Interoperable Data Exchange (GUIDE)
- Implementation Guideline Markup Language (igML)
- BizCodes Initiative
- Universal Data Element Framework (UDEF)
- EEMA EDI/EC Work Group - XML/EDI
- OpenTravel Alliance (OTA)
- Hospitality Industry Technology Integration Standards (HITIS) Project
- Information and Content Exchange (ICE)
- CommerceNet Industry Initiative
- eCo Interoperability Framework Specification
- BizTalk Framework
- eCo Framework Project and Working Group
- Commerce XML (cXML)
- Marketplace XML (mpXML)
- QuickBooks Extensible Markup Language (qbXML)
- ArapXML for General Ledger and Account Receivable/Account

What is XML and Why Should You Care?

- Payable Integration
- SMBXML: An Open Standard for Small to Medium Sized Businesses
- RosettaNet
- Open Catalog Protocol (OCP)
- eCatalog XML (eCX)
- vCARD (Electronic Business Card in XML and RDF)
- Hybrid Mail Language (HML)
- Markup Languages for Names and Addresses
- British Standard BS7666 for Geographical Referencing
- xNAL Name and Address Standard (xNL, xAL)
- Customer Identity/Name and Address Markup Language (CIML, NAML)
- AND Global Address XML Definition
- Whois Export and Exchange Format
- CECA XML Specification for Civil Estate Data
- Historical Event Markup and Linking (HEML)
- iCalendar DTD Document (xCal)
- XML Encoded Form Values
- Capability Card: An Attribute Certificate in XML
- Telecommunications Markup Language (tML)
- Telecommunications Interchange Markup (TIM, TCIF/IPI)
- aecXML Working Group - Architecture, Engineering and Construction
- Building Construction Extensible Markup Language (bcXML)
- MasterBuilder Construction Management and Accounting
- eBuild-XML
- Green Building XML (gbXML)
- Product Data Markup Language (PDML)
- Product Definition Exchange (PDX)
- Electronic Component Information Exchange (ECIX) and Pinnacles Component Information Standard (PCIS)
- ECIX QuickData Specifications
- ECIX Component Information Dictionary Standard (CIDS)
- ECIX Timing Diagram Markup Language (TDML)
- UML eXchange Format (UXF)
- Java Architecture for XML Binding (JAXB)
- XML Data Binding Specification
- XML Localization Interchange File Format (XLIFF)
- Translation Memory Exchange (TMX)
- OpenTag Markup
- P3P Specification: Platform for Privacy Preferences
- Extensible Name Service (XNS)
- Dialogue Moves Markup Language (DMML)
- Scripting News in XML
- InterX.org Initiative
- DESSERT (Document Encoding and Structuring Specification for Electronic Recipe Transfer) (Now renamed RECIPE)
- NuDoc Technology
- Coins: Tightly Coupled JavaBeans and XML Elements
- DMTF Common Information Model (CIM)
- Universal Plug and Play Forum
- XML Transition Network Definition (XTND)
- IPDR.org Network Data Management Usage Specification

What is XML and Why Should You Care?

- Process Interchange Format XML (PIF-XML)
- Multilingual Upper-Level Electronic Commerce Ontology (MULECO)
- DARPA Agent Markup Language (DAML)
- Robotic Markup Language (RoboML)
- Rule Markup Language (RuleML)
- Agent-Oriented Rule Markup Language (AORML)
- Extensible Rule Markup Language (XRML)
- Simple Rule Markup Language (SRML)
- Relational-Functional Markup Language (RFML)
- Ontology and Conceptual Knowledge Markup Languages
- Information Flow Framework Language (IFF)
- Simple HTML Ontology Extensions (SHOE)
- XOL - XML-Based Ontology Exchange Language
- Description Logics Markup Language (DLML)
- Case Based Markup Language (CBML)
- Artificial Intelligence Markup Language (AIML)
- Physics Markup Language (PhysicsML)
- Procedural Markup Language (PML)
- QAML - The Q&A Markup Language
- LACITO Projet Archivage de données linguistiques sonores et textuelles [Linguistic Data Archiving Project]
- Geography Markup Language (GML)
- UK Digital National Framework (DNF) for Geographical Information
- Point of Interest Exchange Language Specification (POIX)
- Exploration and Mining Markup Language (XMML)
- LandXML
- Navigation Markup Language (NVML)
- Extensible Data Format (XDF)
- FITSML for Flexible Image Transport System
- NASA Goddard Astronomical Data Center (ADC) 'Scientific Dataset' XML
- Extensible Scientific Interchange Language (XSIL)
- Object Oriented Data Technology (OODT) and XML
- Astronomical Markup Language
- Astronomical Instrument Markup Language (AIML)
- GedML: [GEDCOM] Genealogical Data in XML
- BannerML
- adXML.org: XML for Advertising
- SPACE XML
- Newspaper Association of America (NAA) - Standard for Classified Advertising Data
- AdMarkup XML DTD for Classified Advertising
- News Industry Text Format (NITF)
- XMLNews: XMLNews-Story and XMLNews-Meta
- NewsML and IPTC2000
- SportsML
- News Markup Language (NML)
- XMLTV
- Notes Flat File Format (NFF)
- Java Help API
- Cold Fusion Markup Language (CFML)
- Edge Side Includes (ESI)

What is XML and Why Should You Care?

- Character Mapping Markup Language (CharMapML)
- WEBDAV (IETF 'Extensions for Distributed Authoring and Versioning on the World Wide Web')
- DAV Searching and Locating (DASL)
- Graphic Communications Association - GCA 'Paper' DTD
- DocBook XML DTD for Technical Documentation
- Apache Cocoon JavaDoc Documentation in XML
- JDoc: XML Format for Sun Javadoc
- JRefEntry DTD
- XML for Publishers and Printers (XPP)
- Job Definition Format (JDF)
- Personalized Print Markup Language (PPML)
- Printing Industry Markup Language (PrintML)
- PML: Markup Language for Paper and Printing
- PrintTalk Consortium
- printcafe eProduction eCommerce eXchange (PCX)
- PostSecondary Electronic Standards Council XML Forum for Education
- IEEE LTSC XML Ad Hoc Group
- Universal Learning Format Technical Specification
- Educom Instructional Management Systems Project (IMS) Metadata Specification
- Shareable Content Object Reference Model Initiative (SCORM)
- Learning Material Markup Language (LMML)
- Schools Interoperability Framework (SIF)
- Tutorial Markup Language (TML)
- International Development Markup Language (IDML)
- Voice Browser Call Control (CCXML)
- Call Processing Language (CPL)
- Call Policy Markup Language (CPML)
- VoiceXML Forum (Voice Extensible Markup Language Forum)
- Speech Application Language Tags (SALT)
- CallXML
- VoXML Markup Language
- Telephony Markup Language (TML)
- DARPA Communicator Project and XML Log Standard
- Multilevel Annotation, Tools Engineering (MATE)
- Computing Environment for Linguistic, Literary, and Anthropological Research (CELLAR)
- Architecture and Tools for Linguistic Analysis Systems (ATLAS)
- TalkBank and the Codon XML-Based Annotation Framework
- ACE Pilot Format DTDs
- Transcriber - Speech Segmentation and Annotation DTD
- Natural Language Semantics Markup Language
- Extensible Telephony Markup Language (XTML)
- SABLE: A Standard for Text-to-Speech Synthesis Markup
- W3C Speech Synthesis Markup Language Specification
- W3C Speech Recognition Grammar Specification
- Java Speech Markup Language (JSML/JSpeech)
- SpeechML
- TalkML
- Project Management XML Schema
- XML for Workflow Management [NIST]

What is XML and Why Should You Care?

- SWAP - Simple Workflow Access Protocol
- XML-Based Workflow [Process Management] Standard: Wf-XML
- Exchangeable Routing Language (XRL)
- Architecture Description Markup Language (ADML)
- Human Markup Language (HumanML)
- Theological Markup Language (ThML)
- XML Scripture Encoding Model (XSEM)
- Open Scriptural Information Standard (OSIS)
- OpenText.org Papyrus Encoding Markup
- LitML: A Liturgical Markup Language
- XML-F ('XML for FAX')
- XHTML-FML: Forms Markup Language
- Extensible Forms Description Language (XFDL)
- XML Forms Architecture (XFA)
- Electronic Form System (EFS)
- Broadcast Markup Language (BML)
- Broadcast Hypertext Markup Language (BHTML)
- IEEE Standard DTD
- Open Settlement Protocol (OSP) - ETSI/TIPHON
- Directory Services Markup Language (DSML)
- DirXML
- DIF Directory Interoperability Proposal
- XML DTD for ACAP - ACAP Data Interchange Format
- WDDX - Web Distributed Data Exchange
- XIOP - XML Corba Environment-Specific Inter-ORB Protocol
- XML-RPC
- Blocks eXtensible eXchange Protocol Framework (BEEP)
- ANTACID Replication Service
- IETF Working Group for Open Pluggable Edge Services (OPES)
- Layered Object Transport Protocol (LOTP)
- XML for Exchange of Structure and Identification of Management Information (SMI)
- WorldOS
- Business Process Modeling Language (BPML)
- Business Rules Markup Language (BRML)
- XML Common Business Library (xCBL)
- Universal Commerce Language and Protocol (UCLP)
- VISA XML Invoice Specification
- NACS XML Data Interchange (NAXML)
- ARTS IXRetail
- First Retail Mark-up Language
- Open Applications Group - OAGIS
- Schema for Object-oriented XML (SOX)
- XMLTP.Org - XML Transfer Protocol
- The XML Bookmark Exchange Language (XBEL)
- Simple Object Definition Language (SODL) and XMOP Service
- W3C XML Protocol
- Web Services Interoperability Organization (WS-I)
- Simple Object Access Protocol (SOAP)
- Universal Description, Discovery, and Integration (UDDI)
- Web Services Conversation Language (WSCL)
- Web Services Description Language (WSDL)
- Web Services for Interactive Applications (WSIA)

What is XML and Why Should You Care?

- Web Services Flow Language (WSFL)
- Web Services User Interface (WSUI) Initiative
- Web Services for Remote Portals (WSRP)
- Web Services Experience Language (WSXL)
- Direct Internet Message Encapsulation (DIME)
- Transaction Authority Markup Language (XAML)
- XML Encoding Rules for ASN.1 (XER)
- Object-Oriented Programming Meta-Language (OOPML)
- FlowML: A Format for Virtual Orchestras
- Clinical Data Interchange Standards Consortium
- National Library of Medicine (NLM) XML Data Formats
- ISIS European XML/EDI Healthcare Pilot Project (XML/EPR)
- Open Healthcare Group 'XChart'
- DocScope: Open Source XML Healthcare Project
- Health Level Seven XML Patient Record Architecture
- ASTM XML Document Type Definitions (DTDs) for Health Care
- The CISTERN Project - Standard XML Templates for Healthcare
- Template Definition Language (TDL)
- Human Resource Management Markup Language (HRMML)
- HR-XML Consortium
- Staffing Industry Data Exchange Standards (SIDES)
- XML-HR Initiative - Human Resources
- Rosetta Group XML Résumé Library
- ECMDATA - Electronic Component Manufacturer Data Sheet Inventory Specification
- Bean Markup Language (BML)
- The Koala Bean Markup Language (KBML)
- Jigsaw XML Format (JigXML)
- MOS-X (Media Object Server - XML)
- FLBC (Formal Language for Business Communication) and KQML
- ISO 12083 XML DTDs
- Electronic Thesis and Dissertation Markup Language (ETD-ML)
- Extensible User Interface Language (XUL)
- The Extensible Bindings Language (XBL)
- User Interface Markup Language (UIML)
- Process Specification Language (PSL) and XML
- Batch Control Markup Language
- SCL Component Test Bed Specification
- AgXML
- American Iron and Steel Institute (AISI) XML Workgroup
- Steel Markup Language (SML)
- Energy Trading Standards Group (ETSG)
- Petrotechnical Open Software Corporation (POSC) XML Related Projects
- PetroXML Initiative
- Partner Interface Process for Energy (PIPE)
- Marine Trading Markup Language (MTML)
- Navy CALS Initiatives XML
- eFirst XML for Scholarly Articles
- XML DTD for Phone Books
- Using XML for RFCs
- Guideline XML (gXML)
- XML Belief Network File Format (Bayesian Networks)

- Predictive Model Markup Language (PMML)
- Data Documentation Initiative: A Project of the Social Science Community
- Triple-s XML Survey Interchange Standard
- SODA2 - An XML Semistructured Database System
- RAX - Record API for XML
- Chess Markup Language (ChessML)
- Mind Reading Markup Language (MRML)

Parts of an XML Application

slide 32

Logical Components of an XML Application

1. XML document (tags and text)
2. DTD or schema (the model of the document)
3. Display Specifications (how formats/behaves)
4. Transformations (from this to that)
5. XML repository or content management system

Software components may include

- XML-aware editors
- XML-authoring templates or forms
- XML composition engines
- Non-XML composition engines
(InDesign, QuarkXPress, HTML browser)

Component 1: 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)

Component 2: The Document Model

DTD (Document Type Definition) or Schema

- Model for one type/class of information (a “document”)
(reference book, bank transfer, journal article, memo, help-topic)
- Set of rules describing how documents of that type can be marked up
- Written in a formal syntax of XML

DTDs / Schemas Express Rules

for example:

- **Reference book** =
Book-level Metadata followed by
Front Matter followed by
Body followed by
Back Matter (which is optional)
- **Purchase Order** =
Order Header followed by
List of Order Detail followed by
Order Summary (which is optional)
- **paragraph** =
data characters and may include any of the following: *Person Names*,
URLs, and/or *Geographic Regions*

Why Use a DTD or Schema?

- DTD or schema 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
- Supports interoperable families of documents
 - ensure that information conforms to model (validation)
 - parties don't have to share software or applications
 - reduces surprises

To Share Information, Share the DTD / Schema

- Publisher communicates to conversion house
- Content provider explains tagging to
 - compositor for typesetting
 - web designer for building website
 - database or repository designer
 - software vendor for customization

Current XML Modeling/Constraint Languages

(so you will have heard the names)

- Formal definitions of XML vocabulary/constraints
- Express XML models and/or constraints for XML documents
- Used in XML validation (called “parsing”)
- Languages
 - XML 1.0 and 1.1 DTD (Document Type Definition) [W3C]
 - W3C XML Schema (XSD) [W3C]
 - RELAX NG [ISO]
 - Schematron [ISO]

Component 3: Formatting (and Behavior)

- Stylesheets
- Display Specifications
- Formatting

Remember What XML Looks Like Without Formatting

```
<?xml version="1.0" encoding="utf-8"?>
<!DOCTYPE collection SYSTEM "CollectionDTD/collection.dtd">
<collection>
<title>Recipe Collection: Breads and Soups</title>

<recipe>

<title>Multi-seed Bread</title>

<class name="type of dish">yeast bread</class>

<component>
<ingredients>
<ingredient>
  <quantity>2</quantity>
  <measure>pkts</measure>
  <foodstuff>active dry yeast</foodstuff></ingredient>
<ingredient>
  <quantity>&frac14;</quantity>
  <measure>c</measure>
  <foodstuff>warm water</foodstuff></ingredient>
<ingredient>
  <quantity>2</quantity>
  <measure>c</measure>
  <foodstuff>warm milk</foodstuff></ingredient>
<ingredient>
  <quantity>&frac34;</quantity>
  <measure>c</measure>
  <foodstuff>sugar</foodstuff></ingredient>
<ingredient>
  <quantity>&frac12;</quantity>
  <measure>c</measure>
  <foodstuff>butter</foodstuff></ingredient>
<ingredient>
  <quantity>1 &frac12;</quantity>
  <measure>tsp</measure>
  <foodstuff>salt</foodstuff></ingredient>
<ingredient>
  <quantity>&frac12;</quantity>
  <measure>c</measure>
  <foodstuff>mixed toasted seeds</foodstuff></ingredient>
```

What is XML and Why Should You Care?

```
<ingredient>
  <quantity>2</quantity>
  <foodstuff>eggs</foodstuff></ingredient>
<ingredient>
  <quantity>7-8</quantity>
  <measure>c</measure>
  <foodstuff>all-purpose flour</foodstuff></ingredient>
</ingredients>

<directions>
<step><p>Disolve yeast in warm water. Add milk, sugar, butter,
salt, eggs, seeds and 3 c flour. Beat until smooth. Stir in
enough remaining flour to form a soft dough ball, and knead.
Let rise until doubled.</p></step>

<step><p>Punch down, divide into halves, shape, and
let rise.</p></step>

<step><p>Bake 350&deg;F for 25-30 min.</p></step>
</directions>
</component>

<source>Becky</source>

<yield>2 loaves</yield>

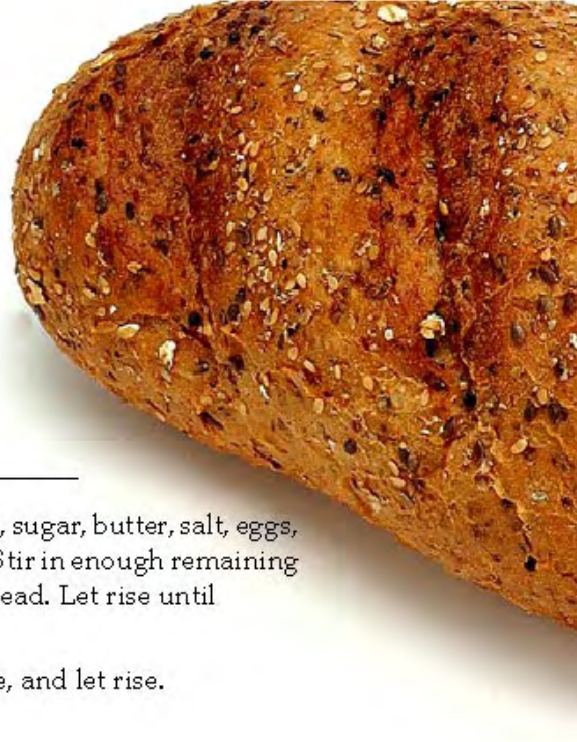
<illustration filename="Graphics/long-loaf.jpg" />

</recipe>
```

What We Would Like to See (Print or Screen)

Multi-seed Bread

2 pkts	active dry yeast
¼ c	warm water
2 c	warm milk
¾ c	sugar
½ c	butter
1 ½ tsp	salt
½ c	mixed toasted seeds
2	eggs
7-8 c	all-purpose flour



1. Dissolve yeast in warm water. Add milk, sugar, butter, salt, eggs, seeds and 3 c flour. Beat until smooth. Stir in enough remaining flour to form a soft dough ball, and knead. Let rise until doubled.
2. Punch down, divide into halves, shape, and let rise.
3. Bake 350°F for 25-30 min.

XML Design Feature

XML separates content from format/behavior

- XML does not say
 - how it looks (16 point Helvetica Bold)
 - what it does (starts a javascript)
- A specification must provide format mapping (usually called a stylesheet, display specification, or output spec)

Therefore Display or Print Needs a Display Specification

- Says what XML data will look like
 - on paper or screen
- Defines an appearance or rendition or behavior
 - for each element
 - in each of its contexts within a document
 - is NOT one tag per one format
(Title of Figure may look different from Title of Chapter)

Format/Behavior is Based on the Tagging (but context dependent)

Display Specifications Give Instructions

(A lot like typesetting specs)

- New chapter ==> New recto page
- New recipe ingredient ==> New list item
- Cross-reference ==> Make it a link
- Citation ==> Add it to the bibliography
(and sort the bibliography before printing it)
- Key term ==> Check that there's a definition for it
- Chapter/title ==>
 - make it bold 18 point Arial *and*
 - put it in the table of contents at 11 point Times Roman *and*
 - put it in the running head at 9 point Arial

One for All or All for One

- One stylesheet, many documents
 - maintains consistency of format (“look and feel”) across documents
 - is easy to develop, maintain, and apply (house style)
- One document, many stylesheets
 - allows for different media types: print, on-line, etc.
 - is easy to produce derivative documents: selections, summaries, indexes, catalogs, ...

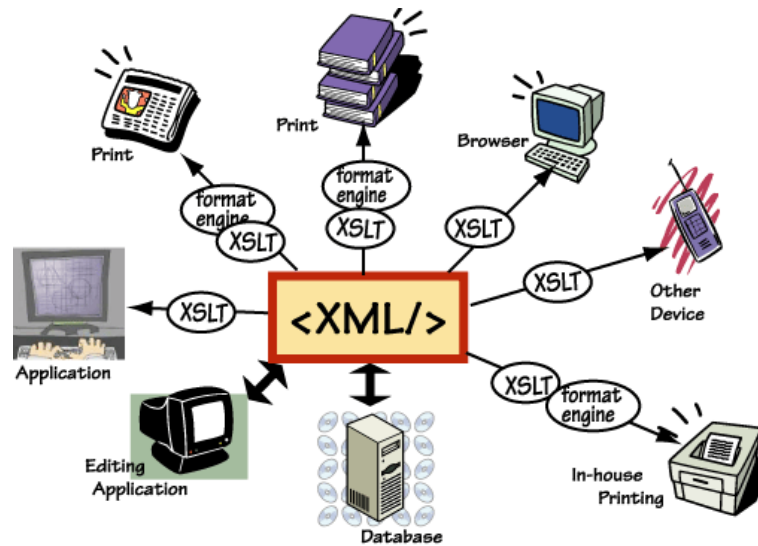
Component 4: XML Transformation

Transforms XML Documents

- 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
- Can transform from one set of tags directly into another

XSLT (Extensible Stylesheet Language Transformation)

XSLT for XML



Why This Is Exciting

- Conventional wisdom was that to make XML useful *we must all use the same tags*
- Transformation means maybe not
- Transform my tags into your tags
- Take a small subset of my tags that you can handle
- We can alias elements and content *when elements are:*
 - semantically the same but different names (purchase-order, po, order, client-purchase)
 - semantically or functionally “close enough” (postal-code, zip-code, parish-number)
 - my elements are recombinations or subsets of yours

Component 5: XML Repository or Content Management System

Native XML Databases

- All the useful usual database features
- Plus,
 - objects stored *are* XML
 - DTD- or schema-aware (knows what valid is)
 - XML context preserved
 - check in and out by element
 - XML search and transformation tools

Content Management Using an XML Repository

- Manage content at many levels of granularity
- Combine data from many sources
- Reuse and repurpose data / Electronic slice and dice data
- Increase searching precision
- Feed many applications from one repository
- Customize output
- Enterprise information portals

XML in Non-XML Databases

- XML is stored in
 - OO databases
 - relational databases
 - object-relational and hierarchical databases
- XML tags not there or hidden in fields or globs
- All the useful database features (storage, search, metadata, inheritance, referential integrity)
- Getting XML out
 - built-in utilities to create XML from relational data
 - requires mapping of XML to fields (maybe non-trivial)

Lots of XML is Managed in File Systems

- Content Management Systems are powerful
- Native XML databases are sexy
- If the cost seems exorbitant
 - not to worry
 - XML can be managed with files

(Everybody needs transforms; some people need CMS/database)

Conclusion

slide 53

The Big News:

XML DOES NOT DO ANYTHING!

XML is a data format.

slide 54

You (and Your Software) Can Do a Lot with XML

- Make formatted pages
- Make web sites
- Make spin-off, sub-set, and superset documents
- Make rich discovery tools
 - RSS and ATOM feeds
 - detailed indices
 - navigable Tables of Contents
- Help your customers archive content for re-use
- Internationalization, localization, custom publishing

The Good News: You Can Do XML and Benefit

- Many good tools are now available
- Software developers are rapidly becoming XML savvy
- XSLT lets you move it all around and change it
- XML browsers are here
- XML databases are here
- Users of electronic products are more sophisticated, demanding better electronic products
- XML really *is* easier than inventing another data format!

Where to Get More Information

The Source for XML and Related Information

- Robin Cover's SGML/XML Web Page,
<http://xml.coverpages.org/>
or
<http://www.oasis-open.org/cover/>

General XML Information on the Web

- W3C's XML page: <http://www.w3.org/XML/>
- XML FAQ (Peter Flynn): <http://www.ucc.ie/xml/>
- XML.com <http://www.xml.com> (O'Reilly's XMLsite)
- XML.org: <http://www.xml.org> (The OASIS XML site)
- Tools
 - XML Tools <http://www.xmlsoftware.com>
 - Free XML Software List (Lars Marius Garshol):
<http://www.garshol.priv.no/download/xmltools/>

Books on XML Concepts (Not So Technical)

- *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 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.

XML Books We Recommend (More Technical)

- *The XML Companion (2nd Edition)*, by Neil Bradley (Addison-Wesley, 2000)
 - Very good basic technical introduction. Clear, concise, thorough, hype-free, for the serious newcomer.

What is XML and Why Should You Care?

- *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.
- *The SGML/XML Cookbook*, by Rick Jelliffe (Prentice Hall PTR, 1998)
 - Balanced, good coverage of broader issues. Oriented towards structured documents and character set issues for human use rather than EDI or systems-integration problems.

slide 60

Other Information Sources

- OASIS Home Page (vendor consortium), <http://www.oasis-open.org/home/index.php>
- Idealliance (formerly GCA) <http://www.idealliance.org>
- A basic newsgroup is `comp.text.xml` (also some on `comp.text.sgml`)
- Useful Lists
 - XSL-List, <http://www.mulberrytech.com/xsl/xsl-list/index.html>
 - XML-L, <http://listserv.heanet.ie/xml-l.html>
 - XML-Developer's List, <http://www.xml.org/xml/xmldev.shtml>

slide 61

Colophon

- Slides and handouts created from single XML source
- Slides projected from HTML which was created from XML using XSLT
- Handouts created from XML
 - source XML transformed to Open Office XML
 - Open Office XML opened in Open Office
 - pagination normally adjusted
 - Saved as PDF
- Slideshow tools available at <http://www.mulberrytech.com/slideshow>