Promotion and Education

XML Fundamentals in B2B e-Commerce

Version 2.2

Agenda

- What is XML?
- XML Benefits
- When and Where XML is used
- XML Example – Valid XML
- XML Standards
- Possible Issues with XML
- XML Terminology
- XML Future

What do we mean by "standard"?

Network Protocol	Transport	Object Invocation	Directory Services	Data Dictionary	Data Format	Presentation/ User Int'face
TCP/IP	FTP	IIOP	LDAP	ASC X12	X12	Graphical
	XML			EDIFACT	EDIFACT	HTML
	SMTP/ Mime	XML	X.500	RosettaNet	SGML	XSL, XHTML, DOM
	HTTP			Others	XML DTD	

Architecture Tools/Programming Languages						
C++	PERL	JAVA	XML, DOM	XQL	CGI	UML
OBI	EDIINT	OBI	VPN	Voice over IP	OTP	ODBC

DTD's, Schema

Where does XML (eXtensible Markup Language) fit into the standards setting framework? XML is a family of technologies that includes things like XSL, XHTML, Xlink, XQL and more

- ***Don't get hung up on all these terms right now.***

Data standards: 3300 BC.

Source: Jon Bosak presentation to BARD, 10/2002

What is XML?

- XML = “eXtensible Markup Language”
- XML originally derived from SGML
  - Brief History
  - Who developed? Why?
  - What organization “owns” the XML standard?
    - See: www.w3c.org/xml
- XML originally designed to improve managing documents (manuals, not POs) on web (publishing content) and has evolved from there
- XML is a meta-language (language to create languages) that allows you/ your company to design a markup language (tags) that describe data
- XML – the “lingua-franca” of B2B

XML Benefits

- Separation of content and presentation
- Development of flexible ecommerce applications
- Open standard – independent of platform, programming language, and media
- Enables more meaningful searches
- Flexible format for different use
  - Transfer format for sharing data
  - Persistent data format for disk storage

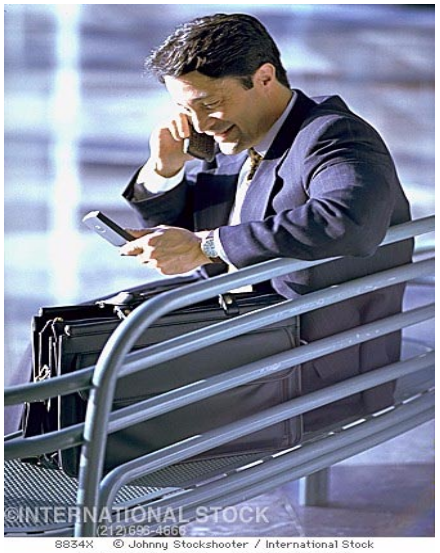
Possible Reasons for Using XML

- Broadens enterprise search capabilities
- Can perform new operations on data once it's in XML form
- May shorten application development time
- XML-enabled business processes will shorten business cycles
- Allows conversion of EDI data to more manageable form
- Ties together multiple internal applications across the company
- Ties applications to those customers
- Ties applications to those suppliers and trading partners
- Joining an e-Commerce trading hub or B2B exchange that uses XML

Zona Research

XML Usage

- Data-driven applications
- Application to web
- Device control
- Application to application
- Computer to computer (data exchange)



XML's Expansion

- Driven by Internet
- Global Interchange
  - Changes in info specifications
  - Potential enabler of ubiquitous connectivity and interoperability for e-Commerce
- Automated machine processing
  - Agents, transaction exchanges, dynamic searches/displays
  - Content processing
- Expansion of Web Documents
  - E-Commerce, E-Content, e-Books
  - Info & Knowledge Management

XML vs. HTML

HTML:

- A text stream is converted into the presentation of a web page. Tags are pre-defined
- Is display oriented

XML:

- A text stream is converted into a data object that may be presented
  - Structures of document are allowed to be nested as necessary
  - Information providers can define new tag and attribute names at will.
- Provides an abstraction layer and allows you to define why you set data apart

HTML/XML

- HTML:

  My Laptop 2000
 Such a Deal
$1438

- XML:

  <product><model>My Laptop 2000</model><dealer>Such a Deal</dealer> <price> $1438</price></product>

HTML vs. XML

```
<OL>
   <LI>HTML allows <B><I>improper nesting</B></I>.
   <LI>HTML allows start tags, without end tags, like the
       <BR> tag.
   <LI>HTML allows <FONT COLOR=#9900CC>attribute
       values</FONT> without quotes
   <li>HTML is case-insensitive
   <LI>White space is not important in HTML
</OL>
```

```
<OL>
   <LI>XML requires <B><I>proper nesting</I></B>.</LI>
   <LI>XML requires empty tags to be identified with
       a trailing slash, as in <BR/>.</LI>
   <LI>XML requires <FONT COLOR="#9900CC">quoted attribute
       values</FONT>.</LI>
   <LI>XML is case-sensitive
   <LI>White space is important in XML
</OL>
```

XML / EDI Comparison

EDI - ASC X12	XML - BizTalk's Semantic
BEG*00*SA*01009575*03641*19991107	<Beginning-Segment-for-Purchase-Order> <Transaction-Set-Purpose-Code Code= "00 Original"/> <Purchase-Order-Type-Code Code= "SA Standalone"/> <Purchase-Order-Number>01009575</Purchase-Order-Number> <Release-Number>03641</Release-Number> <Date>19991107</Date> </Beginning-Segment-for-Purchase-Order>

X12 standard tells us if elements are in the right order, if all mandatory elements are present, and what everything means (semantic interpretation)

Can read this and use it without a DTD or schema but need that DTD or schema if we want to validate it

XML / EDI Comparison

	X12/EDIFACT	XML
birthdate	1960s	1996, W3C standard in 1998 (syntax only)
mapping required	yes	yes
human readable?	Depends on what you call "human"	Depends on how tags are defined
semantic issues	yes	yes
Problems	1. fixed transaction sets 2. repetitive content standards – but there are only 2 of them 3. slow standards evolution 4. non-standard standards 5. highest costs are in back-end integration (80% of project cost) 6. limited penetration?	1. verbose tags can lead to large files 2. 1200+ content standards 3. slow standards evolution 4. non-standard standards 5. still need to do back-end integration 6. limited penetration (so far)
support	Specialized skills needed	Specialized skills needed

Data Driven Applications

- Data transferred using XML may be stored in/retrieved/queried from persistent storage
  - Native XML Databases
  - Major DB Vendors (Oracle, IBM, etc.) enabling storage of XML documents
- XML enables data-driven applications, because XML documents include tags that describe the data as well as the data itself

Application to Web

- XML can be used to transfer data from an application to a user's (client) web browser
  - For example, catalog data stored in XML may be displayed in a web page
    - Presentation of the data may be rendered in the web browser using a style sheet (discussed later)
    - Presentation of the data may be rendered in the web server and sent as HTML (or WML – Wireless Markup Language) to client

Device Control

- XML is increasingly being used to facilitate the display of data on the multitude of different devices connected to the web
  - Problem: Each mobile device, web browser, etc. have somewhat different presentation requirements
  - Underlying data is stored in XML
  - Presentation is controlled through style sheets / server-side XML data manipulation

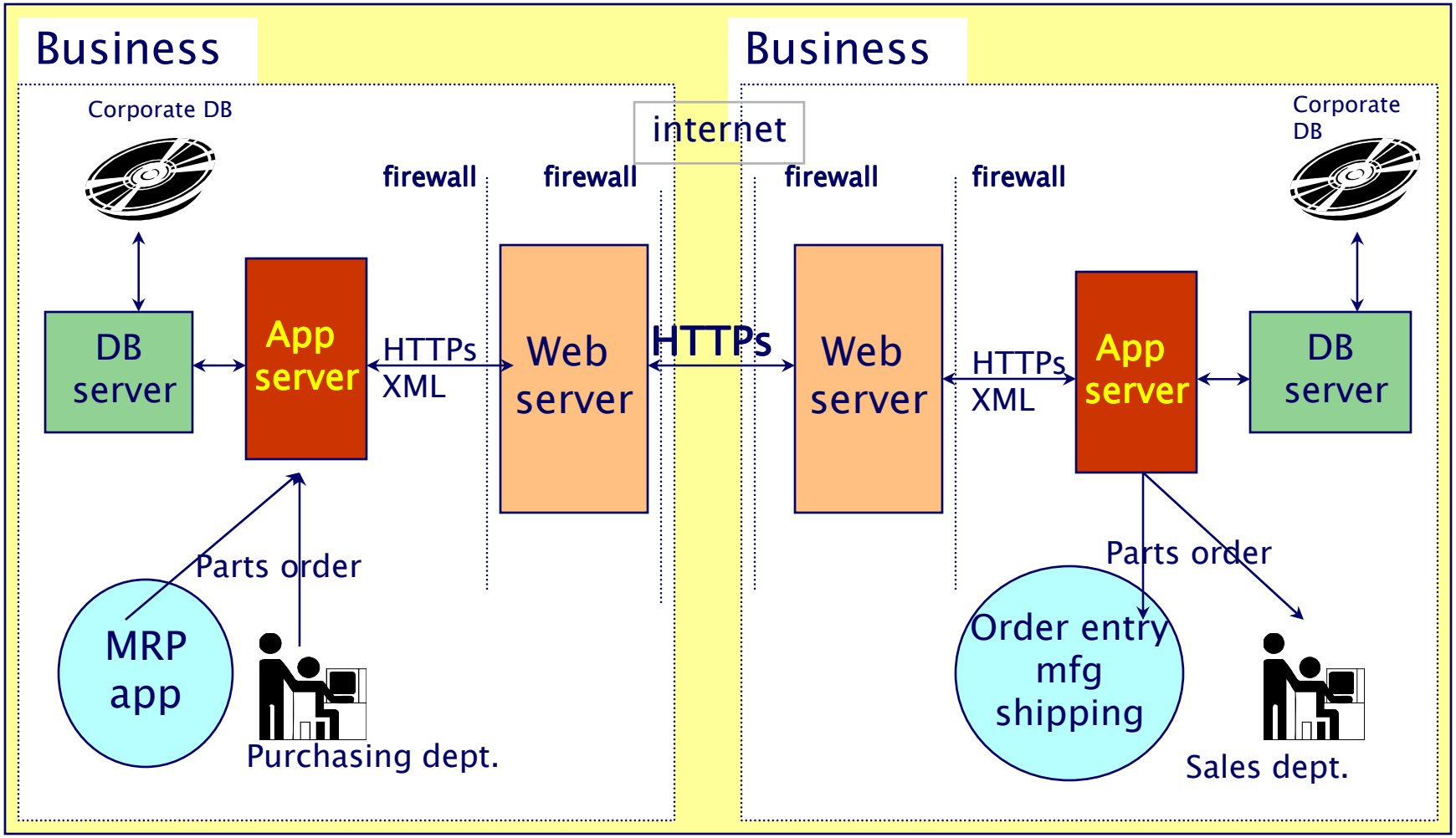
Application-to-Application (A2A)

- Applications may use XML to communicate between themselves
  - For example, an ERP System may transfer transactional data to/from a CRM system using XML
  - A2A is typically “behind the firewall”

Computer-To-Computer Data Exchange

- One server may transfer XML data to another server
  - Server-to-Server XML data exchange is what is envisioned to accomplish B2B exchange of transaction data (in B2B exchanges, this is “outside the firewalls”)
  - Server-to-Server XML data exchange may also happen within in business to transfer data within a company (“within the firewalls”)

B2B E-Commerce Application



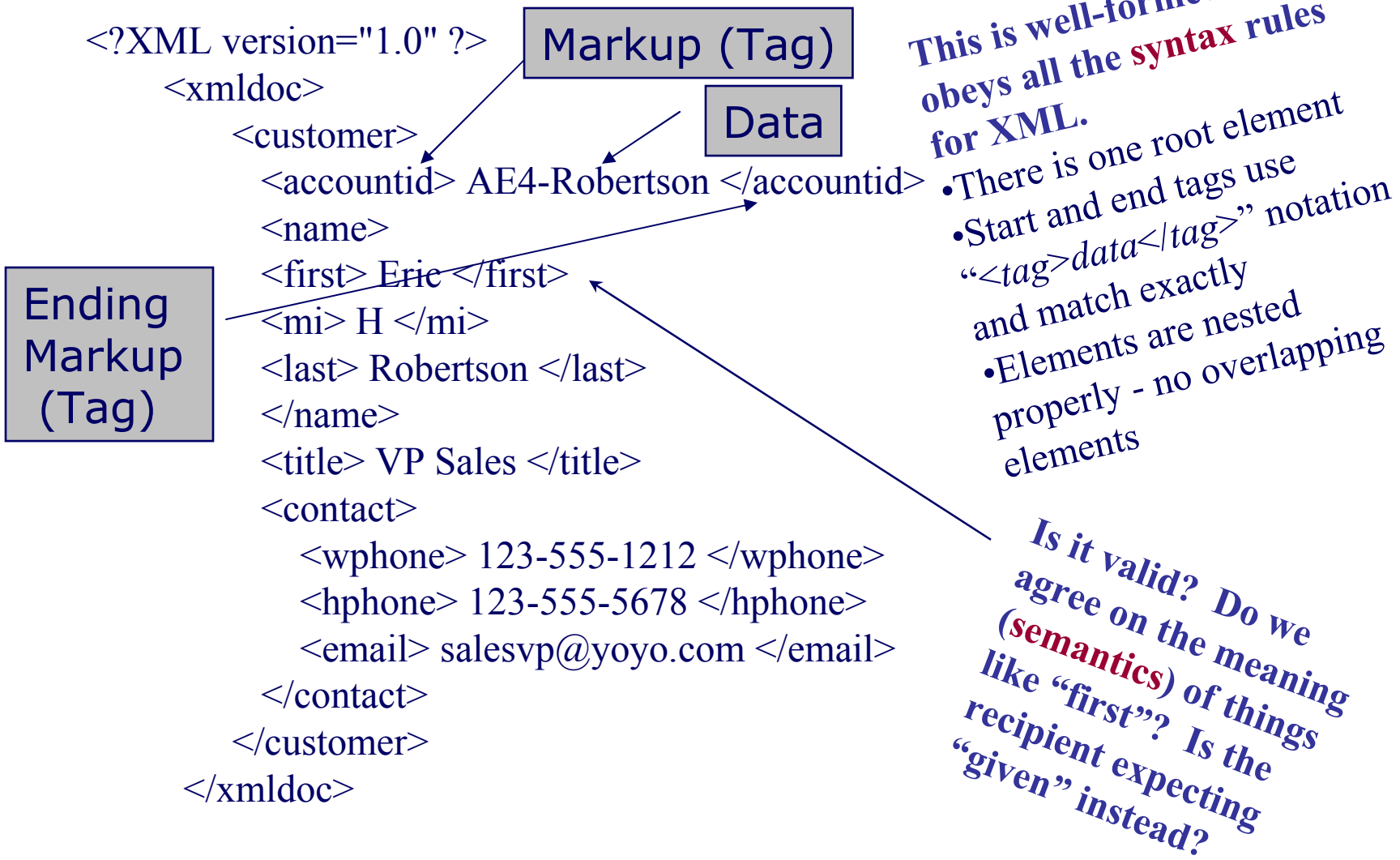
XML & E-Commerce

- XML alone is not a cure for all the ills of e-commerce
  - XML conveys content & structure
  - XML does not convey presentation, behavior or meaning
- To associate presentation, behavior and meaning with the data requires additional components.
  - ➔ Intelligent definition and use of these additional components, along with XML, is key to the next generation of E-Commerce.

Rules for Well-Formed XML

- Some basic rules for XML defined by the W3C in the XML Specification
  - All tags must be balanced - <TAG>...</TAG>
  - Empty tags expressed - <EMPTY_TAG/>
  - Tags must be nested - <B><I>...</B></I>
  - All element attributes must be quoted - <TAG name="value">
  - Text is case-sensitive - <TAG> != <Tag>
  - Comments are allowed - <!-- ... -->
  - Must begin - <?xml version='1.0' ?>
  - Special characters must be escaped
- The W3C specification does not address content - semantics

Well-Formed XML vs. Valid XML



Well-Formed vs Valid XML

Well-Formed: Obeys all the syntax rules

- Syntax rules:
  - 1 root element
  - Start & End Tags match exactly – case sensitive
  - Proper Nesting – No Overlapping Elements

Valid: Must agree on the meaning (semantics)

- Sender / Recipient must agree on what the tags mean
- Sender / Recipient must agree on the structure and sequence of the data
- Sender / Recipient must agree on what's required and what's optional
- Sender / Recipient must agree on datatypes

Valid XML Documents

- XML documents are validated through various means
  - Document Type Definition (DTD)
  - XML Schema
  - Resource Description Framework (RDF)
  - Or a variety of increasingly obsolete methods –
    - Document Content Description (DCD)
    - Schema for Object-Oriented XML (SOX)
    - Document Definition Markup Language (DDML)

Document Type Definition (DTD)

- Building blocks of XML document
- Describes document structure with list of elements
- Example Rules
  - All items in a catalog document must have a unit price
  - Document may have one or more occurrences or an item
- Ensure consistency of tag names and attributes

Document Type Definitions(DTD)

- Pros:
  - Mixed inheritance (use of SGML software)
  - Solid foundation, 10+ years experience
  - Capable but misunderstood
- Cons:
  - Lack Flexibility
  - Seldom kept up-to-date
  - Limited data types for validation
  - Not built on XML syntax

DTD Example

```
<?xml version='1.0'?>
<!Element Submitter (Name, Company)>
<!Element Name (Title, Phone)>
<!Element Name (#PCDATA)>
<!Element Title (#PCDATA)>
<!Element Phone (#PCDATA)>
<!Element Company (#PCDATA)>
```

XML Schemas

- Alternative to DTD's, built on XML syntax
- Defines classes of documents
- Contract to interchange information
- Defines structure, constraints and data types
- Expresses only part of the semantics
- Validation

Why Schemas?

- Common Vocabularies
- Formal Set of Rules
- Building Document Requirements
- Since schemas are expressed in XML, they can be parsed by XML software

XML Standards and Standards Bodies

- Syntax Standards: W3C
- Registry/Repository / Messaging / Transport / Routing Standards
  - BizTalk
  - UDDI
  - ebXML, OASIS, UN/CEFACT
- Semantic Standards
  - CBL
  - CXML
  - OAG
  - RosettaNet

    (note: RNIF, a part of the RosettaNet initiative, is a Messaging Transport Routing Standard)

World Wide Web Consortium (W3C)

- "Owns" the XML specification
  - "thou will use pointy brackets to contain a tag"
  - "thou may'st make up thine own tags as long as they don't collide with a small set of pre-defined tags"
  - "thou may'st use DTD's and stylesheets to define semantics" (until such time as DTD's are no longer supported)
- W3C responsible for 20+ XML specifications
  - Schema, XML Linking, XML Query, Namespaces, Stylesheets, XML Signature, DOM, etc.
- For more info, www.w3c.org

W3C WORLD WIDE WEB consortium

XML Infrastructure Standards

Messaging Routing & Transport Standards

CompTIA
OakBrook Terrace, IL USA

New Member Company
123 Main St.
Somewhere, ST 12345

#

XML Registry/Repository Standards

Databases, or definitions of databases, to retain XML schema/DTD's,business process definitions, and trading partner capabilities

BizTalk

- BizTalk is owned/produced by Microsoft
- BizTalk actually has 3 components:
  - BizTalk Framework – an XML standard for transferring XML messages
  - BizTalk Server – Software for processing XML
  - BizTalk.Org – a repository of XML specifications
- For more info, www.microsoft.com and www.biztalk.org

UDDI

- Universal Description, Discovery and Integration
- Platform independent, open framework for
  - describing services
  - discovering businesses
  - integrating business services
- Directory has “white pages” “yellow pages” and “green pages”
- More Info:

  www.uddi.org

ebXML

- ebXML = "Electronic Business XML"
  - Project sponsored by UN/CEFACT and OASIS
  - Mission: "Enable Enterprises of any size, in any location, to meet and conduct business through the exchange of XML-Based Messages"
- ebXML Project began in late 1999
  - Divided mission into two parts –
    1. Infrastructure Standards – the ebXML Messaging, Routing, Transport Protocol and the ebXML Registry/Repository including Collaboration Profiles and Collaboration Profile Agreements
    2. Semantic Standard known as "Core Components" and "Business Process Specification Schema (BPSS)"
  - Final Specifications for ebXML Infrastructure Standards approved in May, 2001
  - Core Components deferred

ebXML

- What Infrastructure standards does ebXML address?
  - Messaging services
  - XML registries and repositories
  - Collaboration Protocol Profile (CPP)/ Collaboration Protocol Agreements (CPA)
- Business Process Specification Schema (BPSS) provides a bridge between process modelling and specification of eBusiness software components
- For more information: www.ebxml.org

XML Semantic Standards

- Being defined by multiple Consortia
  - "Thou shalt not make up thine own tags" (that's our consortium's job)
  - "Thou shalt give a thing a name, and it will have but one name, and that thing with that name shall have but one meaning and one attribute.
    - Consortium#1: "Thou shalt call its name "<LineItemNumber>" and the name "< LineItemNumber >" shall have but one meaning, which is "unique identifier for each item in a series," and it shall be numeric value between 1 and 9999.
    - Consortium#2: "Thou shalt call its name "<ItemID>" and the name "< ItemID >" shall have but one meaning, be 14 characters, etc.

Common Business Library (CBL)

- CBL is a XML semantic specification spearheaded by CommerceOne
- Semantic definitions available free of charge
  - DTD's, Schema, and SOX definitions available
- CBL to be starting point for UBL (Universal Business Language) – another XML semantic standards effort that is getting underway
- For more info, www.xcbl.org

cXML

- cXML is a XML semantic specification lead by Ariba
- cXML defines a request/ response process for the exchange of transactions by using a set of lightweight XML DTD's (also known as "punchout")
  - DTD definitions available for download free of charge
- For more info, www.cxml.org

Open Applications Group (OAG)

- Consortium of 80+ leading companies defining cross-industry XML B2B and A2A data standards
  - Consortium includes leading ERP & B2B software vendors, automotive, aerospace, and other companies
  - OAG has defined 180+ XML transactions
  - OAG standard is XML messaging "agnostic"
  - For more information: www.openapplications.org

RosettaNet

- A consortium or more than 400 of the world's leading Electronic Components, Information Technology and Semiconductor Manufacturing companies.
- RosettaNet is a self-funded, non-profit organization dedicated to creating, implementing and promoting open e-Business standards.
- These standards form a common e-Business language, aligning processes between trading partners on a global basis.

ROSETTANET
Lingua franca for eBusiness

RosettaNet

- What standards is RosettaNet addressing?
  - Partner Interface Processes (PIPs™) – standardized business processes
  - RosettaNet Implementation Framework (RNIF) – messaging
  - Dictionaries – standardized code values for products, parties (companies)
- For more info, www.rosettanet.org

ROSETTANET
Lingua franca for eBusiness

Possible Issues with XML

- Too flexible
  - Too many “standards”
  - Too easily changed by individuals/companies
- Space Hog
  - Bandwidth
  - Storage
  - Processing
- No VAN’s
  - Security & Privacy issues must be addressed
- Emerging technology – limited experience and implementations

XML Functionality and Terminology

- Extensible Stylesheet Language(XSL)
- Document Object Model (DOM)
- Simple API for XML (SAX)
- XQuery
- XML Mechanisms

Extensible Style sheet Language (XSL)

- An XSL style sheet specifies the presentation of a class of XML documents by describing how an instance of the class is transformed into an XML document that uses the formatting vocabulary
- Supports transformation, presentation of, and interaction with XML documents for use on servers and clients
- Present information visually and non-visually

Document Object Model (DOM)

- DOM is neutral for both platform and language interfaces that will allow programs and scripts to dynamically access and update the content, structure and style of documents.
- Enables applications to be written that work on all browsers, servers, platforms

Simple API for XML (SAX)

- An industry-wide API intended for high performance XML document processing – almost twice as fast
- Only a small portion of document in memory at any given time
- SAX uses event-based processing in order to achieve its goals

XQuery – eXtensible Query Language

- Non-procedural language for querying databases of XML data
  - Standard error conditions
  - Protocol independent
- XML syntax used for query language with readable syntax
- Analogous to SQL, only for XML data
- Version 1 released December 2001

XML Addressing Functionality

- XPointer – supports addressing into the internal structures of XML documents, such as elements, attributes and content
- XLink – allows elements to be inserted into XML documents in order to create links between XML resources
- XPath – language for addressing parts of an XML document. Designed to be used by both XSLT and XPointer
- Xinclude – provides the ability to merge XML documents for application use with an inclusion mechanism
- XML Namespaces - Defines a method for defining element and attribute names used in XML by associating them with URI references

XML Tools

- XML Parsers/Processors

- XSL Editors
- API's
- Browsers
- XML Editors
- DTD Editors
- To name just a few…

XML Resources

- www.xml.com
- www.xmlinfo.com
- www.oasis-open.org/cover
- www.ibm.com/developer/xml
- www.microsoft.com/xml
- www.xmledi-group.org/

XML Future

- XML driving force for B2B and business process standardization
  - Most standards defined using DTD's
  - W3C going the way of XML Schema
- General use of XML internally within companies for A2A data exchange
- XML usage to facilitate data presentation on web browsers and devices
- XML Databases beginning to gain momentum

Thanks.

QUESTIONS?