The Role of XML in eBusiness

ebXML Information Day
Vienna, 9 May 2001

Jon Bosak
Sun Microsystems

XML Basics

XML is........................................................................................ A-2
XML in one slide........................................................................... A-3
XML is "syntax, not semantics" ................................................. A-4
The industrial function of XML.............................................. A-5
The social function of XML ................................................. A-6
XML is...

- A simplified subset of SGML (ISO 8879) developed by a cross-industry group organized and led by Sun Microsystems
  - Builds on 30 years of research and 14 years of standardization
  - Widely deployed (as SGML) in major industrial settings
  - Powerful data modeling -- no limits on namespace or structural depth
  - But small enough for Web browsers

- Not a language but a metalanguage
  - Designed to support the definition of an unlimited number of languages for specific industries and applications
  - All XML languages can be processed by a single lightweight parser

XML in one slide

- Legal XML documents are called well-formed
- A well-formed document describes a logical tree
- If a well-formed document conforms to an optional grammar or schema (e.g., a DTD), it is also valid
  - Well-formedness is a property of the document
  - Validity is a property of the relationship between a document and a grammar or schema
XML is "syntax, not semantics"

- Tags have no predefined meaning
- Unlike HTML, XML by itself conveys only content and structure, not presentation, behavior, or meaning
- The meaning of XML languages must be specified outside of XML itself
  - Operational semantics: programs, servlets, applets, scripts, stylesheets...
  - Definitional semantics: prose, namespaces, ontologies, UML diagrams...

The industrial function of XML

XML is a framework for developing an unlimited number of special-purpose data languages.

XML allows people sharing a common data exchange problem to work out an open solution to that problem.

- Without interference from third parties
- Without dependence on large software vendors
- Without bindings to specific tools
- Without language restrictions
- In a way that lets anyone with a similar problem use the same solution
The social function of XML

Like Linux and Gnome, XML has a political agenda: freedom from vendor control.

- (1996) We knew that HTML could not support data exchange in general
- Without XML, HTML would have been replaced with a binary, proprietary format controlled by a single vendor
- The alternative was SGML (international, open, text-based standard, ISO 8879-1986)
- XML put SGML on the Web. Result:
  - Users can define their own data exchange standards
  - There are many inexpensive, robust tools
  - Data belongs to the people who create it rather than to the software vendors
What's a document?

- A document is data that you can read.
- Document requirements are a superset of data requirements (e.g., recursion)

The basic problem with documents is that we need to display them in many different forms. This is the problem that SGML was originally designed to solve. XML inherits the solution to that problem.

Example: This presentation.

- Written in XML
- HTML generated using a stylesheet for online publishing
- RTF generated using a different stylesheet

Example: an international bookstore

```xml
<?xml version="1.0"?>
<!DOCTYPE 書籍カタログ [  
  <!ELEMENT 書籍カタログ (書籍)+ >  
  <!ELEMENT 書籍 (書名, 著者, 出版社, (定価 | 在庫数)) >  
  <!ATTLIST 書籍 xml:lang CDATA #REQUIRED >  
  <!ELEMENT 書名 (PCDATA) >  
  <!ELEMENT 著者 (PCDATA) >  
  <!ELEMENT 出版社 (PCDATA) >  
  <!ELEMENT 定価 (PCDATA) >  
]>

<書籍カタログ>
  <書籍 xml:lang="JP">
    <書名>XML入門</書名>
    <著者>村田、門馬、荒井</著者>
    <出版社>日本経済新聞社</出版社>
    <定価>2800</定価>
  </書籍>
</書籍カタログ>
```
The Role of XML in eBusiness
Documents and Data

With stylesheet for Japanese

XML入門
著者 村田、門馬、荒井
出版社 日本経済新聞社
定価 2800円

Developing SGML DTDs
著者 E. Maler and J. el Andaloussi
出版社 Prentice Hall
定価 50ドル

With stylesheet for English

XML入門
Author: 村田、門馬、荒井
Publisher: 日本経済新聞社
Price: ¥2800

Developing SGML DTDs
Author: E. Maler and J. el Andaloussi
Publisher: Prentice Hall
Price: $50
What does the document background do for data?

XML standardizes the **concrete syntax** of data exchange in a **text-based notation** designed to be **obvious** to both people and processes.

Deploying XML creates an **open, standardized information infrastructure**.

1. Standardized parsers for putting data into memory
2. Standardized interfaces (DOM and SAX) for processing the data
3. Standardized ways to display data (CSS, XSL-FO)
4. Standardized ways to query data (XPath, XQuery)
5. Standardized ways to link data (XLink, XPointer)
6. Standardized training of people in both publishing and data processing

Separation of data from processing

- The SGML/XML publishing model decouples data from processing
- This isolates changes in large systems, making them more flexible and reliable
- Basing a system on XML makes it well-suited to transactional processing in a heterogenous, asynchronous, distributed environment (like the Web)
The document aspect of XML

XML uses documents as the transfer mechanism for data. XML is text.

- Creates large, standardized collections of data that can be processed by text tools
  - Data mining
  - Long-term retrospective trend analyses
  - Business intelligence (BI)
- Creates an infrastructure in which human beings are still part of the process
  - Troubleshooting
  - Generation of human-readable deliverables (e.g., catalogues)
  - Integration into existing social institutions

Business versus programming

We have been doing business for thousands of years. Now we want to do business electronically. In my opinion, there are two ways this can happen.

1. We can change all of our business and legal practices to optimize electronic data processing.
2. We can structure data processing to model our business and legal practices.

A traditional business transaction is an exchange of documents. XML documents model existing financial and commercial data very well.

I believe that the rapid adoption of XML is evidence that we are choosing ease of data management over ease of programming.
Summary: the XML trade-off

XML trades off

- Performance
- Centralized control
- Uniformity

in order to get

- Persistence
- Distributed control
- Asynchronicity
- A structure obvious to both humans and machines
- A certain kind of readability (like source code)
- Very low cost of entry
- Complete internationalization
What XML does for business

XML promotes a message-oriented view of electronic commerce that isolates business transactions from differences in software, hardware, system architectures, and application programming languages.

Examples of business messages:

- purchase order from a buyer to a seller
- invoice from the seller back to the buyer
- request to make payment through a credit card
- authorization to use credit card
- status reports on success or failure of services

Traditional Electronic Data Interchange (EDI)

EDI and XML examples courtesy of Betty Harvey.

N1*SH*ACE MANUFACTURING*1*987654321*
N2*RECEIVING*N3*234 MAR*N4*
SAN FRANCISCO*CA*94103*US

- Very expensive technology
- Requires special networks
- Implicit structure
- Intended for machines only
The XML version

```xml
<shipper duns="987654321">
  <organization unique-id="aceman">
    <name>ACE MANUFACTURING</name>
    <division>RECEIVING</division>
  </organization>
  <address>
    <street>234 MARKET STREET</street>
    <city>SAN FRANCISCO</city>
    <state>CA</state>
    <zip>94103</zip>
    <country>US</country>
  </address>
</shipper>
```

The XML DTD

The DTD provides a standardized description of the XML document structure.

```xml
<!ELEMENT shipper (organization, address, attention*)>
<!ATTLIST shipper
  DUNS CDATA #IMPLIED
  NAICS CDATA #IMPLIED>
<!ELEMENT organization (name, division*)>
<!ELEMENT name (#PCDATA)>
<!ELEMENT division (#PCDATA)>
<!ELEMENT address (street+, city, state?, country, postalcode)>
<!ELEMENT street (#PCDATA)>
<!ELEMENT city (#PCDATA)>
<!ELEMENT state (#PCDATA)>
<!ELEMENT country (#PCDATA)>
<!ELEMENT postalcode (#PCDATA)>
```
Advantages of XML over EDI

- Explicit structure
- Easier validation
- Can easily use the Internet
- Cheaper to implement
- Can open up electronic commerce to small and medium-size businesses (social agenda again)

XML Standardization

XML agreements .............................................................. D-2
Overview: What needs to be standardized............................. D-3
The XML standards landscape ............................................. D-4
The World Wide Web Consortium (W3C)............................... D-5
W3C XML recommendations ............................................ D-6
OASIS ............................................................................. D-7
XML.org ......................................................................... D-8
xml-dev .......................................................................... D-9
The OASIS TC process ...................................................... D-10
UN/CEFACT ..................................................................... D-11
The UN/EDIFACT Working Group (EWG) ......................... D-12
XML agreements

XML provides a standard framework for making agreements about communication.

- Industry DTDs
- Industry schemas
- Industry namespaces

But it doesn't make those agreements by itself!

Overview: What needs to be standardized
The XML standards landscape

- Proprietary consortia: UDDI, ...
- Vertical industry consortia (hundreds of them, including RosettaNet)
- Horizontal industry consortia: W3C, OASIS, JCP, ...

- Open standards bodies
  - Ad hoc forums: xml-dev, etc.
  - Grass-roots organizations: IETF, OASIS TCs
  - National/regional bodies: ANSI, JIS, BSI, AFNOR, ..., CEN
  - International bodies: ITU, ISO, IEC, JTC1, UN/ECE
- Open source projects: apache, linux, gnome, ...

The World Wide Web Consortium (W3C)

- A project of the MIT Lab for Computer Science in cooperation with INRIA (France) and Keio University (Japan)
- Very broad vendor and customer participation -- funded by over 500 web industry companies and organizations
- Proceedings are confidential, allowing members to avoid premature announcements of product plans
- Employs a staff of technical experts, helping to ensure the quality of the results
- Operates under the management of a Director
  - Ensures a consistent vision
  - Avoids antitrust problems
- http://www.w3.org
W3C XML recommendations

- Core specifications: XML 1.0, XML Namespaces, Infoset, XInclude, XML Base
- XML architecture specifications: DOM, XML Schema, XLink, XPointer, XSLT, XPath, XSL-FO, XML Query
- Specific XML languages: XHTML, SVG, XML-Signature, RDF, MathML, SMIL
- http://www.w3.org/TR

Sun thinks highly of this work and has contributed many resources to it.

OASIS

OASIS (the Organization for the Advancement of Structured Information Standards) is dedicated to the standardized, open exchange of structured data.

- Non-profit corporation founded in 1993 as SGML Open
- Open to everyone, even individuals
- All proceedings are public
- Run by an elected board of directors
- Broad industry support: Sun, IBM, Microsoft, HP, Adobe, Oracle, Software AG, SAP, Boeing, Xerox, ...
- Works on specific applications of XML (e.g., DocBook, Security)
- Provides a neutral venue for cross-industry XML standardization efforts
- http://www.oasis-open.org/
XML.org

XML.org (http://xml.org/) is a free XML registry maintained by OASIS.

- Universally accessible
- Vendor-neutral
- Based on international standards
- Built by Documentum and Sun with help from IBM, Oracle, and SAP
- Not an ebXML registry (implemented a year too early)

xml-dev

- An open mail list hosted by OASIS
- Thousands of participants, no formal organization whatsoever
- Completely open
- Has created at least one very successful de facto XML standard: SAX

There are some big problems with this model.

- Relies on the particular talents of one person
- No clear strategy for maintenance
- Ownership of intellectual property is unclear
- Invites hijacking by big vendors
The OASIS TC process

OASIS allows groups of members to organize and run their own technical committees.

- open
- legally accountable
- democratic
- flexible
- scalable
- language-neutral

The goal is to put the definition of business standards in the hands of the people who will have to use them.

UN/CEFACT

Standards for international commerce are defined by UN/CEFACT, the United Nations Centre for Trade Facilitation and Electronic Business (http://www.unece.org/cefact/)

See the org chart at www.unece.org/cefact/knowlg/structur/struc.ppt
The UN/EDIFACT Working Group (EWG)

- International EDI standards body
- Working group of UN/CEFACT responsible for UN/EDIFACT standards development
- Membership represents many electronic business communities
- ANSI X12 cooperates and contributes to work of EWG
- Significant cross-membership between X12/EWG and the ebXML Core Components (CC) team

E-commerce Architectures

Example: the eCo Architecture................................................... E-2
The ebXML initiative ................................................................. E-3
ebXML project teams ............................................................... E-4
Example: the eCo Architecture

The ebXML initiative

- Joint effort of OASIS and UN/CEFACT
- Goal: "An open technical framework to enable XML to be utilized in a consistent and uniform manner for the exchange of Electronic Business (EB) data in application to application, application to human and human to application environments, thus creating a single global market"
- http://www.ebxml.org/
ebXML project teams

- Business Process
- Core Components
- Transport/Routing and Packaging
- Trading Partners
- Registry and Repository
- Requirements
- Technical Architecture
- Security
- Proof of Concept
- Quality Review
- Marketing

The Future of XML-based Commerce

Levels of implementation ............................................................F-2
Level 1 (immediate future) ..........................................................F-3
Level 2 (near future) ....................................................................F-4
Level 3 (distant future) .................................................................F-5
Example: materials management .................................................F-6
Variation: Vendor-managed model .............................................F-7
Variation: JIT (just-in-time) model ...............................................F-8
Predictions ..................................................................................F-9
Levels of implementation

1. Standardize the infrastructure (messaging, discovery, trading partner agreements)
2. Standardize the messages (business documents)
3. Standardize machine-processable semantics

Level 1 (immediate future)

At this level, the messaging, discovery, and trading partner agreements are standardized and can be automated.

Lowest level: email ad-hoc XML, conducting all other aspects manually using email or phone calls

- This would usually be a one-sided relationship
- But it allows the smallest businesses to participate
- And very large players can engage at this level when it suits them
- Repository supplies messaging specification and commonly used industry tag sets
- This first specification of this layer will be available from ebXML at the end of this week in Vienna
Level 2 (near future)

- Standardized XML messages defined by humans within some standardized semantic framework
  - Enables off-the-shelf software solutions
  - Enables a transition from existing business practices by replacing traditional business documents with human-readable XML equivalents
- Many options for defining the semantic framework
  - UML
  - Prose
  - Prose + DTD
- Repository includes business models and standard XML business document schemas

Level 3 (distant future)

- Standardized messages automatically generated from a single specific standardized semantic framework
  - Modeling done from the top down
  - XML schemas are generated automatically
  - The whole system operates without human intervention
  - Repository includes business models designed for automation
- Advantages:
  - Completely automated
  - Allows dynamic modification of the business models themselves
- Problems to be solved:
  - Entry point for the SMEs
  - Strategy for the transition from traditional business and legal systems
Example: materials management

Variation: Vendor-managed model
Variation: JIT (just-in-time) model

Predictions

- Expect Level One solutions to start appearing after the delivery of the ebXML infrastructure specifications (May 2001).
  - See http://ebxml.org for details
  - This will allow standardized XML messaging for business purposes

- Expect Level Two solutions (standardized business documents) by 2003
  - This will allow small to medium-size enterprises (SMEs) to start to participate in electronic marketplaces

It seems to me that the U.N. is a good place to develop a standard set of XML business schemas.