



XML as a multi-platform solution for electronic publishing.

First a brief history of eXtensible Markup Language (XML). XML is a simplified and subset of the industry standard Standard General Mark-up Language. (SGML). Sir Tim Berbers-Lee in the late 80's - early 90's working at CERN pioneered the Internet and Hyper Text Mark-up Language (HTML) and HTML is a subset of SGML. IBM in the late 60's created General Mark-up Language (GML), and based on IBM's work; in 1983 the American National Standards Institute (ANSI) released its first specification/version of SGML.

Is XML as a cross platform standard?

The XML standard was released in 1998, originally developed by W3C to help overcome the limitations of (HTML), but more importantly XML was seen as more than a mark up or Meta language. XML content can be separated from the formatting structure, hence the content becomes flexible in presentation terms and formats, but more importantly this provides the hub of how XML could be used on different computing platforms. XML format information could be targeted and displayed on many and varied computing platforms, such as mobile phones, televisions and PDA devices. This could signify a major shift in computing to a common data format standard across all platforms.

(Turban 2004) XML standard (and its variants) is used to improve compatibility between systems of business partners by defining the meaning of data in business documents.

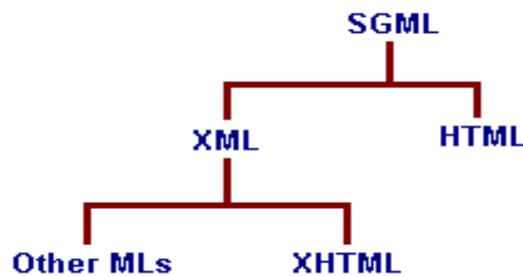


Diagram above indicated family history of XML

Data Structures & Open Systems

Data Type Definitions (DTDs) are used to specify the structure of an HTML and SGML file as in electronic publishing projects.

To address all the different technology platforms XML schemas and vocabularies standards were created. Schema is a W3C standard and is intended as an extended, more powerful and enhanced version of DTDs. Schemas allows XML content to be validated, this is especially useful for database type applications. XML vocabularies are a combination of Schemas and DTDs. There are hundreds of XML vocabularies each designed for specific sectors of industries, such as XMLTV is a set of tools to manage TV viewing. TV listings are stored in XMLTV format. The TV viewer can select favourite TV programmers this then manipulates this stream of XML data. The Open Trading Protocol (OTP) group is defining XML standards for the financial sector and XML/EDI where XML is being used to replace the costly EDI protocol as used in financial transactions.

EbXML is evolving as an XML framework for global enterprises and using HTTP as the backbone for processing transactions. This approach is simpler than existing industry standards such as CORBA and DCOM systems.

These many and varied vocabularies standard documents, which are readily available will allow developers to obtain and adopt these Schemas/DTDs to their own requirements quickly and easily.

XML in Office Products

So is XML being used in everyday office procedures, office applications and for general office computing?

XML vocabulary formats are evolving for Office products and are being supported by Microsoft, Sun Microsystems, and Corel as examples. These formats incorporate include GNOME office, K Office, Microsoft Xdocs, OASIS TC for open office XML format.

These office software applications allow the saving of files i.e. word as an XML format file.

For example:

GNOME office is a set of free productivity tools; developers create products more to the underlying technology, libraries shared and an open systems format.

Koffice is a set of integrated office tools including word processor and spreadsheet; all data files are stored as XML for common interchange of data between applications internally and externally with other software applications.

Microsoft Biztalk and Xdocs: In Microsoft's XML schema, this offers word documents to be saved in XML known as WordML. The XML coding, tagging is simpler and easy word's tagging structure. The problem is Microsoft, as rule never adopts industry standard open systems, which could affect its own proprietary software and market approach.

XML In Digital TV Products

XML is starting to gain a foothold in the Digital TV market. The UK based Digital Terrestrial standards Group (www.DTG.org.uk) and one of their standards groups TV Anytime (TVA) recommends XML as a data format for the broadcasting of Electronic Programming Guide (EPG) information.

XML is also used as part of middleware solution by cable TV broadcasters. Microsoft advocates a form of XML for its software solution for Digital TV.

Publishing & Style sheets

Publishers have long been looking for a single common source of data (i.e. an E-book), which could be manipulated, and with the minimum of re-engineering for distribution to other computing or technology platforms. Is XML the answer?

XML is flexible and multi-purpose hence developers use style sheet to specify how each document and tag should appear for each technology platform. Common stylesheet format include Cascading Style sheet, FOI XSL-FO and XML-HTML as examples. Docbook used in the publishing sector has number of XSL stylesheets. XSL used in the publishing sector is used to manipulate source XML documents. This allows publishers to create new publications from legacy XML material by the selection, arranging, combining of data from different sources. For example a legal publisher could electronically search, select and data mine all local government law cases histories. The publisher could select only cases relating to local government road planning and create a new publication base don this information and using XSL style sheets.

Darwin Information Typing Architecture for Publishing

Darwin Information Typing Architecture (DITA) is an important IBM development used by the publishing sector for use in Scientific, Technical and medical publications. DITA is based on XML and defines a methodology for combining topics into documents, which in the process contain a hierarchy. Therefore DITA provides a modular and adaptable version of XML for the publishers.

XSLT is a powerful Style sheet language for the translating of XML to other formats such as HTML. Data and coding translators are a main requirement within any of the technology sectors, to ease the transfer/porting of data and tagging structures to other data formats or technology platforms.

XSL-FO is a schema for the content management and the tagging format for the appearance of the printed page.

These XML based style sheets, methodologies and processes allows publishers to move towards creating one-source documents, which could be used for multipurpose applications and available for distribution on multi-technology platforms.

Courtesy of <http://www.webreference.com/xml/reference/standards.html> below is a set of XML standards, plus variants of DTDs and Schemas for different business sectors.

Acronym	Name	Description	Status
Basics			
XML 1.0	Extensible Markup Language	The universal format for structured documents and data on the Web	W3C Recommendation
Namespaces in XML	It means what it says, for a change ;-)	A simple method for qualifying element and attribute names used in XML documents by associating them with namespaces identified by URI references	W3C Recommendation
XML Schema		Provides a means for defining the structure, content and semantics of XML documents.	W3C Candidate Recommendation

RDF	Resource Description Framework	A model for representing named properties and property values, also representing relationships between resources.	RDF Syntax: W3C Recommendation
Referencing Documents and Fragments			
XLink	XML Linking Language	Allows elements to be inserted into XML documents in order to create and describe links between resources.	W3C Proposed Recommendation
XPath	XML Path Language	A language for addressing parts of an XML document, designed to be used by both XSLT and XPointer.	W3C Recommendation
XPointer 1.0	XML Pointer Language	"The language to be used as the basis for a fragment identifier for any URI reference that locates an XML document or XML parsed entity."	W3C Working Draft
Styling and Transforming Documents			
XSL 1.0	Extensible Stylesheet Language	XSL is a language for expressing stylesheets. It consists of two parts: XSL Transformations (XSLT), a language for transforming XML documents, and an XML vocabulary for specifying formatting semantics (XSL Formatting Objects).	W3C Candidate Recommendation
Presenting Document Content and User Interface			
XHTML 1.0	Extensible HyperText Markup Language	A reformulation of HTML 4 as an XML 1.0 application.	W3C Recommendation
HDML 2.0	Handheld Device Markup Language Specification	A simple language used to create hypertext-like content for small display, handheld devices.	W3C Submission by phone.com
WAP / WML 1.2.1	Wireless Application Protocol / Wireless Markup Language	An open, global specification that empowers mobile users with wireless devices to easily access and interact with information and services instantly.	WAP Forum Approved Specification
XUL	XML-based User Interface Language	Description language for Mozilla's configurable, downloadable chrome.	mozilla.org
MathML 2.0	Mathematical Markup Language	Describes mathematical notation and captures both its structure and content.	W3C Candidate Recommendation
SMIL 1.0	Synchronized Multimedia Integration Language	Allows integrating a set of independent multimedia objects into a synchronized multimedia presentation.	W3C Recommendation
SVG 1.0	Scalable Vector Graphics	A language for describing two-dimensional vector and mixed vector/raster graphics in XML.	W3C Candidate Recommendation
VML	Vector Markup	Defines a format for the encoding of vector	W3C Note by

	Language	information together with additional markup to describe how that information may be displayed and edited.	Microsoft et al.
WebCGM	Web Computer Graphics Metafile	A profile for the effective application of CGM in Web electronic documents.	W3C Recommendation
PGML	Precision Graphics Markup Language	PGML uses the imaging model of the PostScript language and PDF as its basis in order to provide simple-to-use graphical objects and precise visual fidelity.	W3C Note by Adobe et al.
Authoring Documents			
DocBook 4.1.2	Well, DocBook.	Particularly well suited to authoring books and papers about computer hardware and software (though it is by no means limited to these applications).	OASIS
XML News	Well, news in XML format...		
Metadata: Data about data on the Web			
DCMES 1.1	Dublin Core Metadata Element Set	Qualifiers to describe resources, including but not limited to Web pages. Defines syntax and semantics of commonly used properties like authors, subjects etc.	Dublin Core Metadata Initiative
P3P 1.0	Platform for Privacy Preferences	Providing a simple, automated way for users to gain more control over the use of personal information on Web sites they visit.	W3C Candidate Recommendation
PICS 1.1	Platform for Internet Content Selection	Enables labels (metadata) to be associated with Internet content.	W3C Recommendation
Content Syndication, also a form of metadata			
RSS 0.91	Rich Site Summary	Used to describe content and format of a news channel, which is basically a collection of Web links.	Netscape et al.
CDF	Channel Definition Format	Pretty much the same as RSS.	Microsoft
ICE 1.1	Information and Content Exchange Protocol	Defines the roles and responsibilities of syndicators and subscribers, defines the format and method of content exchange, and provides support for management and control of syndication relationships.	ICE Authoring Group
Data exchange and method invocation across the Web			
XML-RPC	XML Remote Procedure Call	A Remote Procedure Calling protocol that works over the Internet.	UserLand
SOAP 1.1	Simple Object Access Protocol	A lightweight protocol for exchange of information in a decentralized, distributed	W3C Note by Microsoft

		environment.	
WDDX	Web Distributed Data Exchange	Allows Web applications created with any platform to easily exchange data with one another over the Web.	Allaire
WebDAV	Web-based Distributed Authoring and Versioning	A set of extensions to the HTTP protocol, which allows users to collaboratively, edit and manage files on remote web servers.	IETF
DRP	HTTP Distribution and Replication Protocol	A protocol for the efficient replication of data over HTTP.	W3C Note
DSML 1.0	Directory Services Markup Language	Provides a means for representing directory structural information as an XML document.	OASIS
Web Service Definitions			
WIDL	Web Interface Definition Language	A metalanguage that implements a service-based architecture over the document-based resources of the World Wide Web.	W3C Note by WebMethods
UDDI	Universal Description, Discovery, and Integration	The building block that will enable businesses to quickly, easily and dynamically find and transact business with one another using their preferred applications.	
WSDL 1.0	Web Services Description Language	For describing network services as a set of endpoints operating on messages containing either document-oriented or procedure-oriented information.	Ariba, IBM, Microsoft

XML and Databases.

It appears that at this time XML cannot compete with the power of relational databases for data manipulation and the process of databases for data mining on different subject/field variables. For example consumer buying trends data derived from loyalty cards, indicating trends, buying trends, when, where, how, average spends, time of day etc.

The XML: DB initiative formed by SMB GmbH, dbXML group and the OpenHealth Care Group is aimed at resolving some of the limitations, open source licensing, formation of XML database groups for FAQ, and creating standards for database system utilising XML as storage format. XML is seen as the vehicle for portability of data between different computing devices and with major re-engineering of the data to work on another platform.

The structure of XML is more document based than data based. Therefore large databases holding records management type information is a good example for the use of XML. These examples include cataloguing, technical data, and Somerset house data such as: birth, deaths and wills. Native XML databases such as Tamino have been introduced, but these are not relational databases, merely hierarchy that fits with the concept of how XML data is stored.

Business Issues

What are the cost benefits to industry of Standards and of adopting XML?

The topic of Open Standards in this case is to the adoption of XML within the global general computing & technology marketplace. It maybe to use XML as an Enterprise storage format, tagging mobile messaging, or as an ebusiness platform for financial transactions.

In the case of XML versus EDI for financial transaction and processing intercompany documents. There are great advantages to business adopting XML as a standard for data interchange. EDI is expensive, a propriety system therefore difficult for clients to write or improve the software systems independently or to buy third part EDI software. EDI is non-standard and a closed system, plus less than 5% of companies (source Forrester Research), therefore there is not wide use of the system. XML is open systems, inexpensive set up costs and available to all companies. XML is seen as a revolution within the B2B market, one such example for the evolving XML standard to replace EDI is EbXML. Once all companies adopt EbXML for B2B business transactions, document exchanges, and also the nature of XML is Internet based therefore TCP/IP can be used for common communication between incompatible computer systems.

What are the problems with adopting and using XML type products?

XML example limitations can be perceived as the following:

- Evolving worldwide standard.
- Does not have worldwide acceptance
- Time factors, still emerging support for XML by major software suppliers.
- Companies still developing and building XML expertise.
- Schemas and DTDs still evolving for business sectors.
- Not yet tried & tested technology for global markets.
- Integration issues exist.
- Software application support XML

What are the business Applications for XML?

I have highlighted in page 4 of this report some of the office software products being developed with XML in mind as a common storage format. In summary example companies and organisations supporting XML office products are: Microsoft, Corel, Sun Microsystems, Netscape, and Openoffice.org, Oasis Open format XML, RossettaNet and W3C. XML is also a good solution for records management database systems, where the requirement is for document based information to be stored.

Technology Issues with XML.

There is sometimes confusion in the market thinking XML is a database system. It should be noted XML is a text mark up language and not a database management system and XML does not have the feature or function set of a relational database system. XML is good for document storage systems not data based records systems.

Software development tools, code generators and software support libraries will still be changing in line with Industry standards and schemas. Not all XML software tools are available or supported on the many and varied computing platforms.

I explained on pages 4-7 there are many variations of XML on each platform, and an XML standard is dependent on the platform, but it will need a developed standard vocabulary for that platform and which is in line with the global XML standard.

XML Product Roadmap

For XML to become a truly global multi-purpose, multi-platform standard data format it must also incorporate, integrate and adopt other current and developing industry standards software products, plus office applications, tools, protocols, operating systems and all new technology hardware platforms.

There is an momentum behind the world adopting XML due to the following factors these being: economies of scale, simpler technology, increasing businesses adopting XML for global internet based transactions, low set up costs inexpensive, interoperability; wide vocabulary for different markets will help acceptance of XML in many markets, XML & Java integration will also boost the prospects as a world standard, the ability to share data across office applications and across multiple platforms is a must have requirement for business and organisations.

Summary Conclusions.

XML takes the world of computing and especially the publishing sector a step forward to the single source data being multi-platform and multi-purpose. XML could become the global universal data format due in part to its simplicity, worldwide acceptance by major software vendors, business and the substantial costs savings associated with replacing old technology such as EDI with eBizXML.

It remains to be seen if and when XML is widely accepted as a worldwide standard, or will a new technology supercede XML or even confuse the market for XML. Microsoft also has a habit of ignoring standards, or part adopting a standard and even creating their own standards for market areas. It would appear that XML will not at this time take over from relational database structures and systems. For records management and document based database systems, WEB based information systems and global e-commerce transactions based on XML is good solution.

References:

Electronic Commerce. Efraim Turban 2004 Edition

Bibliography

Online access includes:

www.Butlergroup.com

www.xml.org

www.microsoft.com

www.xml-journal.com

www.javaworld.com

www.ehyper.com

www.xml.coverpages.org

www.oasis-open.org

www.w3.org

<http://www.webreference.com/xml/reference/standards.html>

www.xml.com
www.ibm.com
www.xml.com
www.innodata-isogen.com
www.iso.org
www.w3.org
www.oasis-open.org