Impact Of XML On Data Interchange: An XML/EDI Model

Terrence J. Moran, (E-mail: tmoran@sbu.edu), St. Bonaventure University

ABSTRACT

Extensible mark-up language (XML) is emerging as the main standard of representation of structured documents on the World Wide Web (WWW) (Gibb, 2000). An XML/EDI model is presented with advantages and challenges of this model. XML is here to stay and with ever increasing number of big players adopting it, it will become a major component in application and systems design in the near future.

INTRODUCTION

E xtensible mark-up language (XML) is emerging as the main standard of representation of structured documents on the World Wide Web (WWW) (Gibb, 2000). The impact that XML will have in data interchange is outline. In particular two examples are given, the J.B. Hunt transportation company and the North Sea Oil companies. These examples will demonstrate the advantages and impact XML will have in data interchanges.

In addition an XML/EDI model is presented with advantages and challenges of this model. The paper concludes with two other XML applications, one in the magazine publishing industry and one in the banking industry.

BACKGROUND ON XML

First there was SGML (standard generalized markup language) in 1986, then hypertext mark-up language (HTML), now XML. HTML was originally a document type definition (DTD) of SGML, but was later extended to include tags that defined both structural (i.e. the elements contained in a document, such as body, head, title, and paragraph) and presentation features (i.e. how these elements appear, such as bold, italic and color). Aware of the limits of HTML and the desire to avoid the proliferation of proprietary HTML extensions led to the ratification of XML by the World Wide Web Consortium in February 1998. In relation to complexity, XML lies between SGML and HTML and is aimed at providing most of the richness of the SGML command set while remaining easy to learn, implement and use. It is described as a dialect of SGML and is itself a metalanguage (Gibb, 2000).

XML is a markup language, which means in the simplest sense you put codes around the content of your file to tell your software what to do with it. XML is a way to create, maintain, share, and use common information formats that are self-contained, self-defining and allow for the transfer of information over all electronic means-intranets, Internets, and extranets-and all forms of media. XML is extensible, meaning unlike HTML, XML's markup symbols are both self defining and unlimited. Therefore, there is room for growth and specialization, especially when an industry can agree on specific standard document formats described in XML that can be used to communicate and translate data. In addition, XML, like HTML, is a subset of SGML, a universally recognized, platform-independent international standard for defining descriptions of electronic documents (Schmidt, 1999).

HOW XML DIFFERS FROM HTML

XML is a close relative of HTML, the language of the World Wide Web. They both use what is known as markup: symbols that describe the contents of a file (XML) or a page (HTML). While HTML describes the contents of a web page, its text, and its graphics, XML consists of the data itself and how it can be displayed (like HTML), stored (like a file), or immediately processed and used by the application (like data) (Schmidt, 1999).

Both XML and HTML enable people to send messages over the Internet. The essential difference is that HTML lets a sender's browser tell a receiver's web server how to display the format of a message, while XML enable a sender to tell a receiver how to handle the content of a message. This difference is precisely what is needed at this junction in the Internet revolution. XML will enable banks to create knowledge systems and to exchange financial information with customers and such third-party services providers as stockbrokers, mortgage and insurance companies, and mutual funds (Orr, 2000).

Probably the most significant departure from HTML is that XML specifies the structure of a document via its accompany DTD and is robust enough to describe a wide range of abstract structures. Since the write of the document can define this DTD, any customized tags may be defined and used within the document. XML therefore can be used to describe data objects, structure records and many other types of structured data. The industry has been quick to appreciate the potential of this new standard for sharing and exchanging documents on the Internet and intranets and also for sharing a huge variety of structured data using established Internet standards such as hypertext protocol (HTTP) as the mediating protocols (Gibb, 2000, p.217).

WHAT IS GOING TO DRIVE PUSH TO XML OVER HTML

XML was developed by a working committee of the World Wide Web Consortium (W3C), was designed to offer more power and flexibility than HTML while being easier to use than SGML (Beale, 1998).

XML has the ability to take data from any computer, any operating system or application, and share that with computers, Internet, and people. It opens the door for firms and industries to agree on what data should look like and then makes that data plug and play. With XML it doesn't matter if your clients or service centers are using QuickBooks, Great Plains, or SAP. Just hit one button "Send data to XML"- and that file can be used to do budgets and forecasts, tax returns, write-ups, and continuous auditing. Once XML is in place, any business could share its vital information with any other business or group of businesses, in a consistent way. In addition, this format could always be modified to meet future circumstances. This will enable XML to be always current (Schmidt, 1999).

If you have ever wanted to work with data from your business software, whether it is accounting, production, contact management, or any other kind of system, if would probably choose between dBase format, a Lotus or Excel format, or plain or comma-separated text. XML has the potential to make data transfer between two systems or multiple organizations as easily as "Export to XML."

Even searching for information on your local intranet can be significantly improved through XML. Currently, there is only the choice of utilizing an indexer, which looks through all of the words found on a web page, or a categorizer, which lets you search through a manually maintained hierarchical category tree. XML can combine both techniques, searching comprehensively through only pertinent data, since the markup tags define what all that data actually represents. Therefore, XML is not only a technology issue. It is an information, management, and cooperation issue (Schmidt, 1999).

Problems that XML can help solve include:

- Internet searches that return hundreds and hundreds of hits for a simple topic. XML-coded data in a search engine's database would allow users to clearly specify they're searching for information about olive oil, not a petroleum based product and not Popeye's girlfriend.
- The immense task of integrating one application with another.

- The distribution of documents in a variety of standards that may conflict with the tools you use. If different companies all use XML, competing products will still be able to talk to each other.
- The need for multiple copies of the same information: one for your printed reports, one for your Web site and another for CD-ROMs. Today, it takes time-consuming manual effort to turn information in a printed report into a usable format for a digital medium, for instance. However, XML would automatically code the instructions for each format, saving time and reducing the human-error factor.
- But the biggest advantage is the XML is an open standard, not a product of one vendor. It does not hold users hostage to one software company. The more companies agree to a standard series of codes for a given application-like banking- the more powerful XML becomes. And since everyone's product is better with XML features, there are incentives to cooperate.

(Hoffman, 1999)

What is going to drive the push toward XML? XML is getting a boost in many areas related to its use as a way for organizing data. It is also receiving much attention related to Internet commerce as a way that can be sued to just encapsulate the information that has been done for various transactions. For instance, for banking transactions, there has been an XML-based application defined for that purpose (Vizard, 1997).

In addition to the language, XML technology has two other components that enhance its flexibility and power: XML Link and XSL. XML-Link offers linking capabilities that goes far beyond those supported in HTML. A single link can connect to one of several targets, depending on conditions established by the author of the Web. For instance, a link could send a user to introductory or advanced information on a certain software product, depending on that person's level of expertise. XSL, the XML stylesheet language, is a standard that is evolving (Beale, 1998).

IMPACT OF XML ON DATA INTERCHANGE

For decades electronic data interchange (EDI) has been utilized extensively to communicate commercial documents among organizations. Although the benefits of EDI include shorten lead-time, less transaction cost and error rate, the acceptance rate of EDI grows slowly. It was estimated that less than 5% of companies in the United States exchange data electronically (Kalakota, 1996).

The critical factors that constrain the acceptance of EDI are:

- entrance cost is high and thus small and medium-sized enterprises, which constitute large portion of global business, cannot afford it (Iacovou, 1995).
- EDI standards are complex and thus require months or even years to define (Steel, 1996).

In addition, EDI transactions are traditional exchanged via value-added network (VAN) that is very expensive (Lu, 2000).

The following example demonstrates advantages of XML to enable data transmission.

Transportation company J.B. Hunt kicked off a program this past August to move its biggest customers off expensive EDI (Electronic Data Interchange) networks and onto the Web. The deployment, which will cost the firm about $200,000 in software and consulting services, was evaluated with a keen eye on return on investments (ROI). By saving on EDI-based value-added network (VAN) costs and moving new applications t XML and the Web, J.B. Hunt is anticipating a 15-month ROI on its technology and services investment (Karpinski, 2001).

As this example clearly demonstrates, XML has the potential for a large impact on the EDI market. XML will enable the possibility of large amounts of data to be transferred over the Web.

Using enterprise application integration software from IPNet Solutions Inc., J.B. Hunt will move away from EDI formats and private VANs toward open formats, including XML, as well as cheap Web transports, including FTP and HTTP, according to Lana Mangold, J.B. Hunt's application system engineer.

ROI ranked highly among the factors used in scrutinizing the project, said Mangold. "Being in the transportation industry today, if you look at the way diesel prices and other costs are going, it is critical for us to focus on quick ROI. We wouldn't get approval otherwise," said Mangold (Karpinski, 2000).

Web alternatives to EDI are not new; in fact, the technology's maturity makes the projected payback highly likely, according to Shawn Willett at Current Analysis. Right up front you avid the fees from proprietary VAN systems, but you have to be ready to manage it yourself (Karpinski, 2000).

J.B. Hunt, with annual revenue of $2 billion, runs more than 45,000 trailers and 11,000 trucks. Currently, it has EDI links with 250 of its top suppliers. While many EDI-to-Web deployments are focused on brining smaller, less sophisticated vendors online, J.B. Hunt will initially focus on moving its 12 highest dollar customers which is about 30 percent of the company's revenue, from EDI to ensure that the payback is quickly realized. Done the road smaller trading partners will be moved onto the network. By delivery information of about 10 MB of data per day over the Web vs. a VAN, J.B. Hunt and its customers will save about $12,000 per week just on that one application (Karpinski, 2000).

This above example demonstrates clear advantages of XML and real cost savings. The following situation demonstrates how XML is becoming a standard in North Sea oil exploration.

Representatives from seven major operators and contractors working in the UK North Seas (UKNS) agreed this past June to develop and implement standards utilizing XML software for exchanging basic well data over the internet. The deal targets the use of XML on the UK Continental Shelf was reached with the aid of information technology standards body Petrotechnical Open Software Corp. (POSC), industry regulated information utility Common Data Access Ltd., and the U.K. Department of Trade and Industry. BP (British Petroleum), Royal Dutch Shell, ExxonMobil Corp. subsidiary Esso, Conoco Inc., Schlumberger Ltd., Halliburton Co., and Paradigm Geophysical Ltd. Agreed to the project (Snieckus, 2001).

"This is a challenging target but, if successful, it should pave the way for a number of similar projects across a wide range of business processes," said Paul Maton, POSC's European director. "On the face of it, it is a simple thing, but the frustration across the oil industry of not being able to easily and unambiguously identify wells hides a consequentially large cost. The global oil industry, from oil and service companies to data vendors, needs standards in these areas to realize full potential of e-business,' he added (Snieckus, 2000).

Stewart Robinson, director of IT and data services at the DTI Oil and Gas Directorate said that this is not an attempt to push the industry to adopt new standards. The close-knit nature of the UK oil community and the desire of the UK government to fully exploit the resources of the UKCS make the UK an ideal test bed for the emerging standards. Robinson revealed that DTI would work with POSC to globalize the solution and that he is in contact with many other interested governments around the world (Snieckus, 2000).

Since the U.K. oil industry is adopting this standard, will this steam roll into a worldwide standard? They say it is not an attempt to railroad the industry into a worldwide standard, but clearly they would welcome the standard. Oil companies are generally at the top of list in the fortune 500 companies. Their influence cannot be underestimated.

XML/EDI MODEL

To speed the adoption of EDI, an Extensible Markup Language (XML)/EDI framework integrating XML with EDI was proposed by the XML/EDI group founded by Bruce Peat and David Webber in July 1997 (Webber, 1998). The main feature of XML/EDI is to process EDI messages through the Web interface through the Internet (Lu, 2001).

The objective of the XML/EDI framework is to provide businesses, regardless of their sizes, a cheaper and smarter system so that they can conduct transactions electronically with any trading partner worldwide (Lu, 2000).

By utilizing custom tags, XML can easily define structured documents such as EDI transactions. To take advantage of the extensibility of XML, the XML/EDI Group proposed an XML/EDI framework. In addition to EDI and XML, the XML/EDI framework consists of three other main components: templates, agents and global repositories (Webber, 1998).

To put it simply, templates are rules that define how data should be manipulated. Because developers can now define the rules of how the transmitted data should be processed, this gives developers more flexibility and relaxes them from waiting for long and complex processes of standardization. This is especially critical for companies who need to exchange documents with each other and their partnership in a short period of time (Lu, 2000).

Currently, templates are generally defined by utilizing Extensible Stylesheet Language Transformations (XSLT). Templates can be transmitted together with XML message or globally referenced (via global repositories) (Walsh, 1999).

Agents are programs that process XML messages in accordance with the instructions defined in the templates. Agents can be parsers that interpret the XSLT file and then produce the HTML file based on the previous XML message. Agents can also be validators when utilizing Document Type Definitions (DTDs). DTDs define the logical structure or structural schema of the XML document (Goldfarb, 1998, Lu 2000).

Global repositories act like central warehouses so that all reusable objects such as DTDs, agents, and templates can be stored in repositories and can be globally referenced or download (The XML/EDI Group, 1999). XML enables developers to design their own tags. Although this feature provides greater flexibility, many have worried that transaction cost will increase enormously since each organization will have its own set of tags. The idea behind the global repositories is that developers can create a custom tag if one doesn't exist in global repositories. If the need tag is strong, the commonly used tag will finally be created in the industry and saved in the repositories for reference in the future (Lu, 2000).

Presently there are several challenges for the construction of global repositories. For instance, the synchronization and management of objects stored in global repositories are extremely difficult. Suppose that a seller downloaded an XML/EDI document, for example a purchase order (PO), in XML format from a web server. In addition to EDI data, the PO contains its associated template and DTD. Both templates and DTD can be either traveled along with EDI data or references which refer to their location in a local or remote repository. The agents read the DTD and the templates, check the validity of the PO, and then, for example, transform the PO to purchase order confirmation document. Also the agent may communicate with legacy systems for instance MRP (Material Requirement Planning) or ERP (Enterprise Resource Planning) to save XML data for further processing (Lu, 2001).

ADVANTAGES OF XML/EDI AS DATA INTERCHANGE

The XML/EDI has the following advantages:

- Lower costs: This is achieved by using the Internet. Also since most users are familiar with web browsers, the cost of training will be reduced. It is estimated that XML can help cut costs up to 50% (Brooker, 1999).
- Suitable for trading partners of short-term relationship: With the exchanges of commercial documents using XML, it is no longer required to define standards before communications. Consequently, it is suitable for business of project-based partnership (Lu, 2001).
- Improvement of global accessibility: Since the Internet reaches almost everywhere in the world and the cost to access the Internet is low, even a small organization can transact with all institutions on the net.
- Easy to integrate with existing systems: Since XML messages are now carrying the semantic of the data and process rules, this enables the XML messages to become portable and can be used by any system that can

interpret the messages. Also by using modern distributed object techniques, it should be easy to integrate XML/EDI into existing systems (Lu and Hwang, 2001).

CHALLENGES OF XML/EDI

Although XML/EDI eliminates most problems that restraint EDI adoption by small and medium-sized firms, it still faces the following challenges:

- The length of EDI messages significantly increase. It is determines that the size of the EDI messages increases by four to eight times (Ogbuji, 1999).
- XML-related standards are yet to be finalized.
- Security issues have to be considered more fully. XML-Signature, which includes W3C, is proposed to develop syntax and scheme of generating signatures for XML documents (Lu, 2001, p.274).
- Some XML-watchers are worried that it is too easy to make up new schemas - specialized vocabularies for new applications. This could lead to thousands of narrowly defined models that would defeat the main purpose of enabling any business to work seamlessly with numerous trading partners. To meet that challenge, Microsoft announced BizTalk, a broad base language hat could be adapted to many industries. IBM and Hewlett-Packard have followed with similar announcements (Orr, 2000).

IMPACT XML HAS IN TERMS OF WEB PUBLISHING, AND IN PARTICULAR IN MAGAZINE PUBLISHING INDUSTRY

What impact is XML going to have in terms of Web publishing? XML is actually the next step for the Web. But it encompasses more than the Web because it provides a path as well that works for printed output, which customers view as being important. XML overcomes some of the real technology challenges that are showing up on the Web, such as vastly too many hits when searching, to the point where people just give up practically when it comes to trying to find something from scratch. The Web is too slow, too static, and too limited in support for groups working together. In addition, it's too difficult to build non-document applications, where a lot of the potential is on the Web. XML enables building broader solutions to problems that are more scalable (Vizard, 1997).

For example, the type of things that are on the Web now a human can look at and read just fine. But if you want a machine to look at the same eWeb page, all is sees is words. Presently there is nothing to enable machines to work with the information. What customers are facing in addition to just printing information on paper is that they also want to make it accessible on the Web (Vizard, 1997).

With XML it is viable for a corporate IT organization to have one tool to publish on the Web, to publish to print, and to publish, for example a CD-ROM.   That's the whole premise with this technology, and that's the single reason why there is so much excitement with customers in general because they are publishing out of a database. With relational databases, you can have the same relational data and then make it available five hundred different ways with five hundred different reports that subset the information in different ways and present it differently (Vizard, 1997).

An example of XML on the Web application is impact on the magazine industry where they publish online. If XML does achieve its full potential, it will perform two things that expensive production tools like Vignette's V/5 (formerly StoryServer) and OpenMarket's Content Server (formerly FutureTense) have done either poorly or not at all. XML will enable publishers to create content once and then publish it anywhere (cell phone, Palms, the Web), and tie those tools together with traditional magazine production software like Quark's CopyDesk and Adobe's more recent InCopy system (Lindsay, 2001).

XML is designed to be a language that enables publishers to keep their data and the way their data is displayed separate. This is different from traditional publishing or HTML, where the content is entwined with the layouts themselves. In those cases, "We essentially have what we would consider stovepipe workflows," comments Mark Hilton, Adobe's director of product management. "We know at the point of creating the content that we're

going to do this particular thing, and our workflow is oriented around that. The concept of 'create once, publish anywhere' means the decision on how to render your final content gets pushed to the delivery end" (Lindsay, 2001).

Therefore, if you create a version of a magazine in XML, which is designed to read on Palms as well as PCs, the machines themselves can decide how to present it, based on the method they are trained to handle the XML instructions sent with the text. XML is making its presence felt. Firms like Adobe, Quark and Microsoft were blind-sided by the emergence of the Web n 1994 and 1995, and never built the sort of industrial-strength publishing tools that Vignette, OpenMarket and Interwoven stepped in to market. This time Adobe plans to built the ability to read and write documents as XML files into each of its products, including stalwarts such as Photoshop and Illustrator, along with more recent, magazine-oriented creations like InDesign, InCopy and InScope (Lindsay, 2001).

An example of how XML can help magazine publishers is Business Week. At magazines such as Business Week, copy is manually copied from layouts, saved as texts, e-mailed or copied to disks and transferred along to production teams. With XML-based tools, a step or two could be skipped, thus saving time and money (Lindsay, 2001).

"XML is being pitched as this Holy Grail of publishing, but the reality is always different," according to Jurgen Kurtz, Quark's vice president of product management. "Some tools aren't there yet in terms of editorial uses. In 'Avenue' we have a great product, but you need a foundation in place. As we progress, the real adoption will continue until XML, like PostScript, is in the background. When the tools are able to embed technology very much behind the scenes, then the huge wider adoption is possible." That's when the race to build XML tools will turn truly ugly. The hustle to create a standard within a standard – a set of AML commands that will be universal to publishing-is already under way by Adobe, which has partnered with Hewlett-Packard, RealNetworks, Interwoven, Nokia and ATG to build one. Quark professes to be unconcerned by Adobe's arms race, but Microsoft's wider XML plans are already coming under attack by competitors that recognize the importance of this new frontier (Lindsay, 2001).

AN XML BANKING APPLICATION

Banks are late-adopters of XML, but some interesting applications are emerging.

One of the first banks to use XML in a major application was e-banking developer Wells Fargo. Early in 1999, the bank's Institutional Trust & Brokerage unit decided to upgrade its intranet knowledge base to make it more useful for some 1,800 users in the group. The centralized base had to be able to accept information in unstructured form, such as Microsoft Word or e-mail, and from structured databases, for example Excel. Entries would include administrative documents, sales presentations, company forms, corporate-wide communications, and information on competitors and mergers. Proprietary local area networks are able to do some of this, but they reject different applications. XML, by contrast, is software independent, and is more efficient than HTML. Wells decided to convert knowledge base from HTML and use system software from eXcelon Corp., a leader in provider of XML technology. The new system operates as follows:

- A user produces a Word document and saves it as an XML document.
- It is then loaded into eXcelon's dynamic application platform, where it is made available to all users via the intranet's web server.
- Users then can add updates and new information when they want.
- The system then integrates then into the knowledge base immediately.
- Finally, users can extract data for targeted presentations.

(Orr, 2000)

Another major financial application to convert to XML is the latest version of OFX (Open Financial Exchange) developed by Intuit, Microsoft, and CheckFree and supported by some 30 other providers of financial-services technology. Services include personal financial management software, banking, bill presentment and payment, stock purchases, tax preparation, investments, and 401-k plans (Orr, 2000).

CONCLUSION

Extensible mark-up language (XML) is emerging as the main standard of representation of structured documents on the World Wide Web (WWW) (Gibb, 2000).

XML has the ability to take data from any computer, any operating system or application, and share that with computers, Internet, and people. It opens the door for firms and industries to agree on what data should look like and then makes that data plug and play. With XML it doesn't matter if your clients or service centers are using QuickBooks, Great Plains, or SAP (Schmidt, 1999).

If you have ever wanted to work with data from your business software, whether it is accounting, production, contact management, or any other kind of system, if would probably choose between dBase format, a Lotus or Excel format, or plain or comma-separated text. XML has the potential to make data transfer between two systems or multiple organizations as easily as "Export to XML."

This paper outlined the impact that XML will have in data interchange. In particular two examples were given, the J.B. Hunt transportation company and the North Sea Oil companies. J.B. Hunt is looking for a 15-month payback on its service and technology investment. These examples demonstrated the advantages and impact XML will have in data interchanges.

An XML/EDI model was presented with advantages and challenges of this model.

XML is here to stay and with ever increasing number of big players adopting it, it will become a major component in application and systems design in the near future.

REFERENCES

1.   Beale, Stephen, XML ascends on the Web, *Macworld*, February, 1998, pp.28-30
2.   Brooker, E., XML Applications Stand Up TO EDI, *InternetWeek*, April, 1999.
3.   Gibb, Forbes; Colm McCartan; Ruairi O'Donnell; Niall Sweeney; and Rubin Leon, The Integration of Information Retrieval Techniques Within a Software Reuse Environment, *Journal of Information Sciences*, Vol. 26, 2000, pp.211-226.
4.   Goldfarb, C.F. and P. Prescod, *The XML Handbook*, Prentice-Hall, Englewood, NJ, 1998.
5.   Hoffman, Charles; Christopher Kurt and Richard J. Koreta, The XML files, *Journal of Accountancy*, May 1999, pp. 71-77.
6.   Iacovou, C.L., I. Benbasat, A. Dexter, Electornic data interchange and small organizations: adoption and impact on technology, *MIS Quarterly*, Dec. 1995, pp. 465-485.
7.   Kalakota, R. and A.B. Whinston, *Frontiers of Electronic Commerce*, Addison-Wesley, Reading, MA., 1996.
8.   Karpinski, Richard, EDI-To-Web Promises ROI, *InternetWeek*, Aug. 20, 2001.
9.   Lindsay, Greg, XML inches toward the mainstream, *Folio: The Magazine for Magazine Management*, summer 2001, pp.7-8.
10. Lu, Eric Jui-Lin; Ru-Hui Tsai and Shihy Chou, An empirical study of XML/EDI, *The Journal of Systems and Software*, 58 (2001), pp. 271-279.
11. Orr, Bill, Is XML the next big thing? *American Bankers Association Banking Journal*, May 2000, pp.63-64.
12. Schmidt, Walter C. and Eric E. Cohen, XML: Powering the Web into the twenty-first century, *The CPA Journal*, November 1999, pp.20-24.
13. Snieckus, Darius V., UK North Sea to use XML web-based well data exchange, *Oil and Gas Journal*, June 25, 2001, p.38.
14. Steel, K., The Standardization of Flexible EDI Messages. In: *Electronic Commerce: Current Research Issues and Applications.,* Springer, Berlin, 1996.
15. The XML/EDI Group, White paper on XML repositories for XML/EDI, http:www.xmledi-group.org, January, 1999.
16. Vizard, Michael, The XML Web, *InfoWorld*, Dec. 15, 1997, p.62.
17. Weir, John, A Web/Business Intelligence Solution, *Information Systems Management*, winter, 2000, pp.41-46.
18. Walsh, The extensible style language – styling XML documents, *Web Techniques*, vol. 4, 1999.
19. Webber, D.R., Introducing XML/EDI Frameworks, *Electronic Markets*, 8 (1), 1998, pp. 38-41.