XML: Enhancing The Exchange Of Information On The Internet

David H. Olsen, (E-mail: dolsen@b202.usu.edu), Utah State University
David H. Luthy, Utah State University

Abstract

This paper presents the argument that XML (eXtensible Markup Language) is a critical new technology that has important implications for accounting researchers and practitioners because it aims to replace EDI (electronic data interchange) as the primary means organizations communicate business transactions. XML has two primary advantages over traditional EDI that accountants must understand: 1) XML documents are native to Internet applications and thus XML documents can utilize Internet tools and applications and 2) XML documents can be directly incorporated into "backend" database management systems which has the potential to eliminate or dramatically reduce data entry tasks. Not only does eliminating data entry reduce labor costs, it is also likely to reduce data entry errors.

I. Introduction

*T*he connectivity provided by the Internet has facilitated an explosion of information exchange among business organizations. Everything from financial statements to catalogs are available in an electronic format on the Internet. This modern-day phenomenon is possible because of the general acceptance of standard communication protocols such as TCP-IP and computer languages such as HTML, Java, and C++.

HTML (HyperText Markup Language) is a special purpose computer display language that has driven the expansion of today's Internet. Simply stated, HTML controls how the information contained in Web pages is printed or displayed on computer screens. Another special purpose language called XML (eXtensible Markup Language) focuses on describing the content of the information rather than how it is displayed. This difference in focus has had a tremendous impact on how information is exchang-

Readers with comments or questions are encouraged to contact the authors via email.

ed and how business is done on the Internet. A quick look at the structural differences between HTML and XML files illustrates this key point.

II. XML and HTML Differences

Today's Web uses HTML to exchange "pages" that contain data and instructions that tell Web browsers how to present the data. These instructions are called "tags" and are placed in angle brackets like this: <SAMPLE>. One can see a Web page's tags by clicking View on the menu bar of a Web browser and then clicking Source or Page Source.

Figure 1 shows an abbreviated purchase order in the HTML format. The HTML tags provide information about the organization and the format of the document for presentation purposes. For example, the <HEAD> and <BODY> tags enclose the document's heading and the document's main body respectively. Figure 1 includes other tags that indicate paragraphs, underlining, bold font, and other formatting considerations. These tags however, do not provide information

about the content or meaning of the data itself. For example, there is no information in the document that explicitly indicates that a number, for example, is a price or a quantity. In contrast, a document in the XML format provides information necessary to understand the meaning of the data as well as provide for its visual presentation.

Figure 1
Purchase Order Information in HTML

<HTML>

<HEAD>
<P><U>Purchase Order<U/></P>
</HEAD>

<BODY>
<P>XYZ Company
3500 Main Road
Gaithersburg, MD 29999</P>

<P><B>Quantity   Item Description
Price</B></P>

<P>1   Milling Machine M89   235,000</P>

<P>1   Stand S89   6,000</P>
</BODY>
</HTML>

It is true that a clever programmer could write a program to parse the HTML data in Figure 1 and come up with purchase order information. However, considering the myriad formats companies use and the difficulty inherent in accurate parsing, a single computer program that does this type of interpretation for all companies in all situations is simply not possible. For example, when encountering the number "8900" a parsing program would be unable to decide if this is a price, model number, or something else without some external reference.

XML also uses tags, but they are very different from the tags used in HTML. XML tags describe both the structure and the meaning of the

data contained in a document. For example, Figure 2 contains an abbreviated XML purchase order document that contains essentially the same data as the HTML document shown in Figure 1.

Note that the XML tags are also located in <ANGLE BRACKETS>. However, they contain information about the data itself rather than how to present the data. For example, XML explicitly identifies the <NAME> of the <CUSTOMER> issuing a <PURCHASE-ORDER> to buy a <QUANTITY> of "1" <LINE-ITEM> that is a "Milling Machine 8900" at a <PRICE> of "235000." In addition, this customer also wants to buy a <QUANTITY> of "1" of another <LINE-ITEM> that is a "Stand 8900" with a <PRICE> of "6000."

III. Advantages

Two examples illustrate some significant ways XML enhances information exchange. The first example relates to the exchange of transaction documents like the purchase order discussed above. By using XML, trading partners are able to exchange documents over the Internet in much the same way that companies today use electronic data interchange (EDI) over value added networks (VANs). The benefits of using XML are essentially the same as those for using EDI: speed, accuracy, and the ability to integrate data directly into production and accounting systems. However, the benefits of using XML also include cost savings relative to EDI. These savings relate to the use of relatively inexpensive XML-enabled systems rather than specialized or customized systems that are usually needed for EDI. Also, these savings relate to using the relatively inexpensive Internet rather than a VAN. Those who would like to use EDI but cannot afford it can appreciate the magnitude of these benefits.

As a second example, consider the work of a financial analyst who collects and analyzes information about companies for investment purposes. An XML-enabled system could, with several mouse clicks and a few instructions, search the Internet and download financial information from selected companies. For example, specific income statement numbers for companies with a

Figure 2
Purchase Order Information in XML

```
<PURCHASE_ORDER>
    <HEADER>
        <CUSTOMER>
            <NAME>XYZ Company</NAME>
            <ADDRESS>3500 Main Road</ADDRESS>
            <CITY>Gaithersburg</CITY>
            <STATE>MD</STATE>
            <ZIP>29999</ZIP>
        </CUSTOMER>
    </HEADER>
    <LINE>
        <QUANTITY>1</QUANTITY>
        <ITEM>Milling Machine M89</ITEM>
        <PRICE>235000</PRICE>
    </LINE/>
    <LINE>
        <QUANTITY>1</QUANTITY>
        <ITEM>Stand S89</ITEM>
        <PRICE>6000</PRICE>
    </LINE>
    <PURCHASE_ORDER>
```

given SIC code could be downloaded to a spreadsheet or database for further analysis. Using XML avoids complicated and time consuming manual searches and the re-input of data from Web pages. Those who repeatedly do this type of search and analysis can fully appreciate the magnitude of time and cost savings.

Capabilities of XML

The above examples illustrate two of the capabilities provided by XML that are not fully realized on today's Internet. First, the Internet provides the capability to communicate data and integrate applications across organizations. However, computer applications in various companies may not understand each other's data formats unless custom software or services are used to make the translation. As indicated, EDI has been available for many years, but it is very expensive and requires dedicated software and systems. Also, the solution for one company may not fit another.

Two capabilities are needed to remedy this situation. First, there needs to be a way to exchange information over the relatively inexpensive Internet without regard to differences in the hardware or software used by different companies. Second, there needs to be a way to understand what information has been exchanged so it can be processed without translation and re-input. XML provides these capabilities.

Another capability facilitated by XML is the ability to automatically identify and download specific types of data from various Internet sites for comparison and analysis. There is a wealth of information available on the Web in HTML format ranging from financial reports to product descriptions and prices. However, finding and downloading this information and then re-entering it into a spreadsheet or database where it can be combined with other information for analysis can be an inefficient, manual, and error-prone operation. "Information aggregators" and "metabrowsers" provide comparison services in a few areas of business. However, these services are generally inflexible and provide few tools for data manipulation and analysis. XML provides the capability to download data directly into local spreadsheets and databases where extensive tools for data manipulation and analysis can be used.

XML Cost Versus Benefits

XML requires a more complex software environment. However, one of the best things about XML is that it is an "open standard." This means that it does not belong to any one vendor. The full benefit of being an open standard can be appreciated by knowing that HTML (the language of the Internet) and TCP/IP (the communication protocol of the Internet) are both open standards. The general acceptance of XML as the "information-understanding" standard of the Internet should make possible an enhanced information

exchange at a very reasonable cost.

Even as an open standard, there are three general areas of cost including: (1) the cost of developing generally accepted standards, (2) the cost of software in a more complex data environment, and (3) the cost of application-specific software.

Generally Accepted XML Standards

Thanks to the foresight of many who know the benefits of open standards, a World Wide Web Consortium (W3C) was organized in 1994 to promote and guide the development of Web standards. W3C released XML as a formal standard in 1998. Since then a number of industry groups have developed document type definitions (DTDs) that define the structure of data in XML files for their industry. DTDs help systems place XML documents in a proper and consistent format. A DTD also guides systems that receive XML documents on how to interpret them properly. For example, the accounting profession re-
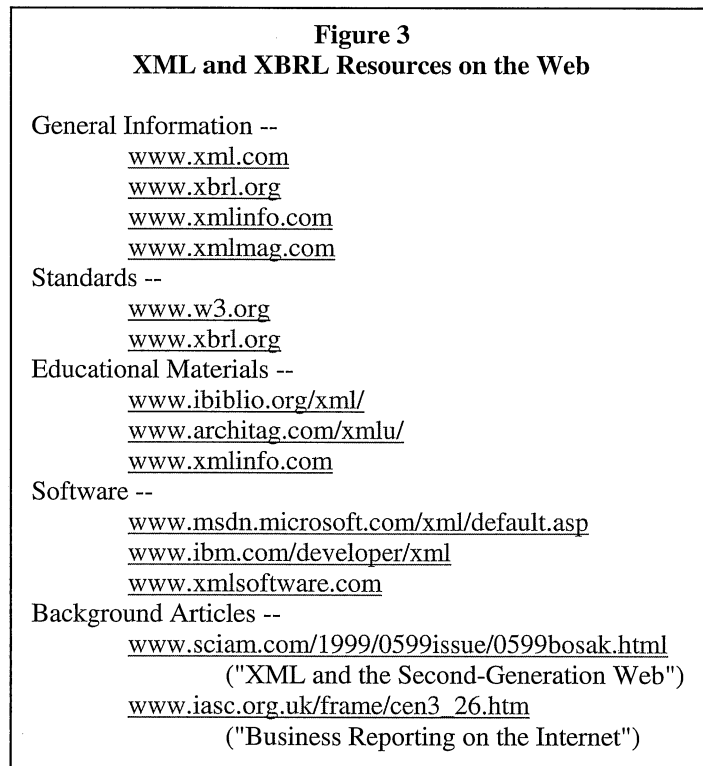
cently released Extensible Business Reporting Language (XBRL) for the exchange of financial information. The purpose of XBRL is "… for the preparation and exchange of business reports and data. The initial goal of XBRL is to provide an XML-based framework that the global business information supply chain will use to create, exchange, and analyze financial reporting information including, but not limited to, regulatory filings such as annual and quarterly financial statements, general ledger information, and audit schedules (www.xbrl.org, 2000)." The willingness of organizations to cooperate in developing standards and the speed with which these standards are being accepted indicate that the benefits of XML will probably outstrip the costs of developing the standards. More information concerning XML and XBRL can be found at web sites listed in Figure 3.

More Complex Data Environment

It's not surprising that providing more information about the contents of a document requires a more complex data environment at higher cost. However, this more complex environment also provides a number of benefits and a great deal of flexibility.

There are three basic files associated with an XML document: (1) the XML data file that contains the basic document, (2) a DTD file that contains information about the contents and structure of the XML data file, and (3) an Extensible Style Sheet (XSL) file that contains information about the formatting and visual presentation of the information in the XML data file. All of these files may be combined into one physical file or they may be located in three separate files depending on the requirements of the application.

Because XSL files deal with the visual presentation of data they may not be needed. For example, XSL files are not needed when docu-

---

Figure 3
XML and XBRL Resources on the Web

General Information --
        www.xml.com
        www.xbrl.org
        www.xmlinfo.com
        www.xmlmag.com
Standards --
        www.w3.org
        www.xbrl.org
Educational Materials --
        www.ibiblio.org/xml/
        www.architag.com/xmlu/
        www.xmlinfo.com
Software --
        www.msdn.microsoft.com/xml/default.asp
        www.ibm.com/developer/xml
        www.xmlsoftware.com
Background Articles --
        www.sciam.com/1999/0599issue/0599bosak.html
                ("XML and the Second-Generation Web")
        www.iasc.org.uk/frame/cen3_26.htm
                ("Business Reporting on the Internet")

---

ments are simply transferred from one computer to another computer and there is no need to have them viewed or printed. However, when visual presentation is important, XSL files provide the information and provide the flexibility to have documents conform to different media specifications and layouts. For example, XSL files can provide formatting for printing on paper, display on a Web page, or recording on a CD-ROM without modifying the basic XML data file. This capability means that XML data is written once but then can be delivered in many different ways.

Even though the XML environment is more complex, the typical user will not deal with the added complexity. Just as the typical user of the Internet does not have to understand the workings of HTML to reap its benefits, the typical user of XML will not have to understand its technical specifications. However, a basic understanding of technologies such as XML provides an opportunity to significantly improve the way business is conducted.

Application-Specific Software

It is unclear what impact XML will have on the cost of application-specific software. However, developments in two areas, business-to-business software applications and financial reporting on the Internet, provide some example indicators. First, a consortium of leading business software companies called the Open Applications Group (OAG) has developed and published over one hundred DTDs that define various business applications. The likely outcome of this type of activity is to make developing business applications in XML easy and reliable while reducing costs and implementation time. The implication is that if business software vendors are not using XML they will soon be out of step with the rest of the business community and may suffer financial consequences. In fact, vendors are adding XML capabilities to their software packages and development tools. IBM, Microsoft, Oracle, and Sun Microsystems are a few of the vendors working to incorporate XML into their technology offerings (Kiely, 1999).

A very specific industry example relates

to the accounting and financial reporting community. The XBRL Project Committee that helped develop and launch XBRL includes representatives of the financial, accounting, software, and governmental communities from around the world. A number of the software companies on the Committee and others have announced XBRL products or their intention to include XBRL in future software releases.

This is an example where the mobilization of the business community and the power and cost savings of open and generally accepted standards is being realized in the adoption of XML.

IV. Summary

XML is a way to structure data and give it meaning so it can be managed and easily manipulated in meaningful ways. It is a sophisticated cousin of HTML, the language of the Web. Both XML and HTML use tags or codes within data files. This enables the exchange of data files among virtually all programs and computers. XML describes the meaning of information and how it should be presented, such as on a Web page or as a printed document. The difference between XML and HTML is that HTML only describes how the information should be presented.

XML makes it possible automate data exchange among companies more easily and at lower cost than with current technologies such as EDI. XML also makes it possible to automate much of the manual and error-prone work of gathering information across the Internet. Thus, resources can be shifted from data gathering and re-entry to data analysis and evaluation.

The use of XML is dependent on the general acceptance of data standards in much the same way that HTML has been accepted as a standard for displaying Web pages. Various organizations and industry groups have cooperated in the development of both XML and the DTDs that define the data structures for specific applications. In addition, a number of software vendors have incorporated XML into their products. Others have announced their intentions to do so. The

indication is that XML will have a tremendous impact as it enhances the exchange of information via the Internet. 📖

References

1. American Accounting Association, "Report of the AAA Committee on Contemporary Approaches to Teaching Accounting Information Systems." May 1986.
2. Chen, P.P., "The Entity-Relationship Model: Toward a Unified View of Data." *ACM Transactions on Database Systems* (January 1976): 9-36.
3. Kiely, D. "XML: More than Just a Quick Fix," *Information Week*, February 8, 1999, page 77.
4. Navarro, A., White, C., and Linda Burman, "Mastering XML", Sybex Publishers, San Francisco CA. (2000).
5. Olsen, D., and S. Kimmell, "Towards Integrating Advanced Database Concepts into Accounting" *Review of Accounting Information Systems* (Summer, 1998).

Notes