

## **EDI Forum: The Journal of Electronic Commerce**

### **Edited Manuscript**

Author: Harvey, Betty

Revision #: 1

Date: 8/25/98

Page 1 of 11

## **XML and EDI—Natural Partners**

XML may give EDI a more substantial structure on the Web

by Betty Harvey

As the Web continues to evolve, the language used to send Web content information must become more powerful. The people who use the Web are growing experienced enough to expect more functionality. Isn't this the technology that was going to change our world?

The world of EDI, too, is anxiously awaiting the next phase in the Web's development, when complexities of a higher order will be contained in the basic data format. Many companies and consultants are looking forward, trying to identify the most likely candidate for a more substantial Web data format. To some, the winner is clearly XML—a cousin of the ubiquitous format HTML.

### **EDI Meets XML**

Methods for performing EDI predate the World Wide Web, but the two have yet to be truly and thoroughly integrated. EDI continues to be used for one-to-one data exchange between businesses, either through proprietary networks or using telephone dial-up. EDI uses proprietary protocols that take time to optimize and, once instituted, are difficult and costly to change. HyperText Markup Language (HTML) does not provide a way to transfer structural information—that is, information about the character of the data being sent. XML, eXtensible Markup Language, which can include such structural information, may prove to be a solution for many EDI shortcomings, providing greater standardization, an intelligent structural framework, and widespread use on the Web.

Both HTML and XML are subsets of SGML, Standard Generalized Markup Language. SGML has been an ISO standard since 1986, and is currently used in every major industry. Although HTML and XML were both conceived in response to requirements of the World Wide Web, HTML lacks that one important attribute—the ability to describe both the structure and the content of data. HTML can describe only the formatting and presentation of a message.

Despite HTML being the *lingua franca* of the Web, this shortcoming precludes HTML's use as a medium for EDI, both on and off the Web. So while electronic commerce is alive and active on the Web, it is not being conducted using standard EDI methods. Instead, these transactions occur through customized programs and databases. Using XML as the standard for EDI transactions could provide standardized protocols that allow for change, and at the same time, preserve existing EDI content. Such standardization and flexibility can only increase the use of EDI, both through existing proprietary networks and over the World Wide Web.

## EDI Forum: The Journal of Electronic Commerce

### Edited Manuscript

Author: Harvey, Betty

Revision #: 1

Date: 8/25/98

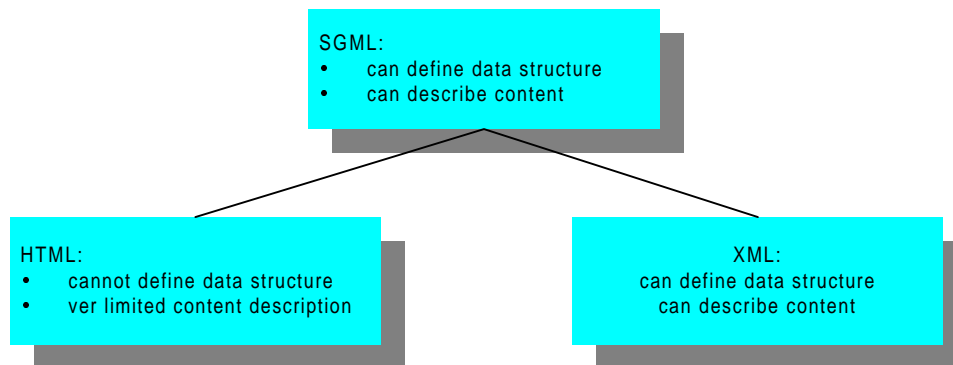
Page 2 of 11

### What is XML/EDI?

XML/EDI is an initiative to adopt the use of the XML data format for delivering traditional EDI transactions and messages. By using EDI dictionaries, XML provides a standard framework to describe any type of data, so that information in any transmission can be searched, decoded, manipulated, and displayed consistently and correctly. The use of XML to encode EDI provides a seamless flow of information through business applications and business transactions. XML presents transmitted information in a hierarchy and describes the relationships between pieces of information. Business applications can then reference and use the information appropriate to the tasks they perform, while ignoring irrelevant information.

### SGML, XML, and HTML: All in the Family

To see the value of XML, and the limitations of HTML, it is helpful to understand the relationship between SGML, XML and HTML. SGML is the parent standard for both HTML and XML. The rapid growth of the World Wide Web can be attributed largely to the ability to easily create and to disseminate information in HTML—a standard structured language. See FIGURE 1.



[Figure 1: HTML and XML both descend from SGML]

SGML defines structure and content using a Document Type Definition (DTD), which defines the rules for defining the structure and content. HTML, however, has only two structural components: <HEAD> and <BODY>, which provide only formatting information. HTML provides no additional information about the structure or character of a message.

Adding content information to HTML in attempts to provide intelligence about data structure have failed. Such attempts fail for three reasons: they do not use a hierarchical information structure (which is essential to SGML/XML), they provide only proprietary solutions, and they cannot be integrated with business applications.

## EDI Forum: The Journal of Electronic Commerce

### Edited Manuscript

Author: Harvey, Betty

Revision #: 1

Date: 8/25/98

Page 3 of 11

In the middle of 1996, a group of SGML professionals and the World Wide Web Consortium (W3C)<sup>1</sup> came together to form the SGML Working Group. W3C is an industry consortium founded in October 1994 to help the World Wide Web evolve by developing common protocols that ensure its interoperability.<sup>2</sup> In January 1997, the W3C recommended the first draft of the eXtensible Markup Language (XML) as a subset of SGML. An important constraint placed upon the Working Group was that the XML specification had to conform to the ISO SGML standard.

The XML standard was developed to help organizations identify important content information within their information systems. XML allows any organization to tag their information based upon structure and content, just as with SGML.

XML is a natural data format for sending and receiving EDI messages and transactions because it is a hierarchy-based information markup language. This hierarchical property allows information to be assigned relationships and placement in the hierarchical tree. Figure 1 shows a hierarchical model of a very simple purchase order defined by the DTD.

Figure 2 demonstrates how the purchase order from Figure 1 can be tagged. Notice that each piece of information can be identified in terms that are semantically understood by both individuals and computer applications, such as browsers, databases, and spreadsheets. This is a powerful concept. The idea that information can be identified, used, processed and understood by both humans and computers provides many opportunities for use and reuse of information.

```
<PO ID="po1">
  <HEADER>
    <ORIGINATOR>
      <ORGANIZATION ID="eccl" DATE="3/07/98">
        <ORNAME>Electronic Commerce Connection, Inc.</ORNAME>
        <ADDRESS id="address101">
          <STREET>13017 Wisteria Drive</STREET>
          <POBOX>333</POBOX>
          <CITY>Germantown</CITY>
          <STATE>MD</STATE>
          <COUNTRY>US</COUNTRY>
          <ZIP>20874</ZIP>
        </ADDRESS>
      </ORGANIZATION>
      <NAME TYPE="REQ">
        <FNAME>Betty</FNAME>
        <LNAME>Harvey</LNAME>
      </NAME>
      <PHONE>301-540-8251</PHONE>
      <FAX>301-540-4268</FAX>
      <EMAIL><A HREF="mailto:harvey@eccnet.com">
    </EMAIL>
  </ORIGINATOR>
  <SHIP-TO>
    <ORGANIZATION ID="eccl" DATE="3/07/98">
      <ORNAME>Electronic Commerce Connection, Inc.</ORNAME>
      <ADDRESS idref="address101">
        </ORGANIZATION>

    <NAME TYPE="REQ">
      <FNAME>Betty</FNAME>
```

## EDI Forum: The Journal of Electronic Commerce

### Edited Manuscript

Author: Harvey, Betty

Revision #: 1

Date: 8/25/98

Page 4 of 11

```
<LNAME>Harvey</LNAME>
</NAME>
<PHONE>301-540-8251</PHONE>
<FAX>301-540-4268</FAX>
<EMAIL><A HREF="mailto:harvey@eccnet.com">
    harvey@eccnet.com</A></EMAIL>
</SHIP-TO>
</HEADER>

<PO-DETAIL ID="po123">
  <ITEM ID="item1" WEIGHT="120" LENGTH="8x11">
    <ITEMNAME>Bond Paper</ITEMNAME>
    <DESCRIPTION>Standard Bond Paper</DESCRIPTION>
    <ITEM-NO>SBP-1</ITEM-NO>
    <QTY>10</QTY>
    <PRICE>$3.99</PRICE>
  </ITEM>
  <ITEM ID="item2">
    <ITEMNAME>Yellow Paper</ITEMNAME>
    <DESCRIPTION>Pale Yellow Paper</DESCRIPTION>
    <ITEM-NO>PYP-1</ITEM-NO>
    <QTY>5</QTY>
    <PRICE>$4.99</PRICE>
  </ITEM>
  <ITEM ID="item3">
    <ITEMNAME> Pink Paper</ITEMNAME>
    <DESCRIPTION>Pale Pink Paper</DESCRIPTION>
    <ITEM-NO>PPP-1</ITEM-NO>
    <QTY>20</QTY>
    <PRICE>$5.99</PRICE>
  </ITEM>
</PO-DETAIL>
</PO>
```

[Figure 2 - Tagged Simple Purchase Order]

XML provides the powerful capability of reusing information objects. An information object is a logical piece of information. A good example of an information object is an address. An individual address can be used in many different contexts within a single transaction. For instance, the same address could be reused within a purchase order for requestor, shipping address, etc. This address can also be used in multiple transactions. XML provides the facility for these reusable objects.

### Why XML for EDI?

Traditional EDI was designed for transactions between two trading partners. It was assumed and required that the trading partners would agree upon both the format and structure of these messages. The messages are field-oriented and contain no semantic meaning intrinsic to the message. The format, once agreed upon, is frozen and cannot be modified unless both partners agree. Custom translators are then built around the transaction structures.

XML still requires that organizations agree upon the information that trading partners want to exchange. However, it allows individual organizations the flexibility of adding or eliminating information to the transaction set without costly negotiations and reprogramming of business systems.

## **EDI Forum: The Journal of Electronic Commerce**

### **Edited Manuscript**

Author: Harvey, Betty

Revision #: 1

Date: 8/25/98

Page 5 of 11

As an example, let's say that a company has an internal part number as well as a vendor part number. The internal part number information could simply be added to the transaction structure without impacting other trading partners.

XML's hierarchical data relationships easily enable the traversal and transformation of information objects within a transaction. Commercial tools are currently available to enable the traversal and transformation of SGML and XML. In August 1997, a joint proposal by Microsoft, Inso, and Arbortext, Inc. was submitted to W3C for a standard style and transformation language called *eXtensible Style Language (XSL)*. This language is based on an ISO standard—IEC 10179 *Document Style Semantics and Specification Language (DSSSL)*. The first draft of the XSL specification was published in August 1998. Now that the XSL specification is finalized, commercial tools for creating XSL should quickly become available. Tools have already been available from Microsoft and Arbortext based on the initial proposal. There is still much debate concerning the use of XSL versus the Cascading Style Sheet (CSS2) for styling XML.

### **Commercial and Industry Support**

XML's momentum and vendor support is phenomenal. The momentum can be mildly compared to HTML. At the second International WWW Conference held in Chicago in the fall of 1993 (4 short years ago) there was a technical track called 'SGML on the Web'. In these early days of the Web, it was apparent that HTML was not going to provide the functionality and flexibility required for real computing applications. The late Yuri Rubinsky, Founder of SoftQuad, International, was a visionary and championed 'SGML on the Web'. Yuri was the Chair of these technical sessions in Chicago. As a result of his early work, 'SGML on the Web' survives today as XML. Because of Yuri's commitment to SGML on the Web, SoftQuad provided Panorama for free. Panorama was the first browser for SGML and later became a plug-in to Netscape.

Organizations who were already committed to SGML relied upon products such as Panorama for delivering 'intelligent documents' on the WWW. As a result several commercial products were developed and are available today for browsing SGML documents on the Web. These SGML browsing products are currently being modified or have already been modified to be XML compliant.

In order for XML to really work, it will require the kind of software and vendor support that we currently have for WWW and HTML development. Even though the XML specification was just accepted in August 1998, we are seeing a tremendous acceptance by major software and tool manufacturers. The support by the vendor community will ensure the success of XML as a WWW information format.

Both Microsoft and Netscape are committed to XML and will allow the viewing of XML in the next release of their Internet browsers. Microsoft Explorer 4.0 currently supports browsing XML files with XSL style sheets using an Active X program. On June 11, 1998, Microsoft announced the development of Microsoft Explorer 5.0, which will have full support for XML and CSS2. Netscape 5.0 will also support XML.

## **EDI Forum: The Journal of Electronic Commerce**

### **Edited Manuscript**

Author: Harvey, Betty

Revision #: 1

Date: 8/25/98

Page 6 of 11

In a surprise move, Netscape recently provided the source code for the Netscape browsers for free. Developers can use and distribute the Netscape browser free of charge. One constraint that Netscape put on the independent applications built around Netscape is that any enhancements or improvement to the code must be given back to Netscape to be redistributed. This bold move will enable the development of many applications for XML/EDI, as well as other vertical applications to be built around the Mozilla (Netscape) browser.

Currently, a number of object-oriented databases are specifically geared towards SGML information storage. These products have either been XML-enabled or are in the process of XML-enabling their tools. Currently, the major SGML object-oriented databases, Chrystal's Astoria, Texcel's Information Manager, and Poet's Content Manager, are all XML-enabled. Other database vendors, such as Oracle and Sybase, are committed to providing XML support for their products. Companies such as Progressive Information Technologies and Xyvision have already created an SGML/XML database capability layered on-top of Oracle.

The Open Financial Exchange (OFX) proposal which Microsoft is developing with Intuit and Checkfree is also based upon XML.<sup>3</sup> The OFX specification is publicly available, as is the DTD. Internet commerce products are being built around this data format. The initial goal of the OFX was to support consumer and small business banking and bill payment, and to allow bill presentment and investment management. Further, more advanced functionality is planned in the future.

In January 1998, the Open Trading Protocol (OTP) Consortium<sup>4</sup> released the first public draft of the proposed OTP specification. The data format for the OTP specification is XML. The DTD is currently being finalized.

The Health Level Seven (HL7) organization develops standards for the health care industry. A special interest group has been established within the HL7 organization to create a standard for use of SGML/XML within all domains of health care. They are currently focusing on the use of SGML/XML for patient medical records.

## **XML/EDI Initiatives**

### **XML/EDI Focus Group**

In July 1997, an ad hoc group of professionals and volunteers from various industries dedicated to the future of XML/EDI standards and products began the XML/EDI initiative. This group started as a volunteer organization and all of their work is still being accomplished through the volunteer efforts of the participants.

The XML/EDI Group currently consists of over 500 individuals that subscribe to their e-mail mailing list. In January 1998, the XML/EDI Group released its first "*Guidelines for using XML for Electronic Data Interchange*".<sup>5</sup>

## **EDI Forum: The Journal of Electronic Commerce**

### **Edited Manuscript**

Author: Harvey, Betty

Revision #: 1

Date: 8/25/98

Page 7 of 11

The XML/EDI Group can be credited as the impetus for other ongoing XML/EDI initiatives. The group's insistence that XML was the best solution for EDI prompted other EDI organizations, X12 for example, to take notice and respond.

### **X12 DISA/CommerceNet/GCA Working Group**

In March 1998, Data Interchange Standards Association (DISA), CommerceNet, and the Graphics Communications Association (GCA) started an ad hoc X12 Working Group to look at the possibility of representing X12 data elements within XML structures. The group will issue a report which is expected to include several recommendations. This group's work is expected to be complete in October, 1998.

Currently, the working group is finalizing a report on mappings and naming conventions of X12 data elements and XML. The **first draft of the report has been completed and has been** sent to DISA, CommerceNet, and GCA this quarter for comment and review.

### **XML/EDIFACT**

In April, 1998, the W3C established an Electronic Commerce Interest Group.<sup>[7]</sup> The mission of the Electronic Commerce Interest Group is to "*Lead the Web to its full potential*". They are planning to encode the UN/EDIFACT EDI standard in XML. In recent days, the W3C XML/EDIFACT working group has still been getting organized.

### **Harmonizing Data Formats**

Disparate data formats have traditionally been problematic for system integration and network design. Harmonization efforts have been on-going since the inception of integrated systems. Most organizations do not operate in a vacuum and, likewise, neither do data formats. XML is being considered as the data transfer format of choice in a variety of environments, not just EDI. Some of these initiatives are:

- STEP/SGML/XML - Standard for Exchange of Product Data (CAD/CAM)
- VML - Vector Markup Language
- CDF - Channel Definition Format
- PSGML - Parsing SGML (Raster Graphics)
- UML/XML - Universal Markup Language

### **XML/EDI Solutions**

There are many different suggested approaches to the implementation of XML use within EDI. When the dust settles and each of these has been applied, there will not be a single feasible solution among them. Organizations have already adopted different EC/EDI methodologies. Some haven't started

## **EDI Forum: The Journal of Electronic Commerce**

### **Edited Manuscript**

Author: Harvey, Betty

Revision #: 1

Date: 8/25/98

Page 8 of 11

down the EDI path, while others are doing full-blown X12 transactions.

### **XML as Data Transfer Vehicle**

Organizations that have already invested heavily in EDI technology are not going to be interested in tossing their investment away and working towards XML/EDI solutions solely because new technology is emerging. These companies may want to decide to use the transport capability of the HyperText Transfer Protocol (HTTP). HTTP is the networking protocol that provides the transport of information over the Web. It enables an affordable and easy-to-use transport vehicle for information. Any organization that has access to the internet has access to the HTTP protocol. XML has the ability to transport different types of information. Traditional EDI transactions can be carried using XML wrappers in the same way as any other Web data type, such as graphics, audio, and video. The use of XML as a wrapper for traditional EDI information is a viable solution for sending and receiving traditional EDI messages.

### **XML for E-commerce/EDI**

Organizations who haven't made an investment into EDI will find that the cost and learning curve for EDI solutions will be dramatically reduced through the use of XML solutions. Using XML for transport, browsing and database functionality reduces the cost of information maintenance. It allows the reuse of the information for multiple purposes.

Today's web applications are already providing this functionality in proprietary ways. Although current web-based applications work very well for individual organizations, they do not allow interoperability and exchange of information between them. There are programs in development as well as some that are available which will allow organizations and industries to work together.

One such program, Junglee's XMLizer ([www.junglee.com](http://www.junglee.com)) was demonstrated at SGML/XML Europe in May. XMLizer provided an interface where the end user can query the Web for a book and receive the book title, book supplier and the cost in XML format. Based on the purchase price and availability the end user could order the book from the supplier of their choice. Amazon.com and Barnes & Noble on-line bookstores are already using this technology. The XML data format provides the apparatus to retrieve, display and order on-line. This application is XML/EDI.

### **EDI: Beyond Messaging and Transaction**

Information is a valuable commodity. The Internet has enabled the dissemination of information at the fingertips. XML provides a multi-functional mechanism for information transfer over the Internet. XML/EDI is an enabling technology for buying, selling, and sending information over the Internet in a single data format. Many companies supply information as a source of revenue and product, and many of these same companies are currently using the Internet for distributing their information products. Providing information over the Internet is cheaper than producing and distributing either CD-ROMs or published books and allows companies to provide 'information on demand'.

## **EDI Forum: The Journal of Electronic Commerce**

### **Edited Manuscript**

Author: Harvey, Betty

Revision #: 1

Date: 8/25/98

Page 9 of 11

These companies currently have their information in either an SGML format or are looking to convert their information to XML. SGML/XML format is currently being used to create information for the Web, CD-ROM, and books.

Data Warehousing is big business. Companies have huge repositories of information that they want to sell. Although paper will not disappear anytime soon, producing paper costs money. Distributing paper costs money. Distributing updates costs money. Organizations are looking for ways to distribute information more economically. The Department of Defense is currently using 'Print on Demand' as a secondary method of distributing technical manuals. Interactive Technical Manuals (IETMs) are currently being distributed on CD-ROM and distribution over the Internet is currently being investigated. Meanwhile, Adobe, Microsoft, Netscape, and many other companies already deliver software and software updates through the internet.

### **XML/EDI Repositories**

The Merriam-Webster Dictionary definitions for *repository* are:

1. A place, room, or container where something is deposited or stored;
2. A person to whom something is confided or entrusted.

Both of the definitions above relate to XML/EDI repositories. They are locations (in most cases, the Internet) where information about EDI can be deposited, stored, maintained and updated. DTDs and their data dictionaries will be stored in the repository. A data dictionary describes the semantics and practical usage of the DTD. Without a data dictionary a DTD is useless. When organizations are in the process of securing an SGML consultant or company to create their DTDs, they should look at the consultant's or company's past performance in supplying data dictionaries and examine the quality. The data dictionary is essential to the DTD's use.

The repositories also supply utilities for manipulating and processing XML, providing style sheets, transformation utilities, and the like.

In the context of all the groups working on XML/EDI, the XML/EDI repository is seen as an essential part of the process. Working groups and others are beginning to put repositories together, even before addressing the requirements of XML/EDI. Whoever holds the key to the repository holds the key to the process. As a result, the second definition from Webster, "*A person to whom something is confided or entrusted,*" is important to remember. Creating and maintaining a repository is not an easy job. It takes dedication, resources, funding and constant maintenance. But it is at the heart of XML/EDI.

Two organizations have already announced the creation of XML/EDI registries and repositories:

CommerceNet (<http://www.commercenet.com>) introduced their eCommerce Registry Service in May. The mission statement says it will be '*An official clearinghouse for all ecommerce*

## **EDI Forum: The Journal of Electronic Commerce**

### **Edited Manuscript**

Author: Harvey, Betty

Revision #: 1

Date: 8/25/98

Page 10 of 11

*documentation and standards. We will either have the data in our database or have a pointer to it. Our objective is to be the place where everyone comes when they want to find ecommerce documentation.'*

GCA (<http://www.gca.org>) issued a press release on May 26, 1998, announcing their intention to 'develop a registration, authentication and accreditation authority...that can satisfy industry requirements for coordinated Internet-based repositories'.<sup>7</sup>

As various industries move forward in the development of XML/EDI, they will develop their own repositories for information. This has already happened in the SGML world. One of the best resources for SGML/XML information is Robin Cover's bibliographic web site devoted to the subject (<http://www.sil.org/sgml.html>). This site gives reference information on the various industry standards and repositories available. Large organizations, such as DoD, ATA, and TCIF, have been working with repositories for many years and have learned valuable lessons along the way. The Navy currently has an SGML repository (<http://navycals.dt.navy.mil>) where standards, DTDs, reports and information can be obtained.

### **Conclusion**

EDI encompasses more than just transaction and messaging systems. EDI spans the entire spectrum of information and commodity services and is integrated into all sectors of an organization. XML/EDI is seen as an enabling tool to ensure that organizations can have access to information in a timely manner but also have a more efficient flow of information through the organization.

There are many smaller issues to be ironed out in the interest of adopting XML/EDI, and there are plenty of groups working hard to do that. The market can be fickle, though, and unless there is a clear directive agreed upon by all sectors, XML/EDI may get stuck on the slow road. However the atmosphere seems favorable and it looks like it will be a fine step in the evolution of electronic commerce. New initiatives to investigate XML/EDI in commerce from all perspectives are multiplying rapidly. The future will begin to take shape when these initiatives come together and the resulting synergy becomes inescapable. Indeed, the future of XML/EDI looks very promising.

### **Table, Figure, and Photo Caption(s):**

Place these elements after the paragraph in which the figure or table is referenced.

Style is FIGURE X: and TABLE X:

### **The following go at the end of the manuscript:**

## **EDI Forum: The Journal of Electronic Commerce**

### **Edited Manuscript**

Author: Harvey, Betty

Revision #: 1

Date: 8/25/98

Page 11 of 11



Betty Harvey is the director of Electronic Commerce Connection, Inc. (ECC), an independent consulting firm specializing in Internet, computer, and information technology.

### **Author Contact:**

Betty Harvey (Harvey@eccnet.com)

Electronic Commerce Connection, Inc

<http://www.eccnet.com>

### **Executive Summary:**

EDI was established before the advent of the World Wide Web. EDI's primary mode was, and continues to be, one-to-one data exchange between businesses, either through proprietary networks or via telephone dial-up. EDI uses proprietary protocols that take time and effort to set up and, once instituted, are difficult and costly to change. These protocols, like HyperText Markup Language (HTML), do not generally permit the inclusion of structural information—that is, information *about* the character of the data included in a message. eXtensible Markup Language (XML), which can include structural information, may prove to be a solution for many EDI shortcomings, providing greater standardization, a framework for intelligent data structuring, and use on the Web.

### **Resources:**

<sup>[1]</sup> World Wide Web Consortium, <http://www.w3c.org>

<sup>[2]</sup> Extensible Markup Language (XML) 1.0, W3C Recommendation 10-February-1998

<sup>[3]</sup> Open Financial Exchange Specification 1.5, March 27, 1998, Intuit, Inc. and Microsoft Corporation

<sup>[4]</sup> Open Trading Protocol, <http://www.otp.org:8080/>

<sup>[5]</sup> Guidelines for using XML for Electronic Data Interchange, Version 0.05, January 28, 1998, <http://www.geocities.com/WallStreet/Floor/5815/guide.htm>

<sup>[6]</sup> XML/EDI-EDIFACT Electronic Commerce Interest Group, <http://www.w3.org/ECommerce/Overview.html>

<sup>[7]</sup> International Cooperation Set to Enable Global XML Use, Press Release, May 26, 1998, GCA