

# Electronic Commerce

## 6. Electronic Data Interchange and XML

*V Rajaraman*



**V Rajaraman is with the Jawaharlal Nehru Centre for Advanced Scientific Research and the Indian Institute of Science, Bangalore. Several generations of scientists and engineers in India have learnt computer science using his lucidly written textbooks on programming and computer fundamentals.**

**B2B e-commerce requires participating businesses to exchange business forms such as purchase order and invoice electronically without human intervention. This is called electronic data interchange (EDI). We will describe in this article the methods used to implement EDI.**

### Introduction

In the first part of this series we stated that in business to business e-commerce, electronic documents are exchanged between business partners by using either a private network or a public switched network. We also stated that in order to interpret them correctly we need a standard notation which is agreed to by both parties. This is called electronic data interchange or EDI for short. EDI is defined as the exchange of business documents between organizations in standardized electronic form, which can be interpreted and used directly by application programs. The major advantages of using EDI are:

1. Handling of paper documents is eliminated.
2. There is no need to manually re-key data in documents such as purchase orders, invoices etc., by participating businesses.
3. Elimination of manual data entry reduces cost, improves accuracy and reliability.
4. Time is saved due to elimination of manual handling and also due to direct application-to-application movement of data at electronic speeds.

### Electronic Data Interchange Standards

We will now describe the steps business A should follow to establish an EDI partnership with business B (see *Figure 1*).

Part 1. What is E-Commerce?, *Resonance*, Vol.5, No.10, 13-23, 2000.

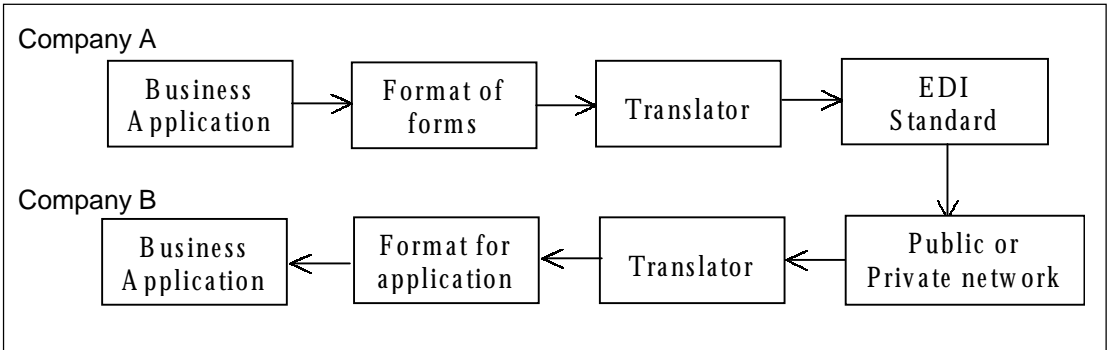
Part 2. E-Commerce System Architecture, *Resonance*, Vol.5, No.11, 26-36, 2000.

Part 3. Secure Messaging, *Resonance*, Vol.6, No.1, 8-17, 2001.

Part 4. Payment Schemes, *Resonance*, Vol.6, No.2, 6-13, 2001.

Part 5. Cash Transactions, *Resonance*, Vol.6, No.4, 8-14, 2001.





**Figure 1. Steps in electronic data interchange.**

1. The first step is to agree on a standard format for commonly used documents such as purchase order, invoice, payment advice, delivery note, etc. The formatting information or data type definition, as it is called, should include description of various fields used such as quantities, price, currency used, delivery date, field lengths, character type, ordering of fields in the document, units used, etc. As companies may transact business with many partners it is desirable to have a universally agreed standard form for all business documents. This realisation led to industry groups such as automobile industry, shipping and transport industry to adopt standards for inter-company transactions. This later evolved into national and international standards. The two standards are ANSI X.12 standard adopted by the American National Standards Institute for electronic transactions in the United States of America and EDIFACT (Electronic Data Interchange For Administration, Commerce and Transport) standardised by the United Nations Economic Commission for Europe.

2. Once an EDI standard is agreed on, company A should send business documents to company B using this format. This would require translation of company A's documents such as purchase order to the EDI format. The EDI messages are text with special characters such as ' , + and : as field separators. There are special tags defined in the EDIFACT dictionary for message header, date, etc. An EDIFACT message is meant to be interpreted by computer programs and is thus not easily understood by people unless they are trained in understanding the

As companies may transact business with many partners it is desirable to have a universally agreed standard form for all business documents.



EDIFACT Form	Meaning
UNH 000002 + ORDERS; DD96A UN: EAN 008	Header
BGM + 220 _ A000512-9'	Order No:A000512
DTM - 137 - 20010415:102'	Message date YR MM DD
NAD + BY Universal Book Traders: 2+N.S.C.Road, Banglore ++560022'	Purchaser's name and address
REF + API : UBT4578'	Purchaser's identity code
NAD + SU+++ PHINC'	Supplier's name
CUX + 2: USD:9'	Order currency: US dollars

**Table 1. A sample EDIFACT message for a book purchase order.**

standard. A purchase order for a book using EDIFACT standard is given in *Table 1*. In fact, EDIFACT standard defines several hundred transaction sets for various types of transactions between organizations and it requires an expert to understand it and convert commonly used documents (which are meant for people to understand) to EDIFACT form using a program.

3. The last decision to be taken is how the data is to be exchanged between the participating businesses. There are three alternatives. One can use the internet, which is a public switched telephone network. The other alternative is to use an extranet (an extranet is a private network connecting the intranets of the participating businesses). The last alternative is to use a private network called value added network provided by some vendors for reliable, secure communications of business data among participating businesses.

We will discuss next the advantages and disadvantages of the three methods.

**Using Internet and Extranet for EDI transactions.**

The major advantage of using the internet is its universal availability. All businesses are now connected to the internet. The cost of exchanging messages using the internet is very small.

An EDIFACT message is meant to be interpreted by computer programs and is thus not easily understood by people unless they are trained in understanding the standard.



The major disadvantages are poor reliability and lack of security. Internet protocol does not provide guaranteed delivery of messages and hackers are a perpetual threat. In B2B e-commerce it is important to ensure reliable, guaranteed and secure receipt of electronic documents by the intended receiver. Acknowledgement of receipt, non-repudiation (i.e. sender cannot deny later that he did not send a document such as a purchase order) and tracing transactions later, if necessary, are required. If internet is used, the appropriate protocol for EDI is called secure multipurpose internet mail extension abbreviated S/MIME. MIME specifies how EDI messages can be sent using simple mail transfer protocol (SMTP) of internet. S/MIME uses a combination of private and public key encryption, public key certification and digital signature. Encryption enhances security, public key certificate and digital signature are used for non-repudiation. If internet is used for EDI the following steps are followed.

1. Agree on EDI standard for the documents to be exchanged.
2. Cooperating businesses should establish e-mail addresses for sending/receiving electronic documents and for other communications related to it.
3. Method of encrypting documents, digital signature standard and acknowledgement of documents should be specified.
4. The computers, which receive electronic documents must always be powered up with a standby system in case of failure. They must be protected from hackers.

Extranet also uses the same method as internet as the protocol used in extranet is also TCP/IP. The main difference is better security as it will be more difficult for hackers to enter an extranet which is a private network connecting cooperating businesses.

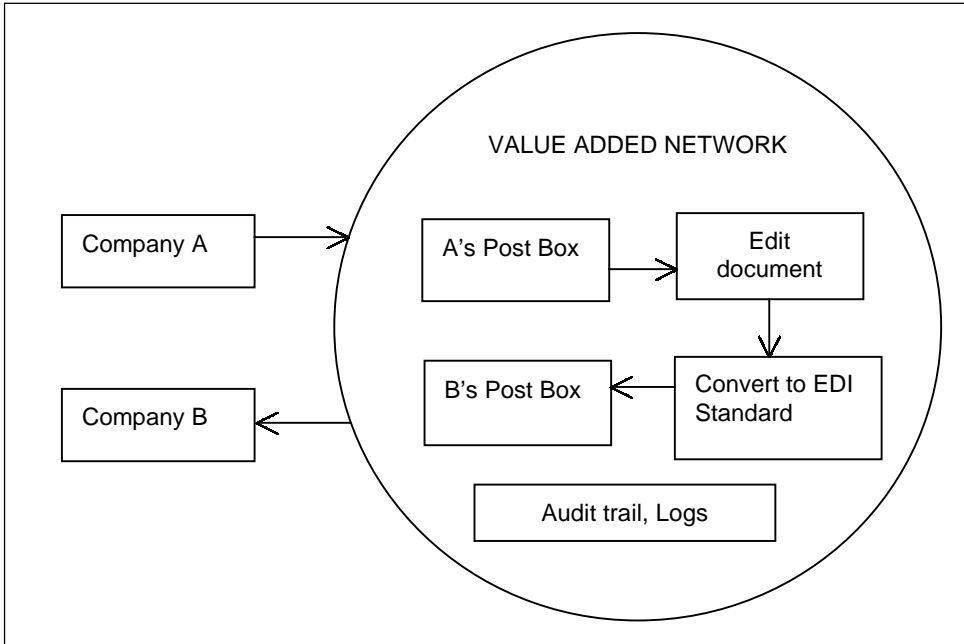
**Value Added Networks**

Value added networks (VAN) are private networks (see *Figure 2*) maintained by vendors such as IBM info exchange and General

In B2B e-commerce it is important to ensure reliable, guaranteed and secure receipt of electronic documents by the intended receiver.

Encryption enhances security, public key certificate and digital signature are used for non-repudiation.





**Figure 2. Value added network.**

VANs guarantee delivery of messages, provide acknowledgement to senders, ensure security of messages, audit trails and non-repudiation.

Electric infoserver which provide EDI services to its customers. VANs provide post boxes for each of its subscribers who want to use their services. A sender wanting to send, say a purchase order will address it to a vendor and deposit it in a 'postbox' maintained by VAN. The VAN service software will receive this, convert it to the required EDI standard format (if requested) and deposit it in a post box which has the recipient's address. VANs operate 24 hours a day, 7 days a week. They have back up systems to provide fail-safe operation. VANs guarantee delivery of messages, provide acknowledgement to senders, ensure security of messages, audit trails and non-repudiation. Logs of all activities are maintained and backed up for a reasonable time to ensure an effective dispute settlement mechanism. Despite all these services offered by VAN, they have not been popular primarily due to their high cost. Only larger businesses can afford to use their services. Internet based document exchange on the other hand, is relatively inexpensive. It also provides connections to all businesses, large and small. Businesses have also found it expensive to implement ANSI X.12 or EDIFACT standard as they are quite complex to learn and use. Thus only



less than 15% of businesses using e-commerce have adopted the EDIFACT/ANSI standards for EDI. Further, EDIFACT as well as ANSI X.12 EDI standards are low-level machine oriented documents. They were developed almost 25 years ago when networks were slow and processors also were slow. With the emergence of networks, which can transfer data at the rate of gigabits/second and processors with 1 GHz clock, speed is no more a concern. Currently the major concern is to enable all businesses, big and small, to participate in B2B e-commerce cost effectively. The electronic business documents to be exchanged must have a flexible structure as businesses find it impossible to adhere to a common format as they have been using their own business documents for a long time and are reluctant to change their formats as it involves costly redesign of systems and also retraining people. Now-a-days businesses across the world transact business with one another. Each country has its own taxation structure, rules and regulations and to expect businesses to adopt a common standard for all business documents is unrealistic. This is the main reason why EDI standards such as EDIFACT and ANSI X.12 are not widely used.

### **XML for EDI**

As was pointed out in the last section, implementing and operating EDIFACT or ANSI based EDI system is inflexible and expensive. Thus most businesses, particularly the small ones, which would like to participate in B2B e-commerce, require a cheaper alternative, which is easy to implement and uses the internet rather than a VAN for communication. The rapid growth of internet with increase in bandwidth and availability of faster processors has made efficiency less important compared to flexibility, ease of understanding and good documentation. This led to the development of a flexible and easily implementable system known as XML (extensible markup language) to describe documents stored in web pages. XML and HTML (hyper text markup language) are both based on what is known as SGML (standard generalised markup language) which gives a standard notation for defining documents. HTML uses pre-

EDIFACT as well as ANSI X.12 EDI standards are low-level machine oriented documents. They were developed almost 25 years ago when networks were slow and processors also were slow.

XML and HTML (hyper text markup language) are both based on what is known as SGML (standard generalised markup language) which gives a standard notation for defining documents.



```

<order>
  <order-no>A000512</order-no>
  <date>
    <year>2001</year>
    <month>04</month>
    <day>15</day>
  </date>
  <purchaser>
    <name> Universal Book Traders </name>
    <address>
      <street> 2 N.S.C. Road </street>
      <city> Bangalore </city>
      <pin-code> 560022 </pin-code>
    </address>
    <purchaser-ID> UBT4578 </purchaser-ID>
  </purchaser>
  <supplier>
    <name> PHINC </name>
  </supplier>
  <currency-type> USD </currency-type>
</order>

```

**Figure 3. XML definition of book purchase order given in Table 1 in EDIFACT notation.**

defined tags to describe the format in which a document will be presented, for example, spacing, headings, italicising, etc. It does not include tags to represent logical structure of data. Thus it is very difficult to isolate and access data from a HTML page and use it in an application. XML, on the other hand, is a logical representation in which we define a structure that directly represents data. A grammar to represent documents called a document type definition is used to define various tags used in XML.

In *Figure 3* we have given the EDIFACT document defined in *Table 1* using XML. Observe the tags used and how easily the XML description can be read and understood. When a company uses XML to describe business documents it also will give a set of statements which define the syntax of the XML program. It is called a document type definition (DTD). This will be published in the company's website so that any application program wanting to use the XML document can download and



```

<!. This is a comment with delimiters ..>
<! ELEMENT order (entry +)>
<!--order is top-level element and is a list of 1 or more entries
..>
<!--an entry is an order-no followed by date, purchaser, supplier
and currency type ..>
<! ELEMENT order-no (# PCDATA)>
<.. #PCDATA means a character string ..>
<! ELEMENT date (year, month, day)>
<! ELEMENT year (# PCDATA)>
<! ELEMENT month (# PCDATA)>
<! ELEMENT day (# PCDATA)>
<! ELEMENT purchaser (name, address, purchaser-ID)>
<! ELEMENT name (# PCDATA)>
<! ELEMENT address (street, city, pin-code)>
<! ELEMENT street (# PCDATA)>
<! ELEMENT city (# PCDATA)>
<! ELEMENT pin-code (# PCDATA)>
<! ELEMENT purchaser-ID (# PCDATA)>
<! ELEMENT supplier (name)>
<! ELEMENT name (# PCDATA)>
< ! ELEMENT currency-type (# PCDATA)>

```

**Figure 4.** The document type definition for order.

interpret the XML document correctly. The DTD corresponding to the XML description of *Figure 3* is given as *Figure 4*. In this definition #PCDATA means that the element contains a text. There are other key words used in DTD which we will not discuss in this article. A reference to where the DTD is available (e.g. a file name) should be given at the beginning of the XML program of *Figure 3*. The two statements which should be placed at the beginning of the XML program of *Figure 3* is given in *Figure 5*. It is assumed that the file order.dtd contains the DTD of order.

XML is gaining popularity because of the following reasons:

1. XML can be used to define the format, and layout of multimedia documents on a web page. The layout is made

```

<? XML version = "1.0"?>
<! DOCTYPE order SYSTEM "order.dtd">

```

**Figure 5.** Statement to be put at the beginning of the XML program of *Figure 3*.





'pretty' (i.e., easy to read) using a language called XML style language (XSL). XML allows use of hyper-links and is thus a good language to design web pages. As it follows a stricter syntax compared to HTML, it is easier to design browsers to retrieve and view XML documents compared to HTML documents.

2. XML has the capability to enforce a common structure for large documents which simplifies editing. The emphasis of structure in XML ensures better stability of documents.

#### **Box 1. Some Applications of XML [4]**

Any document has three features: the raw text, format and structure. Word processors and html use pre-defined fixed tags to format the raw text, whereas XML employs user defined tags to specify the structure of a document, which adds information on what the document contains. These tags are extremely useful to process the documents in diverse applications. Some of these applications are explained below.

- ❖ **Push Technology** – This technology is to send to a customer's browser a time varying information periodically requested by him/her. The required variable information is in what is called a channel and an XML document format is used to specify the information in what is known as channel definition format. Common applications are transmitting prices of shares every hour and sending news flashes such as cricket scores periodically.
- ❖ **On-line Banking** – A standard XML document format known as open financial exchange initiative is used to obtain information such as bank statements, sending standing instructions to banks, etc.
- ❖ **Software Distribution** – Software updates are distributed periodically using a commonly agreed XML definition. A common example is anti-virus software update sent every week.
- ❖ **Database Integration** – Data is delivered as an XML document.
- ❖ **Localization** – Information in web pages can be presented in one of several languages, English, French, Hindi, etc., as the standard character coding used in XML is unicode, which is a 16 bit code for characters.
- ❖ **Use in Scientific Publishing** – Markup languages based on XML are available to describe documents of interest to scientists. For example, CML is a chemical markup language, which is used to describe structure of molecules. Mathematical formulae can be described in what is known as mathematical markup language (MML).



3. Use of XML simplifies electronic exchange of documents as it can define the structure, syntax and semantics of documents. It also supports extending and changing the documents if needed.

4. As an XML document structure is clearly defined, it is possible to write a program to retrieve contents of fields such as item code, quantity ordered, price per unit, etc., from a document such as an invoice received electronically, and use it in an application.

For details of XML and its application in web design and EDI the reader may refer to [2] and [3].

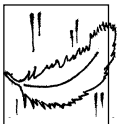
We have given a very brief overview of EDI in this article. Those interested in e-commerce must have a good knowledge of XML and Java (Java is used to access XML documents and process data using it). Apart from EDI, e-commerce also requires publication of price lists on a company's web page, business forms to be filled by customers which are made available on-line and managing customer relations (such as attending to information request, complaints, suggestions, etc.) All these also require use of XML which is more flexible than HTML. Besides this, XML has diverse uses, which are given in *Box 1*. Thus XML is now becoming very popular with web designers and participants in e-commerce.

XML is now becoming very popular with web designers and participants in e-commerce.

### Suggested Reading

- [1] D Minoli and E Minoli, *Web Commerce Technology Handbook*, Tata McGraw Hill, New Delhi, 1999.
- [2] H Maruyama, K Tamura, N Uramoto, *XML and Java*, Addison-Wesley, Reading, MA, USA, (Indian reprint) 2000.
- [3] B Marchal, *XML by Example*, Prentice Hall of India, New Delhi, 2001.
- [4] S McGrath, *XML by Example: Building E-Commerce Applications*, Prentice Hall PTR, Upper Saddle River, NJ, USA, 1998.

Address for Correspondence  
 V Rajaraman  
 IBM Professor of Information  
 Technology  
 JNCASR  
 Bangalore 560 064, India.  
 Email:rajaram@serc.iisc.ernet.in



The exquisite order displayed by our scientific understanding of the physical world calls for the divine.

Vera Kistiakowsky (MIT physicist)

