1-1-2001

A study of eXtensible Markup Language (XML)

Qin Liu
Iowa State University

A study of

eXtensible Markup Language

(XML)

by

Qin Liu

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Major: Business

Major Professor: Sree Nilakanta

Iowa State University

Ames, Iowa

2001

Graduate College
Iowa State University

This is to certify that the Master's thesis of

Qin Liu

has met the thesis requirements of Iowa State University

Signatures have been redacted for privacy

TABLE OF CONTENTS

CHAPTER 1 INTRODUCTION

Extensible markup Language (XML) (W3C, 1998) is a highly functional subset of Structured Generalized Markup Language (SGML) (Goldfarb, 1990). SGML is a way to express structure and content in different types of electronic documents, it was and still is the solution that covers both content and semantic structures. It is the perfect document format for information retrieval. However, it has been considered to be so complex that SGML has never reached widespread public usage.

Together with the World Wide Web (WWW), HyperText Markup Language (HTML) (W3C, 1998) started its successful story as an easy-to-use Internet/Intranet online document format. Today HTML is very popular and is being used for creating Web pages that can be viewed all around the world. Even though it is an SGML application, HTML unfortunately consists of neutral (layout oriented) and not of semantic (content oriented) elements. Therefore, full text search hits are not very helpful.

Unlike HTML, XML is a subset of SGML but limited to the features needed for online documents. It was developed by the WWW consortium to be less complex than SGML and to be able to work in limited bandwidth networks such as the Internet. It also has inherits the most important feature of SGML, namely, the definition of application specific semantic structure elements. XML can be used to define new applications more easily than HTML.

Many industries are in the process of adopting XML as a standard for information exchange. From their new XML industry specific initiatives, many

organizations are keen to make applications working quickly. The benefits of XML are likely to encourage firms to move even more quickly as the maintenance costs, flexibility and performance benefits become realized.

The attention paid by the Internet community to Extensible Markup Language (XML) is impressive. XML has been heralded as the next important Internet technology, the next step following the HyperText Markup Language (HTML), and the natural and worthy companion to the Java programming language.

Recent development in the area of information exchange has clearly shown the emerging importance of XML. Currently there is no systematic examination of the claims of XML being superior to other languages and protocols. Moreover, compiling a list of benefits and costs will be valuable to firms planning to take advantage of this technology.

A brief history of XML is first presented in this chapter, then XML is described in the detail and potential benefits of XML is discussed. In the following chapters, this paper will address several issues including a study of the claims related to XML. In Chapter 2, XML's relationship with current technology of Java and EDI and its potential uses are examined. A study of the benefits claimed by users and vendors of XML will be presented in Chapter 3. Finally the paper presents concluding discussion in Chapter 4.

History of XML

The history of XML is deeply integrated with the history of electronic text processing, document processing, information processing and publishing.

Generalized Markup Language (GML) and GenCode, both invented in 1960s, can be considered as XML's grandparents. IBM developed the (GML) as a solution for handling large amount of documents associated with their systems (Rath, 2000). GML came with a batch processor and allowed reuse of the same source file for different outputs (paper and electronic). The syntax of GML was simple and was instrumental in reducing the cost of data capture of dissemination. GenCode from Graphics Communication Association (GCA) supported generic typesetting codes and was already aware of document types.

In the early 1980s both groups started to standardize the way to specify, define, and use markup in documents with putting together the best of their developments (syntax from GML, semantics from GenCode). The first standardized markup language, Standard Generalized Markup Language (SGML), was published by the International Organization for Standardization (ISO) in 1986 (ISO, 1986). The key component of SGML is the Document Type Definition (DTD), which is used to define application specific document structures that consist of elements and attributes. Each DTD defines an application of SGML. SGML is still a heavily used international standard maintained by ISO. It gives the publishing industry a machine- and process-independent method of separating content from presentation.

Some people at CERN – European laboratory for Particle Physics adopted the SGML syntax for hypertext applications in the late 1980s, which resulted in the HyperText Markup Language (HTML) (Johnson, 1999). Improvements in HTML and the definition of HTML DTD made HTML 1.0 a real SGML application in 1993. To

date, HTML is the most popular application of SGML. It acts as the presentation syntax that Web browsers use to render documents visually.

However, the simplicity of HTML has its price. With its limited small tag set, it can not markup all the information in the world in an appropriate way. HTML supports layout oriented markup but not content oriented markup. Therefore, the coded information presentable to humans is not accessible to machines

The success of HTML came with the success of the World Wide Web (WWW). Both are simple to use, but powerful. There has been no technology in the history of mankind like the Internet that was adopted so quickly by so many people. However, HTML is too specific to represent information generically, and SGML is too overbearing to use in tandem with the web. The difference between SGML and HTML spurred the development of the XML specification. Therefore, the XML language emerged as a simpler, generalized markup for the Web.

Jon Bosak, an engineer at Sun Microsystems and generally regarded as the "father of XML" (Morgenthal, 2000) realized the limitation of HTML early on. He had used SGML extensively for managing technical documents on behalf of two large vendors, first Novell and then Sun Microsystems. [This experience led Bosak to drive higher the expectations for publishing on the web and demand nothing less powerful than SGML as the delivery tool.] In 1996, Jon Bosak was appointed by the World Wide Web Consortium (W3C) to lead a group of people to simplify SGML for the web and develop the eXtensible Markup Language (XML). In February 1998, the World Wide Web consortium (W3C) published the XML specification.

Basic XML

XML is designed to be a common syntax for expressing structure in data. Putting structure in data means to structure the document depending on the content, meaning, or usage of the data. XML documents contain character data and markup. The character data is often referred to simply as content, while the markup provides structures for that content. The distinction between data and markup is drawn more sharply and more visibly in XML than it is in many other systems.

The simplest way to describe the structure of XML is to compare it to HTML, which most people are already familiar with. As in HTML, the structure in XML is built up using markup tags. There is however a very important difference between tags in HTML and tags in XML; unlike HTML tags, XML tags have no predefined meaning. In addition, XML is extensible, structured, and can be validated. XML separates data and the presentation of data, which means that the same XML document can be displayed in a number of ways depending on the media of the presentation without changing the underlying structure of the data.

Elements

The basic building blocks of XML are elements, which are built up by markup tags and character data. Each element represents a logic component of the XML document. All elements consist of one of the following constructs:

- A tag together with the character data that the tag describes.
  Example: <Make> Saab 9000 </Make>.
- A tag together with other elements (which in turn can consist of other tags and elements and so on)

Each XML document therefore consists of elements hanging together in a logical tree structure. There is always one element that contains all the other elements; this element is called the root element. The tree structure of the elements is a vital part of XML and is used when an application wants to read and interpret an XML document, see Figure 1.

Attributes

Elements can also have other information attached to them called attributes. Attributes describe properties of elements. For instance, it is sometimes useful to have an attribute containing the security demands on the person or application that

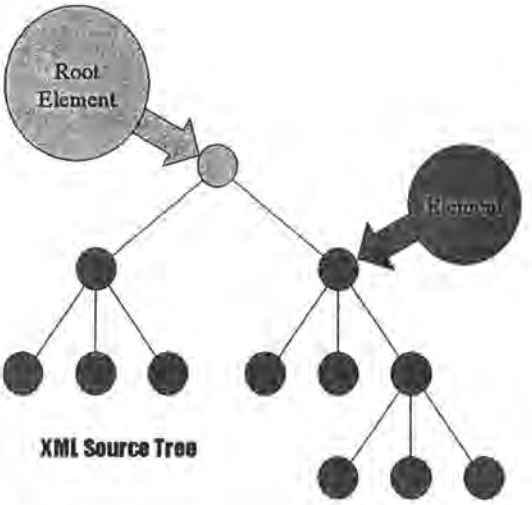

Figure 1: The structure of the XML source tree

wants to access that element. Attributes in XML look very similar to attributes in HTML.

Example: An element with attributes

```
<Employees>
        <!- The Person element has three attributes, email, phone, and fax ->
        <Person email  = "any@where" phone = "111111" fax = "111112">
```

```
        <Name> Kalle Andersson </Name>
    </Person>
    <Person email = "some@where" phone = "222222" fax = "222223">
        <Name> Johan Olsson </Name>
    </Person>
</Employees>
```

Entities

XML is based on the concept of documents that are composed of a series of entities. An entity is a short cut to a set of information. A shorter entity name is typed during data entry, but when it is used, it 'expands' to its full meaning. It works a bit like a macro – it is a set of information that can be used by calling one name. Figure 2 shows how the entities are replaced with the content when they are called through XML parsers.

An entity can be as small as a single character or as large as an entire XML document. An XML document can be broken up into many files on a hard disk or objects in a database and each of them is called an entity in XML terminology. Entities can even be spread across the Internet. Whereas XML elements describe the XML documents' logical structure, entities keep track of the location of the chunks of bytes that make up the document.

Very simply an entity consists of a name and a content. The content is the actual stored data and the name is used to refer to that data. There are several different kinds of entities used for different purposes. If an entity is defined without any separate storage file and the content is given in its declaration, the entity is called an *internal entity*. All internal entities are parsed entities. This means that the XML processor parses them like any other XML texts.

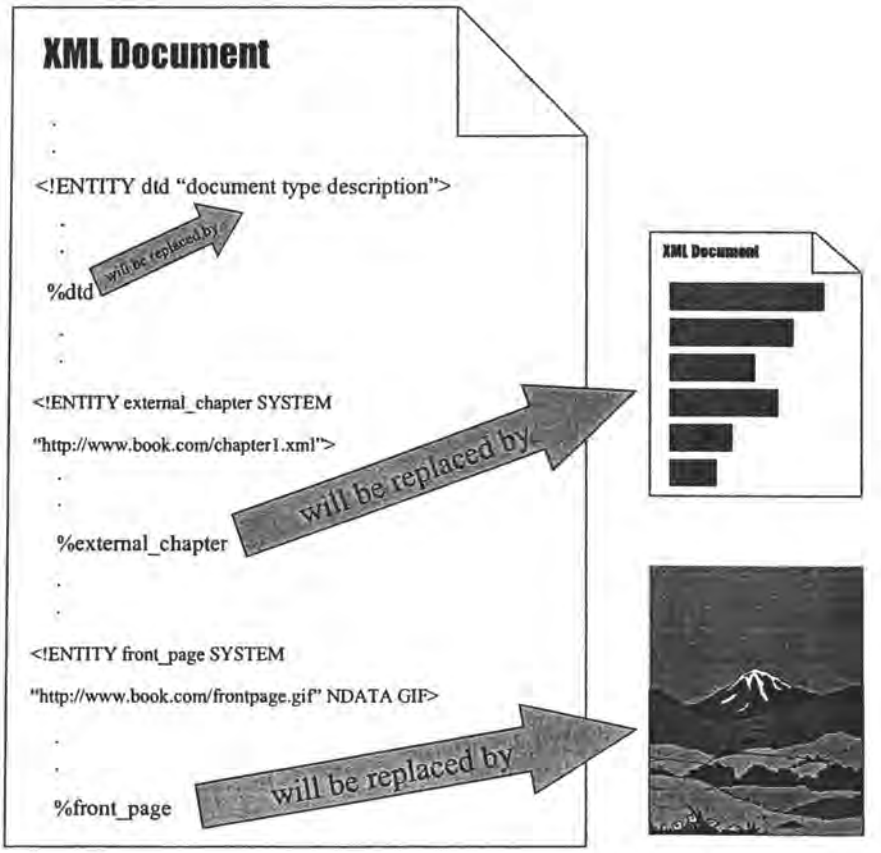

Figure 2: Shows how entities are replaced with their content by XML parser

External entities get their contents from somewhere else in the system. The location of the content is identified using an external identifier; usually this is just the word SYSTEM followed by a Uniform Resource Identifier (URI). External entities are either parsed or unparsed depending on their content. Syntactically, declarations of unparsed entities are differentiated from those of other external entities by the keyword NDATA followed by a notation name.

Document Type Declaration (DTD)

An important part of XML is the ability to store information about what the rest of the document will contain at the beginning of an XML document; this information

is called the Document Type Declaration (DTD). The DTD defines what markup tags can be used in the document, what order they can appear in, what other tags they can contain and so on. In XML it isn't strictly necessary to have a DTD associated with each XML document, but it is considered a bad manner not to have one. DTD also provides XML parsers a set of rules with which they can validate the document. Validation does not imply that the contents of the document are correct, or that certain data fields are numbers or text; rather, it means that all the elements of the document fit into the structure specified by the DTD.

The DTD is used to define the model that the rest of the data in the document must follow. It also allows a document to communicate meta-information to the parser about its content. The DTD can be fetched from an external source or be embedded in the XML document itself. Standardized DTDs for different business areas are in place at several organizations around the world.

Benefits Of XML

XML is a set of tools that allows developers to create web pages on demands and to set standards defining the information that should appear in a document, and in a certain sequence. XML, in combination with other standards, makes it possible to define the content of a document separately from its formatting, making it easy to reuse that content in other applications or for other presentation environments. Most importantly, XML provides a basic syntax that can be used to share information among different kinds of computers, different applications, and different organizations without having to pass through many layers of conversion.

Web developers are the initial target audience for XML, but database developers, document managers, desktop publishers, programmers, scientists, and other academics, such as educational universities and Chemistry Academic Community (ChemXml), are all getting involved. XML provides a simple format that is flexible enough to accommodate diverse needs. Developers performing tasks on different types of applications with different interfaces and different data structures can share XML formats and tools for parsing those formats into data structures that can be used. XML offers its users many advantages as referenced from Microsoft Working Group's documents (Microsoft, 1999) and Coco Jaenicke's paper (Jaenicke, 1999), including:

Simplicity and Ease of Use

XML provides both programmers and document authors with a friendly environment, at least by current computing standards. XML's rigid set of rules helps make documents more readable to both humans and machines. XML document syntax contains a fairly small set of rules, making it possible for developers to get started readily. DTDs can be developed through a standardized process, set by experts, or through experimentation, based on the structures of documents that seem to work well. XML parsers are also reasonably simple to build.

XML documents are built upon a core set of basic nested structures. While the structures themselves can grow complex as layers and layers of detail are added, the mechanisms underlying those structures require very little implementation effort, from either authors or developers. These basic structures can be used to represent complex sets of information, from the full contents of a

document to persistent object state information to a set of commands for a program, without changing the structures themselves.

XML is extremely easy to use. The format is similar to HTML, and the tags can be created in English. An XML document can be created with any text editors, such as Notepad. The result is a data format that is as easy to use as HTML. Easy of use is essential for today's data-intensive applications. Because of XML's short learning curve and convenient format, applications can be deployed and enhanced quickly.

Extensibility

XML is extensible in two senses. First, it allows developers to create their own DTDs, effectively creating 'extensible' tag sets that can be used for multiple applications. Second, XML itself is being extended with several additional standards that add styles, linking, and referencing ability to the core XML set of capabilities. As a core standard, XML provides a solid foundation around which other standards may grow.

Creating DTDs is most likely what the creators of XML had in mind when they called it Extensible Markup Language. XML is a meta-language, a set of rules that can be used to create sets of rules for documents. In a certain sense, an XML document that uses XML-compliant syntax is really using applications of XML with tag sets chosen by its creators for that particular document. XML's facilities for creating DTDs give standard-builders a set of tools for specifying what document structures may or must appear in a document, making it easy to define sets of

structures. These structures can then be used with XML tools for authoring, parsing, and processing, and used by applications as guide to the data they should accept.

By not predefining any tags in the XML Recommendation, the World Wide Web Consortium (W3C) allowed developers full control over customizing their data as they see fit. This makes XML very attractive to encoding data that already exists in legacy databases by using database metadata and other schema information. This extensibility of XML makes it such a great fit when trying to get different systems to work with each other.

Interoperability

XML can be used on a wide spectrum of platforms and interpreted with a wide variety of tools. Because the document structures behave consistently, parsers that interpret them can be built at relatively low cost in a number of languages. XML supports a number of key standards for character encoding, allowing it to be used all over the world in many different computing environments. XML complements Java very well, a finding supported by the considerable amount of early XML development in Java. Also, generic application programming interface (API) for parsers, the Simple API in XML (SAX), is freely available. Parsers are also available in C++, C, JavaScript, TCL, and Python, with more on the way. XML parser development so far has focused on freeware plug-in that provide parsing capabilities to XML applications, generally lowering the cost of building XML-enabled applications.

Open Standards

By making the W3C the keeper of the XML standard, it ensures that no one vendor should be able to cause interoperability problems to occur between systems

that use the open standard. This is assuring to most companies making an investment in this technology, by being vendor neutral, this solution proposes to keep even small companies out of reach of big companies choosing to change to standards on them. For example, if a big company chooses to change the platform at its whim, then most other companies relying on that platform suffer. By keeping all data in XML and using XML in communications protocols companies can maximize the lifetime of their investment in their products and solutions.

International Character Handling

One substantial benefit of using XML is its capability for handling international character sets. Today's businesses are rapidly becoming international, especially when considering Web applications; the Internet is obscuring country borders. It is only natural that business transactions contain street names in Chinese or person names in Arabic. The XML 1.0 specification is defined based on the Unicode character set, so virtually all the characters that are used today all over the world are legal characters.

In addition, XML 1.0 requires that all conformant XML processors must support at least two encodings, 8-bit UCS Transformation Format (UTF-8) and 16-bit UCS Transformation Format (UTF-16). Some XML processors, including the IBM XML for Java Parser, support conversion to/from other locale-specific encodings that are used in different parts of the world. For example, the XML for Java Parser supports 19 encodings including European, Japanese, Chinese, and Korean encodings.

Web Enabled

XML is derived from SGML, and so was HTML. In essence, the current infrastructure available today to deal with HTML content can be re-used to work with XML. This is a very big advantage towards delivering XML content using the software and networking infrastructure already in place today. This should be a big plus in considering XML for use in any projects, because XML naturally lends itself to being used over the web.

Even if clients do not support XML natively, it is not a big hindrance. In fact, Java with Servlets (on the server side0 can convert XML with stylesheets to generate plain HTML that can be displayed in all web browsers.

A Core of Experienced Professionals

Successful application of XML requires data modeling expertise and the building of a new set of tools. Fortunately, the skills of the SGML community are mostly transferable, providing XML with a large group of 'experts' early in its existence. The XML specification was created for the most part by a group of experienced SGML developers, and has received vocal support from many sectors of the SGML community. Vendors are re-purposing their tools, simplifying them for XML. Authors who have been writing SGML texts are focusing on XML as well, bringing markup structure to a wider audience. Companies that need to bring in outside vendors to help with large projects have a pool of firms with established track records to choose from.

CHAPTER 2 IMPACTS AND USES

Impacts

Java

XML in many ways augments Java. XML, however, is also evolving into an object transport protocol that could undermine Java's claim as a ubiquitous platform. XML tags Web-based information for recognition by developers and computers, which is necessary because HTML lacks a way to add meaning to content aside from cryptic URLs. XML aims to add that meaning to Web objects, a task once assigned to Java.

A close relationship between XML and Java has existed since the early days of the XML efforts. One of the first public statements about this relationship came from Sun Microsystems' Jon Bosak, chair of the XML Working Group. He said, "XML gives Java something to do" (About.com, 2000). But it works the other way as well: Java lets XML do something useful.

XML by itself is just a lot of text; a program is needed to manipulate that text and make things happen. Until now, Java has been the language of choice for writing those programs. Java and XML share two very important characteristics that result in XML's success with Java (Fuchs, 1999).

First, both languages were designed being used in distributed systems. XML was conceived as SGML for the Web to take over from HTML. The SGML community had generally been dismissive of HTML from the start and developers were anxious to apply the lessons of HTML's success to SGML as a whole. Java

was not initially intended for the Web but for delivery over networks to a variety of devices. Its feature set was easily adapted to the Web.

Second, both languages are simplifications of powerful beasts whose complexity have gotten out of control. In the case of Java, the beast was C++, which grew from a relatively straightforward object-oriented extension of C to a complex language with support for several complex types of inheritance and templates. In the case of XML, the beast was SGML. XML can be viewed as a subset of SGML; it eliminates several syntactic and lexical quirks in SGML that complicate processing. Many of these SGML features were intended to be used as shortcuts for writing documents. In designing XML, many of these were deliberately thrown away.

The Java technology binding to the W3C Document Object Model (DOM) provides developers with a highly productive environment for processing and querying XML documents. The Java platform can become a ubiquitous run-time environment for processing XML documents. The Java platform's intrinsic support of the object-oriented programming means that developers can build applications by creating hierarchies of Java objects. Similarly, the XML specification offers a hierarchical representation of data. Because the Java platform and XML content share this common underlying feature, they are extremely compatible for representing each other's structure.

XML and Java technologies have many complementary features, and when used in combination they enable a powerful platform for sharing and processing of data and documents. Many developers have come to the conclusion that XML and

Java is the perfect pair because they complement each other so well. While XML can contribute platform-independent data – portable documents and data, Java can contribute platform-independent processing – portable object-oriented software solution. Together, XML and Java technologies allow enterprises to apply Write Once, Run Anywhere fundamentals to the processing of data and documents generated by both Java technology and non-Java technology sources. By extending the Java platform standards to include XML technology, companies will obtain a long-term secure solution for including support for XML technologies in their applications written in the Java programming language.

Electronic Data Interchange (EDI)

EDI is commonly defined as the application-to-application transfer of business documents between computers. Many businesses choose EDI as a fast, inexpensive, and safe method of sending purchase orders, invoices, shipping notices, and other frequently used business documents.

The straight transfer of computer files requires that the computer applications of both the sender and the receiver agree upon the format of the document. The sender must use an application that creates a file format identical to the receiver's computer application. When the sender sends a document, EDI translation software can convert the proprietary format into an agreed upon standard. When the receiver receives the document, the EDI translation software automatically changes the standard format into the proprietary format of his document processing software.

Although the idea behind EDI is a good one, unfortunately it has not worked out so well in reality. One of the problems with the current implementations of EDI is

that they often require a unique solution for each pair of receiver and sender, making EDI costly and time-consuming to implement.

Another problem with traditional EDI is that it is based on the use of rigid transaction sets with business rules embedded in them. These transaction sets are defined by standards bodies, such as the United Nations Standards Messages Directory, for Electronic Data Interchange for Administration, Commerce, and Transport (UN/EDIFACT) and American National Standards Institute's Accredited Standards Committee X12 sub-group (ANSI X12). Transaction sets define the fields, the order of these fields, and the length of the fields. Along with these transactions sets are business rules, which are referred to as implementation guidelines. A fixed transaction set prevents companies from evolving by adding new services and products or changing business processes. The bodies that make the standard transaction sets are ill equipped to keep up with the rapid pace of change in the various business environments they impact. It is also very hard to develop a one-size-fits-all solution.

By using XML to implement EDI, many of the old problems are eliminated. One area where XML can provide immediate value is in establishing a vocabulary and format for the definition of EDI documents (Ogbuji, 1999). This is especially useful when one trading partner for its own internal use has extended the base X12 and EDIFACT documents. Using XML data, trading partners could communicate the schema of their EDI documents. Over the long term, this information may become an automated part of the exchange process, thus simplifying and reducing implementation costs. Besides, XML can maintain the content and structure, but

separates the business rules from the data. By focusing on exchanging data content and structure, trading partners can apply their own business rules. Using XML it is also very easy to extend the communication to support new business processes, EDI will no longer be limited to rigid standards.

The fact that XML documents can easily be distributed using the Internet is another major advantage. Combined with the fact that XML is self-explaining, this provides the entire framework for what the XML/EDI group calls a new "*supply web*". Earlier many EDI implementations only worked within their own Intranets. With XML this is no longer the case since XML provides connectivity through the Internet. All applications are then able to communicate and exchange data, thus the old point-to-point solutions are history. To many people, XML-EDI is also what they call a "politically correct" way to merge ANSI X12 (the US standard) and EDIFACT (the world standard).

Uses

"XML Inside" – A great marketing slogan for an online information provider! (Rath, 2000) XML itself is not a market, it is a technology. This technology is so powerful and flexible that it should be used to improve things and to innovate things. There are a lot of real life applications enabled and enhanced by using XML.

Content Management

The semantic markup in the XML information gives detailed access to the content. Each element works like a semantic content container to provide a handle for that particular information unit and say something about the content (semantic). All this information allows more flexible and more precise handling of data. Content

management with a large amount of data can be supported by an editorial system or document management systems structured with XML.

Metadata Management

Metadata is information about information. It is the technology that makes possible faster, more focussed search and retrieval of information objects. It can be used for retrieval purposes and for management of the information. XML is doing a good job of structuring data. Each relevant part of the data can be searched, accessed and processed. Since metadata is 'only' data, XML is an appropriate format to model and maintain it.

Database Publishing

Well - structured documents in a database open up a wide range of publishing approaches out of the database. All of them can be fully automated: publish only the latest changes as updates, produce personalized edition depending on user profiles or user queries, extract portions out of existing data and compose it to new publications and allow direct access to "Information base". Service providers can add value to information by assigning metadata or link structures. "Information base" becomes a "knowledge base". That is perhaps a major reason why the Web service Yahoo! Is so successful – they added metadata to their information.

For the Web applications, the applications that can not be accomplished within the limitations of HTML drive the acceptance of XML. According to Boye's viewpoints (Boye, 1998), these applications can be divided into four broad categories:

1. Applications that require the Web client to mediate between two or more heterogeneous databases.
2. Applications that attempt to distribute a significant proportion of the processing load from the Web server to the Web client.
3. Applications in which intelligent Web agents attempts to tailor information discovery to the needs of individual users.
4. Applications that require the Web client to present different views of the same data to different users.

The alternative to XML for these applications is proprietary code embedded as "script elements" in HTML documents, and delivered in conjunction with proprietary browser plug-ins or Java applets. It will complement, rather than replace HTML. Whereas HTML is used for formatting and displaying data, XML represents the contextual meaning of the data. They will coexist on the Web for many years to come.

XML separates data format from its content, so creators can have full control over the data. Data format serves the content provider the best by not binding them to particular languages, authoring tools, and delivery engines, but providing a standardized, vendor-independent, level playing field upon which different authoring and delivery tools may freely compete.

CHAPTER 3 THE SURVEY STUDY

XML allows some new things to develop that were not previously feasible. The need for XML is significant because of the technical barriers it overcomes and the powerful business requirements it supports. Since the formal recommendation by W3C of XML1.0 in February 1998, XML has evolved to encompass not just the XML 1.0 specification but also an expanding portfolio of technologies that will complement each other in a variety of environments and domains.

Different industry segments have different vocabularies and contexts, and they need different document schema and metadata. Some industries have adopted XML development currently occurring in the industry. Many groups in different industries are working on drafting XML specifications on various business functions, see the detail list on Appendix A. The purpose of this survey is to show the extent of XML's penetrating into the business's world, identify the XML's benefits in the applications and quantify the benefits users can get from XML.

Sample Selection

The keyword 'XML' was used to search web sites using the popular search engines such as Yahoo!, AltaVista, Google and Lycos Search. Totally 264 documents were received. Some of these documents were written as for product introduction or description of XML product features, which did not claim the benefits from users' side and were excluded from the study.

Within the documents received, further filtering to locate case studies of using XML applications yielded 152 cases. The principal sources for these samples were

Allaire (29 cases), Documentum (20 cases), engima (7 cases), IBM (21 cases), Marimba (19 cases), Microstar (19 cases), Oracle (26 cases), Siebel (5 cases), webMethods (5 cases), and Arbortext (1 case).

Some vendors, however, did not have specific web pages to tell their stories. After a review of trade publications and navigation on the Internet, it was revealed that there were several early vendors and users of XML. These early adopters were selected as a convenient sample for this study.

Most of 152 case studies came from these early XML vendors. Users in these cases had experienced benefits from implementing vendors' products across different industries and countries. The sources of the stories are listed on the Appendix B and the list of the users involved in the study is on the Appendix C, Appendix D shows an excerpt from one case study.

Data Collection

Since XML is not a single product but a new technology, it is difficult to use financial or accounting data to measure the success of the firms through using XML. In addition, there are many qualitative measures to describe the benefits of the XML in the case studies and successful stories and they have to be converted to quantitative measures for the data analysis. Considering these difficulties, the initial research was to summarize the common words or similar meaning of benefit expressions to a benefit list, then to quantify by counting the benefits on the list for each case for each vendor.

The initial data collection is compiled with Excel spreadsheet to get a full picture of vendor, user and benefits. Access database is used to classify the detailed information on categories as vendors, industries and countries. Details including various summary statistics are listed in the following appendixes.

- Appendix E Overall Benefits Counts: displays the cumulative counts on each benefit for the whole sample.
- Appendix F Vendor Source Benefits Counts: displays the cumulative counts on the benefits from each vendor's source.
- Appendix G Industry Benefits Counts: displays the cumulative counts by the users in related industries.
- Appendix H Country Benefits Counts: displays the cumulative counts by the countries in which the users are.

Benefits Analysis

Top ten benefits shown on Table 1 come into being from the overall benefit counts. They reflect the XML's benefits in the first part of the paper and meet the W3C working group's expectation of XML application development as simple, extensible and interoperable.

Data flexibility of XML allows it to describe data contained in a wide variety of heterogeneous applications and the data can be exchanged and processed without having a built-in description of the incoming data. In other words, XML can be easily used to develop new systems, enhance the existing systems, and reduce development cost and short in the development stage in the system life cycle. In the

Table 1 Top Ten Benefits Commonly Mentioned In The Cases

BENEFITS	Counts
Increased productivity	44
Data flexibility	40
Scalability	38
Improved customer services	35
Speed development	35
Data integration from distributed sources	35
Cost savings	34
Improved operation efficiency	31
Easy of use	30
Easy to integrate to the existing systems	30

survey, 40 users (26.5% of total) identified data flexibility as a perceived benefit. It ranks the second in the top ten list, implying that flexibility is one of the common features in the XML applications.

In addition, because XML completely separates the notion of markup from its intended display, authors can embed procedural description of how to produce difference data views within the DTD. It allows individual pieces of data to be altered with only an update notice, greatly enhancing server scalability as a result of a much lower workload. Thirty-eight users contribute this to XML's scalability and consider it beneficial in the larger information exchange.

Because XML is easy to use, maintain and integrate with existing technologies, firms can cut cost on development and operation, as well as save money on training and reduce staff workload. Therefore, XML results in the increased productivity, cost savings and improved operational efficiency on the top ten list, which received 109 cites and 31% of all on this list.

The benefit summarization list is made by counting benefits addressed in the case studies. In the study, most of users listed the benefits they experienced, but

did not rank them. Thus from the benefit counts, it is difficult to identify the degree of importance of each benefit on the list. However, the counts represent the frequency of each benefit appearing in the cases. It is fair to say that above top ten benefits are most common in the business application. In the following part, statistics method is used to test the equality of these benefits.

Statistics Analysis

Inferential statistics, a branch of statistics that employs data to draw inferences or make prediction (Sheskin, 2000), is applied in this study. The population is made up of 152 case studies. From which 78 benefit items are identified. Since the counts of benefits are nominal type of data, non-parametric forms of analyses are performed.

However, it is not practical to have 78 benefit categories in the analyses with a small population, some benefits have only one or few frequency counts and some of these benefits represent the same characteristics of XML. After being further examined, 78 benefits addressed in the case studies are categorized into nine broad categories based on the standard definition of characteristics of XML in the applications (see table 2). This standard classification scheme for describing the benefit of XML was developed based on the Roger's attributes of innovations (Roger, 1983) as well as XML's unique characteristics as an innovation in the information technology fields. Total count for each broad category is aggregated from the individual references to benefits from each case.

Table 2 Broad Benefit Category

Data Quality - the degree to which XML data can provide value to the information

accurate data	customized data format
data control	data recoverability
re-usability	reliability

Compatibility - the degree to which XML technology being used with the existing systems

standard access	device-specific usability
open standards	platform independence
platform neutrality	message protocol independence
standard exchange	easy to integrate to the existing system
compatible with the existing technology	

Cost - cost related issues by using XML in the business process

cost effective	cost savings
reduced development cost	reduced labor cost
reduced operating cost	reduced maintenance cost

Simplicity - the degree to which XML is perceived as relatively easy to understand and use

easy of use	easy maintenance
easy to learn	content easily updated
reduced complexity	

Efficiency - the extent to which XML is beneficial to the process systems

increased productivity	efficient content management
easily tracking transaction	data sharing
increased equipment uptime	high performance
minimized error	speed development

Versatility - the extent to which features of XML are used as build-in functions in the applications

data flexibility	data integration from distributed sources
flexible structure	extensible content
multimedia functionality	multiple functionality
multiple format	scalability
security	stability
rich information	

Responsiveness - the degree to which XML shorts the lead time and response time

easy access	fast data retrieval
speed process	fast transaction and fulfillment cycles
simplifying process	quick response
quick order taking process	

Observability - the degree to which the results of using XML application are visible to the users

dynamic information	improved communication
improved customer service	increased customer base
powerful hyperlink capability	powerful navigation capability
powerful searching capability	improved operation efficiency
real-time information	time saving

Relative advantage - the degree to which XML is perceived as being better than other languages

consistency	data availability
dynamic development	high quality
guaranteed connection	effective customer services
dynamic connectivity	high customer satisfaction
intelligent applications	high data management capability
transaction integration	dynamic control of the content
streamline business process	workflow automation
fast product delivery	flexible to accommodate changes

Studies of classification of new technologies or products often suggest that at the exploratory level the research might want to establish a preference ordering of the groups. The ranking could then be used to examine several causes and effects due to the technology or products. The category of perceived benefits are therefore totaled to see if there is a rank ordering among the groupings. A null hypothesis that there is no rank of the benefit categories is proposed here.

Method

Null hypothesis is a statement of no effect or no difference. Here, for the population in this study, the null hypothesis is that the sample represents, for each of the benefit category, that the observed frequency of a benefit category is equal to the expected frequency of the benefit category. This is tested with χ^2 test using formula Equation 1.

$$X^2 = \sum_{i=1}^{k} \left[\frac{(O_i - E_i)^2}{E_i} \right] \qquad \textit{(Equation 1)}$$

Where O_I = observed number of cases categorized in *i*th category
E_I = expected number of cases in *i*th category under H_0

The technique of χ^2 test is the goodness-of-fit type to test whether the observed frequencies are sufficiently close to the expected ones to be likely to have occurred under H_0. If the agreement between the observed and expected frequencies is close, the differences $(O_i - E_i)$ will be small and consequently χ^2 will be small. If the divergence is large, however, the value of χ^2 computed from Equation 1 will also be large. The sampling distribution of χ^2 under H_0 follows the chi-square distribution with degree of freedom as $df = k - 1$. The expected χ^2 at the given associated probabilities of occurrence under H_0 is taken from the sampling

distribution of chi square. There is a different value of chi square for each *df* at the same associated probability of occurrence.

To use χ^2 in testing hypothesis in this one-sample study, each benefit count is independent of every other; thus, each count is cast into only one of the benefit categories. The total number of counts in all the categories is the number of total benefit counts in the study.

Interpretation of the Test Results

In this study, $k = 9$, $N = 1017$, data and computation is summarized in Table 3. The degrees of freedom employed in evaluating the results is $df = k - 1 = 8$. From the table of the Chi-Square Distribution, as $\chi^2 = 105.35$, the probability of occurrence is $p < 0.001$. Since p is less than $\alpha = 0.001$, the Null hypothesis H_0 is rejected at the significant level 0.001. This indicates that broad benefit categories have order rankings in the XML applications; some benefits are more common with high frequency counts in the study, whereas others have much lower frequency counts than expected. This implies that there is perceived difference among the categories

Table 3 Summary For Sample Data Test

Category	O_i	E_i	$(O_i - E_i)$	$(O_i-E_i)^2$	$(O_i-E_i)^2/E_i$
Data quality	95	113	-84	324	2.87
Compatibility	50	113	-63	3969	35.12
Cost	101	113	-12	144	1.27
Simplicity	81	113	-32	1024	9.06
Efficiency	167	113	54	2916	25.81
Versatility	155	113	42	1764	15.61
Responsiveness	92	113	-21	441	3.90
Observability	144	113	31	961	8.50
Relative advantage	132	113	19	361	3.19
Sum:	**1017**	**1017**	**0**	**11094**	**105.35**

as far as they might be ranked. In order to get more detailed information about benefit commonality, the same χ^2 one-sample test is done on the individual benefits of each broad category. The result is displayed in Table 4.

From Table 4, it indicates that each individual XML benefit in the category does not have the same frequency counts in the study, which implies individual

Table 4 χ^2 Test Summary For Each Broad Benefit Category (α = 0.001)

Category	df_i	$\chi^2_{observed}$	$\chi^2_{expected}$	H_0
Data quality	5	42.55	20.52	reject
Compatibility	8	123.16	26.12	reject
Cost	5	37.12	20.52	reject
Simplicity	4	34.12	18.46	reject
Efficiency	7	67.59	24.32	reject
Versatility	9	170.53	27.88	reject
Responsiveness	6	48.30	22.46	reject
Observability	10	90.44	29.59	reject
Relative advantage	15	79.15	37.70	reject

benefit also has order ranking in each broad benefit category. The cases reviewed in the study cited the XML benefits through using XML built-in applications, and these applications have various specific business functions. Even in the same broad benefit category, different applications bring to users different XML benefits based on the product features and users' business requirement. This also explains that top ten benefits listed in the benefit analyses part are the most popular XML benefits, but not exist in each case study.

Industry Segment Analysis

The samples in the survey represent 19 industries. Since different industries have various specific business requirements for XML, it is difficult to compare the benefit of XML features in the applications across the industries. Same benefits may

not be applied for all industries. χ^2 one-sample test is used here for hypothesis testing.

A frequency tabulation of benefits across the nineteen industries identified 555 separate benefits. Appendix F illustrates these benefits. According to the definition of broad benefit category, each benefit counts are summed up as the frequency counts for each category, data is summarized in Table 5. χ^2 one-sample test method is applied and the result is in Table 6.

The degrees of freedom is $df = k - 1 = 8$. From the table of the Chi-Square Distribution, as $\chi^2 = 46.43$, the probability of occurrence is $p < 0.001$. Since p is less than $\alpha = 0.001$, the Null hypothesis H_0 that each broad benefit category is

Table 6 χ^2 Test Summary On Industries

Category	O_i	E_i	$(O_i - E_i)$	$(O_i-E_i)^2$	$(O_i-E_i)^2/E_i$
Data quality	47	61.67	-14.67	215.11	3.49
Compatibility	33	61.67	-28.67	821.78	13.32
Cost	55	61.67	-6.67	44.44	0.72
Simplicity	43	61.67	-18.67	348.45	5.65
Efficiency	83	61.67	22.67	513.78	7.38
Versatility	77	61.67	16.67	277.78	3.81
Responsiveness	57	61.67	-4.67	21.78	0.35
Observability	75	61.67	14.67	215.11	2.88
Relative advantage	85	61.67	24.67	608.45	8.83
Sum:	**555**	**555**	**0**	**3066.68**	**46.43**

equal for each industry is rejected at the significant level 0.001. This result is consistent with the assumption that not every benefit is applied in each industry. Some benefits are more common with high frequency counts in the study, whereas others have much lower frequency counts than expected.

However, based on the benefits count, it is found that several of the common benefits are shared across a number of industries (see table 7). From table 7, there

Table 5 Data Summary Of Industries For Broad Benefit Category

Data Quality 47
accurate data 9
data control 8
re-usability 8
customized data format 11
data recoverability 1
reliability 10

Compatibility 33
standard access 2
open standards 4
platform neutrality 3
standard exchange 3
compatible with the existing technology 1
device-specific usability 1
platform independence 3
message protocol independence 1
easy to integrate to the existing system 15

Cost 55
cost effective 7
reduced development cost 4
reduced operating cost 14
cost savings 14
reduced labor cost 8
reduced maintenance cost 8

Simplicity 43
easy of use 14
easy to learn 4
reduced complexity 2
easy maintenance 13
content easily updated 10

Efficiency 83
increased productivity 16
easily tracking transaction 10
increased equipment uptime 3
minimized error 6
efficient content management 10
data sharing 13
high performance 10
speed development 15

Versatility 77
data flexibility 16
flexible structure 7
multimedia functionality 2
multiple format 2
security 7
rich information 2
data integration from distributed sources 15
extensible content 4
multiple functionality 3
scalability 14
stability 5

Responsiveness 57
easy access 11
speed process 15
simplifying process 3
quick order taking process 5
fast data retrieval 10
fast transaction and fulfillment cycles 2
quick response 11

Observability 75
dynamic information 9
improved customer service 13
powerful hyperlink capability 1
powerful searching capability 9
real-time information 7
improved communication 1
increased customer base 4
powerful navigation capability 7
improved operation efficiency 13
time saving 11

Relative advantage 85
Consistency 5
dynamic development 3
guaranteed connection 1
dynamic connectivity 3
intelligent applications 4
transaction integration 3
streamline business process 12
fast product delivery 9
data availability 9
high quality 5
effective customer services 3
high customer satisfaction 5
high data management capability 2
dynamic control of the content 8
workflow automation 10
flexible to accommodate changes 3

Table 7 Common Benefits Among Industries

BENEFITS	# of Industries
Cost savings	14
Data flexibility	16
Data integration from distributed sources	15
Easy of use	14
Easy to integrate to the existing systems	15
Increased productivity	16
Scalability	14
Speed development	15
Speed process	15

are nine individual benefits used in at least fourteen industries, which takes up 74% of nineteen industries in the study. It indicates that these benefits are generic among the industries. Listings in Table1 and Table 7 show many similarities. This indicates that user's views of XML benefits can represent the benefits in the industry where it is.

The economic survival of businesses in every industry is directly affected by their ability to quickly deploy and grow a transaction-based corporate business infrastructure (Microstar, 2000). One of the key engines of success is flexible, responsive and standards-based processing systems.

XML, as a flexible content modeling language, coupled with powerful software development tools can reduce implementation and support efforts to meet the demands of customers having different customs, languages, and procedures. With the help of XML, the living transactions can be dynamically constructed from data in relational databases and can include descriptive text and graphics from document management systems.

Moreover, XML addresses the need of a flexible standard that new development tools can reduce the costs associated with creation of new and

modified transactions and with the distribution of supporting applications to end users. Thus XML can minimize development lead times and support costs, and realize economic benefits of transaction-based business services. In the survey, these benefits are expressed in the Table 7 as the common benefits among the industries. It shows that XML is applicable and beneficial in a variety of industries, and the only difference exists in the extent to which these benefits are utilized.

In this study, it is also found that several industries have their unique XML benefits, which confirms the statistic analysis result that not all benefits are equally popular in the case studies. For example, the publishing industry draws benefit from XML's use capability to represent multiple formats and reduced complexity. In publishing industry, challenge is to produce electronic-based information products and offer these in timely and useful ways, which means to serve information from multiple sources in many different formats dynamically to the web. For example, in the case 'PC World Online' from Arbortext, the user mentioned that XML enabled automatic conversion of documents from other formats as greatly reducing the time spent doing manual conversions, and gave users the additional capability of easily offering a variety of contexts for the web. The detailed information for other industries can be found in the Table 8.

From above analysis, it is obvious that industry has its specific XML benefits which is closely related to the characteristics of industry. But they also have some common benefits from using XML built-in applications. As shown in Appendix A, many groups have been working on XML industry specifications and trying to set

Table 8 Unique Benefits In The Industry

INDUSTRY	BENEFITS
Healthcare	data recoverability
Retail	dynamic development
Telecommunication	fast transaction and fulfillment cycles
	rich information
Professional Services	improved communication
Finance and Banking	message protocol independence
Publishing	multiple format
	reduced complexity
Insurance	high data management capability

XML standard for the industry applications. This implies that future XML products will not only provide unique features for the specific industry, but also offer the generic features that can help users easily transit among industries or serve customers from other industries better. There is a large potential XML market for the software vendors if they can differentiate XML product to meet industry needs and build common XML 'communication channel' among industries.

CHAPTER 4 DISCUSSION AND CONCLUSION

Findings

Firstly, one hundred fifty-two cases revealed that seventy-eight XML benefits identified are not common to the firms who have adopted XML applications. For example, some firms claim benefit of easy of use from textual information publishing and messaging, whereas others indicate benefits of flexibility from data integration on data processing.

Secondly, the benefits described in the cases are difficult to measure, such as "improved customer services". Authors did not provide the detail information that they used to conclude the benefits. They may be subjective, instead of objective, in the evaluation of XML benefits. This may not be conceivable to the firms who are considering using XML to replace the existing systems.

Thirdly, benefits vary in the industries based on the industry characteristics and functions, as is evident, for example, when comparing Publishing industry with Telecommunication industry. Because the Publishing industry has to deal with huge amount of textual documents, it prefers multiple formats and reduced complexity in editing to speed to the market. Telecommunication industry, however, focus on online transaction and billing system, it shows rich information and fast transaction and fulfillment cycles as the major benefits in applying XML. The difference across the industries can be found in the part of Industry Segment Analysis in the survey study.

Limitations

A couple of limitations were identified in the review of case studies and benefit counts:

1. Cases in the study were written by the vendors on interviewing or working with their users. It is inevitably that the articles may represent vendor's perception of XML's benefits and overstate its XML product benefits. This can cause benefit counts are overestimated and bring inaccuracy to the study result.

2. There is no comprehensive standard definition of XML broad benefit categories can be used as a reliable reference. It is difficult to evaluate the correctness of definitions and classification of individual benefits.

3. Some of cases are general description of the XML benefits without the detail context, it may lose the detail information of XML impact on the business.

4. Because there is no standardized measurement for the XML benefit, the same word describing XML benefit may have different meanings among vendors' cases.

In the case reviews, it is also noted that sometimes cases from the same vendor cited the same benefits. For example, four of five cases from webMthods identified "easy to integrate the existing systems" and the same 80% cases from Siebel shows the benefit of "quick response". It implies that vendor bias may exist in the case writing.

To address these issues in the study, one solution is to increase sample companies from various industries and vendors, not only restricted to the vendors who have case studies on the web. Another is to have users directly involved in the study; thus, the benefit counts will be more accurate in representing users' views. The standards of measurement can be built to evaluate the different cases based on the same criteria. This may provide the stronger evidence for segment analysis.

Future Research

Several issues related to XML have not been fully discussed in this study. To further explore the potential of XML and examine the impact on the current technologies, some researches are suggested as the followings.

Specifications

Currently, there are over one hundred specifications (see Appendix A) having been developed by different working groups. During the study, a continuous flow of new specifications, based on XML v1.0 Specification, have been sent to the W3C for review. There may be more than one specification in the same industry sector. It is necessary to find the uniqueness of these specifications and possible impacts they may have on the future usage. It is extremely important for the firms to assess the available specification before they move the traditional models to XML applications.

Vendors' XML Efforts

The vendors' efforts have only just begun. The current lack of tools holds back widespread usage of XML. More and more tools are becoming available and it is important to continue and monitor what tools are developed. The research to examine the tools and provide recommendations for specific business area will be much helpful to the companies, who are looking for the correct tools to meet their need. The especially interesting will be to follow what Netscape and Microsoft do with their Internet browsers and what support the database suppliers will present for XML.

Summary

XML technology holds much promise for the future. It is an industry-wide recognized language for representing semi-structured data that could be shared intra- and inter-enterprise. XML is useful for much more than Web pages. Its potential as a universal data transfer format, allowing even applications of different types to exchange data smoothly, holds as much promise as its role as a document management system. Early users focused on its role as specialized browsers. Although allowing users to read an XML document, browsers are only the beginning. XML provides a gateway for communication between applications, which may be distributed across a wide variety of systems. As long as applications can share data through HTTP, file sharing, or any other mechanism, and have an XML parser, they can share structured information that is easily processed. Databases can trade tables, business applications can trade updates, and document systems can share information.

The data from the survey shows that many users have realized improved productivity and saved cost. XML provides a bridge between the publishing, database and web presentation in a way that has not occurred before. It enables the enterprise to link several technology domains that previously spoke different languages to exchange and share data. It is highly likely that XML will have a significant impact on electronic commerce.

APPENDIX A: CATEGORY OF XML SPECIFICATIONS

Business Functions	Organization	XML Specification
Accounting	American Institute of Certified Public Accountancts (ACICPA)	Extensible Financial Reporting markup Language (XFRML)
Advertising	adXML.org	Online Insertion Order
	Newspaper Association of America (NAA)	NAA classified Advertising Satandards Task Force
Architecture and Construction	ConSource.com	Construction Manufacturing and Distribution Extensible Markup Language (cmdXML)
Astronomy and Space	Interface & Control Systems Inc	Spacecraft markup Language (SML)
	NASA	Astronomical Instrument Markup Language (AIML)
		Astronomical Tbles, Images and Spectra datasets
	NASA - The Astronomical Data Center	eXtensible Data Format (XDF)
Automotive	Global Automedia	VehicleExport
	MSR	Standards for information exchange in the engineering process (MEDOC)
	The Society of Automotive Engineers (SAE)	XML for the Automotive Industry -SAE J2008
Banking	Banking Industry Technology Secretariat (BITS)	Interactive Financial Exchange (IFX)
	Financial Services Technology Consortium (FSTC)	Bank Internet Payment System (BIPS)
Bibliographies	Machine Readable Cataloging	MARC
	Ministere de la culture et de la communication, France	BiblioML
Catalogs	MartSoft	Open Catalog Formate (OCF)
	Requisite Technology, Inc	Electronic Catalog XML (eCX)
Communication	Alliance for Telecommunications Industry Solutions (AITS)	Telecommunications Interchange Markup (TIM)
	Wireless Application Protocol Forum (WAP)	Wireless Markup Language (WML)
Computer Graphics	web3D Consortium	Extensible 3D
	World Wide Web Consortium (W3C)	Scalable Vector Graphics (SVG)
Content Syndication	Internet Alchemy	Open Content Syndication (OCS)
	Vignette, et al	The Information and Content Exchange Protocol (ICE)
CRM - Customer Relationship Management	Cognito, Inc	Customer Identity markup Language (CIML)
		Name and Address markup Language (NAML)
Data Mining	The Data Mining Group (DMG.org)	Predictive Model Markup Language (PMML)
Directory services	The DSML Initative	Directory Services Markup Language
	Novell	DirXML
Distributed Management	Distributed Management Task Force, Inc. (DMTF)	Common Information Model (CI M)
Economics	EcoKnowMICS Corp	EcoKnowMICSML
Education	Educom IMS Project	IMS Meta-data Specification
electronic Commerce	CommerceNet	eCo Framework
	Commerce One	Common Business Library (CBL)
	CXML.org	Commerce XML (cXML)
	Electronic Business XML Initiative	ebXML
	IBM	Business Rules Markup Language (BRML)
		Trading Partner Agreement Markup Language (tpaML)
	Joint Electronic Commerce Program Office (JECPO)	Product Data Markup Language (PDML)
	Open Trading Protocol Group (OTP)	Open Trading Protocol (OTP)
	Since Applications International Corporation (SAIC)	Universal Commerce Language and Protocol (UCLP)
	Visa International (VISA)	Visa XML Invoice Specification
EDI - Electronic Data Interchange	data Interchange standards Association (DISA)	ANSI ASC X12/XML

	European Committee for standardization/Information society standardization System (CEN/ISSS)	The European XML/EDI Pilot project
	XEDI.org	XEDI
Energy Enterprise Information	Petrotechnical Open Software Corporation (POSC)	WellLogML
Portals	DataChannel	Portal Markup Language (PML)
ERP - Enterprise resource Planning	Open Application Group (OAG)	Open Applications Group Interface Specification
Financial and Capital Markets	Digital Receipt Alliance	Annotated Digital Receipt
	Financial Information eXchange Protocol (FIX)	FIXML
	FinXML.org	FinXML
	FpML.org	Financial Products Markup Language (FpML)
	Infinity	Network Trade Model (NTM)
	Soliton Associates	Historical Data Markup Language (HDXML)
Food	FormatData	document Encoding and structuring Specification for Electronc Recipe Transfer (DESSERT)
Forms	JetForm Corporation	XML Forms Architecture (XFA)
	UWI.com	Extensible forms description Language (XFDL)
Geography	OpenGIS Consortium Inc	Geography Markup Language (GML)
Healthcare	Phase Forward	Clinical Trial Data Model
Human Resource	DataMain	Human Resources Markup Language (hrml)
	HR-XML Consortium	JobPosting, CandidateProfile, Resume
	Open Application Group (OAG)	Open Applications Group Interface Specification
	Siemens Business Communication Systems	Siemens Time and Attendance System
	Tapestry.Net	JOB Markup Language (JOB)
Industrial Automation	Nacimiento Software Corporation	Virtual Instruments Meta Language (VIML)
Insurace	ACORD	Property and Casualty
		XMLife
	Lexica	iLingo
Legal	U.S. District Court, District of New Mexico	XML Court Interface (XCI)
	Legal XML Working Group	LegalXML
	Xeorx Palo Alto Research Center	Digital Property Rights Language (DPRL)
Middleware	IBM	Unified Customer Reporting Performance Data
Music	Bert Schiettecatte	FlowML
	CifraNet	ChordML
	The Connection Factory	MusicML
	Nacimiento Software Corporation	Virtual Instruments Markup Language
	University of Pretoria	Music Markup Language
News	International Press Telecommunications Council (IPTC)	news Industry Text Format (NITF)
		NewsML
	XMLNews.org	XMLNews-Story
		XMLNews-Meta
Publishing	Open eBook Initiativ	Open eBook Publication structure
	Openly Informatics, Inc	eFirst XML for Scholarly Articles
	Organization for the Advancement of structured Information standards (OASIS)	Exchange subset of the CALS table model
		DocBook
	The Text Encoding Initiative (TEI)	TEI
	VHG.org	Virtual Hyperglossary (VHG)
Real Estate	OpenMLS	Real estate Listing Management System (OpenMLS)
	Real Estate Transaction Standard Working Group (RETS)	Real Estate Transaction standard (RETS)
Retail	First Retail	First Retail Mark-up Language (FRML)

Science	Caltech CACR	Extensible Scientific Interchange Language (XSIL)
	MoDL Project Team	Molecular Dynamics Markup Language (MoDL)
	The OpenMath Society	OpenMath
	Proteometrics	BIOpolymer Markup Language (BIOML)
	Rosetta Inpharmatics	Gene Expression Markup Language (GEML)
	Peter Murray Rust	Chemical Markup Language (CML)
	Visual Genomics	Bioinformatic sequence markup Language (BSML)
	World Wide Web Consortium (W3C)	Mathematical markup Lanugae (MathML)
Software	IBM	Bean Markup Language (BML)
	Flashline.com	Software Component Documentation DTD
		JavaDox
	INRIA	Koala Bean Markup Language (KBML)
	Obejct Management Group (OMG)	XML Metadata Interchange (XMI)
Supply Chain Management	RosettaNet	RosettaNet
Translation	INT'L.com	OpenTag
	Localisation Industry Standards Association (LISA)	Translation Memory eXchange (TMX)
User Interface	Mozilla.org	Extensible User Interface Language (XUL)
	LANSA.Inc	LANSA XML
	UIML.org	User Interface markup Language (UIML)
Voice	Motorola	VoxML
	Sun Microsystems	Java Speech Markup Language (JSML)
	VoiceXML Forum	voiceXML
Weather	MetNet	Weather Observation Definition Format (ODF)
Web Application	Allaire	Cold Fusion markup Language (CFML)
	DevelopMentor, Microsoft, UserLand Software	Simple Object Access Protocol (SOAP)
	Extensible Log Format Project	Extensible Log Format (XLF)
	Hiroshima University Information System Laboratory	Structured Graph Format (SGF)
	Internet Engineering Task Force (IETF)	World Wide Web distributed Authoring and Versioning (WebDAV)
	Microsoft	Channel Definition Format (CDF)
	Python XML SIG	XML Bookmark Exchange Language (XBEL)
	UserLand Software, Inc	XML-RPC
	WDDX.org	Web Distributed Data eXchange (WDDX)
Workflow	Internet Engineering Task Force (IETF)	Simple Workflow Access Protocol (SWAP)
	Workflow management Coalition (WfMC)	Wf-XML

APPENDIX B: LIST OF COMPANIES IN THE STUDY

ActiveBuyer'sGuide.com
Acxiom Corporation
Alamos national laboratory (LANL)
AlintaGas
Alliant Energy
Altantic Mutual Insurance Company
Auracom
Autobytel.com
Automatic Data Processing Inc.
Bank One Corporation
BaseOne Comuters
Bay Networks
Bayer Corporation
Beazley
Beckman Coulter
Bell South
Bell sygma
Bertramooks Ltd
BF Goodrich Aerospace
BigStep.com
BMC Software
Boots Healthcare
Boston Edison
Boundless Corporation
Bridge Information Systems
British Telcom
Brown & Root
Cable & Wireless Communications
Canadian Coast Guard
CARE Canada
Caroline Power & Light Company
Casio, Inc
CB Richard Ellis
Charles Schwad Company
Chicago Blackhawks
CIENA Corporation
Clearnet Communications
Compaq Computer Corporation
Comtrad Industried
Corporate Express
Cotton Incorporated
Critical Path, Inc
Cytec Industries, Inc
Daimler Chrysler
Dallas Gold & Siver Exchange
Deakin Australia
Dell Computer Corporation
Dun & Bradstreet
e-Business eXchange
EarthLink
eHatchery LLC
Epoch Software
Ernerald First Financial, Ltd
eToys
Eurostar
FAA
First Union
FMC Corporation
Ford Motor Company
GE Aircraft
Gentofte Armtssygehus
George Washington University
Giaxo Wellcom, Inc
Grainger
GTE
GTECH
Guinness
Homestead Technologies, Inc
Illinois Power
Informatic Kooperation
InsideDenver
International Rectifier
InterTrust Technologies Corporation
InterVideo, Inc
InterZone Production, Inc
Intuit, Inc
ISO Central Secretarial
J.P.Morgan
Johnson & Johnson
Kvaerner process
LANSA
Linklaters & Alliance
Lockheed Martin
LogicTier's, Inc
London Underground
MatchLogic, Inc
Mature Tymes
Moen, Inc
Molson Breweries
Morgan & bank
Motorola
Murphy Brewery
Nasdaq Stock Market
National Business Review
National Discount Brokers
Navant Corporation
NaviPath, Inc
NAVSEA
NC.Focus
NetClerk, Inc

NetLedger
Newcourt Financial
Nortel Networks
Novartis
OSRAM Gmbh
PayMaxx, Inc
PC World Online
People Soft
PHH Behicle Management Service
Phone-Poulenc Rorer
Pictet & CIE
PointClick.com
Power Technology
PQMS, Inc
Pratt & Whitney
Province of Manitoba
renren.com
Reuters
Rolls-Royce
SABRE
Sanofi Synthelado
SAP
Scudder Threadneedle Investment services Limited
Seagate Technology, Inc
Seals Gmbh
SegaSoft Networks heat.Net
Severn Trent Systems
Siemens
Sprint PCS
Staff Leasing
Sun Microsystems, Inc
SupplierMarket.com
Swiss Army Brands, Inc
SwissAir
Synergistic Media Network, Inc
Telecommunications Managers Association (TMA)
Texas Department of Public Safety
The Home Depot
TransLink
United Center
University of Nottingham & Imperial College
USA.NET, Inc
Versant Corporation
Vervet Logic
VICTORIA-Group
Warranty Net
Wells Fargo
Wickes Inc.
Winstar Communications, Inc
Worldweb.net
XML Solutions

APPENDIX C: WEB LIST OF CASE STUDIES

http://www.allaire.com/casestudies/index.cfm

http://www.arbortext.com/news_and_events/recent_news_releases/case_study

http://www.documentum.com/business/customer/index.html

http://www.engima.com

http://www.ibm.com/developer/xml

http://www.opentext.com/customers/index.html (Microstar)

http://www.marimba.com/solutions/customer.htm

http://success.oracle.com/succst-wwwprd-dcd/plsql/query_success_frame.temp_result

http://www.siebel.com/uk/success/

http://www.webMethods.com/customers/0,1308,00.html

APPENDIX D: EXCERPTS FROM A CASE STUDY

Client: PC World magazine
Vendors: Arbortext
Title: PC World Online: The Fast Track to Dynamic Content Delivery
Web Site: http://www.arbortext.com/Think_Tank/XML_Resources/PC_World/pc_world.htm
Excerpts:

XML Data Delivery

Once the articles are in XML, ***content delivery becomes a much easier, more effective task***. An XML-aware database system stores the XML intact, unlike the template and database system, which can only store lightly coded HTML data in simpler "fields." One advantage of intact XML storage is that ***data retrieval is much more powerful***. Due to the hierarchical structure of XML data, it's just as easy to retrieve a set of articles with the tags intact as it is to retrieve one headline.

Publish to and from multiple formats

XML has enabled ***automatic conversion of documents from other formats, greatly reducing the time spent doing manual conversions***. Converting content into the forms required by licensees was as time-consuming as producing the articles in the first place. Now the same system that produces the articles uses templates and translation tables to produce the licensee content in whatever format is needed. This has given PC World the additional capability of easily offering a variety of contexts, even for the web. For example, a simplified printable web page format, or a version for an affiliated web site, is simply another translation of XML.

Reduced publication time

Rebecca Freed, Managing Editor, reported that ***their XML system has cut the time to go from print to web from two days to two hours***. Additionally, the group estimates that they can publish the old articles in about 10 percent of the time, which they have reinvested in creating higher-quality articles.

Additionally, editors now enter copy directly into the XML system, while server-side programs write out the actual HTML pages. Once the article is created, it is ready to go to the web with little or no manual tagging or reformatting required on the part of the site producers. This is a ***huge savings in time*** that used to be spent updating index pages and pouring text into templates by hand. Editors now handle the entire correction process, **saving time and minimizing additional rounds of corrections**. Granting this direct access to editors has streamlined production 100 percent.

Reduced cost/improved quality

By simplifying the process of creating locating and delivering information, the PC World Online team has ***increased productivity while reducing costs***. The editors and production staff are constantly seeking ways to enrich how their content can be accessed, navigated, presented and used. XML has been the key to fulfilling this goal.

APPENDIX E: OVERAL BENEFIT COUNTS

Benefit Description	Count		Benefit Description	Count
accurate data	29		open standards	4
compatible with the existing technology	1		platform independence	3
content easily updated	22		platform neutrality	4
consistency	7		powerful hyerlinking capability	1
cost effective	12		powerful navigation capability	7
cost savings	35		powerful searching capability	11
customized data format	26		quick order taking process	5
data availability	11		quick response	27
data control	9		re-usability	14
data flexibility	40		real-time information	15
data integration from distributed sources	35		reduced complexity	2
data recoverability	1		reduced development cost	6
data sharing	27		reduced labor cost	12
device-specific usability	1		reduced maintenance cost	10
dynamic connectivity	3		reduced operating cost	26
dynamic control of the content	12		reliability	16
dynamic development	4		rich information	3
dynamic information	16		scalability	38
easily tracking transaction	13		security	13
easy access	13		simplifying process	4
easy maintenance	21		speed development	36
easy of use	30		speed process	26
easy to integrate the existing system	30		stability	7
easy to learn	6		standard exchange	3
effective customer service	4		standardized access	3
efficient content management	22		streamline business process	16
extensible content	5		time saving	21
fast data retrieval	15		transaction integration	3
fast product delivery	21		workflow automation	22
fast transaction and fulfillment cycles	2		**Total**	**1017**
flexible structure	7			
flexible to accommodate change	3		**Total Users**	**152**
guaranteed connection	2		**Total Vendors**	**10**
high customer satisfaction	8		**Total Industries**	**19**
high data management capability	2		**Total Countries**	**13**
high performance	15			
high quality	10			
improved communication	1			
improved customer service	36			
improved operation efficiency	31			
increased customer base	5			
increased equipment uptime	3			
increased productivity	44			
intelligent application	4			
message protocol independence	1			
minimized error	7			
multimedia functionality	2			
multiple format	2			
multiple functionality	3			

APPENDIX F: VENDOR SOURCE BENEFIT COUNTS

Allaire	**29 users**
Benefit Description	**Count**
accurate data	1
content easily updated	10
consistency	2
cost savings	10
customized data format	6
data availability	4
data control	2
data flexibility	10
data integration from distributed sources	3
data sharing	6
dynamic connectivity	2
dynamic control of the content	5
dynamic development	3
dynamic information	11
easily tracking transaction	6
easy maintenance	13
easy of use	13
easy to integrate the existing system	7
easy to learn	5
effective customer service	3
efficient content management	2
extensible content	1
fast data retrieval	5
fast product delivery	1
flexible structure	3
high data management capability	2
high performance	2
high quality	2
improved communication	1
improved customer service	6
improved operation efficiency	4
increased productivity	9
multimedia functionality	1
multiple functionality	2
open standards	4
platform neutrality	1
powerful navigation capability	2
powerful searching capability	5
quick response	6
re-usability	2
reduced complexity	1
reduced development cost	2
reduced labor cost	1
reduced maintenance cost	1
reduced operating cost	1
reliability	1
rich information	1
scalability	8
security	4
speed development	18
speed process	6
stability	3
streamline business process	1
time saving	2
transaction integration	2
workflow automation	4

Arbortext	**1 user**
Benefit Description	**Count**
flexible structure	1
high quality	1
multiple format	1
re-usability	1
reduced labor cost	1
reduced maintenance cost	1
reduced operating cost	1
speed process	1

Documentum	**20 users**
Benefit Description	**Count**
accurate data	10
content easily updated	4
consistency	3
cost effective	1
cost savings	6
customized data format	2
data availability	2
data control	1
data flexibility	10
data integration from distributed sources	9
data sharing	9
dynamic control of the content	3
dynamic development	1
easily tracking transaction	3
easy access	5
easy of use	4
easy to integrate the existing system	5
easy to learn	1
efficient content management	4
fast data retrieval	4
fast product delivery	7
high performance	2
high quality	2
improved customer service	3
improved operation efficiency	7
increased productivity	5
minimized error	1
powerful navigation capability	2

powerful searching capability	4		dynamic information	2
quick order taking process	1		easy access	1
quick response	4		easy maintenance	1
re-usability	7		easy to integrate the existing system	7
real-time information	6		efficient content management	4
reduced labor cost	1		extensible content	1
reduced maintenance cost	3		fast data retrieval	1
reduced operating cost	4		fast product delivery	2
reliability	2		flexible to accommodate change	2
scalability	6		guaranteed connection	1
simplifying process	1		high customer satisfaction	1
speed development	7		high performance	3
speed process	3		high quality	1
stability	1		improved customer service	5
streamline business process	4		improved operation efficiency	9
time saving	6		increased customer base	2
workflow automation	6		increased productivity	3
			message protocol independence	1
engima	**7 users**		minimized error	2
Benefit Description	**Count**		multimedia functionality	1
data integration from distributed sources	3		platform independence	2
dynamic information	1		platform neutrality	3
effective customer service	1		quick response	1
extensible content	1		reduced development cost	3
fast data retrieval	1		reduced labor cost	5
high customer satisfaction	1		reduced maintenance cost	1
improved customer service	1		reduced operating cost	9
improved operation efficiency	2		reliability	1
increased equipment uptime	3		scalability	3
increased productivity	2		security	5
intelligent application	4		speed process	1
powerful hyerlinking capability	1		standard exchange	3
powerful navigation capability	2		standardized access	1
powerful searching capability	1		streamline business process	1
re-usability	1		time saving	1
reduced labor cost	2			
reduced maintenance cost	1		**Marimba**	**29 users**
reduced operating cost	1		**Benefit Description**	**Count**
streamline business process	1		accurate data	2
			content easily updated	3
IBM	**21 users**		cost effective	3
Benefit Description	**Count**		cost savings	5
accurate data	2		customized data format	3
content easily updated	2		data flexibility	6
customized data format	4		data integration from distributed sources	2
data control	1		dynamic control of the content	1
data flexibility	2		dynamic information	1
data integration from distributed sources	8		easy access	1
data sharing	3		easy maintenance	2
device-specific usability	1		easy of use	2
dynamic connectivity	1		efficient content management	3
dynamic control of the content	3		fast product delivery	5

flexible structure	1		reduced operating cost	4
high performance	2		reliability	1
high quality	1		rich information	1
improved customer service	4		scalability	3
improved operation efficiency	2		speed development	2
increased customer base	1		speed process	2
increased productivity	2		streamline business process	3
powerful searching capability	1		time saving	5
quick response	5		workflow automation	2
real-time information	3			
reduced development cost	1		**Oracle**	**26 users**
reduced labor cost	1		**Benefit Description**	**Count**
reduced maintenance cost	2		accurate data	6
reduced operating cost	2		compatible with the existing technology	1
reliability	3		content easily updated	2
scalability	5		consistency	1
simplifying process	3		cost effective	5
speed development	1		cost savings	6
speed process	4		customized data format	3
stability	1		data availability	4
standardized access	1		data control	1
time saving	1		data flexibility	4
workflow automation	4		data integration from distributed sources	3
			data recoverability	1
Microstar	**19 users**		data sharing	1
Benefit Description	**Count**		dynamic information	1
accurate data	5		easily tracking transaction	3
content easily updated	1		easy access	5
cost effective	2		easy maintenance	5
cost savings	7		easy of use	5
customized data format	5		easy to integrate the existing system	4
data availability	1		efficient content management	7
data control	4		extensible content	1
data flexibility	6		fast data retrieval	1
data integration from distributed sources	5		fast product delivery	2
data sharing	8		flexible structure	2
easily tracking transaction	1		high customer satisfaction	1
easy access	1		high performance	5
easy of use	5		improved customer service	8
easy to integrate the existing system	2		improved operation efficiency	5
efficient content management	2		increased productivity	10
extensible content	1		minimized error	1
fast product delivery	4		multiple format	1
high customer satisfaction	4		multiple functionality	1
high quality	2		powerful navigation capability	1
improved customer service	6		quick order taking process	3
improved operation efficiency	2		quick response	2
increased productivity	9		re-usability	3
minimized error	2		reduced complexity	1
platform independence	1		reduced maintenance cost	1
quick response	5		reduced operating cost	3
real-time information	5		reliability	6

scalability	13
security	4
speed development	8
speed process	7
stability	2
standardized access	1
streamline business process	3
time saving	4
workflow automation	5
Siebel	**5 users**
Benefit Description	**Count**
accurate data	1
consistency	1
cost effective	1
cost savings	1
customized data format	1
data integration from distributed sources	1
easy of use	1
easy to integrate the existing system	1
fast transaction and fulfillment cycles	1
flexible to accommodate change	1
high customer satisfaction	1
high performance	1
high quality	1
improved customer service	3
increased customer base	2
increased productivity	3
quick response	4
real-time information	1
reduced labor cost	1
rich information	1
speed process	1
streamline business process	3
time saving	2
workflow automation	1
webMethods	**5 users**
Benefit Description	**Count**
accurate data	2
customized data format	2
data flexibility	2
data integration from distributed sources	1
easy to integrate the existing system	4
fast data retrieval	3
fast transaction and fulfillment cycles	1
guaranteed connection	1
increased productivity	1
minimized error	1
quick order taking process	1
reduced operating cost	1
reliability	2
transaction integration	1
Total	**1017**

APPENDIX G: INDUSTRY BENEFIT COUNTS

Aerospace	**3 users**
Benefit Description	**Count**
data control	1
data flexibility	1
data sharing	1
dynamic information	1
easy of use	1
extensible content	1
fast data retrieval	1
high customer satisfaction	1
increased equipment uptime	1
increased productivity	1
intelligent application	1
reduced labor cost	1
reduced maintenance cost	1
speed process	1
time saving	1
workflow automation	1
Computer & Software Services	**33 users**
Benefit Description	**Count**
accurate data	4
content easily updated	6
consistency	1
cost effective	3
cost savings	4
customized data format	5
data availability	1
data control	2
data flexibility	8
data integration from distributed sources	5
data sharing	2
dynamic control of the content	4
dynamic information	2
easily tracking transaction	2
easy access	2
easy maintenance	3
easy of use	5
easy to integrate the existing system	5
easy to learn	1
effective customer service	2
efficient content management	5
fast product delivery	4
flexible structure	1
high customer satisfaction	3
high performance	2
high quality	3
improved customer service	8
improved operation efficiency	7
increased customer base	2
increased productivity	8
intelligent application	1
minimized error	1
multimedia functionality	1
platform neutrality	1
powerful navigation capability	1
powerful searching capability	1
quick response	8
re-usability	1
real-time information	4
reduced development cost	3
reduced labor cost	3
reduced operating cost	5
reliability	4
scalability	14
security	3
simplifying process	3
speed development	5
speed process	4
standard exchange	1
standardized access	1
streamline business process	2
time saving	3
workflow automation	4
Education	**3 users**
Benefit Description	**Count**
content easily updated	1
customized data format	1
data control	1
data flexibility	1
data integration from distributed sources	1
easy maintenance	2
easy of use	1
easy to integrate the existing system	1
effective customer service	1
extensible content	1
fast data retrieval	1
improved operation efficiency	2
open standards	1
quick response	1
reduced development cost	1
reduced maintenance cost	1
reduced operating cost	1
reliability	1
scalability	1
speed development	2
time saving	1
Finance & Banking	**18 users**

Benefit Description	Count
accurate data	5
compatible with the existing technology	1
content easily updated	2
consistency	2
cost effective	3
cost savings	5
customized data format	4
data availability	1
data control	1
data flexibility	5
data integration from distributed sources	5
data sharing	3
dynamic control of the content	1
dynamic information	1
easily tracking transaction	2
easy access	1
easy maintenance	2
easy of use	4
easy to integrate the existing system	2
easy to learn	1
effective customer service	1
efficient content management	2
extensible content	2
fast data retrieval	1
fast product delivery	3
high performance	4
improved customer service	6
improved operation efficiency	3
increased productivity	2
message protocol independence	1
multiple functionality	1
open standards	1
platform independence	1
platform neutrality	2
powerful searching capability	1
quick order taking process	1
quick response	4
re-usability	1
real-time information	1
reduced development cost	1
reduced maintenance cost	1
reduced operating cost	2
reliability	2
scalability	4
security	3
speed development	5
speed process	4
standard exchange	1
standardized access	2
streamline business process	1
time saving	2
transaction integration	1
workflow automation	3
Government	**5 users**
Benefit Description	**Count**
consistency	1
cost savings	2
customized data format	1
data control	1
data flexibility	2
data integration from distributed sources	1
data sharing	3
easy maintenance	2
easy of use	2
easy to integrate the existing system	2
efficient content management	3
improved customer service	2
improved operation efficiency	2
increased productivity	2
powerful navigation capability	1
powerful searching capability	1
quick response	2
reduced labor cost	1
reduced operating cost	1
security	1
speed development	2
speed process	1
stability	1
time saving	1
Healthcare	**5 users**
Benefit Description	**Count**
accurate data	2
cost savings	1
data availability	1
data flexibility	1
data integration from distributed sources	1
data recoverability	1
data sharing	3
dynamic control of the content	1
easily tracking transaction	1
easy access	2
easy to integrate the existing system	1
efficient content management	1
fast data retrieval	1
fast product delivery	4
high performance	1
high quality	1
improved operation efficiency	4
minimized error	1
powerful navigation capability	1
powerful searching capability	1

quick order taking process	1		Transaction integration	1
re-usability	2			
reduced operating cost	1		**Manufacturing**	**11 users**
speed development	1		**Benefit Description**	**Count**
speed process	1		accurate data	3
streamline business process	1		consistency	2
time saving	2		cost effective	1
workflow automation	1		cost savings	4
			customized data format	1
High Tech	**4 users**		data flexibility	3
Benefit Description	**Count**		data integration from distributed sources	4
content easily updated	2		data sharing	3
cost effective	1		dynamic control of the content	1
data integration from distributed sources	1		dynamic information	1
easily tracking transaction	1		easy access	1
easy access	1		easy of use	2
easy maintenance	1		easy to integrate the existing system	2
easy of use	1		fast data retrieval	3
easy to integrate the existing system	1		fast product delivery	1
efficient content management	1		high customer satisfaction	1
fast transaction and fulfillment cycles	1		high data management capability	1
flexible structure	1		high quality	2
high performance	1		improved customer service	4
increased productivity	1		improved operation efficiency	3
multiple format	1		increased customer base	1
multiple functionality	1		increased equipment uptime	1
powerful searching capability	1		increased productivity	6
reduced operating cost	1		intelligent application	1
reliability	2		powerful hyerlinking capability	1
scalability	1		powerful navigation capability	1
speed process	1		powerful searching capability	2
			quick response	3
Insurance	**3 users**		re-usability	2
Benefit Description	**Count**		real-time information	5
content easily updated	1		reduced labor cost	2
data availability	1		reduced maintenance cost	1
data sharing	1		reduced operating cost	2
data integration from distributed sources	2		reliability	2
dynamic information	1		scalability	1
easy maintenance	1		speed development	2
easy of use	1		speed process	1
fast product delivery	1		stability	2
high data management capability	1		streamline business process	2
improved customer service	1		time saving	2
increased productivity	1		workflow automation	2
platform independence	1			
quick response	1		**Media/Entertainment**	**7 users**
scalability	1		**Benefit Description**	**Count**
speed process	2		accurate data	4
Streamline business process	1		content easily updated	2
Time saving	1		cost effective	2
			cost savings	1

customized data format	1		increased productivity	4
data flexibility	2		minimized error	1
data integration from distributed sources	3		open standards	1
data sharing	2		powerful navigation capability	1
dynamic development	1		quick order taking process	1
easy access	1		quick response	1
easy maintenance	1		reduced labor cost	1
easy of use	2		reduced operating cost	2
easy to integrate the existing system	1		reliability	1
efficient content management	3		scalability	5
fast data retrieval	1		security	2
flexible structure	1		speed development	3
high performance	1		streamline business process	1
improved customer service	2		time saving	2
improved operation efficiency	2		workflow automation	3
increased productivity	3			
powerful navigation capability	1		**Publishing**	**4 users**
quick order taking process	1		**Benefit Description**	**Count**
quick response	1		content easily updated	2
re-usability	1		consistency	1
reduced maintenance cost	1		cost savings	1
reduced operating cost	1		data availability	2
reliability	1		data control	1
scalability	2		dynamic connectivity	1
speed development	3		dynamic control of the content	1
speed process	1		dynamic information	3
stability	2		easy maintenance	2
streamline business process	1		easy of use	1
workflow automation	1		fast data retrieval	1
			flexible structure	1
Professional Services	**8 users**		high quality	2
Benefit Description	**Counts**		multiple format	1
accurate data	2		powerful searching capability	1
cost effective	1		re-usability	3
cost savings	2		reduced complexity	1
customized data format	2		reduced labor cost	1
data availability	1		reduced maintenance cost	1
data flexibility	2		reduced operating cost	1
data sharing	1		speed development	2
easily tracking transaction	1		speed process	1
easy access	1		workflow automation	1
easy of use	1			
easy to integrate the existing system	3		**Real Estate**	**1 user**
efficient content management	1		**Benefit Description**	**Count**
fast data retrieval	1		accurate data	1
fast product delivery	2		cost savings	1
flexible structure	1		data flexibility	1
flexible to accommodate change	1		improved customer service	1
high performance	2		increased productivity	1
improved communication	1		reduced complexity	1
improved customer service	2		reduced operating cost	1
improved operation efficiency	2		speed process	1

workflow automation	1		easy maintenance	1
			easy to integrate the existing system	1
Retail	**10 users**		high performance	1
Benefit Description	**Count**		improved customer service	2
content easily updated	1		increased equipment uptime	1
cost savings	1		increased productivity	1
customized data format	3		reliability	1
data availability	2		scalability	1
data flexibility	1		simplifying process	1
data integration from distributed sources	2		speed development	2
dynamic connectivity	1		speed process	1
dynamic control of the content	1		streamline business process	1
dynamic development	1			
dynamic information	2		**Telecommunication**	**10 users**
easily tracking transaction	1		**Benefit Description**	**Count**
easy maintenance	2		accurate data	2
easy to integrate the existing system	4		content easily updated	1
easy to learn	3		cost savings	2
fast data retrieval	2		customized data format	4
flexible structure	1		data availability	1
guaranteed connection	2		data control	1
high performance	1		data flexibility	3
improved customer service	3		data integration from distributed sources	4
improved operation efficiency	1		device-specific usability	1
increased customer base	1		easily tracking transaction	1
increased productivity	2		easy access	1
minimized error	2		easy maintenance	1
multimedia functionality	1		easy of use	3
platform independence	1		easy to integrate the existing system	1
powerful navigation capability	1		efficient content management	2
quick order taking process	1		extensible content	1
real-time information	1		fast product delivery	4
reduced labor cost	2		fast transaction and fulfillment cycles	1
reduced operating cost	1		flexible to accommodate change	1
reliability	1		high customer satisfaction	2
scalability	1		high performance	1
security	2		high quality	2
speed development	2		improved customer service	1
speed process	3		improved operating efficiency	1
stability	1		increased customer base	1
standard exchange	1		increased productivity	2
streamline business process	1		intelligent application	1
transaction integration	1		minimized error	1
workflow automation	1		open standards	1
			platform neutrality	1
Transportation	**2 users**		powerful searching capability	1
Benefit Description	**Count**		quick response	2
cost savings	2		real-time information	1
data flexibility	1		rich information	2
data integration from distributed sources	1		scalability	1
data sharing	1		Simplifying process	1
easily tracking transaction	1		speed development	1

speed process	1		scalability	1
stability	1		security	1
Streamline business process	2		speed development	1
time saving	2		time saving	2
Travel	**4 users**		**Others**	**13 users**
Benefit Description	**Count**		**Benefit Description**	**Count**
cost savings	1		accurate data	6
customized data format	2		content easily updated	4
data flexibility	1		cost effective	1
data integration from distributed sources	1		cost savings	5
data sharing	1		customized data format	2
dynamic control of the content	1		data availability	1
easily tracking transaction	1		data control	1
easy access	1		data flexibility	4
easy to integrate the existing system	2		data integration from distributed sources	3
fast product delivery	1		data sharing	3
flexible to accommodate change	1		dynamic connectivity	1
improved customer service	1		dynamic control of the content	2
improved operating efficiency	1		dynamic development	2
Increased productivity	1		dynamic information	3
scalability	2		easily tracking transaction	2
security	1		easy access	1
speed development	1		easy maintenance	2
speed process	1		easy of use	4
streamline business process	2		easy to integrate the existing system	1
workflow automation	1		efficient content management	3
			fast data retrieval	3
Utility	**7 users**		fast product delivery	1
Benefit Description	**Count**		flexible structure	1
cost savings	3		high customer satisfaction	1
data flexibility	3		improved customer service	3
data integration from distributed sources	2		improved operation efficiency	2
data sharing	2		increased productivity	4
dynamic information	1		minimized error	1
easy access	1		multiple functionality	1
easy maintenance	1		powerful searching capability	2
easy of use	2		quick response	3
easy to integrate the existing system	2		re-usability	2
easy to learn	1		real-time information	2
efficient content management	1		reduced maintenance cost	3
high performance	1		reduced operating cost	5
improved operating efficiency	1		rich information	1
increased productivity	5		scalability	3
quick response	1		speed development	4
re-usability	2		speed process	2
real-time information	1		streamline business process	1
reduced development cost	1		time saving	2
reduced labor cost	1		workflow automation	3
reduced maintenance cost	1			
reduced operating cost	2		**Total**	**1017**
reliability	1			

APPENDIX H: COUNTRY BENEFIT COUNTS

Australia	**4 users**
Benefit Description	**Count**
cost savings	3
data control	1
data flexibility	4
data integration from distributed sources	1
data sharing	3
easily tracking transaction	1
easy maintenance	1
easy of use	1
easy to integrate the existing system	3
high performance	1
improved customer service	1
improved operation efficiency	3
increased productivity	2
minimized error	1
quick response	1
real-time information	1
reduced development cost	2
reduced labor cost	1
reduced operating cost	1
scalability	1
speed development	3
speed process	1

Belgium	**1 user**
Benefit Description	**Count**
accurate data	1
cost effective	1
customized data format	1
data availability	1
efficient content management	1
quick response	1
re-usability	1
scalability	1
workflow automation	1

Canada	**6 users**
Benefit Description	**Count**
accurate data	2
consistency	1
cost savings	1
customized data format	2
data flexibility	2
data integration from distributed sources	2
data sharing	3
easy maintenance	1
easy of use	1
easy to integrate the existing system	2
efficient content management	1
extensible content	1
fast product delivery	2
high customer satisfaction	1
improved customer service	2
improved operation efficiency	1
increased productivity	1
minimized error	1
powerful navigation capability	1
quick response	2
real-time information	1
reduced labor cost	1
reduced operating cost	1
rich information	1
speed development	1
stability	1
workflow automation	1

China	**2 users**
Benefit Description	**Count**
accurate data	1
fast data retrieval	1
flexible to accommodate change	1
high performance	1
re-usability	1
reduced labor cost	1
reduced operating cost	1
reliability	1
scalability	1

Denmark	**1 user**
Benefit Description	**Count**
data recoverability	1
easy to access	1
speed process	1

German	**6 users**
Benefit Description	**Count**
accurate data	1
cost effective	1
data availability	1
data flexibility	1
data integration from distributed sources	4
data sharing	2
dynamic information	1
easily tracking transaction	1
easy maintenance	1
easy of use	1
easy to integrate the existing system	1
fast product delivery	2
high customer satisfaction	2

high quality	1		fast data retrieval	1
improved customer service	2		high quality	1
improved operation efficiency	2		re-usability	1
increased productivity	2		reduced complexity	1
minimized error	1		speed development	1
open standards	1			
platform independence	1		**Swiss** **4 users**	
platform neutrality	3		**Benefit Description**	**Count**
quick response	1		consistency	1
real-time information	1		cost savings	2
reduced operating cost	2		data flexibility	1
reliability	1		data integration from distributed sources	2
scalability	2		dynamic control of the content	1
security	2		easy of use	2
speed development	1		easy to integrate the existing system	1
speed process	1		easy to learn	2
stability	1		improved customer service	1
standard exchange	2		improved operation efficiency	1
standardized access	1		increased productivity	1
			powerful searching capability	1
Ireland **1 user**			quick response	1
Benefit Description	**Count**		reduced maintenance cost	1
cost savings	1		reduced operating cost	1
increased productivity	1		scalability	2
real-time information	1		security	1
reduced labor cost	1		speed process	1
streamline business process	1		streamline business process	2
			workflow automation	2
Japan **1 user**				
Benefit Description	**Count**		**UK** **14 users**	
content easily updated	1		**Benefit Description**	**Count**
easy maintenance	1		accurate data	3
easy of use	1		content easily updated	2
flexible structure	1		cost effective	2
			cost savings	2
Mexico **1 user**			customized data format	3
Benefit Description	**Count**		data availability	1
data control	1		data integration from distributed sources	6
data sharing	1		data sharing	1
easily tracking transaction	1		dynamic control of the content	1
efficient content management	1		dynamic information	2
			easy maintenance	3
New Zealand **1 user**			easy of use	3
Benefit Description	**Count**		easy to integrate the existing system	3
content easily updated	1		efficient content management	1
consistency	1		extensible content	1
data availability	1		fast data retrieval	1
data control	1		fast transaction and fulfillment cycles	1
dynamic connectivity	1		flexible to accommodate change	1
dynamic control of the content	1		guaranteed connection	1
dynamic information	1		high customer satisfaction	1
easy maintenance	1		high data management capability	1

high performance	3		fast data retrieval	12
improved customer service	5		fast product delivery	17
improved operation efficiency	2		fast transaction and fulfillment cycles	1
increased customer base	2		flexible structure	6
increased equipment uptime	1		flexible to accommodate change	1
increased productivity	3		guaranteed connection	1
multimedia functionality	1		high customer satisfaction	4
powerful searching capability	2		high data management capability	1
quick order taking process	1		high performance	10
quick response	4		high quality	8
re-usability	2		improved communication	1
reduced labor cost	1		improved customer service	25
reduced maintenance cost	2		improved operation efficiency	22
reduced operating cost	3		increased customer base	3
reliability	1		increased equipment uptime	2
rich information	1		increased productivity	34
scalability	1		intelligent application	4
speed development	5		message protocol independence	1
speed process	2		minimized error	4
stability	1		multimedia functionality	1
streamline business process	3		multiple format	2
time saving	4		multiple functionality	3
workflow automation	4		open standards	3
			platform independence	2
United States **110 users**			platform neutrality	1
Benefit Description	**Count**		powerful hyerlinking capability	1
accurate data	21		powerful navigation capability	6
compatible with the existing technology	1		powerful searching capability	8
content easily updated	18		quick order taking process	4
consistency	4		quick response	17
cost effective	8		re-usability	9
cost savings	26		real-time information	11
customized data format	20		reduced complexity	1
data availability	7		reduced development cost	4
data control	6		reduced labor cost	7
data flexibility	32		reduced maintenance cost	7
data integration from distributed sources	20		reduced operating cost	17
data sharing	17		reliability	13
device-specific usability	1		rich information	1
dynamic connectivity	2		scalability	30
dynamic control of the content	9		security	10
dynamic development	4		simplifying process	4
dynamic information	12		speed development	25
easily tracking transaction	10		speed process	20
easy access	12		stability	4
easy maintenance	13		standard exchange	1
easy of use	21		standardized access	2
easy to integrate the existing system	20		streamline business process	10
easy to learn	4		time saving	17
effective customer service	4		transaction integration	3
efficient content management	18		workflow automation	14
extensible content	3		**Total**	**1017**

REFERENCES

About.com, *The Perfect Match: XML and Java*, April 2000, http://xml.about.com/xml/library/weely/aa040100a.htm.

Fuchs, Matthew, *Why XML Is Meant for Java*, June 1999, http://www.webtechniques.com/archives/1999/06/fuchs/.

Goldfarb, F Charles, *The SGML Handbook*, New York : Oxford University Press, 1990.

International Organization for Standardization (1986), *Information Processing – Text and Office Systems – Standard Generalized Markup Language (SGML)*, ISO 8879:1986

Jaenicke, Coco, *Capitalizing on the benefits of XML*, May 1999, http://www.exceloncorp.com/xml_resources/reading_room_links.html.

Johnson, Mark, *XML for the Absolute Beginner – A Guide Tour from HTML to Processing XML with Java*, 1999, http://www.ibm.com/java/education/xmlbeg.html.

Microsoft, *Benefits from XML*, 1999; http://msdn.microsoft.com/xml/general/benefits.asp.

Microstar, *Rapid Internet Transaction Development*, August 2000, http://www.opentext.com/services/content_management_services/xml_services_white_paper.html.

Morgenthal, JP, *Protable Date/Portable Code*, May 2000, http://java.sum.cm/xml/ncfoucs.html.

Ogbuji,Uche, *XML: The future of EDI?* Sunworld, February 1999, Vol. 13 No. 2, http://www.sunworld.com/swol-02-1999/swol-02-xmledi.html.

Rath, Hans H., *XML, Chance and Chanllenge for Online Information Providers*, July 2000, http://www.oasis-open.org/html/xml_chance_challenge.html.

Rogers, Everett M., Diffusion of Innovations, New York: The Free Press, 1983.

Sheskin, J. David, *Handbook of Parametric and Nonparametric Statistical procedures*, Boca Raton: CRC Press LLC, 2000.

World Wide Web Consortium *Extensible Markup Language (XML) Version 1.0*, Feb 10,1998, http://www.w3.org/TR/1998/REC-XML-19980210.

World Wide Web Consortium, *HyperText Markup Language Version 4.0 (HTML),* April 24,1998, http://www.w3.org/TR/1998/REC-html40-19980424.

ACKNOWLEDGEMENTS

I would like to thank Dr. Sree Nilakanta, my major adviser, for his kind support, valuable guidance, and unceasing encouragement during my graduate study.

I would also like to thank the other members of my Program of Study Committee, Dr. Leslie Miller and Dr. Dan Norris, for their effort in reading and improving this thesis.

My special thanks go to my husband, parents and brother for their love and support.