Matthew Bachtell. Advertising Of Extensible Markup Language (XML) Online Public Access Catalogs (OPAC): A Case Study Involving Three Library Vendor's Products. A Master's paper for the M.S. in L.S. degree. April, 2003. 37 pages. Advisor: Jane Greenberg

This case study is an analysis of adversting support for XML by three prominent library OPAC vendors. The three library vendors, Endeavor Information Systems, Ex Libris, and Innovative Interfaces, are all advertising products that use XML to some degree in their OPAC module. Questions focused on in this research revolve around market advertising of the products, XML incorporation, MARC/XML compliance of the MARC XML compliance, and MARC to XML conversion software.

Headings:

OPAC

XML

MARC

MARC XML

LIBRARY VENDORS

# ADVERTISING OF EXTENSIBLE MARKUP LANGUAGE (XML) ONLINE PUBLIC ACCESS CATALOGS (OPAC): A CASE STUDY INVOLVING THREE LIBRARY VENDOR'S PRODUCTS

by
Matthew Bachtell

A Master's paper submitted to the faculty
of the School of Information and Library Science
of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements
for the degree of Master of Science in
Library Science.

Chapel Hill, North Carolina

April 2003

Approved by:

_____

Advisor

## TABLE OF CONTENTS

**Introduction**

As the "Information Age" plunges forward people have grown more accustomed to instantly accessing large quantities of information. This presents an interesting dilemma for librarians who are stereotypically viewed as the keepers/preservers/purveyors of information. How do we, as librarians and information specialists, encourage the general public to utilize the library's website and physical materials when they can simply input a query string into an Internet search engine that will supply acceptable answers to the user? One possible solution to this problem is to make the Online Public Access Catalog (OPAC) a more dynamic and user friendly endeavor that is capable of supplying the user with all possible answers.

How is this shift being achieved? There are several areas of possible focus for this question, the main area of concentration in this paper is the OPAC's transition from a Machine-Readable Cataloging (MARC) format to the Extensible Markup Language (XML). MARC will not be replaced; rather, it will be modified and updated. This new backend of the OPAC will entice people to use a library's catalog before solely consulting the Internet for answers to questions. It will do so because the new XML OPAC will allow for more diverse access to the catalog, via Personal Digital Assistants (PDAs) and web-ready cellular phones, and for more precise searching of the collection.

For the purposes of this paper I will focus specifically on MARC and XML records pertaining to physical books. This is not to say that other resources such as videos, music, and webpages can not be cataloged using MARC or XML it is just easier to focus on a specific resource when discussing it.

The majority of libraries' OPACs are constructed on the MARC format that structures available information about materials into a standard and predetermined but very limited hierarchical format. MARC takes pertinent information from physical resources, i.e. books, and places them into predetermined areas. The MARC format is a markup language, such as the HyperText Markup Language (HTML) or XML, used to capture the essence of an object, in this case a book. For example, the International Standard Book Number (ISBN) is placed into the 020 field. An example of a full MARC record is available in Appendix A, Figure 1. One can see how the information is separated into various sections that represent different areas of the material. These various sections can then be subdivided by using subfields which contain a delimiter code (a symbol such as $) and a letter or number (a, b, 1, 2). This results in "$a Arithmetic / $c Carl Sandburg ; illustrated as an anamorphic adventure by Ted Rand." in the 245 field of the selected MARC record, which is the title/author field. OPACs are designed to search these structured forms for queried information based on the user's input. If the patron were searching for the author Carl Sandburg, they would type "Sandburg, Carl" into the input area and select author to search under. This selection tells the OPAC to look in author subfield $c in the 245 MARC field for the required information. Increasingly OPACs support the option of

searching by keywords, title, author, subject heading, call number, ISSN, ISBN, and various other identifiers.

MARC was designed in the late 1960's for the original Public Access Catalogs (PACs) on computer terminals that were to be used solely in the library. With the growth and widespread use of the Internet, libraries perceived the possibility of a greater user base and expanded the OPAC. This expansion allowed users to access the listing of the library's materials from a remote computer over the Internet. The unforeseen development here was that the user's demands of the library catalog would begin to increase and their search strategies might become more sophisticated and they also might desire a greater sense of flexibility. Searching by the author's last name/first name was no longer sufficient for users and they sought a means of "browsing" the collection via the OPAC.

For libraries to achieve this new type of OPACs they will need to move away from the library's existing dependence on older technology and move towards a format that can be readily accessed and searched. XML may be the foundation tool, or keystone, that will allow libraries to structure their information for more custom and dynamic searches by their patrons. XML brings further potential to the OPAC because it can be used to incorporate photographs, images, and audio *content* into the collection.

This transition from the MARC based OPAC to an XML version will not occur internally within the library but externally via the vendors who supply the catalog management tools. Some library vendors are beginning to develop products that incorporate XML into their OPAC features; however, not all vendors are utilizing

XML. The transition from MARC to XML has not become a standard yet, but some companies have taken a leap forward with their products by supplying XML enabled OPACs. These companies are at the forefront of the next big change in the library environment.

XML has the potential to change the manner in which people interact with the library and its catalog, making the library more relevant to the average user, increasing the use of the library and its resources.

This paper focuses on the advertising of XML capable OPAC features and services by selecting three library vendor products and reviewing them based on a select number of research questions. These research questions seek to draw out various points and highlights of the various products to demonstrate their attributes.

**XML and the Internet**

XML was first released to the public five years ago by the World Wide Web Consortium (W3C)[1]. XML, as a language, has gone thorough several changes since its release and is still touted by many to be the next big innovation on the Internet. With all of the noise made by the XML supporters, limited action has been taken with the language because of unfamiliarity and the time needed to implement it into practical use by the OPAC community.

XML is a structuring language derived from the Standard Generalized Markup Language (SGML) that is used to describe documents and objects. Originally the language was designed to handle large scale electronic publishing but recently has been used to increase the exchange of data and information on the Internet.

---

[1] "World Wide Web Consortium", http://www.w3c.org/, accessed March 14, 2003, 02:37 pm.

A key aspect of XML is that it is a *structuring* language. XML does not possess the same abilities that a computer programming language like Java or C++; rather, it is a metalanguage, that is used to encode information about objects or entities. This metalanguage contains "both content (words, pictures, etc.) and some indication of what role that content plays (for example, content in a section heading has a different meaning from content in a footnote, which means something different than content in a figure caption or content in a database table, etc.)"[2]. Practically all documents and objects have a semblance of structure that can be described utilizing XML. This underlying theory is what gives XML a great deal of its momentum and support in the computing community.

XML does not specify its semantics or a tag set. This freedom from a rigid structure is what allows XML users to develop custom tags. XML provides the facility to define and develop tags and structural relationships, in contrast to HTML, which uses a predefined set of tags to structure information. While HTML is evolving into the Extensible Hypertext Markup Language (XHTML), many of the same elements will carry over. XML uses HTML and XHTML to render its content on the Internet. A transformation process occurs enabling XML to be rendered graphically or aesthically on the Internet, otherwise XML data will remain in a raw data type set, such as Figure 3 in Appendix A. Raw XML data is functional and is technically more powerful than rendered data because it is not forced to go through transformations. Raw data is what the computer prefers to use and can be used by knowledgeable library users. The conversion of XML raw data to HTML or XHTML

[2] "XML.com: What is XML? [Oct. 03, 1998]", http://www.xml.com/pub/a/98/10/guide1.html#AEN58, accessed March 1, 2003, 09:49 am.

is completed through either one of two processes named Extensible Stylesheet Language Transformation (XSLT) or the XML Path Language (Xpath).

To better understand what XML looks like an example is used to illustrate numerous points. Figure 2 in Appendix A is a simple encoded version of a possible conversation between two people marked up in XML. The first tag, <?xml version="1.0"?>, is a declarative statement telling the computer reading the marked data that the version of XML that is being used here is 1.0. The second tag <conversation> is the root tag that is used to describe the overarching theme or subject of the object/document. The third tag, <matt>, is used to describe the first person that is speaking in the conversation. The words that are spoken by Matt are nested following the <matt> tag. To end Matt's part of the conversation the computer must be prompted with a marker that indicates a stopping point. This ending marker is designated with a "/" within the brackets. Therefore Matt's portion of the conversation is ended with </matt>. Kristin's portion of the conversation mirrors the structure created in Matt's part with the tags <kristin> and </kristin> replacing <matt> and </matt>. To indicate to the computer that the conversation has reached its end the tag </conversation> is used. This simple three sentence conversation has been dissected by XML and broken down into its basic elements which are used to describe the content. XML can be used for more robust markups that are far more complex than this example but it offers a simple explanation of the operations of the language.

An example of a MARC record recorded in XML can be found in Appendix A, Figure 3. This XML record is the exact same record as Figure 1 in an XML

format. Notice that the MARC fields, such as 020, are retained but in a different format. The new format of the 020 MARC field is:

```
<datafield tag="020" ind1=" " ind2=" ">
  <subfield code="a">0152038655 :</subfield>
  <subfield code="c">$15.95</subfield>
</datafield>
```

The <datafield tag="020" ind1=" " ind2=" "> tag indicates to the computer that this is the equivalent to the 020 MARC field. Furthermore, subfield indicators, such as <subfield code="a">, replace the need for the delimiter ($) fields in MARC. This example should indicate that all integral information recorded in MARC is retained in the XML format. The appearances of the two may look different but the structure created utilizing XML allows for more complex searching and allows for greater device access. For example by breaking down each element and element component part in each MARC field tag, a search could be performed on these smaller pieces.

Once the essential elements of a document/object are isolated or marked in XML, an XML Schema may be written. An XML Schema could also be written before a document is marked in XML. It is at the discretion of the developer, in the case of this research investigation underlying this paper, the library vendors or the Library of Congress, to decide. Schemas are used to describe and constrain the content of XML documents. They potentially can be the backbone of any XML system because they ultimately supply the structure that the XML is built on. XML Schemas allow other XML documents with similar subjects to interact because they have the same or similar elements as defined by the Schema.

There are a variety of other functions that XML is capable of and a various other components of the XML standard that are relevant but are not elaborated on in this research. For further information on Extensible Stylesheet Language (XSL)[3], XML Linking Language (Xlink)[4], XML Query Language (Xquery)[5], Resource Description Framework (RDF)[6], and Xforms[7] consult pertinent books and websites about XML.

**MARC and XML**

The international standards committee of the Internet, the W3C, has stated that XML is a stable and working standard that companies or organizations can use to develop applications.[8] Based on the foundation of XML, the Library of Congress started to develop an XML project that would begin to modify MARC as the underlying skeleton or framework of library OPACs. This modification will in all likelihood be the next version of MARC released. The Library of Congress is supplying all of the schemas, definitions, and stylesheets without cost to the American public and private industry in the hope that organizations will utilize them in their XML projects concerning libraries and their OPACs. Vendor incorporation of the project will allow for a common thread running amongst various OPACs which will ultimately allow for compatible searching among libraries regardless of vendor.

---

[3] "The Extensible Stylesheet Language (XSL)", http://www.w3.org/Style/XSL/, accessed April 13, 2003, 06:56 p.m.
[4] "XML Linking Language (Xlink) Version 1.0", http://www.w3.org/TR/xlink/, accessed April 13, 2003, 06:56 p.m.
[5] "Xquery 1.0: An XML Query Language", http://www.w3.org/TR/xquery/, accessed April 13, 2003, 06:59 p.m.
[6] "Resource Description Framework (RDF) / W3C Semantic Web Activity", http://www.w3.org/RDF/, accessed April 13, 2003, 07:01 p.m.
[7] "Xforms 1.0", http://www.w3.org/TR/xforms/, accessed April 13, 2003, 07:02 p.m.
[8] "Extensible Markup Language (XML)", http://www.w3c.org/XML/, accessed March 14, 2003, 02:38 pm.

The office responsible for hosting and maintaining this project in the Library of Congress is the Network Development and MARC Standards Office. These two offices have termed the project *MARC XML*[9]. The definition supplied with the explanation of the project is that, "[t]his framework is intended to be flexible and extensible to allow users to work with MARC data in ways specific to their needs"[10].

Those developing the MARC XML project waded through a variety of issues highlighted with the creation of a simple and flexible MARC XML Schema. The two main issues involved were a lossless conversion of MARC to XML and the "roundtripability" from XML to MARC. The XML Schema[11] the Library of Congress developed is at the heart of the entire project. This robust schema takes into account all current MARC fields and allows for a seamless transition to XML because it retains a similar if not exact structure. The conversion process from MARC to XML is equally important because it will retroactively refit a library's existing catalog to work in the XML environment. Conversely the ability to go from XML to MARC is also possibly significant because not every patron is going to have an XML compliant Internet browser. Older browsers must be taken into account because all library patrons must have access to a library's resources. This means that the XML data must be able to convert to MARC.

**Research Objectives**

---

[9] "MARC 21 XML Schema", http://www.loc.gov/standards/marcxml/, accessed March 14, 2003, 02:47pm.
[10] "MARC XML Design Considerations", http://www.loc.gov/standards/marcxml/marcxml-design.html, accessed March 2, 2003, 01:33 pm.
[11] "MARCXML Schema", http://www.loc.gov/standards/marcxml/schema/MARC21slim.xsd, accessed March 14, 2003, 02:40 pm.

The intention of this research is to explore three library OPAC vendors marketing of XML compliant OPACs features and services. Newer XML OPACs are necessary to the growth of libraries that must adapt and integrate change into their system to remain competitive and relevant to a more sophisticated and information-saturated audience. Within this study it is not the researcher's intention to promote one vendor's OPAC over another; rather, the three relevant products are chosen to examine how they function and their potential. There are other vendor products beyond the three selected here, but the three chosen are the most obvious advertisers of their XML OPAC products. All of this serves to support the research question of the transition from MARC OPACs to XML OPACs in libraries.

**Methodology**

A case study approach underlies this research. Three steps were taken to select the cases. The first step involved the gathering of information and an analysis of published materials. Following a detailed Internet search of library vendors 3 library OPAC vendors were selected. The companies chosen for evaluation here are Endeavor Information Systems, Ex Libris, and Innovative Interfaces. These three were selected because all of them use XML in their OPAC products to some degree.

The process of vendor and product selection progressed through three stages. The first step was to do simple searches on the Internet using http://www.google.com for "library vendors", "XML and libraries", "library vendors and XML", and "OPAC vendors". This gave me a fairly comprehensive list of companies that create OPACs for libraries and allowed me to look at their products available for purchase. Their websites were noted and kept for future reference as well as their various products.

The second step involved going to the ALA Mid-Winter Convention in Philadelphia, Pennsylvania which met in January 2003. During this step materials were gathered from all vendors who advertised themselves as OPAC providers in the index of the ALA convention handbook. These materials were then compared to the previously gathered material. The third and final step of the research process involved examining all of the gathered materials and websites of the various companies. Selection criteria revolved primarily around the inclusion of XML in products. Eventually the field was narrowed down to three selections that most obviously advertised XML OPAC products and features. To analyze Innovative Interfaces, Endeavor Information Systems, and Ex Libris' OPACs the researcher developed a series of questions that he sought to answer from their available materials. The full list of questions is available in Appendix B. The questions address various issues surrounding XML incorporation, Library of Congress MARC XML compliance, product specifications, intended audience, and current users.

**Results and Discussion**

**Table 1. What is the name of the product reviewed?**

| Company Name | What is the name of the product reviewed? |
|---|---|
| Endeavor Information Systems | ENCompass Solutions / |
| Ex Libris | MetaLib |
| Innovative Interfaces | Millennium Automated Library System |

There are three variations of ENCompass Solutions[12] from Endeavor Information Systems available for purchase. The first is the ENCompass for Resource Access which "provide[s] instant access to e-resources and expanded search

[12] "ENCompass Solutions",http://encompass.endinfosys.com/, accessed February 23, 2003, 10:00am.

capabilities"[13], supplying access to local resources, such as a library's OPAC and e-resources. The second ENCompass product is called the ENCompass for Digital Collections and it "helps libraries integrate the traditional items in the local OPAC with image, video, audio items –and more – with one search"[14]. The third product, entitled simply ENCompass, is a combination of the two previous versions of the product. This is the product that most large research universities would purchase because it offers them the most flexibility.

Ex Libris' MetaLib[15] is described as "handl[ing] the organization, dissemination, and retrieval of scholarly information in a heterogeneous environment of library catalogs and electronic databases. It serves as an ideal platform for uniting all types of library collections, be they local or remote, traditional or electronic."[16] Key information in this statement indicates that this is a hybrid information system because it states that MetaLib is capable of handling both physical and electronic objects.

Innovative's product discussed in this research is the Millennium[17] automated library system. Millennium incorporates an XML Server, an XML Harvester, and various other XML components that are of interest.

ENCompass, MetaLib, and Millennium all give the sense of something that is new and cutting edge. ENCompass offers the impression of something that is

---

[13] "ENCompass – What is ENCompass" http://encompass.endinfosys.com/whatis/whatisENC2.htm, accessed February 23, 2003, 10:50 am.
[14] "ENCompass – What is ENCompass" http://encompass.endinfosys.com/whatis/whatisENC2.htm, accessed February 23, 2003, 10:56 am.
[15] "Products - MetaLib", http://www.aleph.co.il/MetaLib/index.html, accessed March 6, 2003, 02:58 pm.
[16] "Products – MetaLib Overview" http://www.aleph.co.il/metalib/overview.html, accessed March 6, 2003, 03:22 pm.
[17] "Innovative Interfaces : Products : Millennium : Overview", http://www.iii.com/products/millennium/index.shtml, accessed March 11, 2003, 12:34 pm.

encircling all possible scenarios that a library might experience. This is intended to provide a sense of security because they know that the ENCompass system will take care of the problems. MetaLib plays on the heavy use of the meta- nomer that is attached to many things these days. MetaLib is the "information library" which also plays on a similar encircling feeling as well as injecteing a strong sense of the new with the use of the term. Millennium plays on the recent change in centuries which lends the product a sense of something fresh and new.

**Table 2. How is this product being marketed? As an OPAC, as a Digital Library, as an added service, etc…**

| Product Name | How is this product being marketed? |
|---|---|
| ENCompass Solutions | Digital Library Solution |
| MetaLib | Hybrid Information System |
| Millennium Automated Library System | Digital Collection |

ENCompass is being marketed as "a complete XML-based solution for integrated end user searching across multiple data types and databases"[18]. The product is being defined as a digital library solution for all materials. Endeavor advertises that ENCompass is an added feature product that can be attached to any OPAC. This cross platform compatibility is an interesting and exciting feature because ENCompass does not necessitate purchasing and running Endeavor's main OPAC product Voyager.

Marketing of MetaLib is focused on it being used as a hybrid system that will integrate a variety of functions in the library system. MetaLib can be used as an add-on product because it acts independently of other software that a library may be

[18] "ENCompass - FAQ", http://encompass.endinfosys.com/faq.htm, accessed February 23, 2003, 10:20 am.

utilizing. Libraries may be running Ex Libris or another library vendor and MetaLib will still operate.

The components within Millennium that this study is concerned with are found under the heading of "Digital Collections" on the Millennium website. Millennium is an automated library system which is designed to handle all computer functions needed in a library system. Within this automated library system is the OPAC which is technically a "digital collection" of metadata about the physical resources in the library. The XML Server and Harvester are incorporated products that come packaged with the Millennium product. These XML features are not added services offered by Innovative meaning that they can not be purchased separately or added to existing library software.

The marketing strategies of the three companies offer a look into how they feel the product will best suit a library. All three companies acknowledge in their product descriptions that the future of libraries does not rest exclusively on physical materials. The digital library or hybrid feel that all three of these products have is an example of this physical and electronic integration. All of the products note how XML is a new and integral component of these new OPAC and systems management solutions which is relevant because of the assumable move towards this as the primary method to store MARC data.

**Table 3. Based on the library vendor's language in their literature for whom is this product intended? Is it to be used by public, private, or academic libraries? What sized library is best served by the product?**

| Product | Who is the product predominantly being marketed to? |
|---|---|
| ENCompass Solutions | Various – Primarily College and University Libraries |
| MetaLib | Various - Primarily College and University Libraries |
| Millennium Automated Library System | Various - Primarily College and University Libraries |

Endeavor mainly markets and caters to colleges and universities with their ENCompass product. The ENCompass system appears to be primarily a management solution for a large quantity of data. Most small market libraries do not posses the sheer volume of information and resources that would necessitate such a management tool. Large universities, such as the University of North Carolina at Chapel Hill, would benefit from this because it would allow for the incorporation of materials from other relevant groups and organizations, such as Documenting the American South. Images, sounds, and photographs could be integrated into libraries' OPAC here.

Ex Libris is an international library vendor based in Israel so its consumers often represent a wide variety of institutions. Academic, consortium, corporate, governmental, national, public, research, and special libraries throughout the world use the MetaLib and Ex Libris products. The target audience for this product is therefore widespread. However, the number of libraries who can actually utilize the full potential of the MetaLib system may be limited. Larger research, corporate, and academic institutions would benefit the most because of the volume of data that they posses. These libraries' have a multitude of varying information that could be potentially integrated into one system. Smaller libraries could get by running simply ALEPH or another integrated management software package.

Innovative language to advertise its Millennium project never clearly indicates who this product is geared towards though it is likely planned for academic, university, and corporate libraries. Academic libraries best suited for this product would vary from small college libraries serving 1,200 students to large university

libraries serving 60,000 students. Public libraries could also use the product but they might be better served by using a less powerful management system.

Product marketing is an invaluable to asset to examine because it demonstrates the intended function of the product. All three of these companies are aiming primarily at college and university libraries because of the size and complexity of the materials gathered there. These are high-level management tools that will integrate all materials that a library possesses into a singular interface. Terms such as digital library and hybrid system are great marketing and harken back to that notion mentioned before with their product names.

**Table 4. Has the company incorporated XML into its product's features?**

| Product | Has the Company Incorporated XML into these products? |
|---|---|
| ENCompass Solutions | Yes |
| MetaLib | n/a |
| Millennium Automated Library System | Yes (XML Server, XML Harvester) |

XML is used by ENCompass for "structured requests and receipt of information"[19] within the ENCompass product. The "structured requests" are the patrons inputting queries into the OPAC and the "receipt of information" is entries made to the catalog by library's catalogers. A screen capture of a potential catalog entry can be seen in Figure 4 in Appendix A. The Graphical User Interface (GUI) is a familiar way to add data to an OPAC. The new GUI feature in ENCompass is that the information is embedded into XML immediately and stored in its raw form on the

[19] "ENCompass – What is ENCompass" http://encompass.endinfosys.com/whatis/whatisENC2.htm, accessed March 4, 2003, 03:35 pm.

library's server. This XML data is then reformatted on demand to accommodate a libraries' settings or to a user's personal preferences.

The only statement made concerning XML in the MetaLib product description by Ex Libris is in its "Universal Gateway" portion where it says that MetaLib can "Permit[] simultaneous, unified broadcast searching over a large collection of data sources that are heterogeneous in structure, data syntax (for example, MARC, MAB, XML, EAD, Dublin Core, and TEI)"[20]. It never says that XML is used to store cataloging information, rather, it says that the product can work with data stored in XML. It seems like it is up to the library to choose the integrated management software that will catalog in XML and then MetaLib can then use it. Ex Libris' ALEPH product also does not catalog in XML.

Millennium uses its XML Server to store recently cataloged or converted MARC data in an XML format. (For a screen capture of data stored in the XML Server consult Figure 5 in Appendix A.) Data stored on the XML server can be transformed into other formats such as Dublin Core, Encoded Archival Description (EAD), HTML, and XHTML.[21] The data stored in the XML server could be searched by an OPAC in its raw form, as seen in Figure 4, or it could be converted into HTML or XHTML for a more aesthetic interface. The majority of people using the OPAC will never know about the XML raw data type as the default searching style will likely be set to either HTML or XHTML conversion.

---

[20] "Products – MetaLib Overview" http://www.aleph.co.il/MetaLib/overview.html, accessed March 6, 2003, 04:08 pm.
[21] "Innovative Interfaces : Products : Millennium : Digital Collections" http://www.iii.com/products/millennium/digitalcollections.shtml, accessed March 11, 2003, 01:45 pm.

Millennium uses its XML Harvester to, "leverage[] XML technology to provide an automated cataloging tool which can create library records from metadata records stored on servers anywhere in the world".[22] The "metadata records" that the quote mentions are accessible MARC records on other library sites that this product can access and convert from MARC to XML. Predefined library filters determine where and what the XML Harvester seeks on the Internet. "Metadata records" could also include EAD, Dublin Core, or various other data types. Furthermore, the XML Harvester could convert in-house MARC records to the XML data format. Converted XML data would be placed on the XML Server so that it could be searched by the Millennium's OPAC and given to the library user.

This question gets to the heart of the research. Are companies using XML and if they are using XML are they using it in their OPAC's features? What was found here is that Endeavor's ENCompass and Innovative's Millennium products both use XML in their OPAC functions while Ex Libris' MetaLib can simply understand XML presented to it from various sources.

**Table 5. Is the library vendor advertising that they are following the Library of Congress' XML Schema and MARC XML project?**

| Product | Do these products advertise usage of the Library of Congress' MARC XML Project? |
|---|---|
| ENCompass Solutions | n/a |
| MetaLib | n/a |
| Millennium Automated Library System | n/a |

There is no mention of the ENCompass system ascribing to the Library of Congress's MARC XML Project anywhere in the physical literature or on the product webpage. Because the ENCompass system can be attached to any preexisting OPAC

[22] "Innovative Interfaces : Products : Millennium : Digital Collections" http://www.iii.com/products/millennium/digitalcollections.shtml, accessed March 11, 2003, 02:11 pm.

or competitor's project it is likely that there is some utilization of the Library of Congress's Schema and XML efforts. There is no substantial evidence to this effect and it is merely speculation on the part of the researcher.

There is no mention of Ex Libris utilizing the Library of Congress's MARC XML Project in MetaLib. This statement is founded on the lack of evidence found in the physical and electronic literature gathered about the product.

Innovative Interfaces never mentions that it is following the Library of Congress' MARC XML project. The company does not have to reveal how its products work to the buyer or the general public. However, based on Figure 5, which was obtained from the Innovative Interfaces' website, it would seem that there is not much semblance between the two. Figure 3 is a MARC record recorded in XML which was obtained from the Library of Congress and it does not resemble Figure 5 which is Innovative's XML example used under its XML Server description. This discrepancy points to the Millennium product not working seamlessly with the Library of Congress.

Because the Library of Congress is supplying their MARC XML schema free of charge it was intriguing to see if any of the three companies discuss their use of this project in their various products. The data points to either a lack of compliance or acknowledgement with Endeavor, Ex Libris, and Innovative. There is no rule mandating these vendors use the Library of Congress' schema. However, by using the schema they would be standardizing the MARC XML process which would increase the usability and cross platform potential of a library's OPAC. Of course there could just be a lack of acknowledgement amongst the vendors because this is

not seen as relevant. It is difficult to determine this compliance from the materials gathered for this study.

**Table 6. Does the company offer any conversion software that will convert older MARC records into XML? Is there any software that will take XML records and convert them back to MARC?**

| Product | Is there MARC to XML software available with these products? |
|---|---|
| ENCompass Solutions | No |
| MetaLib | No |
| Millennium Automated Library System | Yes |

ENCompass works with both XML and MARC data in their native forms. Endeavor does not see the need to convert existing MARC data to XML. In support of this statement is the following,

"Your OPAC contains important research materials, collections carefully developed over time to meet the needs of your users. ENCompass provides another layer of access and integration by allowing the library to add materials never before cataloged in the OPAC and makes all of the items searchable from ONE access point."[23].

The inference of this statement is that there is no conversion needed. All new cataloging information added to the OPAC will be in stored in an XML format and all older data will remain in its current MARC format. Queries to the library's OPAC are delivered in their native form based on the material requested.

The company offers no conversion software to convert from MARC to XML or in reverse. There appears to be little emphasis placed on XML in Ex Libris' products including MetaLib.

Innovative offers their XML Harvester which will convert MARC data on the fly into XML marked data and can go in a reversed direction as well. This is an advantage because not everyone will upgrade to an XML compliant Internet browser.

[23] "ENCompass - FAQ" http://encompass.endinfosys.com/faq.htm, accessed March 5, 2003, 12:02 pm.

Because libraries cater to people in all walks of life they will need to possess this ability so that all patrons can be served. The conversion process can occur rapidly and with little delay seen from the user's perspective.

The significance of this conversion software has yet to be measured. Because libraries have not aggressively moved towards XML the current MARC format will retain a measure of value in the future. The interoperability of MARC and XML is key because not all data will migrate to an XML format. Because only one of the three companies polled here see value in supplying this conversion software it does not seem likely that this will be a major area of concern in the future.

Some preliminary testing raises questions about the actual use and status of XML implementation in library OPACs. While only a sampling of each client's OPACs was executed it seemed to indicate that the previous statement was indeed true. The limited sampling range of this case study may restrict the findings here but the three library vendors reviewed are the most prominently marketed on the Internet and were the most impressive at the ALA Mid-Winter Conference 2003 as far as XML is concerned.

**Conclusion and Future Research**

Endeavor's ENCompass, Ex Libris' MetaLib, and Innovative Interfaces' Millennium appear to be as being at the vanguard of XML use in the library environment. These XML products will allow for more dynamic, Web-based, real time access to a library's catalog through their OPACs. Companies, other than the three reviewed here, continue to develop their own XML OPAC products which further push and expand the market as well as drive its advancement.

The findings in this study show that while some library vendors appear to be moving towards XML supported and incorporated products there is no panacea or perfect product. Of the three companies reviewed here none received perfect marks as far as the questioning criteria. This is a result of varying philosophies about the implementation of XML in libraries and especially within the OPAC. It is doubtful that their will ever be a unified approach but this need not matter because of the interoperability of raw XML data. However, the lack of acknowledgement of the MARC XML project from the reviewed vendors is somewhat disconcerting because of the free and open nature of the project. There is no law dictating that these companies follow this project but it would be beneficial to researchers and potential customers to see if the companies are using it or not. Regardless, these three companies' offerings represent the first batch of XML compliant library products and particularly interesting is two of the companies' XML OPACs.

Two limits not mentioned above that need to be addressed in future research on XML OPAC products include increaasing the survey to include more vendors beyond those advertising XML support via print and electronic publication and testing the actual applications of the various vendors products with different colleges and universities OPACs to see if the vendor's marketing statements prove true. By executing these two actions the study would be more robust and not just a schism of the current vendor market of XML products. The testing of the products in a live environment would be a significant step towards evaluation because the XML driven OPAC and its features would be a functional setting which would allow for proper assessment.

Libraries have the potential to push beyond their current boundaries and expand their influences to new and exciting realms because XML is an internationally supported standard in computing. Wireless devices such as PDAs and Internet-Ready Cellular Phones will be able to access the library's catalog from any location in the world, thus, potentially increasing the library's users.

XML is the future of the Internet and it should be incorporated into library OPAC functions to make it a more dynamic and user friendly experience. This will increase the relevance of the library in a user's life and allow them to find more significant and pertinent information for their various needs.

**Appendix A**

010 08cam 2200241 a 4500
001 AMD-8234
005 19930923085101.0
008 920219s1993 caua j 000 0 eng
010 $a 92005291 /AC $o 25508902
020 $a 0152038655 : $c $15.95
040 $a DLC $c DLC $d NOC
050 00 $a PS3537.A618 $b A88 1993
082 00 $a 811/.52 $2 20
099 9 $a C811 $a S213a
100 1 $a Sandburg, Carl, $d 1878-1967.
245 10 $a Arithmetic / $c Carl Sandburg ; illustrated as an anamorphic adventure by Ted Rand.
250 $a 1st ed.
260 $a San Diego : $b Harcourt Brace Jovanovich, $c c1993.
300 $a 1 v. (unpaged) : $b ill. (some col.) ; $c 26 cm.
500 $a One Mylar sheet included in pocket.
520 $a A poem about numbers and their characteristics. Features anamorphic, or distorted, drawings which can be restored to normal by viewing from a particular angle or by viewing the image's reflection in the provided Mylar cone.
651 0 $a North Carolina $x Poetry.
651 0 $a North Carolina $x Juvenile literature.

**Figure 1 – MARC record**

```
<?xml version=”1.0”?>

<conversation>

        <matt>What do you want
        for dinner? </matt>


        <kristin>I don’t know?
        What do you want for dinner? </kristin>

</conversation>
```

**Figure 2 – XML markup of a conversation**

```
<?xml version="1.0" encoding="UTF-8"?>
<collection xmlns="http://www.loc.gov/MARC21/slim">
  <record>
    <leader>01142cam  2200301 a 4500</leader>
    <controlfield tag="001">   92005291 </controlfield>
    <controlfield tag="003">DLC</controlfield>
    <controlfield tag="005">19930521155141.9</controlfield>
    <controlfield tag="008">920219s1993    caua   j      000 0 eng  </controlfield>
    <datafield tag="010" ind1=" " ind2=" ">
      <subfield code="a">   92005291 </subfield>
    </datafield>
    <datafield tag="020" ind1=" " ind2=" ">
      <subfield code="a">0152038655 :</subfield>
      <subfield code="c">$15.95</subfield>
    </datafield>
    <datafield tag="040" ind1=" " ind2=" ">
      <subfield code="a">DLC</subfield>
      <subfield code="c">DLC</subfield>
      <subfield code="d">DLC</subfield>
    </datafield>
    <datafield tag="042" ind1=" " ind2=" ">
      <subfield code="a">lcac</subfield>
    </datafield>
    <datafield tag="050" ind1="0" ind2="0">
      <subfield code="a">PS3537.A618</subfield>
      <subfield code="b">A88 1993</subfield>
    </datafield>
    <datafield tag="082" ind1="0" ind2="0">
      <subfield code="a">811/.52</subfield>
      <subfield code="2">20</subfield>
    </datafield>
    <datafield tag="100" ind1="1" ind2=" ">
      <subfield code="a">Sandburg, Carl,</subfield>
      <subfield code="d">1878-1967.</subfield>
    </datafield>
    <datafield tag="245" ind1="1" ind2="0">
      <subfield code="a">Arithmetic /</subfield>
      <subfield code="c">Carl Sandburg ; illustrated as an anamorphic adventure by
Ted Rand.</subfield>
    </datafield>
    <datafield tag="250" ind1=" " ind2=" ">
      <subfield code="a">1st ed.</subfield>
    </datafield>
    <datafield tag="260" ind1=" " ind2=" ">
      <subfield code="a">San Diego :</subfield>
      <subfield code="b">Harcourt Brace Jovanovich,</subfield>
```

```
    <subfield code="c">c1993.</subfield>
   </datafield>
   <datafield tag="300" ind1=" " ind2=" ">
    <subfield code="a">1 v. (unpaged) :</subfield>
    <subfield code="b">ill. (some col.) ;</subfield>
    <subfield code="c">26 cm.</subfield>
   </datafield>
   <datafield tag="500" ind1=" " ind2=" ">
    <subfield code="a">One Mylar sheet included in pocket.</subfield>
   </datafield>
   <datafield tag="520" ind1=" " ind2=" ">
    <subfield code="a">A poem about numbers and their characteristics. Features
anamorphic, or distorted, drawings which can be restored to normal by viewing from
a particular angle or by viewing the image's reflection in the provided Mylar
cone.</subfield>
   </datafield>
   <datafield tag="650" ind1=" " ind2="0">
    <subfield code="a">Arithmetic</subfield>
    <subfield code="x">Juvenile poetry.</subfield>
   </datafield>
   <datafield tag="650" ind1=" " ind2="0">
    <subfield code="a">Children's poetry, American.</subfield>
   </datafield>
   <datafield tag="650" ind1=" " ind2="1">
    <subfield code="a">Arithmetic</subfield>
    <subfield code="x">Poetry.</subfield>
   </datafield>
   <datafield tag="650" ind1=" " ind2="1">
    <subfield code="a">American poetry.</subfield>
   </datafield>
   <datafield tag="650" ind1=" " ind2="1">
    <subfield code="a">Visual perception.</subfield>
   </datafield>
   <datafield tag="700" ind1="1" ind2=" ">
    <subfield code="a">Rand, Ted,</subfield>
    <subfield code="e">ill.</subfield>
   </datafield>
  </record>
</collection>
```


**Figure 3 – Represents Figure 2 in an XML format (record taken from the Library of Congress' MARC XML project http://www.loc.gov/standards/marcxml/Sandburg/sandburg.xml)**

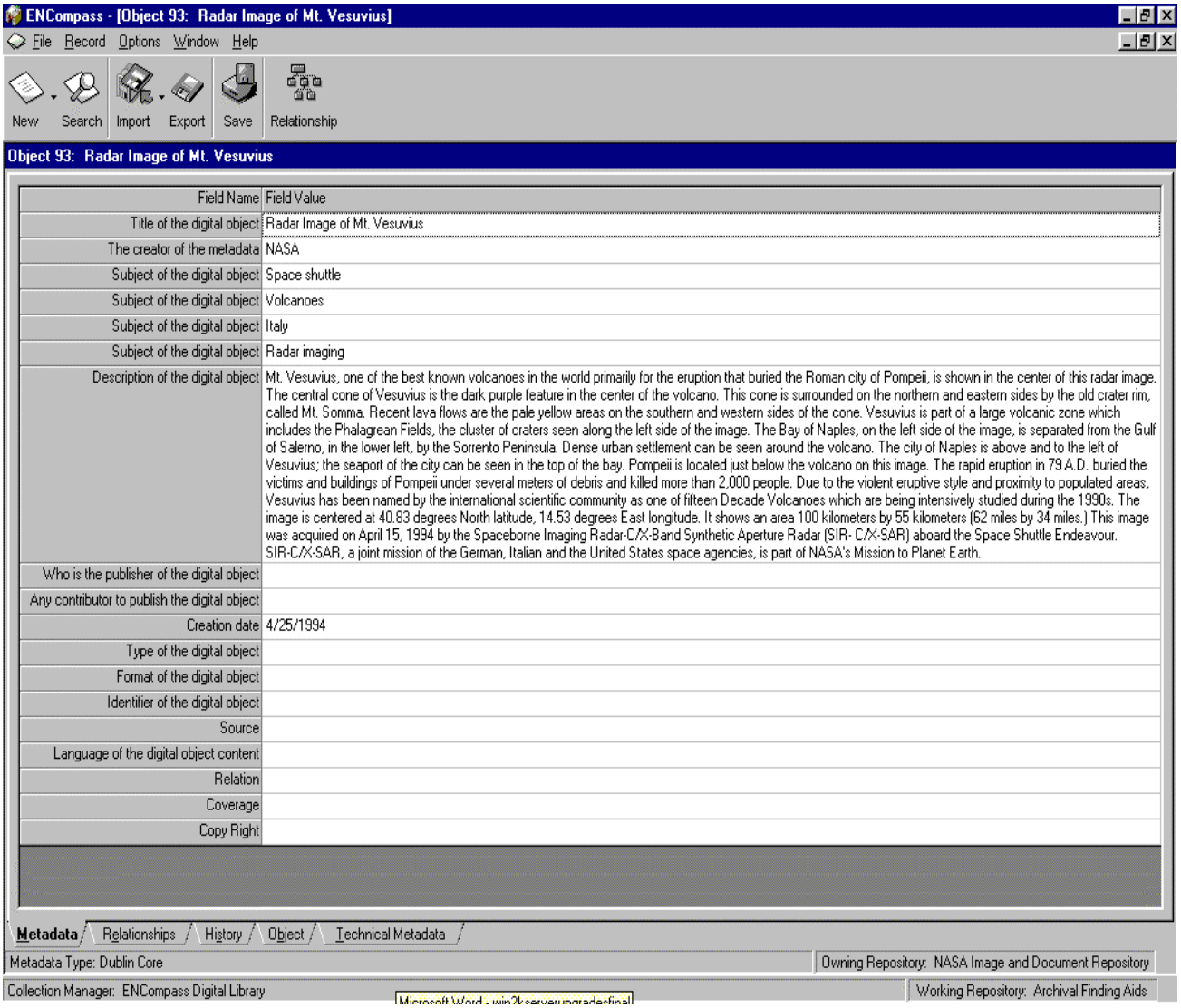


**Figure 4 – Screen Capture from ENCompass Illustrating a Cataloger's Potential Input about an Object (http://encompass.endinfosys.com/screens/objectwebF.gif)**

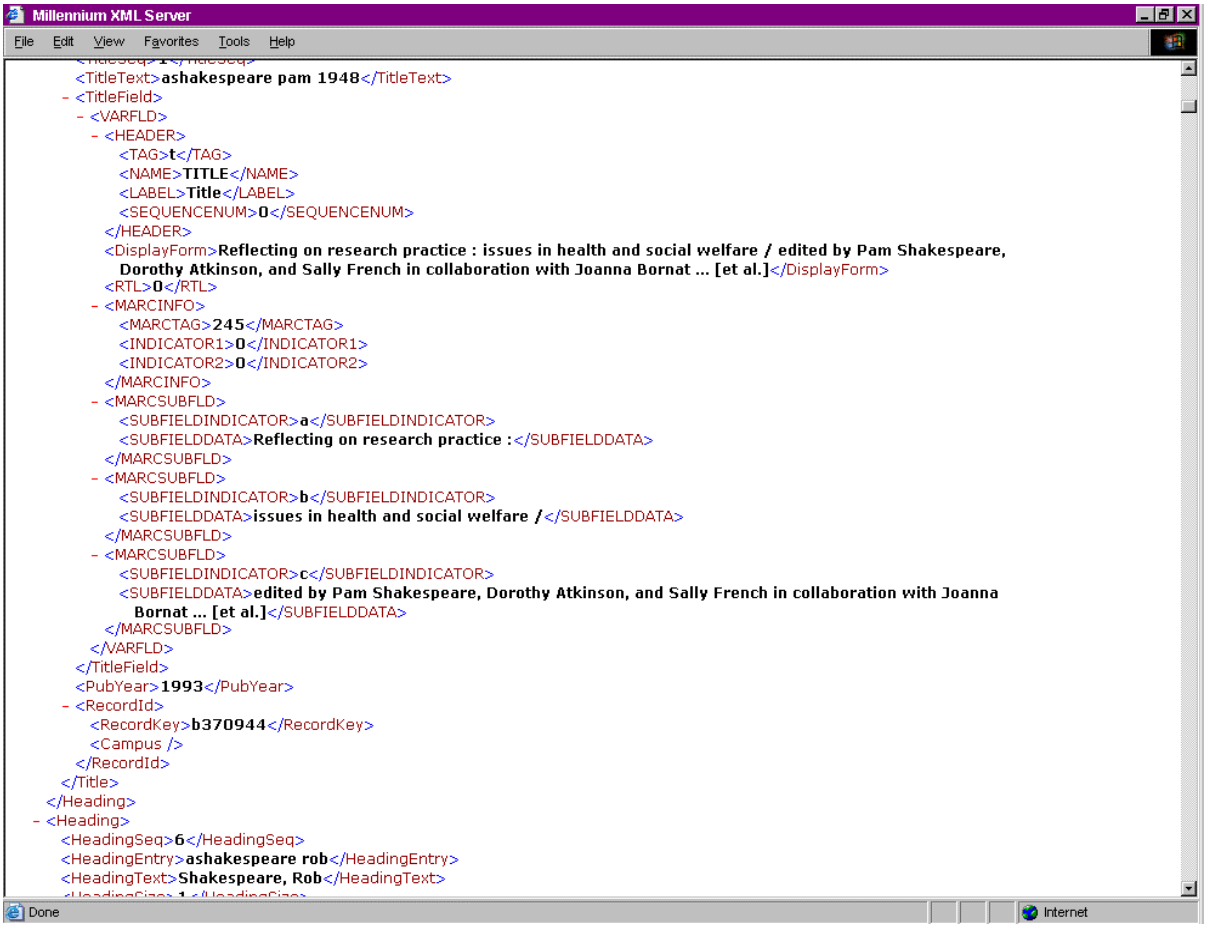


**Figure 5 – Screen Capture of Innovative Interface's XML Server (http://www.iii.com/products/millennium/pop_xmlserver.shtml)**

**Appendix B**

1) What is the name of the product reviewed?

2) How is this product being marketed? As an OPAC, as a Digital Library, as an added service, etc… .

3) Based on the library vendor's language in their literature for whom is this product intended? Is it to be used by public, private, or academic libraries? What sized library is best served by the product?

4) Has the company incorporated XML into its product's features?

5) Is the library vendor advertising that they are following the Library of Congress' XML Schema and MARC XML project?

6) Does the company offer any conversion software that will convert older MARC records into XML? Is there any software that will take XML records and convert them back to MARC?

**Appendix C**

**Endeavor Information Systems**
2200 East Devon Ave. Suite 382
Des Plaines, IL USA 60018-4505
http://www.endinfosys.com/index.htm

**Ex Libris**
46 Sokolov St.
Ramat Hasharon 47235, Israel
http://www.aleph.co.il/
or
1919 N. Sheffield
Chicago, IL 60614-6001

**Innovative Interfaces**
5850 Shellmound Way
Emeryville, CA 94608
http://www.iii.com

**Bibliography**


Banerjee, Kyle. "How Does XML Help Libraries", *Computers in Libraries*, Vol.22, No.8, September 2002, Available at http://www.infotoday.com/cilmag/sep02/Banerjee.htm. Accessed November 26, 2002, 12:56 P.M..

Bray, Tim. "Why XML Doesn't Suck". Available at http://www.tbray.org/ongoing/When/200x/2003/03/24/XMLisOK. Accessed March 28, 2003, 12:36 P.M..

Bray, Tim. "XML Is Too Hard For Programmers". Available at http://www.tbray.org/ongoing/When/200x/2003/03/16/XML-Prog. Accessed March 28, 2003, 12:37 P.M..

Chang, Sheau-Hwang. "XML--the foundation for the next generation library management system". *OCLC Systems & Services* v. 18 no2 (2002) p. 64-6.

Clarke, Kevin S.. "MARC and XML: A Choice or Replacement?", Access2002 Conference, Oct. 22, Windsor, Ontario. Available at http://elane.stanford.edu/laneauth/ksc-access2002/index.html. Accessed November 26, 2002, 13:31 P.M..

Crawford, Walt. *MARC for library use : understanding the USMARC formats*, White Plains, NY: Knowledge Industry Publications, c1984.

Danskin, Alan. "Today MARC Harmonisation, tomorrow the World Wide Web : UKMARC, MARC21, XML and ONIX". *Catalogue & Index*, no.143, (Spring 2002), pp1-3.

Fiander, David J.. "Applying XML to the Bibliographic Description". *Cataloging & Classification Quarterly*, Vol. 33 (2), 2001, pp. 17-28.

Johnson, Bruce. "XML and MARC: Which is "Right" ?". *Cataloging & Classification Quarterly*, Vol. 32 (1), 2001, pp. 81-90.

Klein, Leo Robert. "On the same page. standards, XML and the digital divide; interview with Jeffrey Zeldman and Carrie Bickner" *School Library Journal* Net connect (Winter 2002) p. 12-14

Lam, Ki-Tat. "XML and global name access control", *OCLC Systems & Services*. Vol. 18 (2), 2002, pp.88-96.

Medeiros, Norm. "Liberating online catalog records: Debate on replacing MARC format with XML in online catalog records." *OCLC Systems and Services*, v.16, no.3, (2000) p.100-1.

Miller, Dick. "XML and MARC: a choice of replacement?" presented at the ALA Joint MARBI/CC:DA Meeting, Chicago, 2000. Available on the web at http://xmlmarc.stanford.edu/ALA_2000.htm. Accessed on 11/10/2002 10:22 AM.

Motta, Santo and Giuseppe Ursino. "XML on MOM technology: a new approach for library software. message-oriented middleware application at the University of Catania" *Herald of Library Science* v. 40 no3/4 (July/Oct. 2001) p. 240-1

Tennant, Roy "The importance of being granular: metadata and standards" *Library Journal* v. 127 no9 (May 15 2002) p. 32-4.

Tennant, Roy. *XML in Libraries*. Neal-Schuman Publishers Inc., New York, 2002. Innovative Interfaces, http://www.iii.com/, accessed February 17, 2003, 07:10 am.

Endeavor Information Systems, Inc., http://www.endinfosys.com/, accessed February 17, 2003, 07:26 am.

Ex Libris, http://www.exlibris-usa.com/, accessed February 17, 2003, 07:40 am.

Extensible Markup Language (XML), http://www.w3.org/XML/, accessed February 28, 2003, 03:47 pm.

MARC 21 XML Schema, http://www.loc.gov/standards/marcxml/, Accessed November 26, 2002, 13:00 p.m.

Medlane Project, http://laneweb.stanford.edu:2380/wiki/medlane/overview Accessed November 26, 2002, 13:28 p.m.

World Wide Web Consortium (W3C), http://www.w3.org/, accessed February 18, 2003, 01:40 pm.

XML.com: XML From the Inside Out –XML development, XML resources, XML specifications, http://www.xml.com/, accessed February 28, 2003, 03:51 pm.