2003

Review of XML in Libraries

David Lodwick
Cleveland State University, d.lodwick@csuohio.edu

Publisher's Statement

XML IN LIBRARIES. Roy Tennant (Ed.). *New York: Neal-Schuman Publishers, Inc.,*
2002, 213 pp., ISBN 1-55570-443-3.

HTML had an enormous impact on libraries in the mid-1990s. It was so easy, and
so much fun, that seemingly overnight *everyone* had created a website or had one "under
construction." Then came word that an even bigger blockbuster was on the horizon:
Extensible Markup Language, or XML. With XML, librarians would not be confined to
a limited set of tags governing the display of text and images on the web; instead, they
would be able to define their own tags to create a new world of structured data storage
and transfer options.

XML did not, however, take the library world by storm with the immediacy and
high visibility of HTML. Its slow start may be explained in various ways: prospective
XML users were waiting for standards and consensus to emerge; they were hesitant to
replace existing tools with unproven XML technology; XML is not as easy and forgiving
as HTML. Yet XML, which has attained World Wide Web Consortium (W3C)
Recommendation status (see http://www.w3.org/TR/2000/REC-xml-20001006), has
quietly been gaining an ever-larger foothold in a variety of settings, ranging from the
Library of Congress's MARC-to-XML conversion project to the current Microsoft effort
to build XML support into its Office productivity suite. Those who still remain skeptical
about the usefulness of XML may want to visit a website such as "XML in Vertical
Industries" (http://www.xml.org/xml/industry_industrysectors.jsp), which documents the
use of XML in settings ranging from astronomy to food service (recipes) to oil and gas
drilling.

In <u>XML in Libraries</u>, editor Roy Tennant has assembled thirteen XML project descriptions, each written by a project leader or participant. One purpose of the book is to demonstrate, in Tennant's words, that "XML can be used *now* for the most practical of applications, solve library problems, and create new opportunities." Be forewarned that this is not an XML cookbook – newcomers to XML might want to familiarize themselves with basic XML concepts, acronyms, and jargon before tackling Tennant's collection. And librarians who work in Microsoft shops may be disappointed by the emphasis on open source software.

Most of the projects were born in academic library settings, but public and government libraries are represented as well. Each short -- about 15 pages -- essay is constructed using a uniform structure of subheadings (including "Goals," "Description," "Successes and Failures," and "Lessons Learned"), and each piece closes with contact information about the authors and a list of "Links and Resources" for those who want to learn more about the project software or see an XML project in actual operation. Screen shots and coding samples add an extra dimension to the descriptions. Each essay is placed within one of seven chapters. The chapter headings – Using XML in Library Catalog Records, Using XML for Interlibrary Loan, Using XML for Cataloging & Indexing, Using XML to Build Collections, Using XML in Databases, Using XML for Data Migration, and Using XML for Systems Interoperability -- demonstrate the wide variety of library applications in which XML may be employed.

Many of the contributors were motivated in part by a desire to experiment with XML and the many open-source and commercial software packages that support or interact with XML, and in describing their ventures they acknowledge project

shortcomings as well as victories. In the chapter on data migration, for example, Darlene Fichter of the University of Saskatchewan Library describes the efforts of a small team to convert 550 native law cases from HTML to XML. The team's objectives in moving the cases to XML format were "to ensure long-term preservation, to enhance their utility, and to experiment with search and retrieval systems." Their efforts were aided by the limited number of items to be converted , the existing structure inherent in reported cases, and the fact that the original HTML coders had made an effort to impose a consistent and logical structure when inserting tags into the text of the cases. Although several aspects of their work progressed rapidly, the Saskatchewan team eventually concluded that development of a suitable Document Type Definition (DTD) was taking far too long, and they reluctantly set it aside in favor of an interim DTD that would allow completion of the project on time. Undaunted, Ms. Fichter maintains that the DTD "detour" turned out to be a blessing in disguise, forcing the conversion team to a new and helpful procedural perspective.

Not every library has a collection of HTML law cases to convert to XML, of course, and the selective reader of XML in Libraries may choose to focus only on essays that address a specific library application of interest, such as online publishing or local history databases. Still, reading Tennant's collection from cover to cover is likely to engender in many librarians the desire to emulate an existing project, concoct a new one, or at least learn more about XML and its place in library operations.

David Lodwick
Systems Librarian
Cleveland State University Library