Book Reviews

Margaret Rohdy, Editor

The ABCs of XML: The Librarian's Guide to the eXtensible Markup Language. By Norman Desmarais. Houston, Tex.: New Technology Pr., 2000. 206p. $28 (ISBN 0-9675942-0-0). LC00-8051.

As libraries develop digital collections and make use of digital publications, librarians need to understand the underlying technologies that support those resources. The eXtensible Markup Language (XML) has received attention as a method to revolutionize electronic data interchange, data organization, and information retrieval. In *The ABCs of XML*, Norman Desmarais explains this complex technology and many of its possible uses within libraries. Although this book is not a how-to manual or a reference work, it fills an important need by presenting a concise survey of the technical aspects of XML and of the issues libraries will face in processing, managing, and using XML-based documents and resources. Desmarais's book can serve as a starting point for thinking about XML and thus would be valuable for librarians and students first approaching the topic.

Like many other writers on XML, Desmarais concentrates on the technical structure and underpinnings of the language by describing the logical and physical structures of an XML document. The largest section of *The ABCs of XML* is concerned with technical understanding of how XML is structured. The first four chapters—almost ninety pages—are devoted to detailed technical descriptions of the relationships among XML, SGML (Standard Generalized Markup Language), and HTML (Hypertext Markup Language); the XML document structure;

the display of XML documents; and the creation of links among XML documents and document fragments. Desmarais's descriptions are very clear and will appeal most to readers who want to understand the structure of an XML document. For example, he explains why the declarations at the head of an XML or SGML document are not gibberish, but are a key to understanding the entire document. His discussions of Document Type Definitions (DTDs), linking and pointing, and entity references show the potential of XML to deal with multiple character sets and multiple data types such as digital video and audio. However, readers may have a hard time imagining how a complex XML document, with all the features Desmarais describes, would look. An appendix with a more complex example, as well as an example of a DTD, would have been useful. His discussion (100ff.) of the Resource Description Framework (RDF) is brief, vague, and out-of-character for what is otherwise a detailed work. Desmarais also discusses areas of XML that are under development and expansion—like the eXtensible Style Language (XSL), which will be used to tell software how to display a document, and XLink, Xpoint, and Xpath, which will handle linking among XML documents. His understanding of these complex developments enables him to point out problems that can occur with unnecessarily complex schemes for display and linking. When Desmarais writes that "an XML link isn't exactly simple" (65), the reader understands why.

The remaining approximately sixty pages of text include discussions of processing XML, storing and manag-

ing XML documents, using XML for Electronic Data Interchange (EDI) and e-commerce, and sources of help with content, application, and schema development. This portion of the book is slightly less technically detailed than the first chapters but provides useful insights into how librarians and technology managers will work with XML. It will appeal to readers wanting to know what XML documents will accomplish and how people will manage and work with them. The bibliography includes many of the established publications on XML as well as Web sites cited in the text, although many of the sites are those of companies working on XML product development and thus not helpful sources for additional information about XML *per se*. Readers wanting to create XML documents and systems will need to turn to some of the resources listed, or others like them. The glossary includes the major concepts and technical terms mentioned in the book; the definitions are brief but very helpful for keeping track of this jargon-intensive field. Throughout the book, especially in Chapter 8, "Getting Started," Desmarais provides guidance about software for viewing, producing, and managing XML.

Unlike most writers on XML, Desmarais knows the library audience and focuses his discussions and explanations on the concerns of libraries and librarians. He uses the MARC (MAchine-Readable Cataloging) record structure and cataloging concepts to explain XML. For instance, he employs the concept of "statement of responsibility" to explain groups of elements and container elements (18). Using library concepts helps make a potentially vague topic more concrete

and approachable. Drawing on his understanding of the impact of adherence to standards—or the lack thereof—in libraries, Desmarais provides, especially in the chapter "XML and Its Potential for E-Commerce," great insight into the importance of establishing and adhering to standards to realize XML's full potential. His discussion of EDI is particularly thoughtful, reflecting his understanding of library acquisitions and fund management and the potential of XML for supporting EDI services. Although he understands the needs of libraries, Desmarais focuses almost exclusively on MARC and e-commerce applications. Four appendixes provide examples of SGML and XML used to store and generate MARC records, in particular to support the display of either Chinese characters or the transliterated text of a record. This emphasis tends to shift the reader's thinking away from other current uses of XML (with electronic books and journals, non-MARC metadata formats like the Encoded Archival Description, automated cataloging of electronic resources, etc.) and from developing new uses for XML beyond traditional library considerations. Few references are made to Appendixes 1–3, so perhaps they could have been shortened to make room for other materials. Although Desmarais mentions the development of other XML standards, such as the Chemical Markup Language, he does not discuss how these applications of XML will affect libraries and the work done in them. Desmarais also does not discuss how XML is likely to affect the library as an organization in nontechnical ways, such as workflow and funding.

Desmarais succeeds in explaining why we need to care about "the real goal of the [XML] markup: to clarify the document's underlying structure" (32), how this need differs from simply displaying text, and what advantages there are in using a powerful markup language. By focusing on the possible uses of XML within libraries, Desmarais avoids the vague assertions and abstractions that characterize many descriptions of XML and other markup languages. Although other good XML surveys exist, Desmarais's book is specifically oriented to libraries and will be most useful for readers wanting that perspective.—*Thomas P. Turner (tpt2@cornell.edu), Cornell University, Ithaca, New York*

Digital Libraries. By William Y. Arms. Digital Libraries and Electronic Publishing. Cambridge, Mass.: MIT Pr., 2000. 287p. $45 (ISBN 0-262-01880-8). LC99-14773.

Monographs that deal with the Internet, the Web, or digital libraries are often outdated by the time of publication, but reviewing William Arms's book *Digital Libraries* with this idea in mind would be inappropriate. In the introduction to the chapter on economic and legal issues, Arms describes his approach: "The discussion reflects my own viewpoint, which will probably need revision over time. However, I hope the basic ideas will stand" (99). The basic ideas of this book, contained in a comprehensive historical survey, an assessment of the state-of-the-art, and more importantly, of the accompanying culture should indeed stand the test of time.

The objective of producing a comprehensive review is worthwhile. Arms accomplishes his purpose through personal knowledge and reflections that are based on interaction with well-known individuals and institutions in the digital library community. Arms states that the text reflects his "own experiences and biases" (x) and examples that he knows personally, augmented by information from other individuals, as shown in the list of names in the acknowledgments. The examples may reflect Arms's personal experiences, but his discussion also makes it clear that he has interacted with a variety of digital library projects, researchers, and practitioners.

An examination of the individual chapters within *Digital Libraries* confirms its breadth of coverage. With topics ranging from technological ("The Internet and the World Wide Web") to sociological ("People, Organizations and Change"), the discussion covers the multidimensional aspects of digital libraries. Arms correctly points out that "digital libraries bring together facets of many disciplines, and experts with different backgrounds and different approaches" (1). While there are undoubtedly technical issues associated with digital libraries, his discussion also covers economic, legal, social, and cultural issues. Arms also describes the diverse groups that influence development of digital libraries, including computer scientists, librarians, archivists, lawyers, economists, and publishers. He strikes a balance between praise and constructive criticism of these communities—an approach that is most refreshing. Undoubtedly, each community has much to offer in the development of digital libraries, but each group can also benefit from a broader and greater understanding of the other players. When describing librarians and computer science researchers, Arms states that "Until recently these two communities had disappointingly little interaction; even now it is commonplace to find a computer scientist who knows nothing of the basic tools of librarianship, or a librarian whose concepts of information retrieval are years out of date" (3).

In the chapter "User Interfaces and Usability," Arms reminds us that the library user must assess the effectiveness, utility, and ultimately the success of digital libraries. His discussion might have included more emphasis on educators and instructional designers as an additional community for consideration. While discussions of instruction are included, an explicit treatment of educators' interaction with digital libraries would reflect the