

ON THE DUBLIN CORE FRONT

by **Norm Medeiros**

Coordinator for Bibliographic and Digital Services
Haverford College
Haverford, PA

You're invited: XML's fifth birthday celebration

{A published version of this article appears in the 19:4 (2003) issue of OCLC Systems & Services.}

"The future ain't what it used to be." – Yogi Berra

ABSTRACT

This article looks at XML's first five years of existence. It reviews the original motivation for creating XML, and some of the applications this standard has made possible, specifically Rich Site Summary and Open eBook Publication Structure.

KEYWORDS

XML; Extensible Markup Language; World Wide Web Consortium; Rich Site Summary; RSS; Open eBook Publication Structure; OEBPS

Recently, I was asked by my director to give the library staff a presentation on the changing nature of cataloging. When I suggested that providing a history of the universe might be easier, he reminded me of my interview presentation years prior, which, like the Roy Tennant article he was holding, discussed the marginalizing of MARC in favor of other metadata schemes. After reflecting on my assignment and forming a committee to assist – my typical reaction when faced with anything unpleasant, laborious, or beyond me -- I came to consider this task more an opportunity than a burden. It gave me a chance to take stock of my colleagues' activities; to learn not only how they were using various metadata flavors, but far more importantly, why they had decided on these particular standards. What these schemes, like so many others, have in common is their carrier, XML.

XML TURNS FIVE

Quite by accident, I discovered that XML turned five years old on February 10, 2003. Considering my chief accomplishments at that tender age involved crayons and shoe laces, by comparison XML has had a prodigious first few years. Like an adolescent whom you have not seen since he was in diapers, XML had grown quickly. I remember reviewing the first XML recommendation while on a commuter train bound for Grand Central Station (it may have been a Tuesday), struggling with concepts such as element type declarations, literals, and validity constraint. It was such a

**"XML provides a new versatile structure for tagging and packaging metadata as the rapid proliferation of digital resources demands both rapidly produced descriptive data and the encoding of more types of metadata."
(Guenther and McCallum, 2003)**

divergence from the simple nature of HTML. Dave Holander and C.M. Spergberg-McQueen, two original members of the XML Working Group, offer a sentimental look back at the early years of XML in an article available on the World Wide Web Consortium site

<<http://www.w3.org/2003/02/xml-at-5.html>>. As they note, XML was invented in order to make SGML-encoded papers publishable on the Web – a simple ambition that has since spawned a revolution in the way information is managed across numerous communities (Holander & Spergberg-McQueen, 2003). The earliest reference that I can find to XML in *Library Literature* is from the June, 1997, issue of *Online & CDROM Review*. The article, entitled “A leaner, meaner markup language: simpler form of SGML called Extensible Markup Language (XML),” discusses the embryonic language and its ten guiding principles (A leaner, meaner markup language, 1997):

- XML shall be straightforwardly usable over the Internet
- XML shall support a wide variety of applications
- XML shall be compatible with SGML
- It shall be easy to write programs which process XML documents
- The number of optional features in XML is to be kept to the absolute minimum, ideally zero
- XML documents should be human-legible and reasonably clear
- The XML design should be prepared quickly
- The design of XML shall be formal and concise
- XML documents shall be easy to create
- Terseness in XML markup is of minimal importance

As noted in this early article, “[XML’s] main purpose is to position the Internet or intranet for a far wider range of document markup than just HTML”; this is to say a middle-ground position between the onerous coding required by SGML, and the content insensitivity of HTML.

XML APPLICATIONS TODAY

Numerous XML applications exist, many of which we use without ever knowing that XML is under the hood. A lengthy list of some of these applications is maintained by the Organization for the Advancement of Structured Information Standards (OASIS)

<<http://www.oasis-open.org/cover/xml.html#applications>>. Below I focus on two such applications that respectively have or soon will have an impact on library staff and users.

RICH SITE SUMMARY (RSS)

Rich Site Summary (sometimes referred to as RDF Site Summary) is an XML application that streams channels, commonly in the form of news feeds, to a computer via intermediary software. The end-

“XML is only a tool. Just as a word processor cannot write an interesting article by itself, XML cannot automatically find what people want, present information in an easy-to-read format, or solve problems relating to the content of information resources. XML can’t do your work for you” (Banerjee, 2002).

user experience has been described in recent articles (see for instance Cohen and Notess), but the underlying foundation of RSS, specifically its XML horsepower, is worth describing here. RSS was developed by Netscape in 1999 for use with their MyNetscape portal (RDF Site Summary (RSS) 1.0, 2000). RSS utilizes

the XML namespace feature to parse channel vocabularies. Namespaces are a convenient means of pointing to definitions of elements in order to facilitate context and use of these elements. An RSS document provides a title, link, and description of the channel it is describing. These metadata

provide the framework by which users select channels. The World Wide Web Consortium provides illustrative examples of RSS documents formatted in XML, along with descriptions and usage guidelines for RSS elements such as channel, image, and items <<http://web.resource.org/rss/1.0/spec>>. Use of RSS continues to grow as organizations of all types incorporate channels into their web architecture.

OPEN EBOOK PUBLICATION STRUCTURE (OEBPS)

OEBPS was developed by the Open eBook Forum, the international trade and standards organization for the ebook industry <<http://www.openebook.org/>>. The specification was released in 1999, and has undergone two revisions since then. The purpose of the XML-based specification is threefold (Open eBook Publication Structure Specification 1.2, 2002):

- to give content providers (e.g., publishers and others who have content to be displayed) and tool providers minimal and common guidelines that ensure fidelity, accuracy, accessibility in the presentation of electronic content over various electronic book platforms
- to reflect established content format standards
- to provide the purveyors of electronic book content (publishers, agents, authors et al.) a format for use in providing content to multiple reading systems

The Open eBook Forum is an organization consisting of 85 hardware and software companies, publishers, and authors, including the Association of American Publishers, Random House Inc., the American Library Association, and Palm Digital Media. It recognizes that electronic books will soon need to be rendered on an assortment of readers – devices that range in size, processing power, and operating systems. The OEBPS specification uses stylesheets to render XML-encoded data into suitable formats for particular reading devices. Each ebook described using OEBPS must contain a package file, a group of identifiers that describe relationships within the ebook. The package file, an XML document, consists of six entities that can be broadly defined as technical metadata. These include package identity, metadata, manifest, spine, tours, and guide. The metadata required in the package file is descriptive, and must consist of Dublin Core elements. An example of this metadata is the following (Open eBook Publication Structure Specification 1.2, 2002):

```
<metadata>
  <dc-metadata xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:oebpackage="http://openebook.org/namespaces/oeb-package/1.0/" >
    <dc:Title>Alice in Wonderland</dc:Title>
    <dc:Language>en</dc:Language>
    <dc:Identifier id="xyz" scheme="ISBN">123456789X</dc:Identifier>
    <dc:Creator role="aut">Lewis Carroll</dc:Creator>
  </dc-metadata>
</metadata>
```

Given the important purpose of this specification and the strong organizational backbone affiliated with it, OEBPS will no doubt become an increasingly important standard as electronic books become more popular.

CONCLUSION

XML has come a long way in a remarkably short period of time. It has enabled some innovative tools, including those noted above, and changed the foundation of World Wide Web development. The

amazing accomplishments XML will achieve in its future are left to the imagination, though this phenom's best years surely lie ahead.

REFERENCES

"A leaner, meaner markup language" (1997). *Online & CDROM Review*, vol. 21, no. 3, p. 181-184.

Banerjee, K. (2002). "How does XML help libraries?" *Computers in Libraries*, vol. 22, no. 8, p. 30-34.

Cohen, S..M. (2002). "Using RSS: an explanation and guide." *Information Outlook*, vol. 6, no. 12, p. 6-11.

Guenther, R. & McCallum, S. (2003). "New metadata standards for digital resources: MODS and METS." *Bulletin of the American Society for Information Science and Technology*, vol. 29, no. 2, p. 12-15.

Holander, D. & Spergberg-McQueen, C.M. (2003). "Happy birthday, XML." Available: <http://www.w3.org/2003/02/xml-at-5.html> (Accessed: 28 April 2003).

Notess, G.R. (2002). "RSS, aggregators, and reading the Blog fantastic." *Online*, vol. 26, no. 6, p. 52-54.

"Open eBook Publication Structure 1.2" (2002). Available: <http://www.openebook.org/oebps/oebps1.2/download/oeb12.pdf> (Accessed: 21 May 2003).

"RDF Site Summary (RSS) 1.0." (2000). Available: <http://www.purl.org/rss/> (Accessed: 20 May 2003).