Newsletter

ALA TechSource
www.techsource.ala.org

Library Systems™

An innovative overview of library automation

50 East Huron Street, Chicago, Illinois 60611-2795, USA

December 2002 Volume XXII Number 12

ISSN 0277-0288

XML transforms content

Growing in popularity, extensible markup language (XML) is reshaping the way librarians exchange content on the Web. XML offers flexibility in coding documents and promises longevity—two attractive qualities in today's ever-changing Web environment. A meta language, XML has a self-describing approach that allows coded documents to be migrated forward to future systems to ensure longevity.

XML is powerful because it increases the utility of a document by defining the structure of the content, without limiting the output to a particular format. Like documents tagged in other languages, XML-tagged digital documents can be stored only once and presented in multiple formats such as print, Web, or hand-held devices.

More importantly, XML has a greater number of available tools so custom programming isn't necessary. All major database systems have modules to import and export XML. And XSLT (extensible Stylesheet Language Transformations) can translate XML to HTML with no programming so you can display XML documents directly on the Web. Better yet, Web browsers are incorporating support for native XML so XSLT transformation will someday be unnecessary.

When exchanging information with other libraries, MARC works well because libraries all have it. But to

See **Transforms** *on page 3*

XML application examples from the 2002 LITA Forum

OPAC. The University of Buffalo Libraries created an experimental OPAC comprised of a Web page for each of 2.1 million MARC records on each cataloged item in its collection.

Portal. Auburn University Libraries used Dublin Core and XML to create a portal of finding aids on primary research materials in civil rights collections in libraries, archives, and museums throughout Alabama. North Carolina State University uses XML/XSL to present its list of subject-categorized electronic databases.

Digital archive. The University of Southern Mississippi uses specific markup standards coded in XML for data capture and presentation of items and objects in its digital library.—*JL*

Contact: University of Buffalo web catalog
http://libnet.buffalo.edu/lita
Auburn University
www.transformingamerica.org
NCSU
www.lib.ncsu.edu/eresources/dbfinder.html
Civil Rights in Mississippi Digital Archive
www.lib.usm.edu/~spcol/crda

IN THIS ISSUE

ONE-DAY TABLET PC CONFERENCE

The Open eBook Forum is hosting a one-day program Dec. 5 in New York City to focus on the software, hardware, and applications of the tablet PC in the higher education and professional publishing sectors, as well as with newspapers and magazines.

Contact: Open eBook Forum
www.openebook.org/tabletpc/agenda.asp

GUIDE TO XML SCHEMA FOR LIBRARY

In June the Library of Congress (LC) made the Metadata Object Description Schema (MODS) available for trial use. MODS joins MARCXML, METS (Metadata Encoding and Transmission Standard), and MIX (Metadata for Images in XML) as XML schema on their way toward becoming library standards. Librarians need to distinguish the four schema from one another and know the types of applications appropriate for their use.

MARCXML

MARCXML is a straightforward rendering of MARC 21 in XML, developed by the LC (www.loc.gov/standards/marcxml). MARC fixed and variable fields are treated as XML elements; MARC tags and indicators are given as attributes. Subfields are treated as subelements with the subfield code recorded as an attribute. A simple title field might translate to MARCXML as follows:

```
<datafield tag="245"
ind1="1" ind2="4">
<subfield code="a">The dog
:</subfield>
      <subfield
code="b">its behavior,
nutrition, and health
/</subfield>
</datafield>
```

MARCXML allows a lossless conversion from MARC to XML and back to MARC, known as a round-trip. It is potentially useful in any situation where a full MARC record exists but must be communicated in XML. For example, the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) requires metadata to be harvested in XML format. MARCXML can be used to expose full MARC records to an OAI harvester application.

MODS

MODS is a simplified rendering of MARC semantics in XML (www.loc.gov/standards/mods). MODS elements have simple mnemonic names, and MARC subfielding and indicator values are not explicitly recorded. A title might be represented in MODS as follows:

```
<titleInfo>
<nonSort>The </nonSort>
      <title>dog : its
behavior, nutrition and
health /</title>
</titleInfo>
```

MODS elements are highly compatible with the MARC 21 Bibliographic Format, and MODS documentation includes mappings from MARC to MODS and vice versa. Not all MARC data can be represented in MODS, though, and a MARC record, once converted to MODS, cannot be reconverted to MARC without some loss of specificity—round-tripability is not supported.

MODS can be used in many of the same situations as MARCXML, so long as some loss of content designation can be tolerated. MODS may also have some use as a format for original resource description that is simpler than but generally compatible with MARC data. It may be used by digital library projects as a richer alternative to Dublin Core, with defined elements for place of publication, notes, and other data not included in Dublin Core.

The LC's Digital Audio-Visual Preservation Prototyping Project is experimenting with using METS as a package for storing the digital audiovisual object and its metadata, and using MODS as the descriptive metadata schema (www.loc.gov/rr/mopic/avprot/avprhome.html).

The Web Preservation Project, also known as Minerva, is a collaboration among the LC, the Internet Archive, and university partners to capture and preserve open-access materials on the Web. MODS records are being created for websites in the collection and will be searchable through the Minerva interface. MODS records are expected to be converted to MARC for inclusion in the LC's catalog (www.loc.gov/minerva).

The University of Chicago Press (UCP) is developing the Chicago Digital Distribution Center for its own use and its client presses, with funding from the Mellon Foundation. Electronic books are stored in an online system called the BiblioVault for short-run printing. For each title, the UCP obtains an LC MARC record and converts it to MODS for storage in its own search and retrieval system. The bibliographic data will be distributed to client presses in MARC, MODS, or ONIX formats (http://cddc.uchicago.edu).

METS

Unlike MARCXML and MODS, METS, is not primarily for bibliographic description. It is a structure that includes many different types of meta-

METADATA

data pertaining to a particular digital object. METS has sections for both descriptive metadata and administrative metadata, such as technical characteristics and rights information. But the heart of METS is a set of elements, called the structural map, which defines the hierarchical structure of the object. Such structural metadata is needed, for example, to structure a set of page images into the chapters and sections of a book or to structure an audio file into segments of an interview.

METS itself does not define any elements of descriptive metadata such as author, title, or publisher, but it allows the inclusion of external extension schema. MARCXML and MODS have both been registered as extension schema for use with METS, as has an XML schema for simple Dublin Core.

METS is mostly used as a format for structural metadata in applications that handle the presentation and navigation of digital objects. METS also can be used as a transport syntax for the exchange of digital objects. An important use is as a format for archival information; several institutions are experimenting with METS as a format for submitting, storing, or distributing information within the Open Archival Information System (OAIS) framework.

METS is being developed by the Digital Library Federation and is maintained by the LC (www.loc.gov/standards/mets). An editorial board formed in June oversees METS development (www.loc.gov/standards/mets/news052902.html).

See **XML Schema** *on page 4*

Transforms *from page 1*

exchange data with publishers, vendors, or other nonlibrary parties, XML is more useful because these other groups are more likely to have XML parsers than MARC parsers.

XML has become popular for many reasons:

- XML is extensible. It defines how tags are created rather than defining a tag, so it works for many document types, such as text, voice, film, and those to be developed in the future.
- XML can be parsed by a machine and at the same time read by a person.
- XML is both computer platform and software application neutral, so it supports interoperability.
- XML uses Unicode, making XML able to handle international data with diacritics, special characters, and non-Roman data.

Library applications include:

- XML allows for multiple outputs including CD ROMs and Braille versions that make content accessible to some disabled people.
- XML is used for coding different types of content, enabling libraries to publish collections online and allowing users to select a customized version to view.
- Archival finding aids are being converted from SGML to XML, which is also browsable and searchable.
- With a continuously growing volume of interlibrary loans, XML can help process information among disparate systems.
- XML-encoded messages allow digital library and Web portal applications to exchange information and provide data in different forms, depending on the requestor.
- XML can help in the development of portals through the creation of metadata that can be used to provide navigation and content on a website.

Although XML can support the exchange of data between systems, in itself XML doesn't create a shared vocabulary, and that limitation is still an obstacle in the joint use of disparate systems.

XML was created to streamline the coding provided by SGML and to incorporate the popular syntax of HTML, which prepares documents to be Web accessible. Approved in 1998 by the World Wide Web Consortium (W3C), XML has since spawned a family of specifications including XSL (extensible Stylesheet Language), which addresses the appearance of the output, and Xlink (XML Linking Language), which enables a single link to reference related documents.—*Judy Luther*

Contact: Roy Tennant *XML for Libraries*
www.neal-schuman.com/db/0/290.html
Norm Desmarais *ABCs of XML*
www.newtechnologypress.com/ntp/inprint.html
Listserv *XML4Lib*
http://sunsite.berkeley.edu/XML4Lib/

NISO explores serials subscription information exchange

Increasingly, libraries, content aggregators, publishers, and third-party service providers are exchanging information about serials subscriptions. Several factors may be leading to this increase:

- An increase in the number of electronic journals available
- Libraries are purchasing software that does linking (such as SFX and Linkfinder plus) and need the information about what titles they subscribe to from whom
- Libraries are interested in portals that also need that information
- Interest in local development of e-resource management systems

A standard exchange format would benefit most parties involved, according to "The Exchange of Serials Subscription Information," a study commissioned by the National Information Standards Organization (NISO) with support from the Digital Library Federation. The study suggests that ONIX for Serials, an XML-based metadata scheme under development by publishers' organizations, provides a good foundation for such a format. It also recommends exploring a transaction-based (query/response) model for requesting and receiving serials subscription information.

NISO and EDItEUR, the international organization that maintains the ONIX family of standards, have established a Joint Working Party (JWP) to investigate using ONIX for serials subscription information exchange. The JWP will develop a pilot exchange of serials subscription information by the summer of 2003.—*PC*

Contact: www.niso.org/standards/resources/SerialsWP.html

Accessible learning technologies update

A new set of guidelines, "IMS Guidelines for Creating Accessible Learning Technologies version 1.0," helps librarians learn about online distributed learning. Its comprehensive guide to accessibility issues includes an accessibility primer and a section on using XML for accessibility. The document describes delivery of text, images, audio, video, and multimedia, as well as communication and collaboration tools, interactive interfaces, and testing and authoring tools. An appendix addresses legal issues. This document was issued jointly by the IMS Global Learning Consortium and the CPB/WGBH National Center for Accessible Media (NCAM), the research and development division of the Media Access Group at WGBH Boston, a National Public Radio affiliated station.—*PC*

Contact: CPB/WGBH National Center for Accessible Media
http://ncam.wgbh.org/salt/guidelines

XML Schema *from page 3*

Harvard University uses METS for the framework of an archival Submission Information Package (SIP) as part of a Mellon-funded project to archive scholarly ejournals. The OCLC Digital Archive uses METS for its Dissemination Information Package (DIP) when redistributing objects from the archive. It also intends to use METS as a SIP when it develops a batch ingest function (www.oclc.org/digitalpreservation).

The California Digital Library (CDL) is building a generic METS-based digital object repository. Content profiles specify how the repository should display records for the objects and how to extract information from the METS record for indexing. The repository will provide bibliographic search and display for books published online by the CDL on behalf of the University of California Press. For this project, metadata from MARC catalog records is mapped to MODS and inserted into the METS record, along with additional metadata contributed by the UCP.

In an unrelated project, the University of California at Berkeley is experimenting with converting large numbers of catalog records to MODS and loading them into Software AG's Tamino, an XML database system.

MIX

MIX is a schema developed by the LC as an XML representation of the data elements defined in the NISO draft standard Z39.87-2002 Data Dictionary—Technical Metadata for Digital Still Images (www.loc.gov/standards/mix). MIX does not include descriptive bibliographic information such as an image's creator or title. But it does include detailed technical information,

such as compression, color space, and file size.

MIX can be used as an extension schema for administrative metadata within METS. Its most important use will likely be in digital preservation projects, where precise descriptions of the technical characteristics of images are required.

ONIX

ONIX International, though not on its way to becoming a library standard, should also be included in any discussion of XML metadata for libraries. ONIX is a family of standards developed by publishers for communicating book and serials trade information in XML (www.editeur.org). ONIX may be most useful as a source of information to enrich library catalog records.

The ONIX Product Record format includes descriptive bibliographic metadata, sales data such as price and availability, and related information such as author biographies and book reviews. Although mappings have been created between ONIX and MARC, ONIX data elements are not highly compatible with MARC and do not support AACR2 cataloging.

The LC's Bibliographic Enrichment Advisory Team (BEAT) has extracted tables of contents from ONIX records and linked them to catalog records. Biographies, reviews, cover images, and other forms of enriched content can be derived from ONIX for display in library OPACs. ONIX, though, will not likely supercede MARC in the library environment. MODS and MARCXML are better formats for communicating MARC- and AACR2-compatible information in XML.—*Priscilla Caplan*

PDF ARCHIVING STANDARD IN DEVELOPMENT

Libraries, along with a growing number of international government and industry segments, are concerned about the long-term usability of the vast quantity of content available only in Adobe's portable document file (PDF) format. A new standards effort, referred to as PDF/A, was initiated in August 2002 by NPES, the Association for Suppliers of Printing, Publishing and Converting Technologies, and AIIM International, the Association for Information and Image Management. PDF/A will address the need to electronically archive PDF documents in a way that preserves content over an extended period of time.

The group's goal is to develop a standard for preparing archival PDF documents and associated metadata for archiving that will ensure the documents can be retrieved and rendered predictably into the indefinite future. The International Organization for Standardization (ISO) will develop the proposed standard.—*PC*

Contact: AIIM PDF Archive Committee
www.aiim.org/standards.asp?ID=25013

Technology grants listed in new publications

Librarians can locate technology funding available from corporate technology funders, government, and trade associations through three publications offered by Technology Grant News, a service of Partnerships for Community Inc., to help nonprofit and educational organizations.

The main publication, also called *Technology Grant News,* is printed four times a year at an annual subscription of $85 for nonprofit organizations. It covers grants available to nonprofits, educational institutions, libraries, and museums. It includes grants for digitization, network access, and preservation.

The Supplement of Corporate Technology Funders is a brief guide to more than 110 of the leading high-tech corporate funders, those in business who profited from the market boom of the last 20 years and tend to fund new projects without lengthy review.

The Directory of Current, Ongoing & Open Technology Grants includes government, corporate, and professional association funders.—*PC*

Contact: Technology Grant News
www.technologygrantnews.com

New open-source software available

The number of open-source software products available for library applications continues to increase. The University of Winnipeg Library expects to issue the first release of OpenILL, an open-source interlibrary loan system, this winter. OpenILL will be fully compliant with the ISO 10160/10161 ILL standards and will interoperate with GODOT, an open-source resource-sharing system that provides discovery, linking, and document request functions.

In October, open-source software for portal development became available that requires only a minimum investment in technical resources and expertise. The software, called Scout Portal Toolkit (SPT), includes a Web-based configuration tool that does not require any knowledge of HTML, Unix, or other technical details. The portal interface allows users to search individual metadata fields or sets of fields, and to enter annotations and quality reviews of resources. SPT was funded by the Andrew W. Mellon Foundation and released by Internet Scout Project (beta version 0.9.7).—*PC*

Contact: University of Winnipeg Library
http://cybrary.uwinnipeg.ca/projects/openill
Scout Portal Toolkit
http://scout.cs.wisc.edu/research/SPT

Perception vs. Reality

Study finds gap in students' view

College students' expectations for library service don't match their perceptions about the service they receive for Web-based information, according to a Harris Interactive study conducted for OCLC. The majority of students:

- Prefer remote Web access to library resources
- Would like face-to-face interaction for help, but only 21% would ask a librarian
- Use campus library websites
- Use the library's print-based resources
- Feel advertising doesn't influence their information choices
- Aren't ready to pay for information access
- Want to search other libraries' collections

Their perceived barriers include:

- The inability to access databases remotely
- Difficulty searching and navigating within the library and its website
- Costs of copying and printing
- A shortage of knowledgeable librarians
- A lack of customer orientation

In June 2002, OCLC released these study results in a white paper, "Information Habits of College Students looking at How Academic Librarians Can Influence Students' Web-based Information Choices." The paper suggested how librarians might reach more students, such as:

- Create access points using popular Web locations, such as search engines, portals, and the campus bookstore.
- Deliver websites that directly present and connect users to information from many databases and sources.
- Grant reciprocal borrowing and enable remote circulation with nearby libraries.
- Provide easier remote access.
- Deliver personalized, interactive service.
- Create a customer-friendly experience to match merchant and consumer websites.—*JL*

Contact: OCLC
www2.oclc.org/oclc/pdf/printondemand/informationhabits.pdf

WRITEABLE TABLET PC INTRODUCED

Released in November with much fanfare, Microsoft's Windows XP Tablet PC Edition, a superset of Windows XP Professional that includes Windows Journal, features pen-based computing. Windows Journal software allows users to store handwritten notes directly on their tablet PC using a stylus and to enter data directly into Word, Excel, PowerPoint, or Outlook documents.

Handwriting recognition is a debated feature and opinions vary about how well this feature works. Critics say the machine trains the user to write clearly instead of adapting to the user's handwriting. This software release has prompted the design of a new variety of laptops with electromagnetic pens and pen-sensitive screens that generate a greater flow of digital ink with increased pressure on the screen but don't respond to human fingerprints.

Acer TravelMate 100, Fujitsu Lifebook P-1000, and Viewsonic Tablet PC V1100 lead the development of fully functional tablet PCs. Acer is introducing an ultralight convertible PC with a screen that rotates and snaps flat over the keyboard. But Dell, Gateway, and IBM aren't producing tablet PCs; they aren't convinced the market is strong enough. Sony and IBM withdrew their touch screen PCs earlier in 2002.—*JL*

Contact: Microsoft
www.microsoft.com/windowsxp/tabletpc
Acer
http://global.acer.com/products/notebook/tmc100/index.asp
Fujitsu
www.fujitsupc.com/www/penworld/technology.shtml?tabletpc
Viewsonic
www.viewsonic.com/products/tablet_pc_tabletpcv1100.htm

Desired usage statistics DETERMINED

More than 600 librarians, the majority from academic libraries in North America, voted on desired usage statistics in COUNTER's Web-based survey that ran from July to mid-August 2002. COUNTER is the U.K.-based initiative, Counting Online Usage of Networked Electronic Resources, that is working to create a usage statistics code of practice for publishers and librarians by December 2002.

According to the survey results, the most important statistic is the number of requests for full-text articles by journal and month, followed by the total number of searches and sessions. The majority of libraries want to know the number of journal titles requested that are both subscribed and not subscribed by each library.

Data are needed for each institution in a consortium, as well as by consortia. Although the majority has not correlated individual departments with IP ranges, those libraries that have done so want statistics reflecting that level of detail. For the majority, though, distinguishing onsite and offsite usage is more important.

Librarians want to be notified by e-mail each month as reports are ready and want to download them into Excel. They prefer to have one to two years' worth of data stored on the publisher's website.

In the comments field of the survey, librarians indicated they would like to know the price of each journal and the publisher as part of the data presented.—*JL*

Contact: COUNTER
www.projectcounter.org

Correction
The description of the USA PATRIOT Act included in the November issue was missing some information. The final bill, H.R. 3162, incorporated provision of the two earlier anti-terrorism bills: H.R. 2975, which passed the House on Oct. 12, 2001, and S. 1510, which passed the Senate on Oct. 11, 2001. Provisions of H.R. 3004, the Financial Anti-Terrorism Act, were also incorporated into H.R. 3162, as Title III. So this act ultimately became the Uniting and Strengthening America Provide Appropriate Tools Required to Intercept and Obstruct Terrorism Act—or, the USA PATRIOT Act.

Library Systems Newsletter
50 East Huron Street
Chicago, IL 60611-2795 USA

ALA TechSource
www.techsource.ala.org

NON PROFIT
US POSTAGE
PAID
PERMIT 3226
CHICAGO, IL

All systems go @ your library™

Library Systems Newsletter delivers hard data and innovative insights about the world of library technology, every month.

Publisher
Miriam Wuensch
800-545-2433, ext. 3252
312-280-3252
mwuensch@ala.org

Editor
Chris Santilli
630-495-9863
chris@wordcrafting.com

Contributing Editors
Priscilla Caplan
352-392-9020, ext. 324
pcaplan@ufl.edu

Judy Luther
610-645-7546
judy.luther@informedstrategies.com

Andrew Pace
919-515-3087
apace@unity.ncsu.edu

Administrative Assistant
Judy Foley
800-545-2433, ext. 4272
312-280-4272
jfoley@ala.org

Production and design by Angela Hanshaw, American Library Association Production Services.

Library Systems Newsletter is published monthly by ALA TechSource, a unit of the publishing division of the American Library Association.