

Catalogers Group Minutes
July 3, 2003

Present: S. Benamou, V. Bross, B. Feinberg, S. Layne, J. Lopear, N. Norris, T. Marra, J. Matthieson, D. McGarry, L. Mendes, C. Miller, J. Morehead, H. Phan, J. Riemer, A. Riggio (recorder), R. Stumps
Guest Speaker: Howard Batchelor

Subject of Presentation:
METS (Metadata Encoding & Transmission Standard)

METS is sponsored by the Library of Congress. An excellent way to get information about METS is to visit the METS home page, hosted by LC:
<http://www.loc.gov/standards/mets/>

Two essential facts about METS:

1. If you don't know XML, you don't know METS
2. MoA2 project drove the development of METS: a need for standards was expressed by the major digital players (Harvard, UC, MIT, etc.) when non-text based digital objects began to be transmitted over the Web. There became a clear need to include administrative metadata with the digital object, and to have accompanying standards for that metadata.

METS, like Dublin Core, is now at the point where it is growing and becoming more and more complex.

Batchelor anecdotally mentioned two statements that he overheard at the DLF Spring Forum last May:

1. From David Seaman of DLF: "We all have a digital library, and we all have the same one."

Comment: people create digital objects, but we have no central repository—the objects are not accessible

2. From Steven Abrams, Harvard: "It will be easier to teach catalogers XML than to teach programmers cataloging."

Comment: If we could do more cataloging in digital libraries, there would be fewer problems

Batchelor then discussed the concept of digital libraries as "silos." Often digital files and their associated metadata are being stored, but not shared or transferred.

Some associated issues:

1. What standards are being used when creating digital objects?
2. What standards are being used when storing digital objects?
We need to store digital information on non-specific platforms, such as: TIFF files instead of JPEG files; WAV files instead of Real Audio files.
3. Can we afford the high price of metadata? How can we afford the digital future?

SCORM (Sharable Courseware Object Reference Model) has recently come into the digital picture at UCLA for courseware management. According to the SCORM Web site, "SCORM is a suite of technical standards that enable web-based learning systems to find, import, share, reuse, and export learning content in a standardized way."

<http://projectinterconnect.org/scorm/scindex.htm>

It remains to be seen how/if METS and SCORM will work together.

METS is effective in the sharing of metadata via the OAI (Open Archives Initiative) standard.

METS also complies with OAIS (Open Archival Information System). METS records can exist as part of SIP (Submission Information Package), DIP (Dissemination Information Package), and AIP (Archival Information Package) files.

End user accessibility to digital files can be based on the views enabled by programs written "on top" of METS records.

Batchelor talked in general terms about "metadata." The core of the METS record centers on structural metadata, which is the only *required* metadata in a METS record.

He briefly explained the history of XML (eXtensible Markup Language), which serves as the structural foundation of Web applications and information and storage. He compared HTML (Hypertext Markup Language), which describes data in terms of its presentation on a Web page, with XML, which describes content based on data. XML must be validated against a DTD (document type definition), unlike HTML. The Web has little trouble accepting HTML files (even miscoded data), as they do not require that sort of validation.

There are several advantages and disadvantages to using HTML in a library setting. However, XML offers flexibility, structure, its acceptance as an international standard, and its ability to be read by other machines.

An XML file consists of a "root element," which describes a source; "entities," which are characteristics of that source; and "attributes," which describe aspects of the characteristics.

The Document Object Model (DOM) allows application program interfaces to be written. The display of XML relies on style sheets (called "XSL") which reside separately from the XML files.

As a real-life example, Batchelor showed an MoA2 object, the Patrick Breen Diary. METS carries metadata separately, and allows sharing of files. Disparate files can be assembled and used on the fly, while metadata describes the digital object location, features, etc.

MARC can be incorporated into an XML structure, as can Dublin Core and VRA Core. METS records can point to MARC data, can ingest MARC, or reinterpret MARC data.

A METS record is comprised of six sections: the header, descriptive metadata, administrative metadata, files, structural map, and behaviors.

Batchelor ended his presentation, and some discussion followed. We returned to the first slide of his presentation:

What is the difference between

1. A cataloger
2. A librarian
3. A digital library employee

Batchelor sees the cataloger as the producer of metadata, the librarian as one who serves the customer, and the digital library employee as the one who creates digital content. It was left for the audience to ponder this interpretation ...