Editorial

METS and the metadata marketplace

Michael Seadle

The author

Michael Seadle is Editor of *Library Hi Tech*.

Keywords

Libraries, Information, Data, Marketing

Abstract

One purpose of the Metadata Encoding and Transmission Standard (METS) is to deal with the multiplication of metadata types in recent years, and especially metadata that deal with non-paper materials, including audiovisual sources and their digital representations. In that sense, it is a kind of meta-metadata. But is it needed? Market forces may decide.

Electronic access

The research register for this journal is available at
http://www.emeraldinsight.com/researchregisters

The current issue and full text archive of this journal is available at
http://www.emeraldinsight.com/0737-8831.htm

Introduction

A friend asked me recently to take part in a panel discussion about metadata formats for audio preservation, and particularly to talk about the Metadata Encoding and Transmission Standard (METS) (http://www.loc.gov/standards/mets/METSOverview.html), since one of my research partners on the Digital Library Initiative-funded National Gallery of the Spoken Word participates in its development. My own staff use Encoded Archival Description (EAD), Text Encoding Initiative (TEI), Dublin Core (DC) and an eXtensible Mark-up Language (XML) Document Type Definition (DTD) developed locally for a project to encode nineteenth century US cook-books (http://digital.lib.msu.edu/cookbooks). We also work closely with our technical services department, which uses MAchine Readable Cataloging (MARC) records. One more metadata type seemed like overkill.

Yet to some extent the purpose of METS is to deal with the multiplication of metadata types in recent years, and especially metadata that deals with non-paper materials, especially audiovisual sources and their digital representations. METS has four sections: descriptive metadata, administrative metadata, file groups, and structure map. The first of these allows external references to other forms of metadata, or their physical incorporation inside a METS "wrapper". The other three sections include information that other metadata types may, but do not necessarily, encode. In other words, METS represents a kind of meta-metadata.

METS was developed for a purpose. It is an integral part of the Open Archival Information System (OAIS) reference model (http://ssdoo.gsfc.nasa.gov/nost/isoas/ref_model.html), and serves to facilitate the exchange of digital objects from one repository to another. But is it needed?

Transformations

A number of projects use cross-walks to transform one form of metadata to another. For

255

example, at Michigan State University we use an Innovative Interfaces XML Harvester tool (http://www.iii.com/html/products/p_web.shtml#xml), that we helped develop to convert the collection level information in our EAD records into MARC communication format. This lets us load records into our online catalog that then point back to the more detailed EAD record. Other projects have developed cross-walks that convert DC records into MARC, and take the header information from TEI into MARC. None of these transformations is easy. EAD, for example, handles the date information for personal names differently from the MARC 100 field, and it has a very different philosophy for subjects. DC simplifies the basic set of MARC tags greatly, which tempts coders to enter a broader range of information than would be acceptable in many well-regulated MARC databases.

METS avoids the transformation issue by taking descriptive metadata in whatever format and with whatever local rules or specific standards metadata follow. This liberal approach makes sense in an environment where standards are evolving too rapidly for cross-walk development, and where the point is to ingest some form of description without misleading implications from overly simplistic copying of one field to another in a format with different standards.

For many of the same reasons, it seems ill-suited to searching a large repository with any precision. The definitions of the fields across the various metadata types remain incommensurate. The degree to which that is a genuine problem remains untested, since METS databases are still in the prototype phase.

Information gaps

The information that METS collects about administrative metadata, file groups, and structural mapping represents areas that are often neglected, because the originating archive has local sources, or the information seems obvious, or the information takes extra work to gather. Including explicit sections for this information in the METS specification should help remind people about its importance for files with a life outside their place of origin. Of course it does not guarantee that the information will be complete, or it will be right, or it will agree with the EAD, DC, or MARC version of the same information, which could well be neatly wrapped inside (or referred to by) the descriptive portion of the METS record.

The latter problem need not occur if people are careful, and if standards-makers develop appropriate cross-walks that keep the information in sync. That may, however, be too much to expect. Even within fairly tightly regulated databases like an Anglo-American Cataloging Rules, 2nd ed. (AACR2) MARC catalog, encoding practices show the usual human tendency toward inventiveness. Without a relatively mechanical means for copying, say, the intellectual property information from an EAD record in the descriptive section to the administrative section of the METS record, the contents will tend to diverge, especially after multiple updates. Then the question of which encoding is more authoritative begins to arise.

There are some easy solutions, such as not encoding information elsewhere that can better be represented in METS, and perhaps even removing it from other metadata types. From a METS viewpoint, that seems logical. But an EAD or MARC proponent might balk, and perhaps argue in favor of enhancing the already-existing options for administrative, file, and structural information in the EAD or MARC specification.

Conclusion

To a large extent, the metadata marketplace will answer the question of whether METS is really needed. That market has certainly changed in the last decade. MARC remains dominant among libraries, but DC has acquired a substantial following, EAD has attracted interest outside archives, and TEI is widespread among text-encoding projects. That degree of change suggests that METS has a chance, and it certainly has some powerful backers, including the Library of

Congress. But the metadata market may also be saturated.

In part, the success of METS will depend on the success of the OAIS reference model which uses it. That model has some well established competitors, in particular the Open Archives Initiative (OAI) metadata harvesting, though OAI could in principle use METS too.

If you want to know more about METS and other metadata standards, including EAD, consult the official Web site at the Library of Congress (http://www.loc.gov/standards/).