What Should We Teach about METS in a Digital Preservation Course?

Jim Waters and Robert B. Allen
College of Information Science and Technology, Drexel University
James.Waters@drexel.edu rba@drexel.edu

Abstract
We describe a METS (Metadata Encoding and Transmission Standard) class assignment which is part of a Digital Preservation course. The assignment requires students to both critically evaluate this important framework and apply it practically to metadata management for digital objects. The results indicate that it was a valuable assignment for LIS and IS students who could conceptually grasp METS readily, but that many have trouble with integrating external metadata schemes and with XML syntax. These results are informing a redesign of the assignment.

Background
The Drexel Master's-level Digital Preservation course attempts to provide a comprehensive overview of digital preservation for both Librarians and IS professionals. This is newly developed course upper division course and the students should have a general foundation in both metadata and XML.

The course covers the underlying preservation principles common to both physical and digital materials but expands on the different strategies, approaches and technologies required for dealing with digital materials. This course covers a broad range of repository issues ranging from toolkits (e.g., Archivist Toolkit, DSpace), to multimedia (e.g., MPEG-7), to service specification (e.g., LOCKSS and iRODS), but also for metadata which is applicable to digital repositories. In particular, we introduce METS (Metadata Encoding and Transmission Standard). METS is important as part of the OAIS framework both of which are growing in popularity. METS was an outcome of the Making of America 2 initiative [2].

The Metadata Encoding and Transmission Standard (METS) occupies a middle ground between the abstract package definitions derived from the OAIS framework and the low level data definitions found in the NISO extension scheme as illustrated in Figure 1.. It provides a critical nexus between a conceptual framework and implementation issues. Because the schema is implemented in XML, it also provides practice for students with high-level technical validation issues.

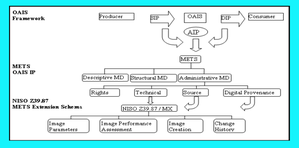

Figure 1:Library of Congress standards framework showing the relationship of OAIS, METS and the NISO protocol Z39.87 from [1]

Assignment
Since we aim to help educate both future researchers and practitioners, we are concerned that they should be equipped with the willingness to challenge accepted practices. Thus, we believe that students should be familiar not just with METS and OAIS on a conceptual basis but that they also have experience with the METS schema and practice with creating realistic METS documents in XML, a valuable skill. By using a METS framework for a realistic task we hoped students would gain perspective as to what is and is not useful about METS. However, we still need to avoid the assignment being too broad or too detailed.

We decided to focus this assignment on three sections. (Administrative Metadata, File Group and Structural Map) The Descriptive Metadata section is frequently used in conjunction with external schemas such as MARC and the Behavior section is seldom used in practice and refers to interface issues beyond the scope of this course, thus these two sections were omitted from the assignment though they are covered in the course.

For this assignment we will concentrate on:
- Administrative Metadata Section,
- File Group Section, and
- Structural Map Sections,

Create a simplified METS document containing only the three sections mentioned above (Administrative, File, Structure Map) to describe a specific (named) Digital object. This object can be derived from a Book, Historical Document, Movie, Video, Painting, Photograph, Sculpture, other work of art, or Person.

Don't get too worried about the precise XML syntax for this exercise, but make sure the document describes the object. Since many METS elements/attributes are optional use your judgment and only include those that would be useful for your object

Discuss the METS scheme with respect to:
How well does the METS scheme allow you to describe **your** target object?
How useful would your document be for retrieval or playback applications with respect to your chosen object?
Are there are any elements/attributes you would add to the METS scheme to make it more useful?
Are there are any elements/attributes you would remove from the METS scheme?

Finally choose a specific example of a second (different) object type from the list above and **briefly** (1 to 2 paragraphs) discuss how well the METS scheme deals with that type of object.

Fig 2. Text of the assignment

Moreover we hoped to encourage the student to critically evaluate the value of such standards. METS imposes a structural framework for metadata creation, we wanted students to reflect on how useful was **this** structure in particular and structured approaches in general? Similarly, we chose to concentrate on the overarching conceptual structure. The schema for a METS document is comparatively complex and while our students are familiar with XML they are not all equally technically competent and we have found that those with greater technical ability will chose to be more technically accurate anyway and those with weaker technical skills should not be distracted by syntax.

Results
Nine students completed the assignment. Six of those nine students understood the ability of METS to use wrappers for other metadata schemata to overcome limitations in the base scheme. One however did not define the external metadata scheme. While most students understood the relationship between the **structmap** and the **file section** on a conceptual level there were notable weaknesses found here. Some students created file pointers that did not agree with files in the file section. Some students incorrectly nested Divs, others created maps which simply listed a structure without relating it back to the file section. Many students had trouble understanding the use of the area element including confusion over how to label, begin and end areas, sometimes having literals for time segments. Other students created area elements representing the entire object or had beginning and end points that were inappropriate labels such as literal labels for parts of an image not coordinate pairs

Several students remarked that the assignment was difficult and the XML syntax tricky. Most students felt the METS schema was intrinsically flexible and powerful due to the ability to pull in external schemata as required. Some though failed to understand that many METS elements were optional; despite being informed so explicitly. Similarly METS' strength for representing complex or multi-representational objects was noted but sometimes METS was criticized for being difficult to apply to simpler objects. Overall, students tended to be moderately positive towards METS.

Some students though failed to understand that many METS elements were optional; despite being informed so explicitly. Similarly METS' strength for representing complex or multi-representational objects was noted but sometimes METS was criticized for being difficult to apply to simpler objects.

Comments from Students

The METS scheme looks like it is set up as a basic outline for metadata storage. There is little standardization on what information can be used to describe the target. This is a double-edge sword, as there is great flexibility in what can be entered, but the lack of standard wording may cause retrieval problems and potentially interoperability problems

I do not see any advantage for retrieval that standard title, author, and subject metadata fields couldn't achieve. The METS scheme does have an advantage for playback, because the structural map allows the metadata to identify and point to specific movements within a single composition.

Sometimes there are more subtle interconnections so I wish to have better choices how to incorporate longer description into the document (like scope note in EAD but with relation to other materials or maybe "the readers who like this book may also like" link).

METS might be improved by adding more finding aid-like information like location of the physical objects, the history of the creation of the document, and other providence information. Overall, it could be improved (or be more readable to human eyes), if there were more free text sections, although some sections can be adaptable to this.

We consider that the assignment was a moderate success. While students had problems implementing the METS schema they generally grasp how it worked conceptually. It seems that METS which was part of one week's teaching required rather more coverage and this will be reflected in future versions of the assignment.

IMPLICATIONS AND EXTENSIONS
This assignment represents one facet of our broader effort to introduce archival repositories and their metadata issues. METS is based on OAIS. Future plans may include the use of JHOVE [4] for validating METS documents, specifically METS-ALTO documents and an exercise relating to Z39.87. The Z39.87 exercise will require students to both create metadata records for digital still images and critically evaluate the usefulness of the framework with respect to organization, and retrieval issues.

ACKNOWLEDGMENT
This project was supported by the Institute of Museum and Library Services (IMLS) Grant RE-05-05-0085-05 on developing a Model Curriculum for the Management of Digital Information.

References
[1].Waibel, G. *Like Russian Dolls: Nesting Standards for Digital Preservation*. RLG DigiNews 2003 10/10/2008]; Available from: www.rlg.org/preserv/diginews/diginews7-3.html.
[2].LOC. *METS and Overview and Tutorial*. 2003 10/10/2008]; Available from: http://www.loc.gov/standards/mets/METSOverview.html.
[3].LOC. *NISO Metadata for Images in XML Schema*. 2008 10/14/2008]; Available from: http://www.loc.gov/standards/mix/.
[4].JSTOR_Harvard_College. *JHOVE - JSTOR/Harvard Object Validation Environment*. 2007 10/13/2008]; Available from: http://hul.harvard.edu/jhove/.