University of Nebraska - Lincoln
DigitalCommons@University of Nebraska - Lincoln

Library Philosophy and Practice (e-journal) | Libraries at University of Nebraska-Lincoln

8-1-2011

METS: A Survey of Recent Literature and Applications

Lindsey Wagner
lnwagner@gmail.com

Follow this and additional works at: http://digitalcommons.unl.edu/libphilprac

Part of the Library and Information Science Commons

Wagner, Lindsey, "METS: A Survey of Recent Literature and Applications" (2011). *Library Philosophy and Practice (e-journal)*. Paper 642.
http://digitalcommons.unl.edu/libphilprac/642

StatCounter.com RSS Add to Google

http://unllib.unl.edu/LPP/

Library Philosophy and Practice 2011

ISSN 1522-0222

METS: A Survey of Recent Literature and Applications

Lindsey Wagner

Introduction

METS is a less than ten-year-old metadata standard which can be used for both preservation and transmission of digital objects and their metadata. It uses XML to describe descriptive, administrative and structural information about the object. It is well known for its flexibility which can be both a positive and a negative. It allows institutions using METS to insert the metadata they want and may already be using into the METS document, but it also is a detriment to interoperability since institutions are implementing METS in very different ways. This makes METS documents created by one institution unusable for another institution that implements METS in a different way. While METS is a tool which can be used for transmission of metadata, the fact that there is no standard implementation actually impedes the transmission of metadata. This paper will include a discussion of METS as well as a summary of the research that has been done.

METS History

METS began as a project called Making of America II (MoA II). MoA II built on the work of the University of Michigan and the Cornell University libraries which had created a collection of digital books and serials on the theme of the Making of America (The Digital Library Foundation, 2009). At the same time, the Making of America project attempted to reach a consensus on how best to handle digitized materials (Cornell University Library, 2010). The Digital Library Federation (DLF) and the University of California at Berkeley led Cornell University, the New York Public Library, Pennsylvania State University, and Stanford University in the creation of a testbed which would produce a system for looking into metadata issues and finding solutions ultimately leading to a standard or standards (Hurley, Price-Wilkin, Proffitt, & Besser, 1999).

The MoA II project created an XML DTD which specified encoding for descriptive, administrative, and structural metadata for textual and image-based objects (Library of Congress, 2009). Testbeds were then created using that DTD. Several shortcomings were identified and the MoA II community decided to create a new version which would instead be an XML schema and would not prescribe any descriptive or administrative metadata. In 2001, Jerome McDonough, the primary author of the MoA DTD, finished the schema which then became METS.

The Library of Congress Network Development and MARC Standards Office (NDMSO) became the maintenance agency and a METS editorial board was created with McDonough as its chair (Cundiff, 2004). McDonough is still on the board but is no

longer chair. Currently, Nancy Hoebelheinrich is the Administrative Chair and Rick Beaubien is the Technical Chair (Library of Congress, 2010).

What is METS?

METS is a standardized way for libraries and other institutions to create and share administrative, structural and descriptive metadata. METS itself is a container which links the different kinds of metadata together. METS does not define the XML syntax and vocabulary of the administrative and descriptive metadata. Instead, an XML document can be inserted into the proper place in the METS document or a link to external metadata in any format can be placed within the METS document. Any XML schema can be used but some have been endorsed by the METS editorial board (Cantara, 2005).

XML is commonly used for encoding metadata in both traditional and digital libraries. This is because it is not a proprietary standard; it is an open standard registered with the International Standards Organization which means that using it does not force an institution to use a specific software application. It is also simple and flexible making it easy to make it do what is wanted or needed (Gartner, 2008). Another reason why XML is a good fit for METS is that the use of namespaces in XML allows for inputting metadata schema other than METS into the METS document (Cundiff, 2004). Using namespaces allows two schemas with elements of the same name to coexist in the same document since they will be prefixed with that namespace (Gartner, 2008).

METS Structure

The first element in a METS document is the <mets> root element. A root element is required in XML. The METS root element has five optional XML attributes: two, ID and OBJID (Object Identifier), identify the METS document; TYPE specifies what type of object is being documented; LABEL identifies the title of the object; and PROFILE which specifies which, if any, registered METS profile is being used. The root element also contains namespace declarations which declare all of the metadata schemas used in the METS document along with the associated namespace prefixes.

There are seven subsections in a METS document and only one, the structural map, is required. The first subsection is the METS Header (<metsHdr>) which contains mainly information about the creation of the digital object. The METS Header is followed by the Descriptive Metadata (<dmdSec>) subsection which is self-describing. The METS schema does specify the XML vocabulary and syntax for this section, but the METS editorial board has endorsed Dublin Core, Metadata Object Description Schema (MODS), MARCXML MARC 21 Schema (MARCXML) and VRA Core (Library of Congress, 2008).

The third section is Administrative Metadata (<amdSec>) which has four subsections: Technical Metadata (<techMD>), which contains information about creation, format and use; Rights Metadata (<rightsMD>), which contains information about intellectual property rights and licensing; Source Metadata (<sourceMD>), which contains information about the analog source of the digital object; and Digital Provenance Metadata (<digiprovMD>), which contains information about file relationships, migration, transformation and other administrative decisions affecting the digital object. Like the Descriptive Metadata section, there is no prescribed syntax and vocabulary for this section, but there are endorsed schemas: textMD (Schema for Technical Metadata with Text), NISO Technical Metadata for Digital Still Images Standards Committee (MIX), and Preservation Metadata (PREMIS) (Library of Congress, 2008).

The File Section (<fileSec>) follows and contains an inventory of all files which are a part of the digital object. The Structural Map (<structMap>) comes next and is the only required section of METS. The Structural Map outlines the hierarchical structure of the object by using a series of nested elements. The sixth section is the Structural Map Linking Section (<structLink>) which provides links between any two div elements

in the Structure Map. Last is the Behavior Section (<behaviorSec>) which contains information about executable code that is associated with the object (Cantara, 2005).

Flexibility

The flexibility of METS is a great benefit, but it is also a major drawback. Because METS does not prescribe the XML syntax and vocabulary or even the content of the administrative and descriptive metadata, it can fit the needs of many different institutions. It allows the metadata created before an institution begins using METS to be input into the METS document without having to transform the data from one standard to another (Seadle, 2002).

Its flexibility also means that METS documents are less interoperable than documents created with a more rigid standard. Because of the different ways METS is put into practice, an institution cannot import just any METS document for the digital object in their collection. The exporting institution would have to use the same implementation of METS or the importing institution would have to transform the data so that it works with their system.

Because of this, institutions have begun creating institutional profiles which contain an institution's implementation of METS (i.e., which METS sections they use in their documents and the metadata schemas they use for the descriptive and administrative metadata), so that METS documents can be exchanged between digital collections in that institution and any other institution that decides to use that particular METS profile (Pearce, Pearson, Williams, & Yeadon, 2008). METS profiles can be registered with the Library of Congress Network Development and MARC Standards Office, so that they can be shared with other institutions (Library of Congress, 2007). This allows other institutions to use the same profile as another institution which they would like to import METS documents and digital object from.

Because the flexibility of METS allows for the use of any metadata schema for its administrative and descriptive metadata, redundancies can occur. For instance, some information can be placed in both the administrative metadata and the structural metadata sections. The institution must decide whether to keep those redundancies, and if they decide not to, they must decide on which section the information belongs in (Gartner, 2008). There are some benefits to keeping redundancies. A repository may want to leave them in since the institutions importing digital objects and metadata from them may only use METS as a way to transmit metadata. Other institutions may want use METS to preserve their objects and would want to keep the data within METS.

Federated Searching

Gartner (2008) believes that digital library technology could lead to databases where people can search the collections of more than one institution at a time. It would be similar to a union catalog, but the user would be able to access the digital object instantly. METS can allow for federated searching, but in order to do so all of the collections being searched will most likely have to use the same external metadata schemas and content standards. It may not be possible for a system to process the different types of administrative and descriptive metadata which can be used with METS. Creating a more rigid standard for METS would allow for federated searching across all digital collections that would want be a part of this database (Gartner, 2008).

Using PREMIS in METS

PREMIS is one of the metadata schemas endorsed by the METS for the administrative metadata subsections, but there are many choices regarding exactly what to put where. PREMIS has four types of entities: object entities, which contain mostly technical metadata; event entities, which contain the actions done to the object; rights entities, which contain information about the intellectual property rights of the object;

and agent entities, which contain information on the agents which are related to the object in some way.

The rights entities fit neatly into the METS Rights Metadata, but the other entities do not fit as well into one section of METS. The object entities can include information which would fit into both the Technical Metadata portion of METS and the Digital Provenance section. The events entities would for the most part go under the Digital Provenance section, but some would also fit into Technical Metadata. The agent entities can be associated with either the Digital Provenance or Rights section depending on whether the agent is associated with an event or with the rights of the object. Some institutions may prefer to keep all of the PREMIS metadata together in the METS document which is possible, but the institution will have to decide which subsection of the Administrative Metadata section the PREMIS metadata will be placed in.

Another issue in using PREMIS with METS is redundancy. Most of the redundancies are PREMIS technical entities which are also attributes within the METS Structure Metadata. There can be some advantages of using one over the other because, in some cases, one will provide more information than the other with additional XML elements or attributes. Another redundancy issue involves using Metadata for Images in XML Schema (MIX) along with PREMIS. MIX is input into the Technical Metadata subsection and can be used with PREMIS in a second Administrative Metadata subsection or MIX can be added to the PREMIS metadata by using the element <objectCharacteristicsExtension>. Each institution will have to decide in which section they would like to place these redundancies.

The PREMIS Maintenance Activity is working on guidelines for using PREMIS with METS. The group, which consists of METS and PREMIS implementers from a variety of institutions, has had difficulty in balancing flexibility with stricter standards which would promote interoperability. They are also taking into consideration how METS is used since it can be both a method of delivery and a way in which to store and preserve a digital object. If METS is only to be used as a method of delivery, it can change which PREMIS elements should be used since the METS Structural Metadata may not be kept once the document has been imported (Guenther, 2008).

Example of a METS Profile

The National Library of Australia has a METS profile which is registered with the Library of Congress which uses PREMIS for its administrative metadata. They created a three level model in order to create metadata for different sorts of objects. The highest level is a generic profile which describes the basic rules. The second level is content specific and clarifies the rules for different types of objects. The third level is for local registration of Australian METS profiles. It will include the top level generic profile as well as the second level content specific profiles, but it will also refine requirements for local needs.

The Australian METS profile requires each part of an object to have its own METS document. This allows for dissemination of only a part of the object, for instance an image from an article, instead of having to disseminate the entire object which is not needed.

The National Library of Australia also decided that METS documents may also describe analog objects which may one day be digitized. This could allow for finding of an item which has not yet been digitized and even requesting the item to be digitized (Pearce, Pearson, Williams, & Yeadon, 2008).

Exporting Metadata using METS and OAI-PMH

A project by the University of Wales Aberystwyth required a "bridge" be built which allowed automatically exporting an item from one repository to a separate repository.

The "bridge" would have to connect repositories which use different software to store their digital materials, in this case theses. They decided to use METS since even though the repositories used different software; all of the software was able to use METS. They used the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) as the standard for harvesting the records.

Because of the flexibility of METS, they were able to use two descriptive metadata sections, one using MODS and one using qualified Dublin Core, as the different systems used different descriptive metadata standards. An external program was used to import metadata which sends OAI-PMH requests to the contributing institutions. Since the program only imports metadata, once the METS documents are imported into the repository system, the system uses the METS file identifiers to download the actual objects from the original institution (Bell & Lewis, 2006).

Using METS for Archiving eJournals

eJournals can be difficult to work with for a digital repository. Articles are submitted by publishers in a variety of file formats, with different metadata formats and vocabularies. To deal with this, the British Library created five different linked METS documents: one each for journals, issues, articles, manifestations (all files needed for one rendition of an article) and submissions.

They use MODS for their descriptive metadata and PREMIS for their administrative metadata. MODS is also used to show the parent/child relationships between journal, article and issue. The "child" links to the parent instead of the other way around so that the "parent" metadata document is not being constantly updated every time there is a new submission. MODS also links an old version of an article to a corrected or enhanced version. MODS is not used in the manifestation and submission METS documents since it is not necessary to describe them. The British Library also uses MODS to display their rights metadata because the data is of a descriptive nature since the library does not use it for a repository function. PREMIS is used to link manifestations to article and submission objects. It also links derived files – for instance, when a file is transformed from one format to another, the new file is considered a derived file.

The British Library has decided to redundantly store information in METS and PREMIS since the METS portion of the document may be accessed separately from the PREMIS portion. This makes sure that all of the needed information is present even when not accessing the entire METS document (Dappert & Enders, 2008).

Crosswalk from METS to IMS Content Packaging

IMS Content Packaging (IMS-CP) is similar to METS; it is an XML specification which can exchange content, but, instead of being based in the digital library community, it is a part of the educational technology community. The educational community can make use of documents from research libraries which are encoded using METS. Using a crosswalk, items can be brought from the digital collection of a research library into, for instance, a learning management system where it can be shared with students.

There are a number of similarities between METS and IMS-CP. Neither specifies descriptive and administrative metadata, instead institutions using the two standards must use external metadata standards for those sections. Both standards also specify the structure of the object and list the files which are a part of the object. Neither dictates how the content should be presented, although IMS-CP documents are often organized in a way that the HTML file which controls how the object is displayed becomes a part of the content.

Yee & Beaubien (2004) create a crosswalk using a XSLT engine to transform a METS document into an IMS-CP document. It is limited, though, in that it does not transform the administrative and descriptive metadata to standards that are used by IMS-CP.

This is incredibly challenging as any metadata standard can be placed in a METS document and it would require an incredible amount of coding to do. This problem, of course, does not just exist when creating a crosswalk between IMS-CP and METS. Creating a crosswalk between METS and any other metadata standard would be difficult unless the other format was just as flexible in the use of external descriptive and administrative schemas as METS.

Conclusion

METS is a useful tool for the transmission of digital objects and their metadata. It was created with an incredible amount of flexibility which may lessen as both Gartner (2008) and Pearce et al. (2008) both believe that the creation of profiles is more of an interim solution and that METS may soon end up with more rigid standards. This would allow more exchange of digital objects and metadata since they would be able to be exchanged with any institution using METS instead of just the institutions using the same profile. Similar to what MARC has done to library cataloging, the standardization of METS would also create less work since the metadata could be shared leading to less duplication of efforts. Gartner also believes that creating many different institutional profiles is a wasted effort because they are essentially 'reinventing the wheel.' It would also make the creation of crosswalks easier as there would be one standard instead of many for the descriptive and administrative metadata.

The problem, of course, lies with picking an implementation. Since there are so many choices, it would be difficult to narrow it down to just one that would work for all institutions. Each digital collection has separate needs based on the objects in the collection and the software they use. There is also the problem of legacy data which would have to be converted to the new standard implementation of METS. This will probably cause problems with many institutions and may cause them to stop using METS. Part of the reason MoA II was replaced with METS was because it had specifications for administrative and descriptive metadata. The flexibility of METS was a huge selling point, so disposing of that flexibility may cause some problems.

Works Cited

Bell, J., & Lewis, S. (2006). Using OAI-PMH and METS for exporting metadata and digital objects between repositories. *Program, 40*(3), 268-276.

Cantara, L. (2005). METS: The Metadata Encoding and Transmission Standard. *Cataloging & Classification Quarterly, 40*(3/4), 237-253.

Cornell University Library. (2010). *Making of America: About the project.* Retrieved from http://dlxs2.library.cornell.edu/m/moa/about.html

Cundiff, M.V. (2004). An introduction to the Metadata Encoding and Transmission Standard (METS). *Library Hi Tech, 22*(1), 52-64.

Dappert, A., & Enders, M. (2008). Using METS, PREMIS and MODS for archiving eJournals. *D-Lib Magazine, 14*(9/10). Retrieved from http://www.dlib.org/dlib/september08/dappert/09dappert.html

Gartner, R. (2008). *Metadata for digital libraries: State of the art and future directions.* Retrieved from http://www.jisc.ac.uk/whatwedo/services/techwatch/reports/horizonscanning/hs0801.aspx

Guenther, R. (2008). Battle of the buzzwords: Flexibility vs. interoperability when implementing PREMIS in METS. *D-Lib Magazine, 14*(7/8). Retrieved from http://www.dlib.org/dlib/july08/guenther/07guenther.html

Hurley, B.J., Price-Wilkin, J., Proffitt, M., & Besser, H. (1999). *The Making of America II testbed project: A digital library service model.* Retrieved from http://www.clir.org/pubs/reports/pub87/contents.html

Library of Congress. (2007). *METS profiles*. Retrieved from http://www.loc.gov/standards/mets/mets-profiles.html

Library of Congress. (2008). *METS extenders: External schemas for use with METS*. Retrieved from http://www.loc.gov/standards/mets/mets-extenders.html

Library of Congress. (2009). *METS: An overview and tutorial*. Retrieved from http://www.loc.gov/standards/mets/METSOverview.v2.html

Library of Congress. (2010). *METS editorial board*. Retrieved from http://www.loc.gov/standards/mets/mets-board.html

Pearce, J., Pearson, D., Williams, M. & Yeadon, S. (2008). The Australian METS Profile – A journey about metadata. *D-Lib Magazine, 14*(3/4). Retrieved from http://www.dlib.org/dlib/march08/pearce/03pearce.html

Seadle, M. (2002). METS and the metadata marketplace. *Library Hi Tech, 20*(3), 255-257.

The Digital Library Federation. (2009). *The Making of America, Part 2.* Retrieved from http://www.diglib.org/standards/dlfmoaii.htm

Yee, R., & Beaubien, R. (2004). A preliminary crosswalk from METS to IMS content packaging. *Library Hi Tech, 22*(1), 26-81.