Journal of Library Metadata

ISSN: 1938-6389 (Print) 1937-5034 (Online) Journal homepage: http://www.tandfonline.com/loi/wjlm20

Providing Metadata for Compound Digital Objects: Strategic Planning for an Institution's First Use of METS, MODS, and MIX

Michael Dulock & Christopher Cronin

To cite this article: Michael Dulock & Christopher Cronin (2009) Providing Metadata for Compound Digital Objects: Strategic Planning for an Institution's First Use of METS, MODS, and MIX, Journal of Library Metadata, 9:3-4, 289-304, DOI: 10.1080/19386380903405199

To link to this article: http://dx.doi.org/10.1080/19386380903405199

Published online: 10 Dec 2009.

Journal of Library Metadata, 9:289–304, 2009

ISSN: 1938-6389 print / 1937-5034 online
DOI: 10.1080/19386380903405199

Providing Metadata for Compound Digital Objects: Strategic Planning for an Institution's First Use of METS, MODS, and MIX

MICHAEL DULOCK
University of Colorado at Boulder Libraries, Boulder, Colorado, USA

CHRISTOPHER CRONIN
University of Chicago Library, Chicago, Illinois, USA

The University of Colorado at Boulder recently engaged in a grant-funded pilot project to use Metadata Encoding & Transmission Standard (METS), Metadata Object Description Standard (MODS), and NISO Metadata for Images in XML Schema (MIX) for a collection of digitized Sanborn fire insurance maps of the state. This article will draw on this experience to outline the processes and decision making required to implement new metadata structures, and will offer some insights on planning strategically for an institution's first use of these increasingly important metadata standards.

KEYWORDS metadata standards, compound digital objects, strategic planning, METS, MODS, MIX

INTRODUCTION

Sanborn maps are historical fire insurance maps of towns and cities at the detailed level of individual blocks, streets, and buildings. Beginning in the 1860s, the maps were created by the Sanborn Map Company to aid in assessing the risk of fire damage and, therefore, in setting insurance premiums. Today, the historical maps are valuable resources for the study of architecture, urban planning and development, demographics, environmental conditions, and even genealogy.

In 2006, the University of Colorado at Boulder Map Library received an Institute of Museum and Library Services (IMLS) grant through the Colorado

Address correspondence to Michael Dulock, University of Colorado at Boulder Libraries, Cataloging & Metadata Services Department, 184 UCB, 1720 Pleasant Street, Boulder, CO 80309-0184, USA. E-mail: michael.dulock@colorado.edu

State Library under the provisions of the Library Services and Technology Act (LSTA) to digitize a portion of the maps that were no longer under copyright restriction. The segment of the collection selected comprised 346 maps on 2,385 sheets, detailing 79 cities in 52 counties in Colorado covering the years 1883–1922.

Sanborn maps consist of multiple sheets that together form the entire map, with an "index sheet" showing the relationships between each sheet in the map. Within the context of the maps selected for this project, the number of sheets per map ranged from one to 98. In a digital environment these types of resources are often referred to as compound (or complex) digital objects, wherein the object itself comprises multiple discrete components that are usually organized in a specific way to form a cohesive whole. The compound structure of the map sheets, combined with a goal of standardizing technical metadata capture, resulted in the project team's decision to explore implementing Metadata Encoding & Transmission Standard (METS, available at http://www.loc.gov/standards/mets/) as part of the digitization plan for the collection.

One of METS' important strengths is its ability to support metadata provision for compound digital objects beyond just descriptive metadata. In contrast to other metadata schemes—like MARC, Dublin Core, VRA Core, and EAD—METS facilitates the association, organization, and collocation of technical, structural, descriptive, provenance, and behavioral metadata for digital objects within a single METS wrapper. To this end, METS addresses some of the challenges inherent in the distributed environment in which digital objects—and the metadata that describe them—are managed by the organization, as well as found and used by patrons. METS packages together, or points to, all the information about a resource's archival master, access, preview, and metadata files, thereby reducing reliance on the institutional memory of a few to recall the various locations of those files. In thinking ahead to possible future scenarios, METS' standardization will allow the University Libraries to access, reuse, recondition, or repackage digital objects and metadata (e.g., migrating to different delivery or storage systems, preservation reformatting of digital objects, integrating collections with those of other institutions).

When the project team was originally exploring METS, there were few commercially or freely available METS creation tools in the marketplace. The project team explored the development of a homegrown tool that generates METS records from existing but dispersed metadata sources. Digitizing the Sanborns also provided an opportunity to explore the use of an established standard for technical metadata: the Metadata for Digital Still Images in XML (MIX) standard, maintained by the Library of Congress and NISO. Following this standard allows for consistent recording of specific details on the process of digital imaging, including equipment and software used for digitization. This article will discuss how the team developed various metadata creation tools and how it built technical metadata capture into the automated process of digital imaging, as well as some of the lessons learned in how to plan strategically for an institution's first implementation of Metadata Encoding & Transmission Standard (METS), Metadata Object Description Standard (MODS), and NISO Metadata for Images in XML Schema (MIX).

CASE STUDY

The process of evaluating whether and how to implement METS revealed that the personnel involved did not possess all of the skills required to build an application that could create METS records. Team members from Digital Resources Cataloging had knowledge of the metadata schemes and standards involved (XML, METS, MODS, MIX, MARC, MARCXML); representatives from the Map Library had knowledge of both the Sanborn maps and their users; the Digital Initiatives Librarian had scripting experience and expertise with the University Libraries' digital asset management system, Luna Insight. No one, however, had the programming capabilities required to build an application that could generate METS records.

The team allocated a portion of the grant funds to hire a contractor to develop a METS generation application. The amount of funding available was limited, as was the timeframe for the grant, so there was a desire to work with a programmer who was local. The team pursued hiring a graduate student from the Computer Science program on campus. A qualified doctoral student was identified, with whom the team met to discuss deliverables, available funds, and timelines. Including the student programmer, the project team totaled six individuals.

Requirements for the functionality of the tool were created collaboratively according to the needs of all the departments involved. Staff from the Map Library addressed patron needs and collection concerns. Digital Resources Cataloging provided input on descriptive and technical metadata standards, as well as application functionality from the perspective of catalogers who would actually use the tool. The Digital Initiatives Librarian provided direction related to local information technology systems and sought to ensure a broad understanding of how various facets of the project would, or would not, integrate with existing University Libraries infrastructures. These requirements were discussed at the first meeting of the entire group and were readdressed as needed throughout the duration of the project.

Because METS, MODS, and MIX were new to everyone on the team, taking time to educate ourselves on the standards was an important part of the process. A group-study approach was adopted to facilitate broad understanding of what all aspects of the project would entail. While the catalogers on the team took primary responsibility for learning about and communicating the implications of adopting MODS for descriptive data, the process of learning about both MIX and METS were shared by everyone in the group. To ensure that whatever tools were developed fit all of our needs, the Computer Science student also familiarized himself with all METS documentation, and joined discussion and planning meetings. It was decided that all documentation for the process, the product, and the METS packages specifications, would be written collaboratively.

With the requisite skills in place for the creation of a METS generation application, the project team assessed what metadata already existed, and how it could be repurposed to create the desired METS records. Following the University Libraries' policy to create separate records for different formats, electronic resource MARC records, containing descriptive metadata for each digitized map, were derived from the print records in OCLC.

File-naming conventions were devised at the outset of planning, both for the process of developing the METS tool and for communicating deliverables to the digitization vendor. Continuity of file naming was essential so that each subordinate metadata unit would be uniquely identifiable and matched with the correct image file for the map. Base file names were constructed using the first three letters of the city and the last two digits of the year represented by the map (e.g., a map of Erie in 1893 became "eri93"). While there was discussion of including all four digits of the year (in order to avoid conflicts across centuries), it was felt that the digital collection would probably not expand beyond 1923, making such conflicts unlikely.

Base file names were added to a local 9XX field in each MARC record; the METS application was designed to use the contents of this field to automatically establish file names for the METS records (e.g., eri93.xml), as well as names for the constituent parts, or sections, of each METS file. For example, the MODS descriptive metadata section for Erie 1893 would be "DMDeri93," and the administrative metadata section would be "AMDeri93." The vendor also provided image filenames for each sheet according to these conventions (e.g., "eri93001," "eri93002," and so forth) as well as the images' corresponding MIX technical metadata files, (e.g., "TMDeri93001," "TMDeri93002," and so forth). A truncated outline of the METS XML structure for the 'Erie 1893" example appears in Figure 1.

The basic framework for the METS records was based on examples available at the METS Web site. While the Sanborn maps constitute compound digital objects, they do not exhibit some of the more complex features that more complicated METS implementations have exploited (i.e., multiple languages, parallel files, etc.). Thus, a model was created for the application that includes a simple structural map, descriptive metadata for each object, and technical metadata for each component image. The <structMap> section of the METS record, which defines the hierarchical order of the compound object for presentation, uses the ORDER attribute along with filenames to represent the relatively simple, "flat" nature of the collection. Ninety-five percent of the map sheets follow simple numerical order, so that the lowest-numbered

```
<mets TYPE="map" LABEL="Erie 1893" OBJID="eri93">

    <!-- descriptive metadata section -->
    <dmdSec ID="DMDeri93">
        <mdWrap MDTYPE="MODS" LABEL="MODS record">
            <xmlData>
                <mods xmlns="http://www.loc.gov/mods/v3" version="3.2">
                    <titleInfo>
                        <title>Erie, Weld Co., Col. [insurance map]</title>
                    </titleInfo>
                    <titleInfo type="alternative">
                        <title>Erie, Weld County, Colorado</title>
                    </titleInfo>
                    <titleInfo type="uniform">
                        <title>Sanborn fire insurance maps</title>
                        <partName>Colorado</partName>
                    </titleInfo>
                    <name type="corporate">
                        <namePart>Sanborn-Perris Map Co</namePart>
                        <role>
                            <roleTerm authority="marcrelator"
                            type="text">creator</roleTerm>
                        </role>
                    </name>
                    <typeOfResource>cartographic</typeOfResource>
                    <genre authority="marcgt">map</genre>
                    .
                    .
                    .
                </mods>
            </xmlData>
        </mdWrap>
    </dmdSec>

    <!-- administrative metadata section -->
    <amdSec ID="AMDeri93">

        <!-- technical metadata -->
        <techMD ID="TMDeri93001">
            <mdWrap MDTYPE="NISOIMG">
                <xmlData> …  </xmlData>
            </mdWrap>
        </techMD>
    </amdSec>

    <!-- section containing constituent files -->
        <fileSec>
            <fileGrp>
                <file ID="eri93001" MIMETYPE="image/tiff" ADMID="TMDeri93001">
                    <FLocat LOCTYPE="URL />
                </file>
            </fileGrp>
        </fileSec>

    <!-- structural relationship section -->
    <structMap TYPE="physical">
        <div TYPE="map" DMDID="DMDeri93" LABEL="Erie 1893">
            <div TYPE="sheet" ORDER="1" LABEL="sheet 1 (index map)">
                <fptr FILEID="eri93001" />
            </div>
        </div>
    </structMap>

</mets>
```

FIGURE 1 Truncated METS XML record for a Sanborn map sheet can be assigned ORDER = 1, and so on, in sequence. For those few maps that have unusual numbering (i.e., some maps have one or more sheets designated "unnumbered"), the value of the ORDER attribute must be manually manipulated to represent the correct sequence. There are other ways within METS to handle sequence, but given limitations on time and funding, and the relatively small number of outliers requiring manual intervention, using ORDER was determined to be the best option at the time.

The MIX records, by the definition of the standard, were already in XML format, and thus could be incorporated into a METS record without further manipulation. Additional work was required, however, to ready the existing MARC records for incorporation into METS. The team could have chosen to take advantage of METS flexibility to simply link out to the descriptive metadata record already in the library catalog, rather than embed it in the METS file, but "flexibility is often the enemy of interoperability, and METS is no exception to this rule" (McDonough, 2006, p. 151). Pointing to external metadata would not allow the University Libraries to easily export descriptive metadata later, if needed, and would inhibit searching across that metadata in whatever system the METS files were eventually housed.

The target descriptive metadata schema for inclusion in the METS records was MODS. While the team could have simply used the existing MARC record within the METS structure, the project provided an opportunity to gain experience with MODS, which retains most of the information contained in a MARC record, but its XML tags utilize common-language namespaces rather than numeric fields. The application under development was intended to work with XML-based metadata, so the MARC records first had to be converted to MARCXML, an expression of the MARC standard in XML format.

MarcEdit, a freeware application created by Terry Reese at Oregon State University, was used to convert the MARC records into MARCXML (Reese, 2009). All of the MARC records for this collection were extracted from the University Libraries' ILS, which exports multiple records in a single file. MarcEdit was then used to crosswalk the single MARC file into a single XML file that contained all 346 MARCXML records. As a result, the application developed by the project needed to include functionality to create individual MARCXML files from a single collective document before they could be crosswalked to MODS for ultimate inclusion in METS packages. At the time, MarcEdit was not using the current version of MODS, and the METS generation tool was developed to perform that MARCXML-to-MODS transformation in a single step, using a style sheet of the operator's choosing. Should the University Libraries ever require MODS files on their own, separate from the METS wrapper, a function was built to perform just the MARCXML-to-MODS conversion. This functionality could also be performed by MarcEdit, but including it in the new application was seen as a way to develop a unified suite of MODS and METS generation services in a single tool.

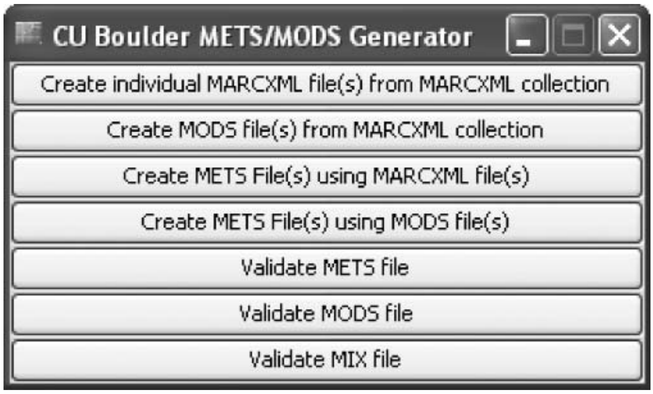

FIGURE 2 Main menu screen of the METS generation application

As a part of the project, the developer of the METS tool created user and technical documentation, the latter of which includes instructions for modifying the tool for future uses. The ability to edit the tool for other projects besides the Sanborn maps was an important consideration, and was a requirement from the beginning. All documentation produced by the programmer was reviewed, tested, and revised by members of the project team before the end of the grant period to ensure it was accurate, usable, and written in common language. The user documentation was augmented with instructions for using MarcEdit to transform MARC records into MARCXML format. The application and documentation are available for download from the University Libraries' Web site (Ahmad, 2007). The main menu screen of the application is shown in Figure 2. In addition to the functions detailed below, the tool also validates METS, MODS, and MIX files against their respective XML schemas.

Each menu button on the application leads to an input screen for that particular function. The first button allows the user to split a composite MARCXML file into individual files and collect them in a folder (Figure 3).

The second menu button creates MODS files, either from a folder of individual MARCXML files or from a collective MARCXML file (Figure 4). This flexibility was added during the testing process so that, in future projects,

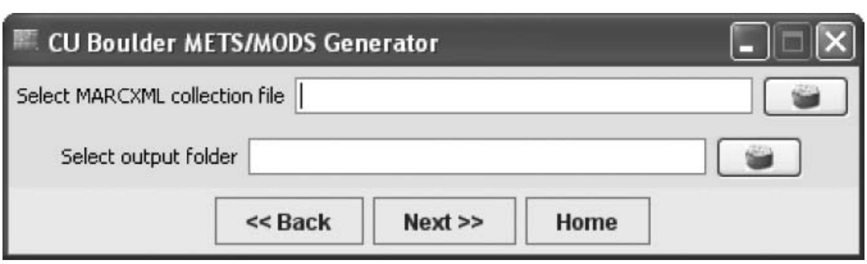

FIGURE 3 Divide a composite MARCXML file into individual files

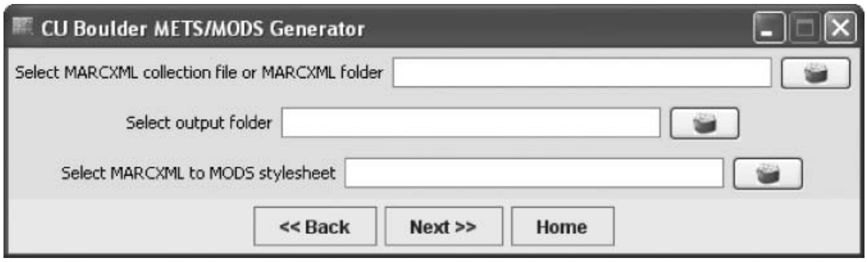

FIGURE 4 Convert MARCXML to MODS

the application could create MODS from existing metadata in either form. Functionality was also added to allow the user to choose the desired XSLT style sheet for the transformation, in anticipation of updated standards.

The third menu button creates METS files using MARCXML and MIX files, incorporating the MARCXML-to-MODS transformation in a single screen with the METS creation (Figure 5). For quality assurance and audit management purposes, the "METS agent name" field is used to record the metadata creator's name in each METS record. The "METS agent type" pull-down menu allows the metadata creator to select "INDIVIDUAL," "ORGANIZATION," or "OTHER," so that the University or other institution can be credited with

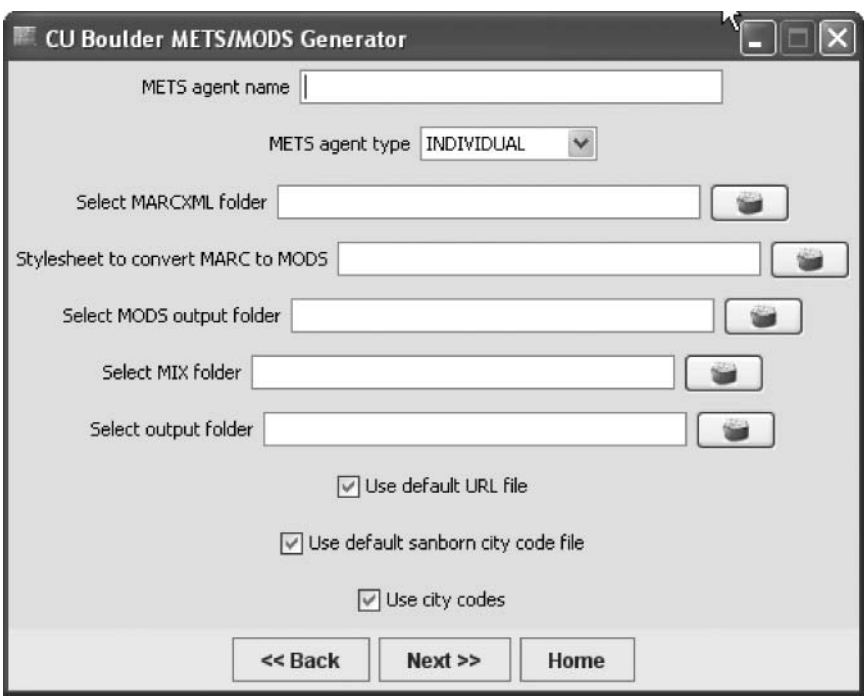

FIGURE 5 Create METS from MARCXML and MIX, Sanborn project defaults checked

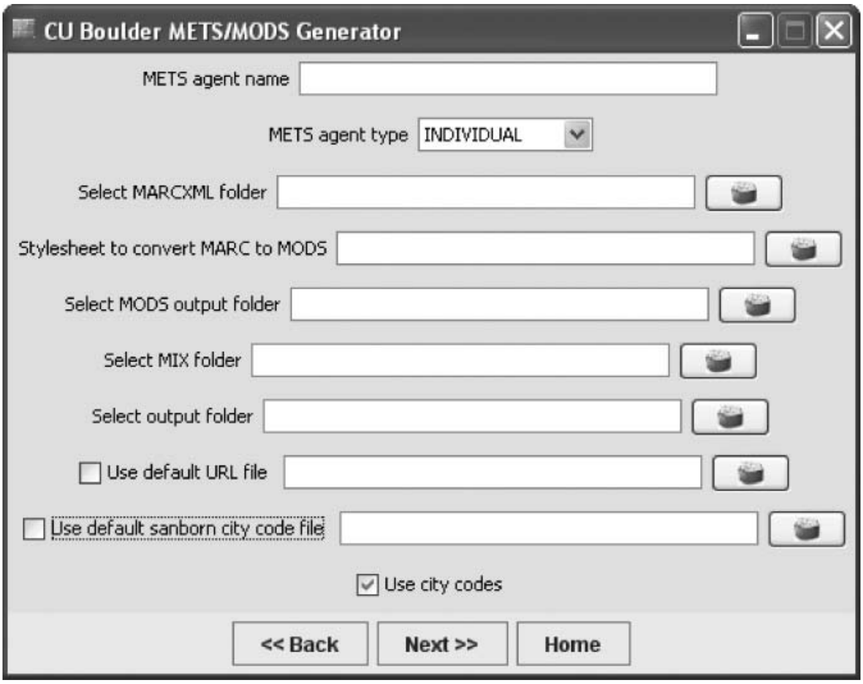

FIGURE 6 Create METS from MARCXML and MIX, Sanborn project defaults unchecked

record creation, in addition to named individuals. A field was included for a separate output folder for MODS records so those records could be individually accessed, instead of only existing within the METS record. Check boxes were added for specifications that were tied to the Sanborn project: "Use default URL file"; "Use default sanborn city code file"; and "Use city codes." The URL file is used to populate the fileSec portion of the METS records for the image files. The city code file is used to create intelligible labels in the METS records, a Sanborn-specific requirement that may not be relevant to other projects. "Use default sanborn city code file" is only active when "Use city codes" is checked. These specifications were needed to create METS records containing valid associations with the embedded MODS records, and to construct valid URLs that would lead to the digital files in Luna, the University Libraries' digital asset management system.

Using the check boxes, a user can bypass those Sanborn-specific files and use the tool for other projects. Unchecking "Use default URL file" allows the user to specify another URL configuration file. Similarly, the user can select some other city code or other naming configuration file by unchecking the "Use default sanborn city code file" check box (Figure 6). Unchecking "Use city codes" disables the use of any city code file.

The fourth menu button also creates METS records, but without the MARCXML transformation step (Figure 7). This was added during testing in case future projects utilized existing metadata that was already in MODS

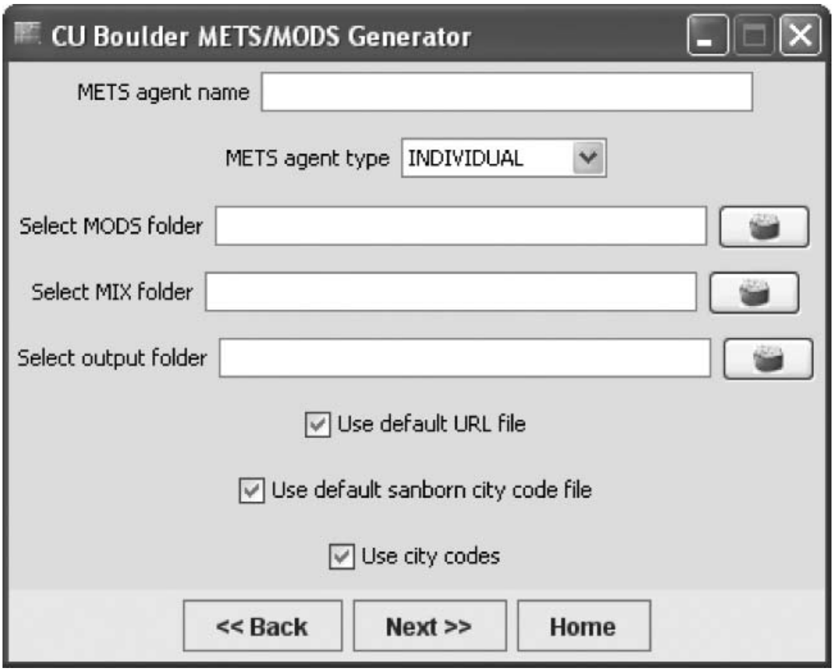

FIGURE 7 Create METS from MODS and MIX

format and did not need to be transformed. This screen otherwise includes the customizable features mentioned above.

To date the tool has been used only in conjunction with the Sanborn map project. The Luna metadata display for each map includes a link to the raw XML METS file, which opens in a separate Web browser. Luna, however, does not have native METS support, and so does not exploit the structural metadata included in the METS records. Each sheet is viewable as a discrete image, but the relationships between sheets and maps are only discernable by file names. The University Libraries may consider implementation, in the near future, of an XML gateway that would be able to take advantage of the display possibilities afforded by METS. There are additional collections that, once digitized, would consist of compound digital objects, and thus could benefit from METS' ability to encode structural data. In the meantime, the METS packages are not serving a role beyond that of long-term preservation.

STRATEGIC PLANNING

As with many things we do in libraries, the metadata planning process is as important—or even more important—than implementation. The decision to implement complex metadata standards like METS, MODS, and MIX requires commitment from across the organization. The staff time required for key players to familiarize themselves with both the syntax and semantics of a standard is significant. Such an investment should take into consideration the broader scope of what the institution wishes to accomplish for its digital library program as a whole, as opposed to as an individual project. For most organizations, the philosophy for applying these metadata standards should extend beyond application to a single project or collection. What digital resources or collections does the institution have that could also benefit from incorporation into a METS environment? Looking ahead, what categories of resources are envisioned as candidates for these kinds of metadata?

Time devoted to learning and using these complex standards, even in limited implementations, can pay dividends at a later time. Knowledge of METS may help prepare staff for future standards that share the same purpose of handling several types of metadata in one package (Pearce, Pearson, Williams, & Yeadon, 2008). It is also possible that another standard may grow out of METS that will incorporate technical, descriptive, structural, and other types of metadata into a single schema. In addition, XML architecture is central to metadata integration and interoperability efforts at present and may remain so for the immediate future (Gartner, 2008). Facility with XML provides a foundation upon which staff may build knowledge of specific schemas and eases the learning curve when exploring new schemas expressed on the same platform. In addition to familiarity with the specifics of complex metadata schemas, the experience gained through the use of these rich standards benefits future projects in areas such as workflow and process planning (Pearce et al., 2008; Breytenbach & Groenewald, 2008) and development of best practices (Verheusen, 2008). Lessons learned about which facets of a given standard or schema worked, and which did not, may be applicable to future initiatives (Breytenbach & Groenewald, 2008).

Identifying and analyzing the technical expertise required to implement these schemas is critical to an institution's ability to carry it out successfully. Subsequent matching of the required expertise with existing staff, or recognizing staffing gaps that need to be filled prior to implementation, will also be crucial. Institutions with substantial technical-systems staff on hand may be able to create robust and complex applications and interfaces in house (Guerard & Chandler, 2006). Institutions lacking sufficient technical expertise may seek to hire consultants or other outside experts to work on specific aspects of an implementation (Schwartz & Iobst, 2008). When the grant proposal for the Sanborn project was submitted, it was the first planned implementation of METS in the state. The project team successfully highlighted METS as an innovative component of the project, but it also required that some funds go toward hiring someone with the programming skills necessary to accomplish various aspects of the process—skills that were not already represented in existing staff. Any funds that went to programming necessarily meant fewer funds going toward digitization and, therefore, less actual content being made accessible. However, this was an initiative grant. The project team considered the process of building a framework for ongoing creation of metadata as integral to ongoing sustainability by providing the capacity to both continue postgrant digitization and extend the work accomplished through the grant to other collections. The group-study approach, as well as the collaborative documentation process, allowed for broad understanding of all parts of the project and reduced the chance of creating silos of responsibility and knowledge. In the case of the Sanborn project, involving the programmer in the group education, the planning, and the implementation stages of the project afforded a higher degree of understanding across all members of the team.

While part of the planning process for adopting new metadata standards requires an inward assessment of institutional capabilities, another aspect of the planning should look outward. Both METS and MODS have implementation registries, within which institutions that have already employed these standards have described their projects and provided links to documentation. A review of some of these implementations, as well as a broad scan of the literature, can provide useful information on lessons learned, on how others have approached similar issues, and on solutions that have succeeded and failed. This "investigation stage" should be a focus for early-stage teamwide planning, with results shared broadly to promote common understanding amongst all team members.

What have others learned from their own METS implementations? In working toward establishing a "data commons," Pearce et al. (2008) used METS in combination with Preservation Metadata Implementation Strategies (PREMIS) as part of a series of profiles of increasing specificity in anticipation of use with other projects. METS was found to be the best available schema for encoding different types of metadata for a variety of complex digital objects. Potential problems were identified, however, in the possibility of duplication and redundancy between METS elements and elements from extension schemas. Relationships in METS between different types of metadata can also be awkward and cumbersome. The creation of project-specific profiles was also identified as a less-than-desirable process, which Pearce and colleagues hope to discard once they have established robust best practices.

Dappert & Enders (2008) used METS and PREMIS, along with MODS for descriptive metadata, as part of an archive for electronic journals. This combination of schemas was again found to be suitable for use with complex objects due to its flexibility and extensibility. Redundancy and duplication were also recognized, but in this case were found to be positives, in that the institution had choices regarding where and how to store certain metadata. At the same time, this high level of flexibility was seen as a potential hindrance, as the various options necessitated careful consideration of which element from which schema was best for a given data point, when to choose one over the other, and when to duplicate metadata.

West, Llona, Gerontakos, and Biggins (2007) found the structural capabilities of METS to be ideal for the complexity inherent in a collection of digitized architectural photographs. Guerard and Chandler (2006) expressed a similar sentiment, noting that METS allowed one to express "the structure of multipart, multilevel documents (e.g., multipaged letters, manuscripts and pamphlets, photo albums, scrapbooks)" (p. 53).

Performing an environmental scan to identify and analyze tools for metadata creation is an important part of the investigation phase of planning. Once the institution has clarified its metadata needs, the next step is deciding whether to adopt existing tools from the marketplace or to develop tools in house. Implementations at institutions with substantial technical resources utilized sophisticated METS generation tools tied to robust, multipart application systems created in house (Guerard & Chandler, 2006; Pearce et al., 2008; West et al., 2007). At the time the Sanborn project was being planned, there were few METS record-generation tools available. The team had to balance the resources that would be required to create such a tool against the needs of the rest of the project (e.g., creating as many digital surrogates as possible). System development is, by necessity, something that should be evaluated with care and caution, particularly when a grant-funded project is primarily intended for end-product digitization. In the end, the team decided to integrate a staffing model that would accommodate the development of such a tool but also committed to planning a system that would be as extensible as possible.

Any tool developed locally would have to be usable and useful for postgrant continuation of the Sanborn project, but also for other future digitization efforts. The team at the University of Colorado Libraries has used the METS generation application to create valid METS records for the digital Sanborn maps. These XML files are available through the Luna digital library. Luna does not, however, presently make use of the structural information encoded in the METS documents. The XML files are essentially available for viewing but are not functioning to their potential. The metadata records accompanying each image are a slightly modified form of the MARC records for the maps in the online catalog. There is a desire to integrate an XML gateway into the digital library that would allow the system to take advantage of the structural information afforded by METS. Plans for this, unfortunately, must wait for appropriate staffing and/or funding for that stage of development.

The role of metadata in long-term and sustainable digital preservation cannot be overstated and is one of the great benefits facilitated by METS and MIX: "Recording this information documents the custodial history of the data and will assist scholars in the future in determining how accurately the data in front of them reflects the digital information originally captured"

(McDonough, 2006, p. 154). And as one digital image project team concluded: "There will be an ongoing need for long-term preservation of the digitized images, an area in which the relevant technologies are likely to develop considerably in coming years" (West, et al., 2007, p. 115). Preservation metadata answers a multitude of needs: (1) recording provenance information; (2) documenting ownership, rights, and licensing information; (3) providing technical metadata necessary for format migration and future access; and (4) providing technical metadata needed to verify the authenticity and validity of the digital object.

If the project involves in-house digitization, mechanisms for technical metadata capture will have to be developed. Outsourcing digital imaging, however, also requires advance planning for the outsourcing of technical metadata creation. In the case of the Sanborn project, it was determined that the capture of technical metadata needed to be automated and done at the same time as the capture of the digital image itself in order to acquire high-quality and robust data about the digitization process. MIX was identified as the standard of choice because of its compliance to ISO standards, its packaged integration into METS, its specific focus on digital still images, and the fact that it is maintained and supported by the Library of Congress. The planning for this technical metadata component of the project requires a relatively thorough knowledge of MIX prior to outsourcing the digitization. In preparation for the RFP/RFI process, the team needs to know enough to communicate its needs to prospective digitizing agencies, and to request all the information it needs to select the appropriate vendor. In other METS implementations, technical metadata capture has been done both manually (West et al., 2007; Breytenbach & Groenewald, 2008) and by automated process (Guerard & Chandler, 2006; Dappert & Enders, 2008). Guerard and Chandler's workflow allowed for both automated capture and manual entry of technical metadata, if needed.

Establishing a methodology for quality assurance is also essential to the planning process, particularly when outsourcing metadata provision. In a scenario in which technical metadata is being automatically recorded at the time of image capture, the timeline for testing metadata quality should ideally parallel the timeline for assessing the quality of the vendor's digital imaging. In the case of the Sanborn project, the project team identified certain MIX elements that it felt, while not labeled "required" or "mandatory" by the standard itself, were nevertheless essential data points for long-term needs. The RFP announcement included a complete list of all elements the University Libraries was requiring as part of its MIX profile, and vendors were asked to address how they would accommodate those needs. Prospective vendors were also informed of the need for them to deliver MIX XML documents that they had already validated against the schema. This way, if errors occurred, the vendor could identify and fix them during the digitization process, to avoid the University Libraries' receiving malformed or noncompliant records after the fact. Documenting the metadata requirements in this way allows the project team to fully understand in advance what it is really asking vendors to do, it allows the vendors to understand the same, it facilitates accurate cost and time estimates, and it provides a framework against which the final products can be tested for quality (both by the vendor and by the library).

CONCLUSION

METS, MODS, and MIX provide standardized encoding for managing the various metadata associated with compound digital objects and for documenting the relationships between the objects and their metadata. Implementing new forms of metadata provision will always require a certain amount of planning, particularly for those metadata that bring with them a high threshold for knowledge acquisition and system readiness. Procedures and best practices should be documented to establish a baseline against which success and failure can be measured, as well as to provide a basis for future endeavors. Like all aspects of digital projects, metadata planning and provision is collaborative. Stakeholders from affected units, such as systems personnel and collection managers, must provide input when appropriate. While metadata managers may make the initial argument for moving to a digital library program that uses METS, the adoption of such a complex standard necessitates broad institutional support and involvement. These factors can contribute to the success of institutions planning their first foray into the adoption of these metadata standards.

ACKNOWLEDGMENTS

The authors would like to thank Kathryn Lage (Map Librarian), Holley Long (Digital Initiatives Librarian), and Faisal Ahmad (Computer Science doctoral student) for their contributions to the project.

REFERENCES

Ahmad, F. (2007). METS generator [Computer software and documentation]. Retrieved February 7, 2009, from http://ucblibraries.colorado.edu/systems/digitalinitiatives/tools.htm

Breytenbach, A., & Groenewald, R. (2008). The African elephant: A digital collection of anatomical sketches as part of the University of Pretoria's institutional repository—a case study. *OCLC Systems & Services: International Digital Library Perspectives*, *24*(4), 240–251.

Dappert, A., & Enders, M. (2008). Using METS, PREMIS and MODS for archiving eJournals. *D-Lib Magazine*, *14*(9–10). doi:10.1045/september2008-dappert

Gartner, R. (2008, April). Metadata for digital libraries: State of the art & future directions. *JISC Technology & Standards Watch*. Retrieved March 1, 2009, from http://www.jisc.ac.uk/media/documents/techwatch/tsw_0801pdf.pdf

Guerard, G., & Chandler, R. L. (2006). California cultures: Implementing a model for virtual collections. *Journal of Archival Organization*, *4*(1/2), 45–67.

McDonough, J. P. (2006). METS: Standardized encoding for digital library objects. *International Journal of Digital Libraries*, *6*(2), 148–158.

Pearce, J., Pearson, D., Williams, M., & Yeadon, S. (2008). The Australian METS profile—A journey about metadata. *D-Lib Magazine*, *14*(3–4).

Reese, T. (2009). MarcEdit (Version 5.1) [Computer software]. Retrieved February 7, 2009, from http://oregonstate.edu/˜reeset/marcedit/html/index.php

Schwartz, L. M., & Iobst, B. (2008). Planning for the integration of the digital library, clinical decision support, and evidence at the point of care. *Medical Reference Services Quarterly*, *27*(2), 146–157.

Verheusen, A. (2008). Mass digitisation by libraries: Issues concerning organisation, quality and efficiency. *Liber Quarterly: The Journal of European Research Libraries*, *18*(1), 28–38. Retrieved February 27, 2009, from http://liber.library.uu.nl/

West, J. D., Llona, E., Gerontakos, T., & Biggins, M. (2007). Implementing an image database for complex Russian architectural objects: The William Brumfield Collection. *Slavic & East European Information Resources*, *8*(2/3), 109–126.