Aligning METS with the OAI-ORE Data Model

Jerome P. McDonough
Graduate School of Library & Information Science, University of Illinois at Urbana-Champaign
501 E. Daniel Street, MC-493
Champaign, IL 61820
+1 217-244-5916
jmcdonou@illinois.edu

ABSTRACT
The Open Archives Initiative – Object Reuse and Exchange (OAI-ORE) specifications provide a flexible set of mechanisms for transferring complex data objects between different systems. In order to serve as an exchange syntax, OAI-ORE must be able to support the import of information from localized data structures serving various communities of practice. In this paper, we examine the Metadata Encoding & Transmission Standard (METS) and the issues that arise when trying to map from a localized structural metadata schema into the OAI-ORE data model and serialization syntaxes.

Categories and Subject Descriptors
E.2 [**Data Storage Representations**]: Object representations

General Terms
Design, Standardization

Keywords
METS, OAI-ORE, structural metadata, aggregation, modeling

1.INTRODUCTION
Open Archives Initiative - Object Reuse and Exchange (OAI-ORE) defines a set of standards for the description and exchange of aggregations of Web resources. As noted in the ORE User Guide – Primer 5, we create and use sets of resources all the time, from card files of recipes we may be keeping in our kitchen to special collections residing in an academic library. Despite this, the basic standards and architecture for the World Wide Web provide no mechanism for identifying and describing a set of resources on the web. OAI-ORE provides a solution to this problem, in the form of an abstract model for identifying a set of web resources along with specific guidelines for producing serialized representations of such descriptions in a variety of formats. Together, these specifications provide a comprehensive solution to the problem of identifying and describing aggregations of web resources in a way that makes such aggregations available both for human consumption and for machine processing.

Problems with identifying and describing sets of resources did not originate with the World Wide Web, of course, nor did solutions. Even if we consider only the case of identifying and describing sets of networked electronic resources, numerous solutions to the problem of aggregation identification and description have emerged over past quarter century, from HyTime 5 and TEI 5 through more contemporary standards such as XFDU 5 and the Open Publication Structure 5. While this proliferation of approaches can be problematic for those concerned with interoperability, it is both to be expected, and from the point of view of each individual community, desirable. Different communities of practice have different reasons for wanting to identify and describe aggregations of resources, and different operations they wish to perform with those sets. The ability to custom-tailor a mechanism for aggregating resources to the needs of a community is a great benefit.

This benefit comes at a price, however, and the price is increased difficulty in exchanging information about sets of resources with those who do not share the standards of your local community. Moreover, just as the web has expanded the pool of resources available for us to aggregate and describe, it has also expanded the opportunities for different communities of practice to exchange information. One of the great potential benefits of OAI-ORE is its ability to serve as a lingua franca for exchange of aggregation descriptions between communities.

Realizing OAI-ORE's full potential as an exchange syntax will require careful consideration of the mapping between localized schema for aggregation description and the OAI-ORE data model and serialization syntaxes. This paper will examine some of the issues that arise in mapping from the Metadata Encoding and Transmission Syntax (METS) to OAI-ORE, and make suggestions for possible best practice in METS implementation to insure OAI-ORE compatibility.

2.OAI-ORE Requirements
The ORE Specification – Abstract Data Model 5 details the data model that underlies all serialization formats for OAI-ORE and sets forth a small number of structural constraints for OAI-ORE conformance. I will not recapitulate the full data model here, but as any mapping from METS to OAI-ORE must respect OAI-ORE's constraints, a brief review of certain aspects of the data model and the requirements they impose on the description of an aggregation are in order.

The foundational notion in the OAI-ORE abstract data model is the idea of an aggregation, a resource that consists of a set of other resources. Aggregations are described via a resource map; any OAI-ORE resource map must describe a single aggregation, and enumerate all of the constituent aggregated resources. The OAI-ORE specifications draw a clear distinction between an aggregation, which is an abstract entity, and a resource map, a concrete document that provides a serialized description of the aggregation. Both the aggregation and the resource map that describes the aggregation are resources, and both must have a protocol-based URI assigned to them. A resource map cannot be assigned the same URI as the aggregation it describes.

Resource maps have several other constraints beyond the requirement that they describe a single aggregation. A resource map may make a variety of assertions about the aggregation and the aggregated resources, but these assertions must be extractable from the serialized resource map as a set of triples that form an RDF Graph that fulfills certain minimal structural requirements:

- The graph must express the relationship between the resource map and the aggregation described by the resource map (using the ore:describes predicate);
- It must express certain minimal metadata about the resource map: the authoring authority for the resource map (expressed using the dcterms:creator predicate) and the last modification date/timestamp for the resource map (expressed using the dcterms:modified predicate);
- It must express the relationship between the aggregation and one or more aggregated resources constituting the aggregation (using the ore:aggregates predicate); and
- It may assert other properties about the resource map, the aggregation, aggregated resources, related resources or literals, but the RDF graph produced by all assertions must be connected.

A resource map may express a variety of other relationships between the aggregation and other similar resources, and between resources in the aggregation RDF Graph and other resources and types, but it must fulfill the above minimal requirements to conform with OAI-ORE.

The OAI-ORE Vocabulary specification 5 also imposes requirements on the content of certain pieces of metadata recorded within an OAI-ORE resource map. An agent designated as the creator of a resource map must be a human being, and the contents of the dcterms:creator element identifying that individual should be a URI. The date and time of creation of a resource must be listed in ISO 8601 format.

Because an aggregation is itself a resource, and any resource can be an aggregated resource, it is possible for an aggregation to be composed of other aggregations. In the case of such recursive nesting of aggregations, a resource map that asserts another aggregation as an aggregated resource may also assert an ore:isDescribedBy relationship between that aggregated resource and another resource map describing that other aggregation. This allows client software processing the resource map to know that the aggregated resource is itself an aggregation, and how to find the resource map that describes that aggregation.

3.METS and OAI-ORE Alignment

3.1Metadata Element Considerations

METS is a specification developed and maintained by the digital library community to enable the management of digital library objects within a repository and the exchange of such objects between repository systems. It defines an XML syntax for identifying both the content files constituting the object and metadata describing the object and content files, and asserting a variety of relationships between the various component content files and metadata. A METS file thus serves a very similar purpose to a resource map; it identifies the various resources that constitute the object, expresses a variety of relationships between those resources, and enables librarians to associate a variety of metadata with the constituent resources and the digital library object itself.

Because METS files and OAI-ORE resource maps exist for similar purposes, they have certain similar features. Both standards, for example, have facilities for recording information regarding the date of creation and last modification of an instance document, as well as the authoring agency for the serialized file. They differ, however, in terms of the information that is considered mandatory and optional, as well as in how particular types of information should be expressed. METS authors wishing to insure a METS file can be readily converted into OAI-ORE syntax must insure that their METS files provide the minimal information required by the OAI-ORE specifications, and that the structure of the METS file lends itself to translation into OAI-ORE's RDF graph structure. [1]

A resource map must provide protocol-based URIs for both the aggregation it describes and the resource map itself. Within the METS schema, the closest equivalent to an identifier for the aggregation would be the OBJID attribute on the root <mets> element, which records "the primary identifier assigned to the METS object as a whole."(5, p. 29) In migrating from METS to OAI-ORE serialization syntaxes, the value for the OBJID attribute should be recorded as the URI for the OAI-ORE aggregation. However, this does impose additional requirements on the METS file, both that the OBJID attribute be present (it is optional in the METS schema) and that the OBJID attribute must contain a protocol-based URI to identify the digital library object. The URI for the resource map presents a somewhat more interesting question. METS does not provide a specific element or attribute to record a URI for the METS file itself, and even if it did, clearly a URI for the METS file would need to differ from the URI assigned to an OAI-ORE resource map for the same digital library object. Probably the simplest solution for mapping between a METS-encoding of a digital library object and an OAI-ORE version would be to simply generate the URI for the OAI-ORE resource map dynamically at the time of its creation. However, if maintaining some association between the METS encoding and the OAI-ORE encoding is desirable, METS does provide a header element (<metsHdr>) intended to record information about the METS file. A subelement within the METS header area is the alternative record identifier (<altRecordID>), which records identifier values for the digital

[1] The METS initiative has created a separate XML schema to specify profiles of the METS format that may impose additional requirements and constraints on the use of METS. A METS profile document specifying the additional requirements set forth in this paper is available at http://hdl.handle.net/2142/9571.

object represented by the METS document which are alternatives to the primary identifier recorded in the <mets> element's OBJID attribute. The <altRecordID> could be used to record a protocol-based URI to be assigned to an ORE resource map generated from the METS file. The <altRecordID> element has an attribute of TYPE that is used to record the type of identifier contained within the body of the element. By recording the TYPE for the identifier as "ore:ResourceMap", processing software can identify that URI as the one that should be assigned (or has previously been assigned) to a resource map for the object.

In addition to URIs for the aggregation and the resource map, the OAI-ORE abstract data model insists that two pieces of metadata be recorded about any resource map: the identity of the authoring agent, and the last modification date-timestamp of the resource map. METS does provide the ability to record this type of information. The <agent> element within the <metsHdr> portion of a METS file can be used to record the creator of the METS file (in which case the element's ROLE attribute should have a value of "CREATOR"). And the CREATEDATE and LASTMODDATE attributes on the <metsHdr> element record the date of creation and of last modification of the METS file. However, the mapping between this information in the METS file and an OAI-ORE resource map generated from information in the METS file presents a few intellectual problems. It may not necessarily be the case that the agency responsible for creating the resource map is the same as the agency that authored the METS file, and obviously a resource map created from a pre-existing METS file does not share the same date of creation with the METS file. It would make more sense for a process generating a resource map from a METS file to record the actual date on which the resource map was created. If this is the case, however, copying the date last modified from the METS file into the resource map is a questionable act. We can be relatively certain that many metadata managers would be given pause upon seeing a date of last modification that preceded the date of creation.

Mapping a METS file's <agent> element and LASTMODDATE attribute to a resource map's authoring entity and date of modification must be a context-sensitive act, and part of the context to consider is whether the resource map is maintained as its own static file or generated dynamically upon request. With respect to the authoring entity, if the translation into an OAI-ORE resource map involves no significant intellectual effort (e.g., it is carried out by an XSLT process), then mapping a METS <agent> with a CREATOR role to the resource map's authoring entity is appropriate, and presumably this would be the case in most situations. However, if production of the resource map from the METS file involves significant intellectual work, or if the resulting resource map differs significantly in terms of loss of information or change in structure, than recording a new authoring entity for the resource map would be preferable. In those cases where a resource map, once generated from a METS file, is to be maintained as a separate entity, then the resource map should have its own date of creation and last modification recorded independently of any information in the METS file. However, in some situations a resource map might be generated dynamically in response to a request. One could envision, for example, an OAI data provider creating resource maps dynamically from METS files in response to queries from harvesting agents. In those cases, it would probably be more appropriate to have the creation date and last modification date for the resource map be copied from the matching attributes in the METS file, so that harvesting agents can reliably determine when real changes to the underlying data have occurred.

Those translating METS files into the OAI-ORE data model should note one final point about the <agent> element in METS and the corresponding information in an OAI-ORE resource map. In addition to requiring that the agent identified as responsible for the creation of a resource map must be human, the OAI-ORE abstract data model specifies that the identity of the authoring entity must be expressed using the dcterms:creator predicate. It also specifies that the object of that predicate should be a URI, but may be a blank node serving as the subject of triples using the foaf:name and foaf:mbox predicates. The METS <agent> element has two subelements, <name> and <note>, both of which are defined as having a content type of string. In order to facilitate mapping to the OAI-ORE resource map, the <note> subelement in METS should be used to record the e-mail address of the agent, if available, and nothing else. Processing software can examine the contents of the <name> element and determine whether it is a valid URI. If it is, it can be used as the object for the dcterms:creator predicate in a resource map; if not, it can be assumed that the contents of the <name> element should be mapped to the object of the foaf:name predicate, and the <note> to be mapped to the object of the foaf:mbox predicate. This is an inelegant solution, but the METS schema as currently constructed does not allow for mixed content or the use of elements from arbitrary namespaces within the <agent> element. Revisions to the METS schema to allow greater flexibility in the use of elements outside the METS namespace within the METS header area would facilitate interoperability with OAI-ORE and perhaps with other schema as well.

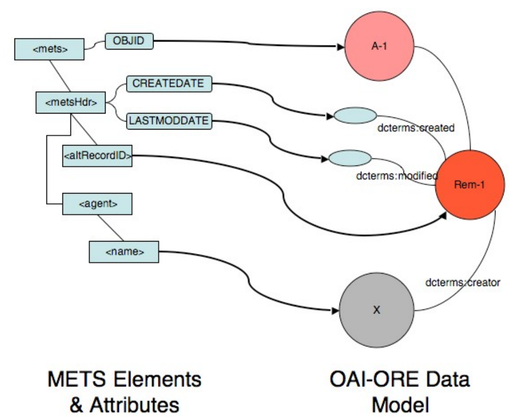

Figure 1 - Basic Metadata Mapping

3.2Structural Considerations

In addition to requiring that specific metadata elements be recorded, the OAI-ORE abstract data model also imposes certain structural requirements on a resource map. The requirements include that a resource map must describe a single aggregation, that it must assert the relationship between the aggregation and aggregated resources using the ore:aggregates predicate, and that it must insure that the triples contained within a resource map form a connected RDF graph. METS encodes must adopt certain authoring practices in creating METS documents in order to insure that it is possible to fulfill these requirements in mapping from METS to an OAI-ORE resource map, and the design of METS effectively precludes mapping a single METS file to a single OAI-ORE resource map serialization in all but the most trivial instances of METS.

The heart of a METS files is the structural map element (<structMap>); it is the only major component of a METS file that is required by the METS schema. The structural map sets forth an abstract, hierarchical model of the digital library object and then links various content files (or portions of content files) to portions of the tree structure within the structural map. In mapping from METS to an RDF graph representation conforming with the OAI-ORE model, the structural map must serve as our starting point.

The tree structure within a structural map is recorded as a nested series of division (<div>) elements. A given <div> element may contain further <div> elements, one or more <mptr> elements (which link to another METS file representing the content for the <div> element), one or more <fptr> elements (that link to content files identified in the <fileSec> portion of the METS document representing the content of the <div>), or any combination of these. The tree structure in a METS structural map may be arbitrarily deep, and it is possible to have more than one structural map within a single METS file.

While OAI-ORE allows for recursive nesting of aggregations, it also requires that any given resource map describe only a single aggregation; recursive nesting of aggregations is achieved by having any subsidiary aggregation identified by its URI and also including an ore:isDescribedBy relationship providing the URI for the resource map for the subsidiary aggregation. Any <div> element within a structural map in METS may potentially aggregate one or more resources (either via <mptr> links to other METS files or <fptr> links to content files within the METS <fileSec> element). Given this, any <div> element within a METS file must be seen as constituting an aggregation, and mapped to its own OAI-ORE resource map. A complex structural map in METS will therefore map into a set of recursive aggregations.

As mentioned, any given <div> element may contain subsidiary <div> elements, <mptr> elements, and <fptr> elements. All of these subsidiary elements constitute aggregated resources. In the case of subsidiary <div> and <mptr> elements, they represent aggregated resources that are themselves aggregations. The <fptr> element is a more complicated case, due to both the content model defined for the <fptr> element itself and the nature of the information recorded in any <file> elements identified by an <fptr> element. In the simplest case, an <fptr> provides a link to a single <file> element in the METS <fileSec> area; however, an <fptr> may also containing subsidiary elements (<par> and <seq>) which in turn identify <file> elements within the <fileSec> via subsidiary <area> elements. That is, an <fptr> element may also serve to aggregate resources, and may in fact contain a tree structure (using nested <par> and <seq> elements).

Complicating this situation further is the fact that resources in a resource map must be identified via a protocol-based URI. While METS can record a URI for a content file within the xlink:href attribute on the <FLocat> subelement within the <file> element, the content model for the <file> element is complicated and allows for a variety of other possibilities. A <file> element can embed the content file directly within the METS file itself (either as Base64 encoded binary or in XML format), and it may also identify subsidiary byte streams within a single content file. It may even indicate the presence of hierarchically nested files (to describe situations such as Unix tar archives, for example). In this latter case, a <file> element describes an aggregation, requiring its own resource map if mapped to the OAI-ORE data model.

Given the above, those mapping a METS file into the OAI-ORE data model must adhere to the following rules if they wish to produce a conformant OAI-ORE serialization:

- In the case of a METS document with multiple <structMap> elements, each <structMap> such be considered to describe a separate aggregation, with the root <div> element for each <structMap> serving as the starting point for mapping from the METS document into the OAI-ORE data model;
- All <div> elements must be considered to describe an aggregation, and will require their own resource map;
- A subsidiary <div> element within a <div> must be considered to describe an aggregation, and should be treated as an instance of recursive aggregation;
- A subsidiary <mptr> element within a <div> identifies a separate METS file which must be considered to describe an aggregation, and should be treated as an instance of recursive aggregation;
- A subsidiary <fptr> element which identifies a <file> element containing a subsidiary <FLocat> element with a protocol-based URI in the xlink:href attribute may be considered to describe a single resource and be treated as a simple case of aggregation, with the xlink:href attribute value on the <file> element serving as the object of a ore:aggregates predicate.
- A subsidiary <fptr> element which employs either <par> or <seq> subsidiary elements to link to multiple <file> elements must be considered to describe an aggregation and should be treated as an instance of recursive aggregation.
- A <par> or <seq> element must be considered to describe an aggregation and be treated as a case of recursive aggregation;
- An <area> element that does not employ the SHAPE, COORDS, BEGIN, END, BETYPE, EXTENT or EXTTYPE attributes, and that is linked to a <file> element containing a single subsidiary <FLocat> element with a protocol based URI in the xlink:href attribute, may be considered to describe a single resource and be treated as a simple case of aggregation, with the xlink:href attribute value serving as the object of a ore:aggregates predicate;
- An <area> element that does employ the SHAPE, COORDS, BEGIN, END, BETYPE, EXTENT or EXTTYPE attributes, and that is linked to a <file> element which contains a single subsidiary <FLocat> element with a protocol-based URI in the xlink:href attribute, may be considered to describe a single resource. However, METS does not provide a single attribute or element to contain a URI to identify such a subsidiary component of the data within a file so identified. In these cases, those mapping from METS to OAI-ORE should consider manufacturing a URI using the URI contained within the <FLocat> element's xlink:href attribute as a base, and using the <area> element's attribute names and attribute values as an

appended query, as a means of identifying the resource within a resource map;[2]

- An <area> element which is linked to a <file> element containing more than one <FLocat> element, a subsidiary <file> element, or a subsidiary <stream> element must be considered to describe an aggregation and be treated as a case or recursive aggregation;
- A <file> element containing a single <FLocat> element with a xlink:href attribute may be considered a resource;
- A <file> element containing more than one <FLocat> element must be considered to describe an aggregation and be treated as a case of recursive aggregation;
- A <file> element containing a subsidiary <file> element must be considered to describe an aggregation and be treated as a case of recursive aggregation;
- A <file> element containing a subsidiary <stream> element must be considered an aggregation and be treated as a case of recursive aggregation.

An example of mapping a simple case of METS to OAI-ORE aggregations and aggregated resources is provided in Figure 2 below. This mapping assumes that each <file> element contains a single <FLocat> element with an xlink:href attribute containing a protocol-based URI and no additional subelements.

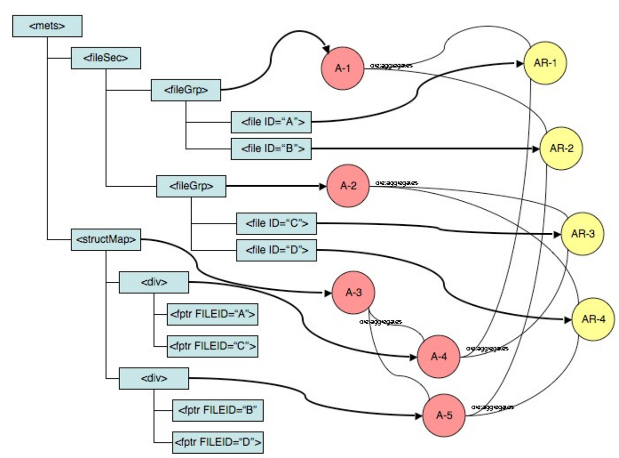

Figure 2 - Mapping METS Structures to OAI-ORE Aggregations & Aggregated Resources

The use of <FContent> elements within a METS file is obviously problematic when trying to map into the OAI-ORE data model, as neither the <FContent> element nor the <file> element provides a separate subelement or attribute to associate a URI with the embedded content file. If a METS document containing <FContent> elements needs to be mapped into the OAI-ORE abstract model, the embedded content files should be extracted, stored externally to the METS document, have URIs assigned to them, and new <FLocat> elements employing those URIs should be inserted into the METS file to replace the previous <FContent> elements.

As the above list makes clear, METS is, if anything, a bit too fond of the concept of aggregation in its design. Nor is METS' use of the concept of aggregation of resources limited to the above examples. Several other types of aggregation should be mentioned, as those mapping from METS into OAI-ORE may want to account for these additional groupings when converting from METS into an OAI-ORE serialization.

The most prominent example is the use of <fileGrp> elements to divide <file> elements up into sets. <fileGrp> elements are used to group classes of content files comprising part of a digital library object (e.g., to group all Master image files together). In cases where a METS file uses a single <fileGrp> element, considering this grouping as an aggregation might be technically correct, but it adds little useful information to reproduce this in an OAI-ORE serialization. However, in cases where a METS file employs multiple <fileGrp> elements, retaining knowledge of these aggregations might be beneficial. In those cases, each <fileGrp> should be considered to describe an aggregation, with the subsidiary <file> elements providing the needed information about the aggregated resources (such as their URIs), and a separate resource map produced for them. A resource map generated from the structural map information within a METS file may wish to include, for any <file> element mapped to a resource within the OAI-ORE aggregation, an ore:isAggregatedBy predicate to assert that the resource is a constituent of the aggregation originally described by the <fileGrp> element.

METS has another mechanism for indicating associated groups of files, the GROUPID attribute on the <file> element. The GROUPID attribute is used to indicate that files located within different file groups are equivalent (for a locally-defined meaning of 'equivalent'). For example, the GROUPID attribute might be used to indicate that a master TIFF image described in one <fileGrp> is equivalent to a JPEG derivative prepared from the TIFF described in a separate <fileGrp>. Since METS' designers intentionally left the semantics of the GROUPID attribute's use vague, appropriate mechanisms for mapping these groupings of files may vary depending on local use of the GROUPID attribute. If GROUPID is used to indicate that files described in separate <fileGrp>s are versions of one another, then adding dcterms:isVersionOf assertions linking all of aggregated resources in the resource map with information derived from <file> elements which share a GROUPID attribute value might be an adequate means of recording this information. Creating a new resource map to describe these same resources as an aggregation of their own would be an alternative approach.

The METS standard does not only aggregate content files, however. It can also aggregate metadata records. In fact, any METS record may be viewed as an aggregation of metadata records. One particular segment of the METS format, the administrative metadata section (<amdSec>), deserves special mention, as it has can have a relationship with the <fileGrp> elements of a METS document that may complicate mapping into an OAI-ORE serialization.

METS allows any <div>, <file>, <stream> or <fileGrp> element to be linked to administrative metadata records contained within a <amdSec> element (including any of the <amdSec> subelements <techMD>, <rightsMD>, <sourceMD> and <digiprovMD>)

[2] An <area> element describing a byte stream within an audio file using time codes to indicate the beginning and end points within the file might look like this: <area FILEID="F1" BEGIN="00:02:45" END="00:06:38" BETYPE="TIME">. Assuming that the <FLocat> element for the linked <file> element has an xlink:href attribute value of http://mets.uiuc.edu/test.wav, then a generated URI for the resource identified by the <area> element might be http://mets.uiuc.edu/test.wav?BEGIN=00:02:45&END=00:06:38&BETYPE=TIME

using XML ID/IDREF attribute matching. A single <amdSec> element can contain multiple metadata records of different types, and a <fileGrp> element can link to a root <amdSec> element containing multiple subsidiary elements. The METS Editorial Board has clearly defined the semantics in this case (5, p. 29). In the case of a link between a <fileGrp> element and any record within an <amdSec>, or to the <amdSec> record itself, the linked administrative metadata records are assumed to describe the individual files contained within the file group, not the file group itself. That is to say, a link between a <fileGrp> element and an <amdSec> does not associate the aggregation represented by the <fileGrp> with the administrative metadata, but the individual aggregated resources with the administrative metadata. ID/IDREF Linking between <fileGrp> and <amdSec> elements is simply a convenient shorthand to record the existence of links between a large number of files and administrative metadata records. In converting from METS to OAI-ORE serialization then, the information within administrative metadata records needs to be asserted as properties of the appropriate aggregated resources, and not of the aggregation itself.

One final point should be made regarding authoring practices for METS documents and mapping to OAI-ORE's data model. OAI-ORE insists that the assertions contained within a resource map must be extractable as a set of triples that when combined form a RDF graph that is connected. We can look at a METS document as a directed graph structure where elements represent vertices and arcs are represented by either containment (an arc travels from a parent element to its children) or by ID/IDREF associations (an arc travels from the element possessing an IDREF attribute to the element with the matching ID attribute). When converting from METS into OAI-ORE, we need to insure that all of the information translated constitutes a connected graph. As the structural map identifies the resources to be contained within an aggregation, it should serve as the starting point for exploring the graph structure, and elements that are to be mapped into the OAI-ORE data model must be reachable via a traversal starting from the root <div> node of the structural map.[3]

It is possible to place a great deal of information within a METS file that is not accessible via such a traversal. A <dmdSec> does not necessarily have to be associated with a <div>, <stream>, <file> or <fileGrp>, nor does an <amdSec>. However, <dmdSec>, <amdSec>, <techMD>, <rightsMD>, <sourceMD>, <digiprovMD>, and <file> elements in a METS file which are not referenced via an ID/IDREF link do not form part of a graph, and therefore cannot be mapped successfully into the OAI-ORE data model. METS authors should make certain that any information that they intend to map into an OAI-ORE data model is included within such a graph by employing the appropriate ID/IDREF associations.

It should be noted that employing a rule which states that METS information that is to be mapped into an OAI-ORE model must be accessible via a traversal over a directed graph originating with the root <div> node in the structural map effectively eliminates both the <structLink> and <behaviorSec> sections of a METS document from inclusion within any resource map produced from a METS document. While this may seem unfortunate, it is probably the most appropriate way to handle these sections. Software behaviors in a <behaviorSec> should operate on a METS <structMap> or <div>, and the translation from a METS file into an OAI-ORE serialization would in all likelihood render those behaviors useless for the OAI-ORE incarnation. The <structLink> section describes hyperlink structures between <div> nodes in a METS <structMap>, usually as a means of replicating hyperlink structures in underlying content (such as web archives). It is debatable whether carrying this information from METS into OAI-ORE is appropriate, but if it is deemed essential, such a move can be supported. The <smlink> elements in a <structLink> section link two <div> elements in the <structMap>. As each <div> should become its own aggregation when mapped into OAI-ORE, the information contained in each <smlink> XLink attributes could be recorded as an assertion linking the two aggregations. Note that the W3C recommends that when harvesting XLink information for RDF statements, the value of the xlink:arcrole attribute must be employed as the predicate for the relationship being expressed 5, so if METS documents are to be mapped into the OAI-ORE model, any <smlink> elements must include an xlink:arcrole attribute.

4.Conclusion

The mapping between METS and OAI-ORE outlined above demonstrates the flexibility of the OAI-ORE data model and its ability to accommodate the complex relationships between data and metadata within a community standard such as METS. It also demonstrates that certain forms of information are simpler to map from METS into OAI-ORE than others. Aggregation relationships inherent in METS, such as the recursive nesting of <div> elements, map into the OAI-ORE data model relatively transparently. Other forms of structural metadata within METS, such as the ability to identify geometric subregions of image files or segments of audio/video data streams as the target of a <div>, are somewhat more problematic to map into OAI-ORE's data model (although by no means impossible). Institutions interested in mapping from METS into OAI-ORE will need to decide what needs they are trying to address with this mapping and how much of the information contained within METS needs to be translated into an OAI-ORE serialization syntax. For some cases, a relatively simple transformation of METS which converts <file> elements into aggregated resources and then indicates the presence of a METS file for those wishing more comprehensive metadata regarding a resource map suffice. In others, a complete transformation of all metadata residing within a METS document into OAI-ORE serialization syntax may be desirable. The mapping provided here attempts to outline a relatively complete mapping between the data models and syntaxes of the respective standards, but the manner and extent of mapping between METS and OAI-ORE descriptions of an aggregation will obviously depend on local needs and practices.

This mapping also shows that conversion between METS and OAI-ORE can be automated, although there may be a certain amount of loss of information in the conversion process, depending on the construction of the original METS document. METS documents that contain metadata sections not linked via ID/IDREF attributes from the METS structural map or file sections can only be partially converted into OAI-ORE. This mapping also indicates that a certain amount of caution needs to be employed when engaging in an automated conversion from METS to OAI-ORE serializations. A METS agent with a role of creator can be automatically mapped to a dcterms:creator element in an OAI-ORE serialization, but a computer will not be able to determine whether the agent in the METS document is a human being or a machine agent. Given OAI-ORE's rules on the nature

[3] The exceptions are those metadata elements that relate directly to the description of the aggregation represented by the METS file: the OBJID attribute, and the information contained in the METS header element.

of creators, it would be easy to produce an OAI-ORE version of a METS document that was syntactically correct, but semantically invalid.

Mappings such as those outlined within this paper have greater significance than simply easing problems of interoperability between contemporary communities of practice. As Ashley 5 recently noted, we have little experience to date with problems arising in digital curation from the migration of digital objects across schemas and ontologies over time. By considering the requirements of an exchange syntax such as OAI-ORE, document authors employing localized formats such as METS can improve their chances of interacting successfully with other contemporary communities of practice, as well as those in the future.

5.REFERENCES

[1] Ashley, K. (April 3, 2009). Panel Presentation on "Gaps and Persistent Challenges" at *DigCCurr 2009: Digital Curation Practice, Promise and Prospects, April 1-3, 2009, Chapel Hill, NC*.

[2] Consultative Committee for Space Data Systems (January 2008). *XML Formatted Data Unit (XFDU) Structure and Construction Rules*. CCSDS 661.0-B-0. Washington, DC: CCSDS Secretariat. http://sindbad.gsfc.nasa.gov/xfdu/pdfdocs/xfduspec.pdf

[3] International Digital Publishing Forum (Sept. 11, 2007). *Open Publication Structure (OPS) 2.0 v1.0*. http://www.openebook.org/2007/ops/OPS_2.0_final_spec.html

[4] ISO/IEC (1997). *Information technology – Hypermedia/Time-based structuring language (HyTime)*. ISO/IEC 10744:1997.

[5] METS Editorial Board (2007). *Metadata Encoding and Transmission Standard Primer and Reference Manual*. Washington, DC: Digital Library Federation. http://www.loc.gov/standards/mets/METS%20Documentation%20final%20070930%20msw.pdf

[6] Open Archives Initiative (Oct. 17, 2008). *ORE Specification – Abstract Data Model*. Carl Lagoze et al. (Eds.). http://www.openarchives.org/ore/1.0/datamodel.html#Aggregation

[7] Open Archives Initiative (Oct. 17, 2008). *ORE Specification – Vocabulary*. Carl Lagoze et al. (Eds.). http://www.openarchives.org/ore/1.0/vocabulary.html

[8] Open Archives Initiative (Oct. 17, 2008). *ORE User Guide – Primer*. Carl Lagoze et al. (Eds.). http://www.openarchives.org/ore/1.0/primer.html

[9] TEI Consortium (Oct. 31, 2008). *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. Lou Burnard and Syd Bauman (Eds.). TEI Consortium. http://www.tei-c.org/release/doc/tei-p5-doc/en/Guidelines.pdf

[10] World Wide Web Consortium (Sept. 29, 2000). *Harvesting RDF Statements from XLinks*. Ron Daniel Jr. (Ed.). http://www.w3.org/TR/xlinks