

METS: Implementing a metadata standard in the digital library

Introduction

The digitization of library materials is now an established technology, and has been instituted throughout the world in programmes which convert holdings to digital form in order to allow wider access and better preservation. Although the techniques of conversion are now well understood and have been implemented successfully on a wide scale, the question of metadata for administering and describing these materials has not yet been satisfactorily addressed in the way in which it has long been for more traditional library materials: no standard container for these descriptions, analogous to the MARC record in the standard library environment, has yet been adopted universally. The consequences of this lack of standardization are poor cross searching abilities, limited facilities for the interchange of records and in many cases metadata tied to proprietary packages which constantly threaten obsolescence and the consequent future costs of conversion.

The metadata needed for the adequate description and maintenance of a digital object is potentially far more complicated than that needed for its equivalent in a traditional library: in addition to the standard descriptive metadata which one might find in a catalogue record, administrative metadata, necessary for the handling, archiving and maintenance of the digital object, is also an essential feature of the digital environment. In addition, digital objects are usually far from the unitary items that monographs constitute, and so require often complex descriptions of their internal structure (structural metadata) in addition to descriptions of their intellectual content.

Various attempts have been made to define standards for the handling of this heterogeneous metadata, with varying degrees of success. One scheme which has rapidly achieved recognition as the most widely applicable standard is METS (“Metadata Encoding and Transmission Standard”)¹, an XML schema devised by the Library of Congress and the Digital Library Federation. METS attempts to provide a logical and tightly structured framework within which all metadata, descriptive, administrative and structural, for a digital library object can be contained.

How METS works

A METS file contains five sections, each of which records a type of metadata relevant to a given facet of a digital object (see Appendix 1 below for a sample record). The core of a METS file is called the *structural map*; this maps out the internal structure of the object by using a series of nested <div> elements, which mirror the major sub-divisions within the object and their relationships to each other. For a monograph, for instance, this may indicate divisions into chapters, and within those into sub-chapters, sections, sub-sections etc. Each <div> element uses a pointer to reference the files that contain the images, texts or other digital surrogates of its components: these are themselves contained in another section, the *file group section*, where the files in all their variant forms (for instance thumbnails, delivery images, and archival images in the case of still image files) are listed together, each separately marked with a unique identifier so that they can be referenced from within the structural map.

Descriptive and administrative metadata for the object are held in sections of the METS file called <dmdSec> and <admSec> respectively. Both follow a similar pattern, either allowing the metadata in question to be embedded directly in the METS document (in an <mdWrap> element), or to be held in a remote file and referenced by an <mdRef> element. The metadata itself may be held in any relevant scheme, identified by the MDTYPE attribute of its parent element: in the example given, qualified Dublin Core fields are used for descriptive metadata within an <mdWrap> container, while the administrative metadata is held in remote files referenced by URLs.

A final section, <behaviourSec>, not shown in the example, allows the encoding of information on how the object should be rendered to the user (for instance, by specifying software packages necessary for its display).

A fuller description of METS and its functions may be found in a paper by the author written for the UK's Joint Information System's Committee's Technology and Standards Watch reports².

METS in the Oxford Digital Library

The Oxford Digital Library, a heterogeneous collection of digital objects taken from the holdings of Oxford University's libraries, has implemented METS as the core of its metadata system. The metadata needs of such a varied set of materials are complex, and a number of essential criteria were soon established that any viable metadata system would have to meet:-

- ⊙ it would have to be written in XML to provide archival robustness, freedom from being tied to a proprietary system, and ready interchangeability with other systems and collections
- ⊙ it would have to be able to cope with a heterogeneous variety of metadata but would nevertheless have to be as rigorous as metadata systems currently in place in the more traditional library environment
- ⊙ it would have to handle structural metadata, which is almost completely neglected in the world of the MARC record
- ⊙ it would have to be easily automated and maintained, and allow the conversion of legacy data from existing projects.

METS fulfills many of these requirements: it is written in XML, is very logical (if verbose) and so easily automated, and handles the internal structure of objects very clearly. It is, however, only a container for metadata, analogous to the MARC record, and does not specify the *content* of the metadata itself: to implement it usefully, therefore, required the employment of cataloguing rules (analogous to AACR2) to govern the content that populates its structures.

A rigorous set of guidelines were therefore drawn up to cover the content of the descriptive metadata used within the ODL³. Dublin Core fields form the main standard for content: these use a strict set of qualifiers, which are designed as far as possible to map to AACR2 practices, so allowing ODL procedures to conform as closely as possible to those followed in standard cataloguing. All names (personal, geographic and corporate) used in ODL records follow standard name authority procedures so that they can be incorporated into pre-existing indexes, and MARC relator codes are used to qualify these names so that an ODL record may readily be integrated with traditional catalogues.

The successful implementation of METS depends to a major extent on the application of a coherent and easily maintained system of identifiers for all digital objects, their components and their metadata. METS functions essentially by an extensive system of references, which relate digital files to their metadata (descriptive and administrative), and to their place in the structure of the digital object. A clear identifier is required for every component of a METS file from the ID for the file itself to the divisions within the overall structure of the document: substantial effort must be put into devising a well thought-out system of these identifiers before METS can be implemented.

The ODL has implemented a strict hierarchy of identifiers which are based on a nine-character code for each collection. This is used to generate identifiers for each digital object, each of its component files, the metadata files which describe it and each component within these metadata files. The overall ID scheme is shown in diagram 1 below, and its application for components within the METS file may be seen in the sample METS file in Appendix 1.

Edit munahi010-aab description

Title	<input type="text" value="<main>Études sur les glaciers#<supplied>atlas"/>
Subject	<input type="text" value="<lcsch>Glaciers"/>
Type	<input type="text" value="Image.Graphic.Map"/>
Description	<input type="text" value="<summary>Plates accompanying a study of glaciers by 19th century glaciologist Louis A"/>
<hr/>	
Creator	<input type="text" value="<aut>Agassiz, Louis, 1807-1873"/>
Date	<input type="text" value="<issued>1840"/>
Identifier	<input type="text"/>
Rights	<input type="text"/>
Format	<input type="text" value="<dimensions>480 x 320#<extent>18 plates: ill."/>
Source	<input type="text" value="OUME. 52"/>
<hr/>	
Contributor	<input type="text" value="<tgt>Bettannier, Joseph"/>
Publisher	<input type="text" value="Neuchâtel (Switzerland): Jent et Gassmann"/>
Language	<input type="text" value="fre"/>
Coverage	<input type="text" value="<spatial>Alps"/>
Relation	<input type="text" value="<isRequiredBy>ODL:munahi010"/>

Diagram 2: Input screen for Dublin Core metadata

Using METS with other standards

METS provides a viable metadata framework for a digital object, but will often prove inadequate by itself as a way of expressing all of its relevant facets, and is certainly incapable of expressing information on the whole collection of which the object forms a part. To do so requires METS to be used in conjunction with other standards which can provide this functionality: the ODL therefore uses it in conjunction with other XML-based standards as one component in a wider system.

The ODL is a collection-based repository, and attempts to subdivide its components into meaningful collections which represent their diversity of format and content while still allowing full cross-searching. To encode the metadata for its component collections, the ODL uses the *Encoded Archival Description (EAD)*,⁴ the most widely-used standard within the archival community for collection level descriptions: this ensures compatibility with traditional archives and allows the content of the ODL's virtual collections to be searched in tandem with that of its non-electronic equivalents.

The EAD file for each collection follows the same system of identifiers imposed on all components of the ODL, so allowing easier navigation and also the automatic generation of this often complex

set of identifiers. EAD allows the digital components of a collection to be divided rationally into series, sub-series etc, which makes the browsing of its contents easier and more helpful, while still allowing searching across the collection as a whole. The information encoded within the EAD file on the digital objects themselves is the minimum necessary to allow browsing and the location of the METS file which contains the full metadata for each item: this includes the Dublin Core title element for the digital object and the identifier necessary to locate its METS file.

Electronic texts will be an important part of the ODL, and so an accepted standard for encoding them is essential: the Text Encoding Initiative (TEI)⁵ is the most widely accepted DTD for marking up texts, and so it will be used to encode all such items held within the ODL collections. The TEI is very flexible in its use, and so a standard for its application has to be followed: the ODL will be using the TEI in Libraries Guidelines⁶ issued by the Library of Congress and other bodies, which are currently in the process of revision.

Linking the METS file to the TEI files that contain textual information is straightforward: from within the <div> elements in the structural map, the TEI file itself is referenced by a <fptr> element, within which is contained an <area> element: within this, the beginning and end of the part of the TEI which corresponds to the <div> is referenced by BEGIN and END attributes. A reference to a section within a TEI file may, therefore, look like this:-

```
<div ID="munahi010-aaa-div.1.4" LABEL="Chapter 1">
  <fptr>
    <area FILEID="munahi010-aaa-aaa" BEGIN="munahi010-aaa-aaa.2"
END="munahi010-aaa-aaa.3"/>
  </fptr>
</div>
```

The most important consideration here, as in all uses of METS, is the implementation of a coherent and logical system of identifiers which allows the construction of METS files to be automated and their maintenance rendered much easier. Once such a system has been mapped out, the application of METS becomes more straightforward and its functionality as a logical container for a wide variety of complex metadata is more fully realized.

Conclusion

Despite its recent provenance, METS has already demonstrated that it can function well in live working environments that make considerable demands on it: the ODL project alone has demonstrated how readily it can handle complex metadata for a wide variety of source materials.

Two key factors have to be carefully considered and decided upon before METS can be used effectively, the design of a coherent system of identifiers and the choice of standards for the *content* of the metadata contained within, or referenced from, the METS file. METS is only a container for this metadata, and cannot by itself ensure either that a usable system of IDs is adopted, or that the content held within its framework is of sufficient quality to make it useful. The ODL has given much thought to both of these considerations in a successful METS implementation, and has taken care to ensure that they allow the widest possible compatibility with other library collections, electronic and otherwise.

After many years in which the digital library community has tried a bewildering variety of approaches to handling the complexities of its metadata, METS does offer the prospect of establishing a coherent but flexible standard which will bring to the digital library some of the advantages that the MARC record brought to its traditional counterpart. Its advantages will become more apparent as a critical mass of METS-encoded metadata is created, which is itself contingent on its widespread adoption: once such a mass is generated, and more software becomes available to generate and exploit METS files, the possibilities of the digital library in terms of wider access to

materials and greater co-operation between projects will be more genuinely realized.

Appendix 1: Sample METS file

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<mets xmlns="http://www.loc.gov/METS/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns:xlink="http://www.w3.org/TR/xlink"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:odldc="http://odl.ox.ac.uk/odl-dc-extensions/1.0/"
  xsi:schemaLocation="http://www.loc.gov/METS/ ../schema/mets.xsd http://www.w3.org/TR/xlink
  ../schema/xlink.xsd http://purl.org/dc/elements/1.1/ ../schema/simplifiedc20021212.xsd
  http://odl.ox.ac.uk/odl-dc-extensions/1.0/ ../schema/dc-odl.xsd"
  ID="munahi010-aaa">
<metsHdr RECORDSTATUS="interim"></metsHdr>
<dmdSec ID="munahi010-aaa-dmd-0001">
  <mdWrap MIMETYPE="text/xml" MDTYPE="DC" LABEL="Dublin Core Metadata">
    <xmlData>
      <odldc:title.main>Études sur les glaciers</odldc:title.main>
      <odldc:title.supplied>atlas</odldc:title.supplied>
      <odldc:creator.aut>Agassiz, Louis, 1807-1873 </odldc:creator.aut>
      <odldc:subject.lcsh>Glaciers</odldc:subject.lcsh>
      <odldc:description.summary>Plates accompanying a study of glaciers by 19th century glaciologist
      Louis Agassiz</odldc:description.summary>
      <odldc:description.note>Dessinés d'après nature et lithographiés par Jph. Bettannier 1840.
      Neuch,tel, Lithographie de H. Nicolet</odldc:description.note>
      <dc:publisher>Neuch,tel (Switzerland): Jent et Gassmann</dc:publisher>
      <odldc:contributor.ltg>Bettannier, Joseph</odldc:contributor.ltg>
      <odldc:date.issued>1840</odldc:date.issued>
      <dc:type>Image.Graphic.Map</dc:type>
      <odldc:format.dimensions>480 x 320</odldc:format.dimensions>
      <odldc:format.extent>18 plates: ill.</odldc:format.extent>
      <dc:source>OUM:E. 52</dc:source>
      <dc:language>fre</dc:language>
      <odldc:coverage.spatial>Alps</odldc:coverage.spatial>
      <odldc:relation.isRequiredBy>ODL:munahi010-aaa</odldc:relation.isRequiredBy>
    </xmlData>
  </mdWrap>
</dmdSec>

<amdSec ID="munahi010-aaa-amd-0001">
  <techMD ID="munahi010-aaa-tmd-0001-0">
    <mdRef MDTYPE="OTHER" OTHERMDTYPE="ODL Admin Metadata Scheme" LOCTYPE="URL"
    xlink:href="file:/export/home/odl/munahi010/admMetadata/aaa/0/munahi010-aaa-0001-0.xml"/>
  </techMD>
  <techMD ID="munahi010-aaa-tmd-0001-3">
    <mdRef MDTYPE="OTHER" OTHERMDTYPE="ODL Admin Metadata Scheme" LOCTYPE="URL"
    xlink:href="file:/export/home/odl/munahi010/admMetadata/aaa/3/munahi010-aaa-0001-3.xml"/>
  </techMD>
  <techMD ID="munahi010-aaa-tmd-0001-6">
    <mdRef MDTYPE="OTHER" OTHERMDTYPE="ODL Admin Metadata Scheme" LOCTYPE="URL"
    xlink:href="file:/export/home/odl/munahi010/admMetadata/aaa/6/munahi010-aaa-0001-6.xml"/>
  </techMD>
</amdSec>

<amdSec ID="munahi010-aaa-amd-0002">
  <techMD ID="munahi010-aaa-tmd-0002-0">
    <mdRef MDTYPE="OTHER" OTHERMDTYPE="ODL Admin Metadata Scheme" LOCTYPE="URL"
    xlink:href="file:/export/home/odl/munahi010/admMetadata/aaa/0/munahi010-aaa-0002-0.xml"/>
  </techMD>
  <techMD ID="munahi010-aaa-tmd-0002-3">
    <mdRef MDTYPE="OTHER" OTHERMDTYPE="ODL Admin Metadata Scheme" LOCTYPE="URL"
    xlink:href="file:/export/home/odl/munahi010/admMetadata/aaa/3/munahi010-aaa-0002-3.xml"/>
  </techMD>
  <techMD ID="munahi010-aaa-tmd-0002-6">
    <mdRef MDTYPE="OTHER" OTHERMDTYPE="ODL Admin Metadata Scheme" LOCTYPE="URL"
    xlink:href="file:/export/home/odl/munahi010/admMetadata/aaa/6/munahi010-aaa-0002-6.xml"/>
  </techMD>
</amdSec>

<amdSec ID="munahi010-aaa-amd-0003">
  <techMD ID="munahi010-aaa-tmd-0003-0">
    <mdRef MDTYPE="OTHER" OTHERMDTYPE="ODL Admin Metadata Scheme" LOCTYPE="URL"
    xlink:href="file:/export/home/odl/munahi010/admMetadata/aaa/0/munahi010-aaa-0003-0.xml"/>
  </techMD>
  <techMD ID="munahi010-aaa-tmd-0003-3">
    <mdRef MDTYPE="OTHER" OTHERMDTYPE="ODL Admin Metadata Scheme" LOCTYPE="URL"
    xlink:href="file:/export/home/odl/munahi010/admMetadata/aaa/3/munahi010-aaa-0003-3.xml"/>
  </techMD>
  <techMD ID="munahi010-aaa-tmd-0003-6">
    <mdRef MDTYPE="OTHER" OTHERMDTYPE="ODL Admin Metadata Scheme" LOCTYPE="URL"
    xlink:href="file:/export/home/odl/munahi010/admMetadata/aaa/6/munahi010-aaa-0003-6.xml"/>
  </techMD>
</amdSec>

<fileSec>
  <fileGrp ID="munahi010-aaa-fgrp-0001">
    <file GROUPID="0" ID="munahi010-aaa-0001-0" MIMETYPE="image/tiff" ADMID="munahi010-aaa-tmd-0001-
    0">
      <FLocat LOCTYPE="URL"
      xlink:href="file://hfs.ox.ac.uk/data/odl/munahi010/digObjects/aaa/0/munahi010-aaa-0001.tiff"/>
    </file>
    <file GROUPID="6" ID="munahi010-aaa-0001-6" MIMETYPE="image/jpeg" ADMID="munahi010-aaa-tmd-0001-
```

```

6">
    <FLocat LOCTYPE="URL" xlink:href="http:odl/munahi010/digObjects/aaa/6/munahi010-aaa-0001-
6.jpg"/>
    </file>
    <file GROUPID="3" ID="munahi010-aaa-0001-3" MIMETYPE="image/jpeg" ADMID="munahi010-aaa-tmd-0001-
3">
        <FLocat LOCTYPE="URL" xlink:href="http:odl/munahi010/digObjects/aaa/3/munahi010-aaa-0001-
3.jpg"/>
        </file>
    </fileGrp>
    <fileGrp ID="munahi010-aaa-fgrp-0002">
        <file GROUPID="0" ID="munahi010-aaa-0002-0" MIMETYPE="image/tiff" ADMID="munahi010-aaa-tmd-0002-
0">
            <FLocat LOCTYPE="URL"
xlink:href="file://hfs.ox.ac.uk/data/odl/munahi010/digObjects/aaa/0/munahi010-aaa-0002.tiff"/>
            </file>
            <file GROUPID="6" ID="munahi010-aaa-0002-6" MIMETYPE="image/jpeg" ADMID="munahi010-aaa-tmd-0002-
6">
                <FLocat LOCTYPE="URL" xlink:href="http:odl/munahi010/digObjects/aaa/6/munahi010-aaa-0002-
6.jpg"/>
                </file>
                <file GROUPID="3" ID="munahi010-aaa-0002-3" MIMETYPE="image/jpeg" ADMID="munahi010-aaa-tmd-0002-
3">
                    <FLocat LOCTYPE="URL" xlink:href="http:odl/munahi010/digObjects/aaa/3/munahi010-aaa-0002-
3.jpg"/>
                    </file>
                </fileGrp>
                <fileGrp ID="munahi010-aaa-fgrp-0003">
                    <file GROUPID="0" ID="munahi010-aaa-0003-0" MIMETYPE="image/tiff" ADMID="munahi010-aaa-tmd-0003-
0">
                        <FLocat LOCTYPE="URL"
xlink:href="file://hfs.ox.ac.uk/data/odl/munahi010/digObjects/aaa/0/munahi010-aaa-0003.tiff"/>
                        </file>
                        <file GROUPID="6" ID="munahi010-aaa-0003-6" MIMETYPE="image/jpeg" ADMID="munahi010-aaa-tmd-0003-
6">
                            <FLocat LOCTYPE="URL" xlink:href="http:odl/munahi010/digObjects/aaa/6/munahi010-aaa-0003-
6.jpg"/>
                            </file>
                            <file GROUPID="3" ID="munahi010-aaa-0003-3" MIMETYPE="image/jpeg" ADMID="munahi010-aaa-tmd-0003-
3">
                                <FLocat LOCTYPE="URL" xlink:href="http:odl/munahi010/digObjects/aaa/3/munahi010-aaa-0003-
3.jpg"/>
                                </file>
                            </fileGrp>
                            <fileGrp ID="munahi010-aaa-fgrp-0004">
                                <file GROUPID="0" ID="munahi010-aaa-0004-0" MIMETYPE="image/tiff" ADMID="munahi010-aaa-tmd-0004-
0">
                                    <FLocat LOCTYPE="URL"
xlink:href="file://hfs.ox.ac.uk/data/odl/munahi010/digObjects/aaa/0/munahi010-aaa-0004.tiff"/>
                                    </file>
                                    <file GROUPID="6" ID="munahi010-aaa-0004-6" MIMETYPE="image/jpeg" ADMID="munahi010-aaa-tmd-0004-
6">
                                        <FLocat LOCTYPE="URL" xlink:href="http:odl/munahi010/digObjects/aaa/6/munahi010-aaa-0004-
6.jpg"/>
                                        </file>
                                        <file GROUPID="3" ID="munahi010-aaa-0004-3" MIMETYPE="image/jpeg" ADMID="munahi010-aaa-tmd-0004-
3">
                                            <FLocat LOCTYPE="URL" xlink:href="http:odl/munahi010/digObjects/aaa/3/munahi010-aaa-0004-
3.jpg"/>
                                            </file>
                                        </fileGrp>
                                    </fileGrp>
                                </div>
                            </div>
                        </div>
                    </div>
                </div>
            </div>
        </div>
    </fileGrp>
</structMap>
</div ID="munahi010-aaa-div.1" LABEL="Section 1">
    <div ID="munahi010-aaa-div.1.1" LABEL="Plate 1: Panorama des glaciers du Mont Rose, partie
orientale de la chaine">
        <fptr FILEID="munahi010-aaa-fgrp-0001"/>
    </div>
    <div ID="munahi010-aaa-div.1.2" LABEL="Overlay 1a: Panorama des glaciers du Mont Rose,
partie orientale de la chaine">
        <fptr FILEID="munahi010-aaa-fgrp-0002"/>
    </div>
    <div ID="munahi010-aaa-div.1.3" LABEL="Plate 2: Panorama des glaciers du Mont Rose, partie
occidentale de la chaine">
        <fptr FILEID="munahi010-aaa-fgrp-0003"/>
    </div>
    <div ID="munahi010-aaa-div.1.4" LABEL="Overlay 2a: Panorama des glaciers du Mont Rose,
partie occidentale (sic) de la chaine">
        <fptr FILEID="munahi010-aaa-fgrp-0004"/>
    </div>
</div>
</structMap>
</mets>

```

¹ <http://www.loc.gov/standards/mets/>

² http://www.jisc.ac.uk/index.cfm?name=techwatch_report_0205.

³ Available at <http://www2.odl.ox.ac.uk/guidelines/>

⁴ <http://www.loc.gov/ead/>

⁵ <http://www.tei-c.org/>

⁶ <http://www.indiana.edu/~letrs/tei/>