METS Awareness Training

An Introduction to METS
Digital libraries – the need for a metadata standard

- Digitisation technology now well established and well-understood
- Standards for digitisation processes have settled down and are widely recognised
- Still a disparity in approaches to metadata - no 'MARC standard' for the digital library
The lack of a standard – what it means...

- poor cross-searching
- limited interchange facilities
- metadata tied to proprietary packages
- consequent obsolescence and costs of conversion
- little chance of a 'hybrid library'
What is needed?

- A standard for metadata content: analogous to AACR2
- A standardised framework for holding and exchanging metadata: analogous to the MARC record
- METS is designed to fulfil the latter function
METS to the rescue!

- Produced by Library of Congress Standards Office and Digital Library Federation
- Written in XML
- Provides framework for holding all types of metadata for digital object
- Does not prescribe content of metadata, but recommends a number of schemes for this
METS and OAIS

- METS was designed to function within the OAIS framework.
- It can act as a Submission Information Package (SIP) Delivery Information Package (DIP), providing a standardized transfer syntax.
- It acts as an Archival Information Package (AIP) for storage and preservation.
An overview of the METS file

- Generally one METS file corresponds to one digital object (which may incorporate many files)
- All metadata (descriptive, administrative and structural) encoded in single document
- Each type is held in a separate section, linked by identifiers
- All metadata and external data (eg. images, text, video) is either referenced from METS file or can be held internally
The inside of a METS file

- METSHeader
- fileSec
- dmdSec
- admSec
- behaviorSec
- structMap

- file inventory
- descriptive metadata
- administrative metadata
- behaviour metadata
- structural map
<structMap>

<div ID="munahi010-aaa-div.1" LABEL="Section 1">
  <div ID="munahi010-aaa-div.1.1" LABEL="Plate 1">
    <fptr FILEID="munahi010-aaa-fgrp-0001"/>
  </div>
  <div ID="munahi010-aaa-div.1.1" LABEL="Plate 2">
    <fptr FILEID="munahi010-aaa-fgrp-0002"/>
  </div>
</div>

</structMap>
The inside of a METS file

- **METSHeader**
- **fileSec** → file inventory
- **dmdSec** → descriptive metadata
- **admSec** → administrative metadata
- **behaviorSec** → behaviour metadata
- **structMap** → structural map
<fileSec>

fileSec

fileGrp

file

file

file
<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"
    </file>

</fileGrp>
The inside of a METS file

- METSHeader
- fileSec: file inventory
- dmdSec: descriptive metadata
- admSec: administrative metadata
- behaviorSec: behaviour metadata
- structMap: structural map
Descriptive and administrative metadata

- Descriptive and administrative metadata may be handled in two ways:
  - embedding directly within the METS file within an `<mdWrap>` element
  - being held in an external file and referenced from the METS file using an `<mdRef>` element
Cobbett's parliamentary history of England, from the Norman Conquest, in 1066 to the year, 1803: from which last-mentioned epoch it is continued downwards in the work entitled, "The parliamentary debates".

Cobbett's Parliamentary History - volume 2

<amdSec ID="munahi010-aaa-amd-0001">
  <techMD ID="munahi010-aaa-tmd-0001-0">
    <mdRef MDTYPE="MIX" LOCTYPE="URL"
      xlink:href="../munahi010-aaa-0001-0.xml"/>
  </techMD>
</amdSec>
IDs and METS

- METS uses IDs to express the relations between its component parts
- A coherent system of identifiers is therefore essential

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Project ID</td>
<td>munahi010</td>
</tr>
<tr>
<td>Item ID</td>
<td>munahi010-aaa</td>
</tr>
<tr>
<td>Technical metadata</td>
<td>munahi010-aaa-tmd-0001</td>
</tr>
<tr>
<td>File groups</td>
<td>munahi010-aaa-fgrp-0001</td>
</tr>
<tr>
<td>File IDs</td>
<td>munahi010-aaa-0001-3</td>
</tr>
<tr>
<td>divs</td>
<td>munahi010-aaa-div.1</td>
</tr>
</tbody>
</table>
What to put in a METS file?

- METS does not prescribe the content (particularly the descriptive metadata) which it can contain

- However, the METS board does endorse some schemas as recommended for use with METS:-

**Descriptive Metadata**
- Dublin Core
- MODS (Metadata Object Description Schema)
- MARCXML MARC 21 Schema (MARCXML)

**Administrative Metadata**
- Schema for Technical Metadata for Text (NYU)
- Library of Congress Audio-Visual Prototyping Project
- NISO Technical Metadata for Digital Still Images (MIX)
- METS Schema for Rights Declaration
METS Profiles

- METS is very flexible in its application – there are multiple ways of encoding everything:
  - metadata and data can be embedded or referenced
  - any scheme can be used for this metadata
  - file inventory can be organised in multiple ways (by referenced object, by type of file etc)
  - This all reduces interchangeability of METS records.
METS Profiles (cont.)

- This can be countered by METS Profiles:
  - XML documents describing application of METS in a given project/institution
  - follows METS Profile schema and each profile has to validate against it
  - registered with central repository at Library of Congress

- Profiling is essential for interoperability of OAIS packages