
Managing Content Diversity with METS Profiles

California Digital Library

Brian Tingle

Prepared for METS Opening Day West Coast; Stanford University, April 7-8, 2004:

Copyright © 2004 The Regents of the University of California

The California Digital Library maintains a growing collection of Metadata Transmission and Encoding Standard (METS) Documents. These METS Documents are representations of digital library objects that are made available through a variety of management and end user services. Extant METS Documents are mostly the results of conversions of non-METS representations of a diversity of digital objects, of various types and formats, from various sources. METS Profiles provide a systematic way of identifying species of METS Documents, and recording their important characteristics.

Table of Contents

Profile source files	1
Narrative Descriptions of Profiles	1
Machine Readable Descriptions of Profiles	2
BASE Profiles and XML Includes	2
Profiles from outside CDL	2
Files Generated from source files	2
Related URLs	3
Standards	3
Human user web services "powered by" METS Profiles	3
CDL Profiles	3

Profile source files

For each METS Profile supported in production on ark.cdlib.org, there is a master XML source file describing the profile. That master file has the extension '.profile.xml.' That master, when any XML Includes are expanded, and any resulting 'xml:base' attributes removed, will validate against the METS_Profile schema.

Narrative Descriptions of Profiles

The schema is expressive enough to record detailed specifications that must be met in order for a METS instance to conform to that PROFILE. There are two audiences that we hope would be very interested in this content; people programming systems to create METS objects that match a profile, and people programming systems that will have to do things with the objects.

As of April 2004, all METS used in production access services at CDL have been created at CDL, usually through an iterative process. The METS Profiles that exist now are more like field notes on observations of object behaviour in the wild than formal specifications that someone could use to create new objects that meet a profile. Time and resources permitting, they will be modified to increase their formality

and utility as specifications.

Machine Readable Descriptions of Profiles

ark.cdlib.org uses two different XSL Transforms on METS documents. One extracts a Dublin Core record from the object. The other is used to create the HTML index page for the object in the display service.

The tool section of the METS_Profile schema provides a URI element. For a profile to get "picked up" in the system, the METS_Profile needs to have a tool element with a tool/description/p[@ID='toQDC'] and a tool with tool/description/p[@ID='toHTML']. The tool/URI of these tools must reference an XSLT stylesheet that extracts a Dublin Core record and an XSLT to create an HTML page from the METS, respectively.

```
<tool>
  <agency>California Digital Library</agency>
  <URI>http://ark.cdlib.org/xslt/extract-dc/kt3v19p5bk.dc.xslt</URI>
  <description>
    <p ID="toQDC">This XSLT is used to extract a
      Dublin Core record from the object</p>
  </description>
</tool>
<tool>
  <agency>California Digital Library</agency>
  <URI>http://ark.cdlib.org/xslt/mets-page/kt3v19p5bk.html.xslt</URI>
  <description>
    <p ID="toHTML">This XSLT is used to create an
      HTML index page for the object.</p>
  </description>
</tool>
```

BASE Profiles and XML Includes

Lots of profiles are similar in certain ways. Common elements can be included from a Base.

```
<structMap>
  <xi:include
    href="./BASE-DynaXML2003.profile.xml
    #xpointer(/METS_Profile/structural_requirements/structMap/requirement)"/>
</structMap>
```

Profiles from outside CDL

While all the profiles and profile URIs in the system as of April 2004 are CDL assigned and maintained, this scheme will accommodate profiles from other sources. The published METS_Profile will be saved to the master source directory. If the required <tool> sections do not exist, the 2 XSLTs will be created and the sections added to the profile. Then, that profile will be ready for use in the system.

Files Generated from source files

Right now, everything that has been assigned an ARK at ark.cdlib.org has a METS Document representation acting as its binding record. Since modern CDL profiles have been assigned ARKs, they also need to be transformed into METS files for their ARKs to work. To enable this, there is a METS Profile for

METS Profiles. The METS Profile for METS Profiles includes an XSLT to generate an HTML representation of the METS Profile instance.

Multiple systems at CDL are profile aware. The METS Profiles are the authoritative source of information for these applications. They either access the files directly, or batch programs process the directory of profiles into a configuration file.

The official list of CDL profiles will be generated by batch processing METS Profiles.

Related URLs

Standards

- METS Official Web Site [<http://www.loc.gov/standards/mets/>]
- HTML version of METS Profile Schema [http://ark.cdlib.org/mets/profile_schema_documentation/]
- Rich Text Format version of METS Profile Documentation [http://www.loc.gov/standards/mets/profile_docs/METS.profile.requirements.rtf]
- Archival Resource Key (ARK) [<http://www.cdlib.org/inside/diglib/ark/>] page on Inside CDL.

Human user web services "powered by" METS Profiles

- The Online Archive of California [<http://www.oac.cdlib.org/>]
- University of California Press: eScholarship Editions [<http://texts.cdlib.org/ucpress/>]
- www.californiadigitallibrary.org [<http://www.californiadigitallibrary.org/>] (portal to UC websites with public content in CDL's METS collection)
- Japanese American Relocation Digital Archives [<http://jarda.cdlib.org/>]
- MOAC [<http://www.bampfa.berkeley.edu/moac/>] California museums working with libraries and archives to increase and enhance access to cultural collections
- The California Heritage Collection [<http://sunsite.berkeley.edu/CalHeritage/>]
- The CalCultures Project [<http://calcultures.cdlib.org/>] will have content in production Summer 2004. This will be the first project where CDL will be ingesting METS from a third party into production. All the METS for this project will be generated by GenX at UC Berkeley. The exact details are not resolved yet, but I imagine that we will have one profile set up at CDL for all METS that are generated with GenX, unless we find out we need more.

CDL Profiles

<http://ark.cdlib.org/mets/profiles/> will be maintained with an up-to-date listing of profiles in production, with links to available documentation.

BEPRESS repository export (<http://ark.cdlib.org/ark:/13030/kt200014dk>)
This profile is under development. It is not active at this time.

DC OAC image (OAC-LSTA-DC) (<http://ark.cdlib.org/ark:/13030/kt4g5012g0>)

Image objects created for LSTA from Dublin Core source.

DC OAC text (OAC-ETEXT) (<http://ark.cdlib.org/ark:/13030/kt7j49p867>)

Profile for OAC texts with Dublin Core metadata. Structure is optimized for dynaXML.

DDI Table (<http://ark.cdlib.org/ark:/13030/kt1g5010zb>)

DDI object for Counting California

EAD DAO* extracted object (<http://ark.cdlib.org/ark:/13030/kt3q2nb7vz>)

Encoded Archival Description provides a mechanism to define objects in is dao and dapgrp tags. This profile is for METS object created by a batch extraction process from EAD Finding Aids in the Online Archive of California.

EAD Finding Aid (<http://ark.cdlib.org/ark:/13030/kt0t1nb6x7>)

This profile is used internally by CDL during the ingest of EAD encoded Finding Aids.

MODS OAC image (<http://ark.cdlib.org/ark:/13030/kt400011f8>)

Images created for LSTA project from MARC source use this profile.

MODS OAC text (<http://ark.cdlib.org/ark:/13030/kt5k40135s>)

Profile for OAC texts with MODS metadata. Structure is optimized for dynaXML.

MODS eSch text (oceans) (<http://ark.cdlib.org/ark:/13030/kt5z09p6zn>)

Same as the OAC MODS profile, but with different branding. We do not really want to trigger branding with a profile, so this is just temporary until we have better branding.

Profile for Profiles (<http://ark.cdlib.org/ark:/13030/kt8s20152f>)

A Profile for METS documents describing profiles of METS documents

UCPEE netlib book (<http://ark.cdlib.org/ark:/13030/kt3v19p5bk>)

Objects created by the UC Press eScholarship Editions project.

pre MODS (crs reports) (<http://ark.cdlib.org/ark:/13030/kt667nb8wm>)

This is a place holder profile to assign ARKs to these Items. Now that these have ARKs, SCP can create MARC records, and then we can create a MODS based profile for these.

submission package profile (<http://ark.cdlib.org/ark:/13030/kt4k40124g>)

place holder for profile in development.

BASE Profile for DynaXML 2003 (BASE-DynaXML2003)

This is not a Full Profile. It is just a BASE to build on. It documents how DynaXML 2003 (not XTF) requires METS. This will change in 2004 with XTF.

BASE Profile for LSTA images (BASE-DynaXML2003)

This is not a Full Profile. It is just a BASE to build on. It documents how simple/moderately complex image objects designed from LSTA are structured.