Creating METS Application Profiles using METS and MODS

Morgan Cundiff

Network Development and MARC Standards Office

Library of Congress

Outline

- Quick overview of METS
- Quick overview of METS Profiles
- Quick overview of MODS relatedItem element
- Description of LC Compact Disc Profile (especially structural features)
- Demo of a CD object (I Hear America Singing)

What is METS?

METS is an XML Schema designed for the purpose of creating XML document instances that express the hierarchical structure of digital library objects, the names and locations of the files that comprise those objects, and the associated metadata. METS can, therefore, be used as a tool for modeling real world objects, such as particular document types.

METS

What are the 7 Sections of a METS Document?

```
<mets>
  <metsHdr/>
  <dmdSec/>
  <amdSec/>
  <fileSec/>
  <structMap/>
  <structLink/>
  <behaviorSec/>
</mets>
```

The Structure Map

```
<mets>
  <structMap>
    <div>
      <div></div>
    </div>
  </structMap>
</mets>
```

Example 1

The File Section (fileSec)

```
<mets>
  <fileSec></fileSec>
  <structMap></structMap>
</mets>
```

Example 2

METS

The Descriptive Metadata Section (dmdSec)

```
<mets>
  <dmdSec></dmdSec>
  <fileSec></fileSec>
  <structMap></structMap>
</mets>
```

The Descriptive Metadata Section with mdRef

```
<mets>
  <dmdSec>
    <mdRef/>
  </dmdSec>
  <fileSec></fileSec>
  <structMap></structMap>
</mets>
```

Example 3

The Descriptive Metadata Section with mdWrap

```
<mets>
 <dmdSec>
  <mdWrap>
   <xmlData>
       <!-- insert data from different namespace here -->
   </xmlData>
  </mdWrap>
 </dmdSec>
 <fileSec></fileSec>
 <structMap></structMap>
</mets>
```

Example 4

The Descriptive Metadata Section with MODS as extension schema

```
<mets:mets>
  <mets:dmdSec>
    <mets:mdWrap>
      <mets:xmlData>
        <mods:mods></mods:mods>
      </mets:xmlData>
    </mets:mdWrap>
  </mets:dmdSec>
  <mets:fileSec></mets:fileSec>
  <mets:structMap></mets:structMap>
</mets:mets>
```

Example 4

The Descriptive Metadata Section with MODS and relatedItem elements

```
<mets:mets>
 <mets:dmdSec>
  <mets:mdWrap>
   <mets:xmlData>
    <mods:mods>
     <mods:relatedItem type="constituent">
      <mods:relatedItem type="constituent"></mods:relatedItem>
     </mods:relatedItem>
    </mods:mods>
   </mets:xmlData>
  </mets:mdWrap>
 </mets:dmdSec>
 <mets:fileSec></mets:fileSec>
 <mets:structMap></mets:structMap>
</mets:mets>
```

MODS relatedItem element

1. Child element to MODS
2. relatedItem element has same content model as mods (titleInfo, name, subject, physicalDescription, note, etc)
3. The relatedItem element makes it possible to create very rich analytic records for contained works within a MODS records
4. relatedItem element is repeatable and it can be nested recursively (thus making it possible to build a hierarchical tree structure)

```
<mods:mods>
 <mods:titleInfo>
  <mods:title>Bernstein conducts Beethoven and Mozart</mods:title>
 </mods:titleInfo>
 <mods:name>
  <mods:namePart>Bernstein, Leonard</mods:namePart>
 </mods:name>
 <mods:relatedItem type="constituent">
  <mods:titleInfo>
   <mods:title>Symphony No. 5</mods:title>
  </mods:titleInfo>
  <mods:name>
   <mods:namePart>Beethoven, Ludwig van</mods:namePart>
  </mods:name>
  <mods:relatedItem  type="constituent">
   <mods:titleInfo>
    <mods:partName>Allegro con moto</mods:partName>
   </mods:titleInfo>
  </mods:relatedItem>
  <mods:relatedItem type="constituent">
   <mods:titleInfo>
    <mods:partName>Adagio</mods:partName>
   </mods:titleInfo>
  </mods:relatedItem>
 </mods:relatedItem>
</mods:mods>
```

METS document with two hierarchies (relatedItem and div elements)

```
<mets:mets>
 <mets:dmdSec>
  <mets:mdWrap>
   <mets:xmlData>
    <mods:mods>
     <mods:relatedItem>
      <mods:relatedItem></mods:relatedItem>
     </mods:relatedItem>
    </mods:mods>
   </mets:xmlData>
  </mets:mdWrap>
 </mets:dmdSec>
 <mets:fileSec></mets:fileSec>
 <mets:structMap>
  <mets:div>
   <mets:div></mets:div>
  </mets:div>
 </mets:structMap>
</mets:mets>
```

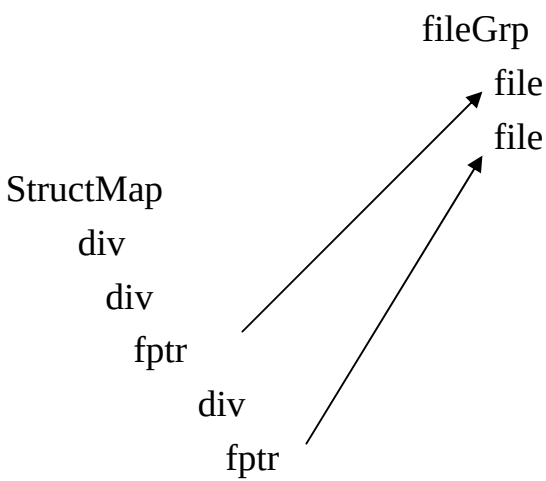
Linking in METS Documents
(XML ID/IDREF links)

DescMD
mods
relatedItem
relatedItem
AdminMD
techMD
sourceMD
digiprovMD
rightsMD
fileGrp
file
file
StructMap
div
div
fptr
div
fptr

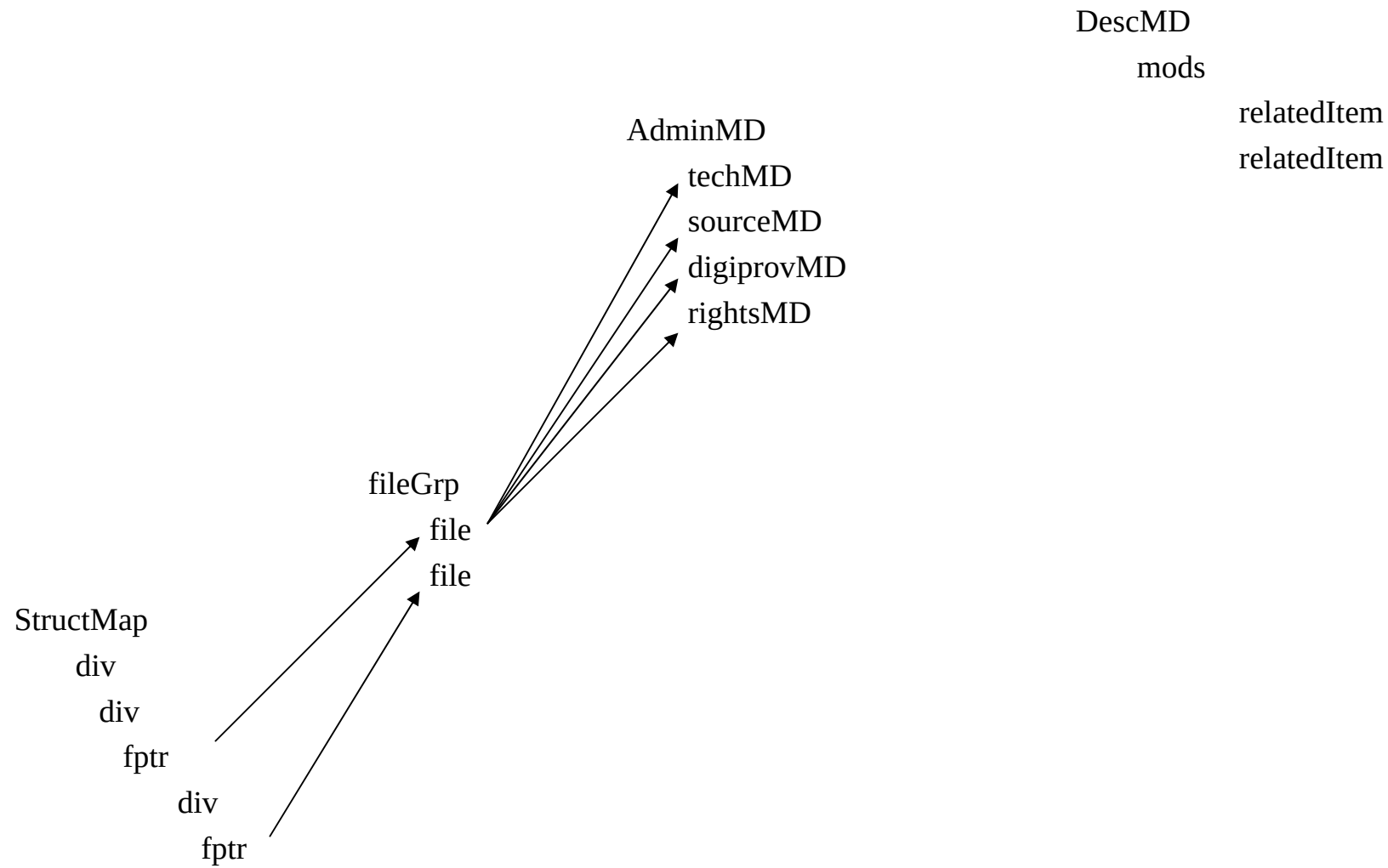
Linking in METS Documents
(XML ID/IDREF links)

DescMD
mods
relatedItem
relatedItem

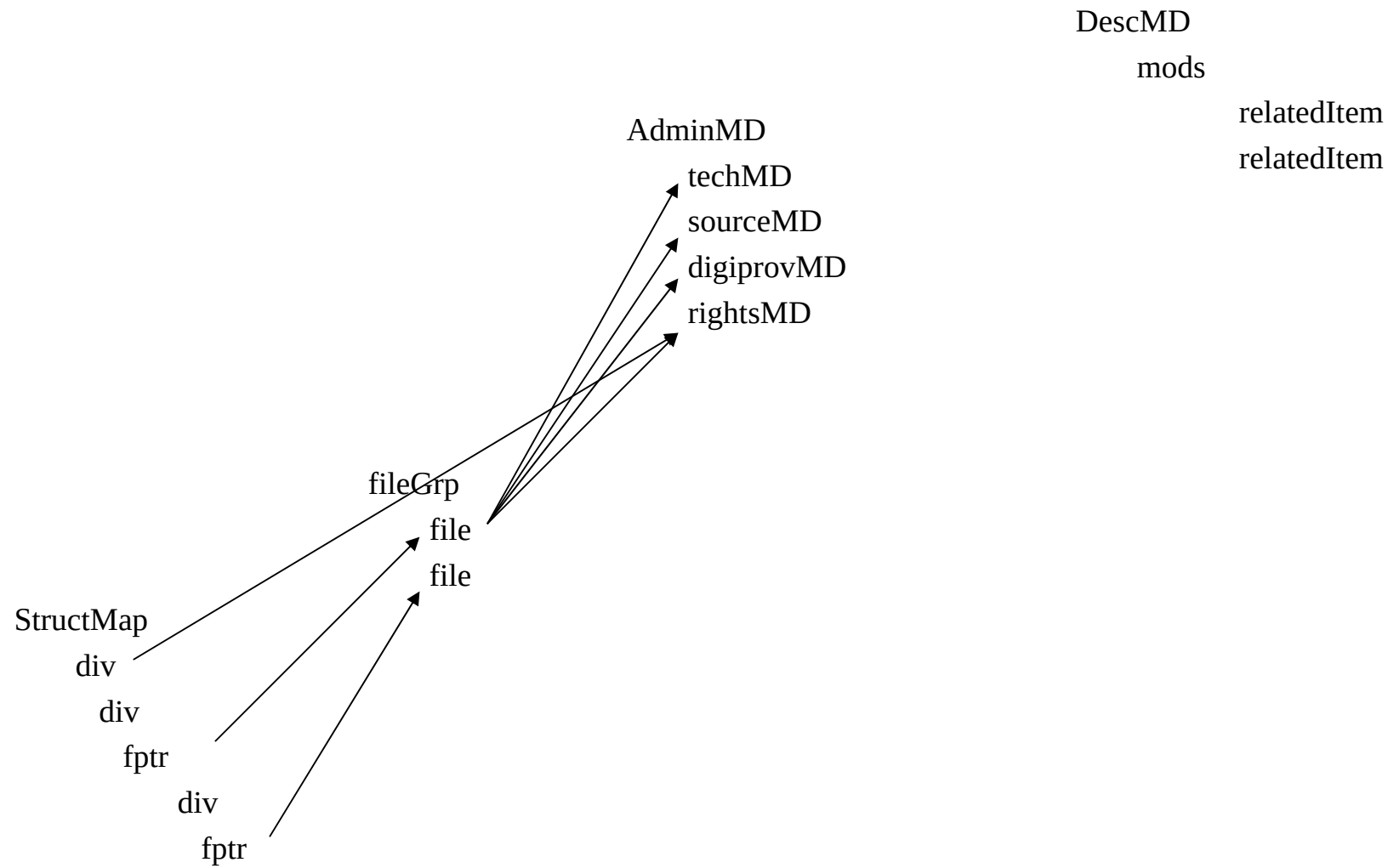
AdminMD
techMD
sourceMD
digiprovMD
rightsMD



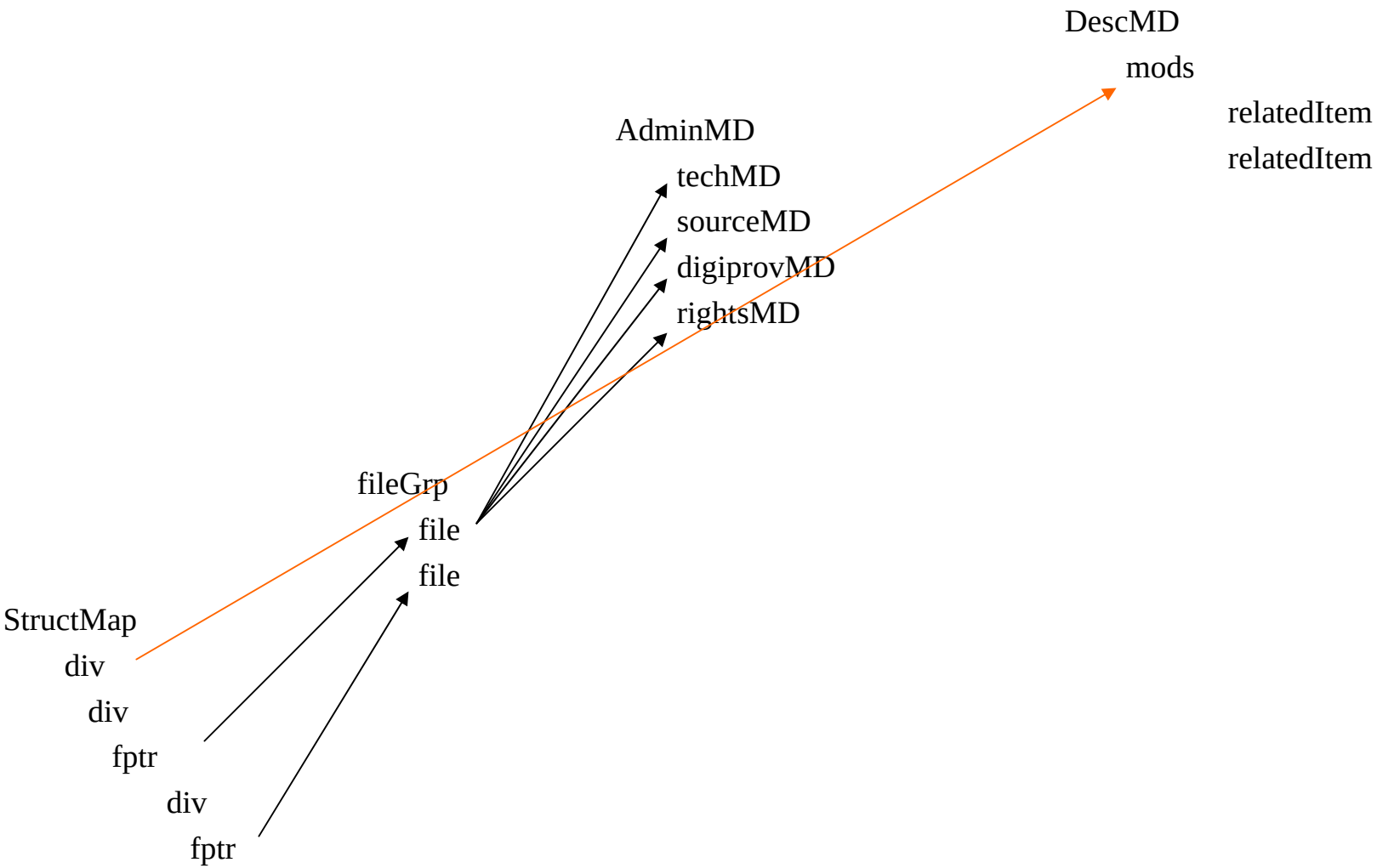
Linking in METS Documents
(XML ID/IDREF links)



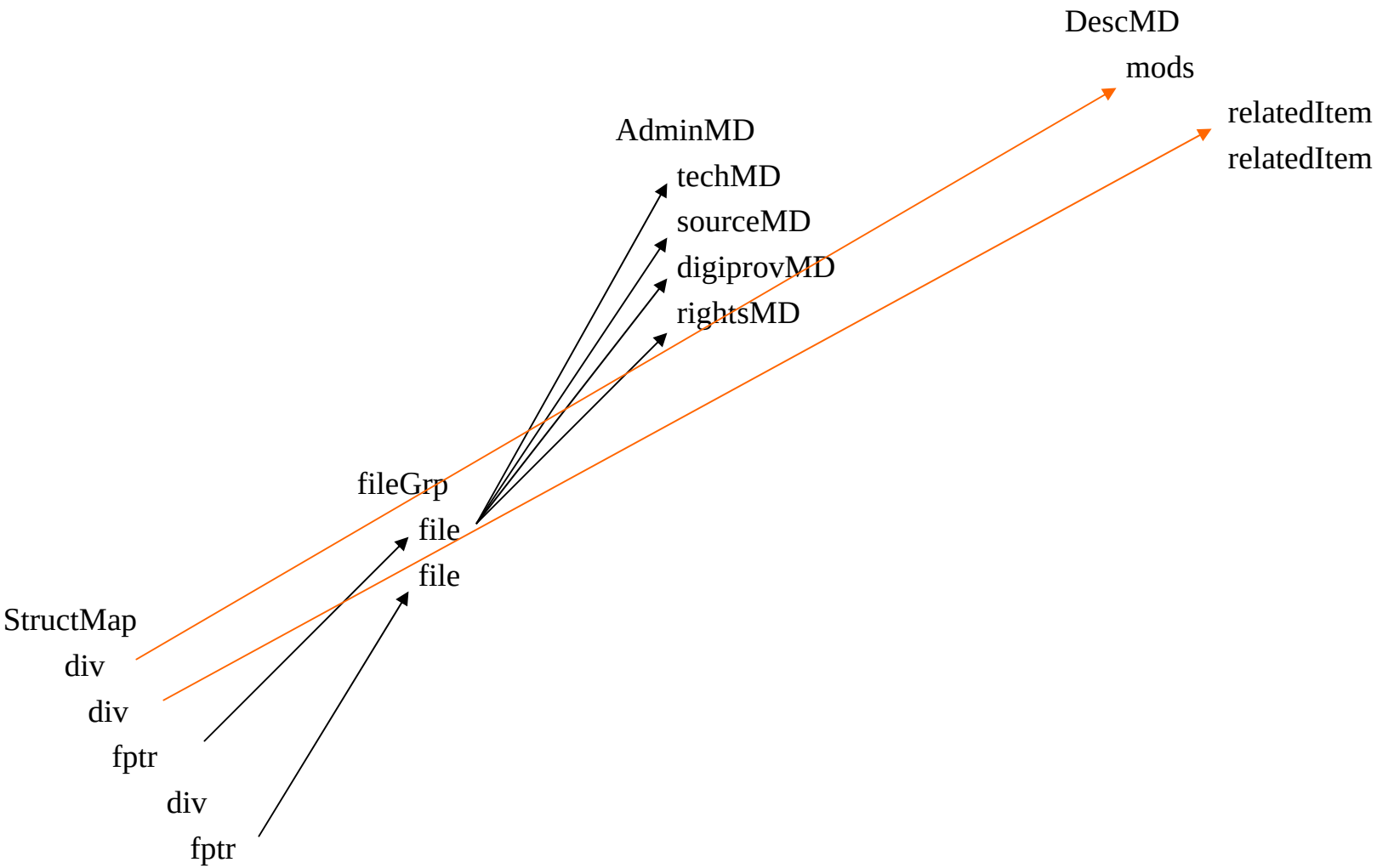
Linking in METS Documents
(XML ID/IDREF links)



Linking in METS Documents
(XML ID/IDREF links)



Linking in METS Documents
(XML ID/IDREF links)



What is a METS Application Profile?

"METS Profiles are intended to describe a class of METS documents in sufficient detail to provide both document authors and programmers the guidance they require to create and process METS documents conforming with a particular profile."

A profile is expressed as an XML document. There is a schema for this purpose.

Note: A METS Profile is a human-readable prose document and is not intended to be "machine actionable".

What is a “class of documents”?

- An open question
- At LC we are making a one-to-one relationship between a document class, or “object type” and a typical library item, e.g. a book, a photograph, a compact disc, etc., and…
- “Collection objects”

How does one get started with Profiles?

http://www.loc.gov/mets

- Download the documentation
- Download the XML Schema for Profiles
- Six registered profiles to date (UC Berkeley, Oxford University, Greenstone)

What are the 13 components of a Profile?

1. Unique URI
2. Short Title
3. Abstract
4. Date and time of creation
5. Contact Information
6. Related profiles

What are the 13 components? (cont'd)

7. Extension schemas
8. Rules of description
9. Controlled vocabularies
10. Structural requirements
11. Technical requirements
12. Tools and applications
13. Sample document

Summary: Application Profiles

- Prose document to provide guidance for the creation on METS documents
- A step towards machine processing of METS documents (software tool building)
- A step towards interoperability of digital libraries
- METS “frontier”

A METS Profile for audio compact discs

- Draft version Profile document:
  http://www.loc.gov/standards/mets/profiles/LC_profiles/compactDisc/
- Example document:
  http://lcweb2.loc.gov/cocoon/ihas/loc.natlib.ihas.200003790/default.html
- Comments welcome:
  mcundiff@loc.gov

Why a METS Profile for Audio Compact Discs?

- Needed CD Profile for “I Hear America Singing”
  - Digital Library project (NDMSO Tech Team)
  - XML/Standards/Open Source
  - Uses METS and MODS for all objects
  - All METS documents are based on a Profile
  - http://www.loc.gov/standards/mets/profiles/LC_profiles/compactDisc/

Attributes of Audio Compact Discs

Complex object

- Multiple data/media types (audio, images, text)
- Complex logical structure (anthology of multiple works)
- Complex physical structure
- Many variations (all CD's are not alike!)

CD Profile: Design Goals

- Express the logical structure of the object using MODS
- Express the physical structure of the object using METS
- Provide a correlation between the two hierarchical structures
- Allow for a reasonable amount of variation (i.e. some options in the form of occurrence indicators: mandatory, optional, repeatable, etc.)
- But not infinite variation (leading to too much complexity that defeats the purpose).
- Balanced, “90% solution”
- Develop profile methodology that can be used for other profiles

CD Profile: What it does not address

- The profile is completely agnostic about administrative metadata. It can be present or not.
- Profile is not specific about how to use the File Section. The profile only states that the fptr elements in the structMap must point to the appropriate file elements in the fileSec.

Logical Structure

- If there is more than one work recorded on the compact disc, the MODS element must contain one relatedItem TYPE=constituent child element for each work. All relatedItem TYPE=constituent elements must be given an ID attribute and must have a titleInfo child element.

- If a work has logical sub-parts, those sub-parts may be represented with relatedItem TYPE=constituent child elements. Examples of logical sub-parts of a work are the movements of a musical work, the acts and scenes of a dramatic work, etc. The logical structure of each work will thus be represented as a hierarchy of relatedItems.

CompactDiscObject (Logical structure)

- MODS
  - work
    - part
      - part
      - part
    - part
    - part
  - work
  - work

Logical (MODS)

```
<mods:mods>
 <mods:titleInfo>
  <mods:title>Album Title</mods:title>
 </mods:titleInfo>
 <mods:relatedItem type="constituent" ID="DMD_disc01_tr001">
  <mods:titleInfo>
   <mods:title>Work 1</mods:title>
  </mods:titleInfo>
 </mods:relatedItem>
 <mods:relatedItem type="constituent" ID="DMD_disc01_tr002_003">
  <mods:titleInfo>
   <mods:title>Work 2</mods:title>
  </mods:titleInfo>
  <mods:relatedItem type="constituent" ID="DMD_disc01_tr002">
   <mods:titleInfo>
    <mods:partName>Part Name 1</mods:partName>
   </mods:titleInfo>
  </mods:relatedItem>
  <mods:relatedItem type="constituent" ID="DMD_disc01_tr003">
   <mods:titleInfo>
    <mods:partName>Part Name 2</mods:partName>
   </mods:titleInfo>
  </mods:relatedItem>
 </mods:relatedItem>
</mods:mods>
```

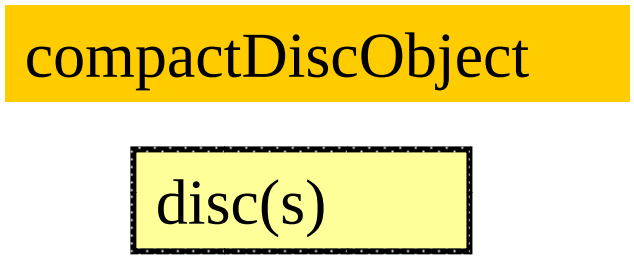
Logical (MODS)

```
<mods:mods>
 <mods:titleInfo>
  <mods:title>Album Title</mods:title>
 </mods:titleInfo>
 <mods:relatedItem type="constituent" ID="DMD_disc01_tr001">
  <mods:titleInfo>
   <mods:title>Work 1</mods:title>
  </mods:titleInfo>
 </mods:relatedItem>
 <mods:relatedItem type="constituent" ID="DMD_disc01_tr002_003">
  <mods:titleInfo>
   <mods:title>Work 2</mods:title>
  </mods:titleInfo>
  <mods:relatedItem type="constituent" ID="DMD_disc01_tr002">
   <mods:titleInfo>
    <mods:partName>Part Name 1</mods:partName>
   </mods:titleInfo>
  </mods:relatedItem>
  <mods:relatedItem type="constituent" ID="DMD_disc01_tr003">
   <mods:titleInfo>
    <mods:partName>Part Name 2</mods:partName>
   </mods:titleInfo>
  </mods:relatedItem>
 </mods:relatedItem>
</mods:mods>
```

Notes

Note that the profile makes no further descriptive cataloging requirements either in terms of content (i.e. particular descriptive data) or content designation (particular MODS elements and attributes). That being said, it should be obvious (given that the mods element and the relatedItem element both have the same content model) that there is the possibilty of providing more extensive description for each relatedItem. If, and to what extent, this possibility is expoited, is considered to be outside the scope of the profile. Rather, it is an issue of local cataloging practice.

It should also be noted that because the mods element and the relatedItem type="constituent" child elements represent the logical hierarchy of the compact disc (i.e. as an XML tree structure), then it becomes possible to consider that information can be inherited from one level of the hierarchy to another. For instance, in the example, it is clear that the partName with value “Part Name 1" is a part of the work with title “Work 2" by virtue of the child-to-parent relationship in the XML tree. Again, if, and to what extent, this kind of potential for inheritance is exploited is beyond the scope of the profile and is a matter of cataloging practice.



(Physical structure:
terms and hierarchy)

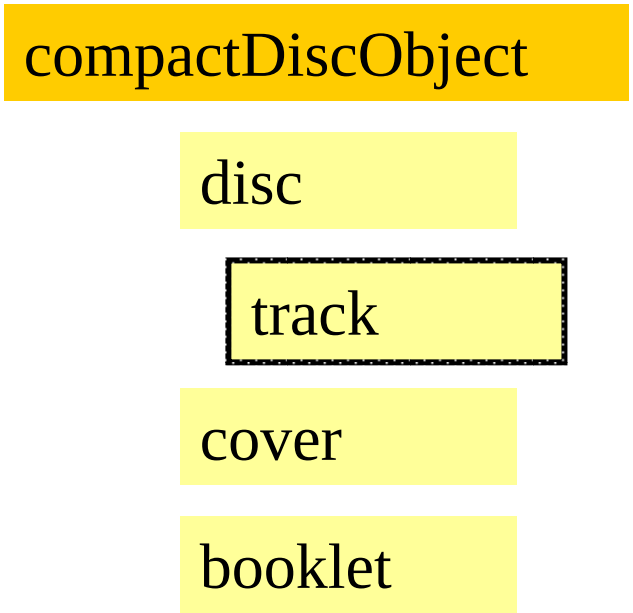
compactDiscObject (Physical structure)

disc(s)

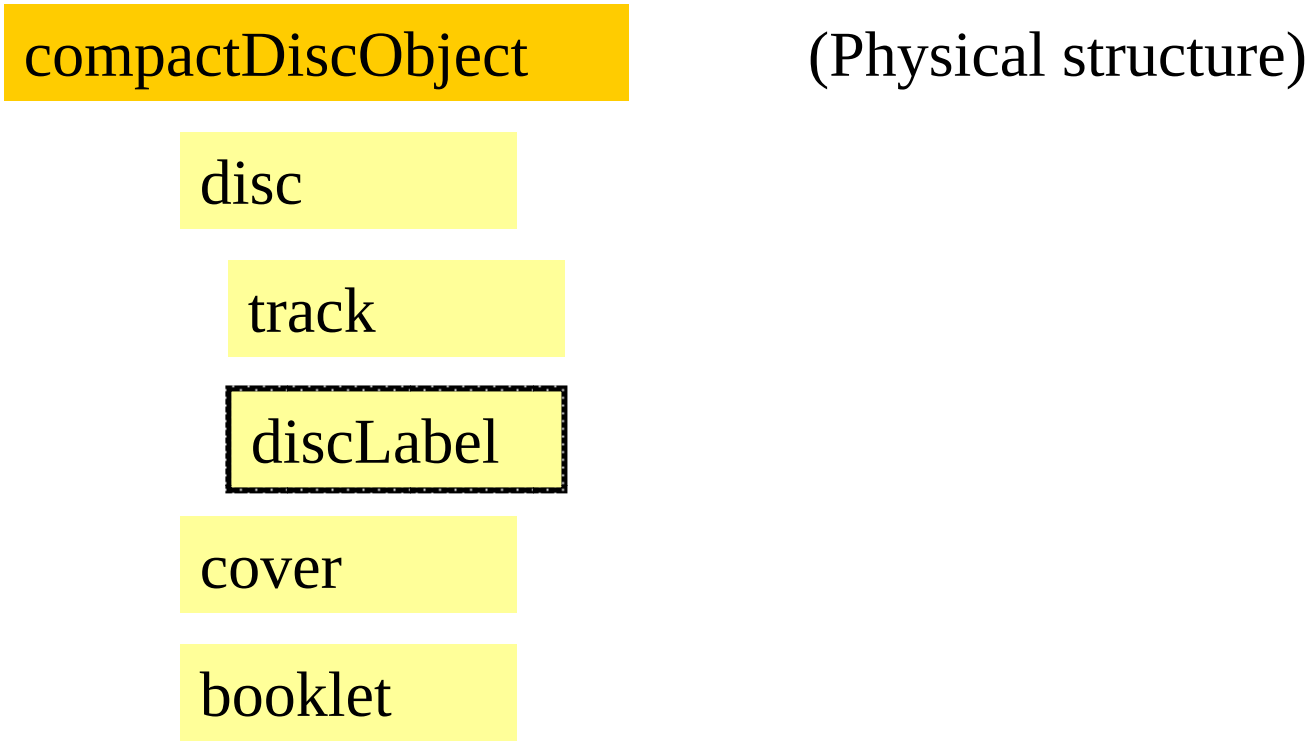
cover (container)

compactDiscObject (Physical structure)

- disc(s)
- cover
- booklet



(Physical structure)


compactDiscObject
(Physical structure)
disc
track
discLabel
cover
booklet

compactDiscObject (Physical structure)

- disc
  - track
    - trackSegment
  - track
  - discLabel
- cover
- booklet

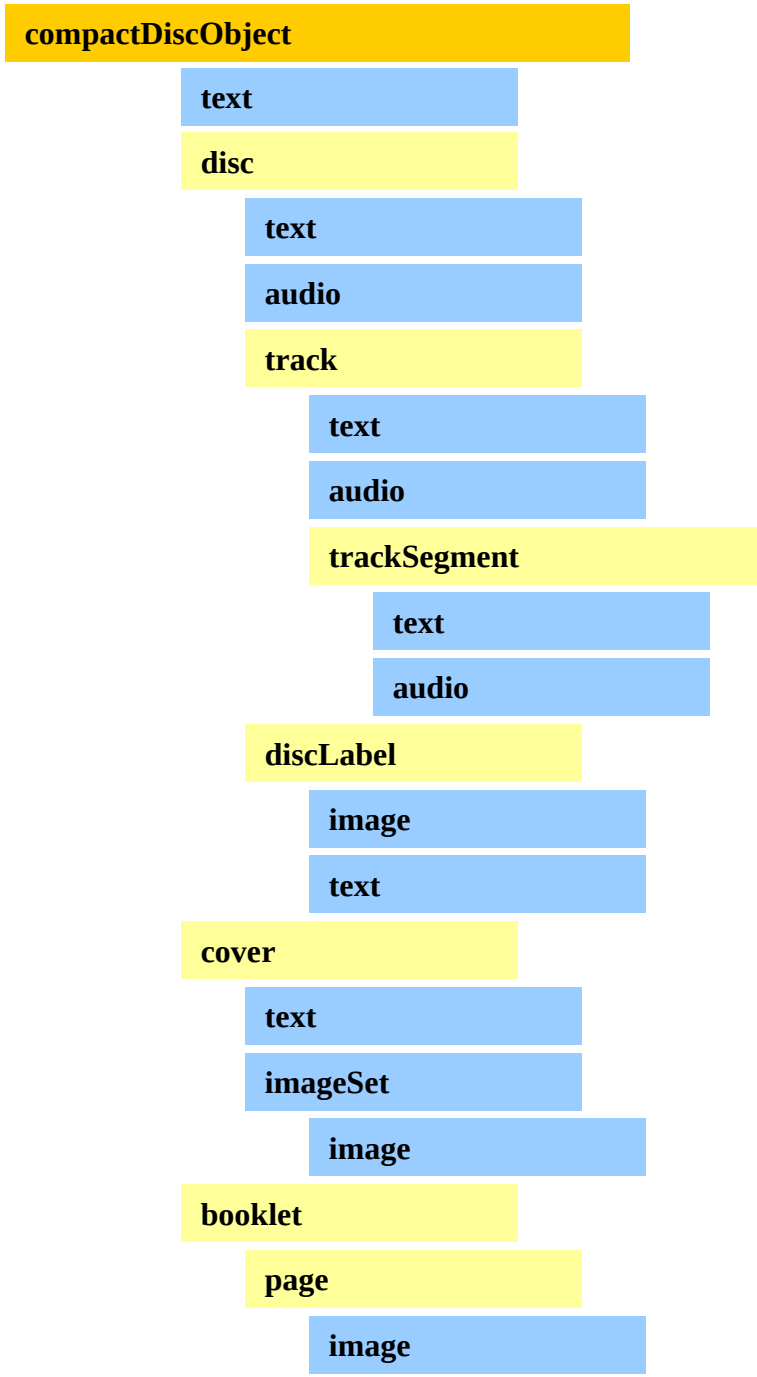
compactDiscObject (Physical structure)

- disc
  - track
    - trackSegment
  - track
  - discLabel
- cover
- booklet
  - page

- compactDiscObject
  - disc
    - track
      - trackSegment
    - track
  - discLabel
  - cover
  - booklet
    - page

(Physical structure)

“physical components”



“physical components”

“content components”

CompactDiscObject

- **disc**
  - **track**
    - **audio**
  - **track**
    - **audio**
  - **track**
    - **audio**
- **booklet**
  - **page**
    - **image**
  - **page**
    - **image**

(Physical structure)

Typical, simple example

```
<mets:div TYPE="cd:compactDiscObject>
    <mets:div TYPE="cd:text>
    <mets:div TYPE="cd:disc>
        <mets:div TYPE="cd:text>
        <mets:div TYPE="cd:audio>
        <mets:div TYPE="cd:track>
            <mets:div TYPE="cd:text>
            <mets:div TYPE="cd:audio>
            <mets:div TYPE="cd:trackSegment>
                <mets:div TYPE="cd:text>
                <mets:div TYPE="cd:audio>
        <mets:div TYPE="cd:discLabel>
                <mets:div TYPE="cd:image>
                <mets:div TYPE="cd:text>
    <mets:div TYPE="cd:cover>
        <mets:div TYPE="cd:text>
        <mets:div TYPE="cd:imageSet>
            <mets:div TYPE="cd:image>
    <mets:div TYPE="cd:booklet>
        <mets:div TYPE="cd:page>
            <mets:div TYPE="cd:image>
```

These terms are used as div TYPE attribute values in the form of qualified names (“prefix:localName”)

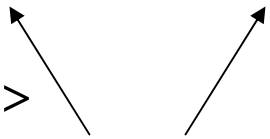
Becomes a formal, standardized, controlled “div typing” vocabulary for use with the profile

div TYPE prefix “cd” is associated with namespace URI

```
<mets:mets PROFILE="lc:compactDisc"
           xmlns:mets="http://www.loc.gov/METS/"
           xmlns:mods="http://www.loc.gov/mods/v3"
           xmlns:cd="http://www.loc.gov/mets/profiles/compactDisc">

 <mets:structMap>
  <mets:div TYPE="cd:compactDiscObject">
   <mets:div TYPE="cd:disc">
    <mets:fptr/>
   </mets:div>
  </mets:div>
 </mets:structMap>

</mets:mets>
```



Logical (MODS)

```
<mods:mods>
 <mods:titleInfo>
  <mods:title>Album Title</mods:title>
 </mods:titleInfo>
 <mods:relatedItem type="constituent" ID="DMD_disc01_tr001">
  <mods:titleInfo>
   <mods:title>Work 1</mods:title>
  </mods:titleInfo>
 </mods:relatedItem>
 <mods:relatedItem type="constituent" ID="DMD_disc01_tr002_003">
  <mods:titleInfo>
   <mods:title>Work 2</mods:title>
  </mods:titleInfo>
  <mods:relatedItem type="constituent" ID="DMD_disc01_tr002">
   <mods:titleInfo>
    <mods:partName>Part Name 1</mods:partName>
   </mods:titleInfo>
  </mods:relatedItem>
  <mods:relatedItem type="constituent" ID="DMD_disc01_tr003">
   <mods:titleInfo>
    <mods:partName>Part Name 2</mods:partName>
   </mods:titleInfo>
  </mods:relatedItem>
 </mods:relatedItem>
</mods:mods>
```

Logical (MODS)

```
<mods:mods>
 <mods:titleInfo>
  <mods:title>Album Title</mods:title>
 </mods:titleInfo>
 <mods:relatedItem type="constituent" ID="DMD_disc01_tr001">
  <mods:titleInfo>
   <mods:title>Work 1</mods:title>
  </mods:titleInfo>
 </mods:relatedItem>
 <mods:relatedItem type="constituent" ID="DMD_disc01_tr002_003">
  <mods:titleInfo>
   <mods:title>Work 2</mods:title>
  </mods:titleInfo>
  <mods:relatedItem type="constituent" ID="DMD_disc01_tr002">
   <mods:titleInfo>
    <mods:partName>Part Name 1</mods:partName>
   </mods:titleInfo>
  </mods:relatedItem>
  <mods:relatedItem type="constituent" ID="DMD_disc01_tr003">
   <mods:titleInfo>
    <mods:partName>Part Name 2</mods:partName>
   </mods:titleInfo>
  </mods:relatedItem>
 </mods:relatedItem>
</mods:mods>
```

Physical (structMap)

```
<mets:structMap>
 <mets:div TYPE="cd:compactDiscObject" DMDID="MODS1">
  <mets:div TYPE="cd:disc">
   <mets:div DMDID="DMD_disc01_tr001"  TYPE="cd:track“ >
    <mets:div TYPE="cd:audio">
     <mets:fptr FILEID="FN10081"/>
    </mets:div>
   </mets:div>
   <mets:div DMDID="DMD_disc01_tr002"  TYPE="cd:track“ >
    <mets:div TYPE="cd:audio">
     <mets:fptr FILEID="FN10090"/>
    </mets:div>
   </mets:div>
   <mets:div DMDID="DMD_disc01_tr003"  TYPE="cd:track“ >
    <mets:div TYPE="cd:audio">
     <mets:fptr FILEID="FN1009F"/>
    </mets:div>
   </mets:div>
  </mets:div>
 </mets:div>
</mets:structMap>
```

Logical (MODS)

```
<mods:mods>
 <mods:titleInfo>
  <mods:title>Album Title</mods:title>
 </mods:titleInfo>
 <mods:relatedItem type="constituent" ID="DMD_disc01_tr001">
  <mods:titleInfo>
   <mods:title>Work 1</mods:title>
  </mods:titleInfo>
 </mods:relatedItem>
 <mods:relatedItem type="constituent" ID="DMD_disc01_tr002_003">
  <mods:titleInfo>
   <mods:title>Work 2</mods:title>
  </mods:titleInfo>
  <mods:relatedItem type="constituent" ID="DMD_disc01_tr002">
   <mods:titleInfo>
    <mods:partName>Part Name 1</mods:partName>
   </mods:titleInfo>
  </mods:relatedItem>
  <mods:relatedItem type="constituent" ID="DMD_disc01_tr003">
   <mods:titleInfo>
    <mods:partName>Part Name 2</mods:partName>
   </mods:titleInfo>
  </mods:relatedItem>
 </mods:relatedItem>
</mods:mods>
```

4 relatedItems

Physical

```
<mets:structMap>
 <mets:div TYPE="cd:compactDiscObject" DMDID="MODS1">
  <mets:div TYPE="cd:disc">
   <mets:div DMDID="DMD_disc01_tr001"  TYPE="cd:track“ >
    <mets:div TYPE="cd:audio">
     <mets:fptr FILEID="FN10081"/>
    </mets:div>
   </mets:div>
   <mets:div DMDID="DMD_disc01_tr002"  TYPE="cd:track“ >
    <mets:div TYPE="cd:audio">
     <mets:fptr FILEID="FN10090"/>
    </mets:div>
   </mets:div>
   <mets:div DMDID="DMD_disc01_tr003"  TYPE="cd:track“ >
    <mets:div TYPE="cd:audio">
     <mets:fptr FILEID="FN1009F"/>
    </mets:div>
   </mets:div>
  </mets:div>
 </mets:div>
</mets:structMap>
```

3 divs TYPE=cd:track

Logical (MODS)

```
<mods:mods>
 <mods:titleInfo>
  <mods:title>Album Title</mods:title>
 </mods:titleInfo>
 <mods:relatedItem type="constituent" ID="DMD_disc01_tr001">
  <mods:titleInfo>
   <mods:title>Work 1</mods:title>
  </mods:titleInfo>
 </mods:relatedItem>
 <mods:relatedItem type="constituent" ID="DMD_disc01_tr002_003">
  <mods:titleInfo>
   <mods:title>Work 2</mods:title>
  </mods:titleInfo>
  <mods:relatedItem type="constituent" ID="DMD_disc01_tr002">
   <mods:titleInfo>
    <mods:partName>Part Name 1</mods:partName>
   </mods:titleInfo>
  </mods:relatedItem>
  <mods:relatedItem type="constituent" ID="DMD_disc01_tr003">
   <mods:titleInfo>
    <mods:partName>Part Name 2</mods:partName>
   </mods:titleInfo>
  </mods:relatedItem>
 </mods:relatedItem>
</mods:mods>
```

Low-level relatedItems (i.e. relatedItems with no relatedItem child elements)

Physical

```
<mets:structMap>
 <mets:div TYPE="cd:compactDiscObject" DMDID="MODS1">
  <mets:div TYPE="cd:disc">
   <mets:div DMDID="DMD_disc01_tr001" TYPE="cd:track“ >
    <mets:div TYPE="cd:audio">
     <mets:fptr FILEID="FN10081"/>
    </mets:div>
   </mets:div>
   <mets:div DMDID="DMD_disc01_tr002" TYPE="cd:track“ >
    <mets:div TYPE="cd:audio">
     <mets:fptr FILEID="FN10090"/>
    </mets:div>
   </mets:div>
   <mets:div DMDID="DMD_disc01_tr003" TYPE="cd:track“ >
    <mets:div TYPE="cd:audio">
     <mets:fptr FILEID="FN1009F"/>
    </mets:div>
   </mets:div>
  </mets:div>
 </mets:div>
</mets:structMap>
```

divs that are parents of div type=“cd:audio”

Logical (MODS)

Physical

```
<mods:mods>
 <mods:titleInfo>
  <mods:title>Album Title</mods:title>
 </mods:titleInfo>
 <mods:relatedItem type="constituent" ID="DMD_disc01_tr001">
  <mods:titleInfo>
   <mods:title>Work 1</mods:title>
  </mods:titleInfo>
 </mods:relatedItem>
 <mods:relatedItem type="constituent" ID="DMD_disc01_tr002_003">
  <mods:titleInfo>
   <mods:title>Work 2</mods:title>
  </mods:titleInfo>
  <mods:relatedItem type="constituent" ID="DMD_disc01_tr002">
   <mods:titleInfo>
    <mods:partName>Part Name 1</mods:partName>
   </mods:titleInfo>
  </mods:relatedItem>
  <mods:relatedItem type="constituent" ID="DMD_disc01_tr003">
   <mods:titleInfo>
    <mods:partName>Part Name 2</mods:partName>
   </mods:titleInfo>
  </mods:relatedItem>
 </mods:relatedItem>
</mods:mods>
```

```
<mets:structMap>
 <mets:div TYPE="cd:compactDiscObject" DMDID="MODS1">
  <mets:div TYPE="cd:disc">
   <mets:div DMDID="DMD_disc01_tr001" TYPE="cd:track“ >
    <mets:div TYPE="cd:audio">
     <mets:fptr FILEID="FN10081"/>
    </mets:div>
   </mets:div>
   <mets:div DMDID="DMD_disc01_tr002" TYPE="cd:track“ >
    <mets:div TYPE="cd:audio">
     <mets:fptr FILEID="FN10090"/>
    </mets:div>
   </mets:div>
   <mets:div DMDID="DMD_disc01_tr003" TYPE="cd:track“ >
    <mets:div TYPE="cd:audio">
     <mets:fptr FILEID="FN1009F"/>
    </mets:div>
   </mets:div>
  </mets:div>
 </mets:div>
</mets:structMap>
```

Low-level relatedItems (i.e. relatedItems with no relatedItem child elements)

Sequence of 3 nodes

Divs that are parents of div type=“cd:audio”

Corresponding sequence of 3 nodes linked to logical sequence by ID/IDREF

Demo IHAS CD

http://lcweb2.loc.gov/cocoon/ihas/loc.natlib.ihas.200003790/default.html

Parting thoughts…

Profile places emphasis on “object modeling” and structural characteristics of the digital object

Whether one uses METS or another “wrapper” or “aggregator” (such as MPEG-21) one will still have to go through the object modeling process for complex objects

This approach requires a logical analysis of the program content of the content disc and that that analysis be reflected in the cataloging of the item (i.e. there is an emphasis on “embedded analytic records”). This is new and it is labor intensive but is has big payoffs in the subsequent usability of the object.