



METS Tutorial

iPRES 2021 - October 19th 2021

Presentation available for download at:

https://github.com/mets/METSboard/tree/master/iPres2021Tutorial

19 October 2021



Welcome!

Presenters:
Karin Bredenberg
Juha Lehtonen
Tobias Steinke

19 October 2021 2



Content

- Introduction to METS
 - Short background (brief history and rationale of METS)
 - Benefits of implementing METS
 - METS in detail
 - Core elements
 - METS profile
- Case study, Finland
- The case of using PREMIS in METS
- Support and the METS community
- Conformance
- Simple example to build familiarity
- Wrap up



INTRODUCTION TO METS

19 October 2021 4



BACKGROUND

19 October 2021 5



What is METS?

- METS = Metadata Encoding and Transmission Standard
- Maintained by the METS Editorial Board
- Schema is hosted at the Library of Congress
- Current version 1.12.1



What is METS? (cont'd)

- An XML schema-based specification for encoding "hub" documents for materials whose content is digital.
 - Hub doc draws together dispersed but related files
 - METS uses XML to provide a vocabulary and syntax for identifying the digital pieces that together comprise a digital entity, for specifying the location of these pieces, and for expressing the structural relationships between them.
 - Content files
 - Descriptive metadata
 - Administrative metadata



METS Editorial Board

 The METS Editorial Board is an international group of volunteers committed to maintaining editorial control over METS, its XML Schema, the METS Profile XML Schema, and official METS documentation. The Board promotes the use of the METS specification, maintains a registry of METS Profiles, and endorses best practices in the use of METS as they emerge. Members represent important communities of interest for METS, including members of the Digital Library Federation, its initial sponsor, and the Library of Congress, its maintenance agency.



METS History

- Originates in Making of America II initiative
 - Making of America II (MOA2) was a Digital Library Federation sponsored initiative that started in 1997. Participants included UCB (lead), Stanford, Penn State, Cornell, and NYPL.
 - GOAL: to create a digital object standard for encoding structural, descriptive and administrative metadata along with primary content
 - RESULT: MOA2.DTD (an XML DTD)

METS History (cont'd)

- UCB Library and CDL adopt MOA2
- Other institutions (LC, Harvard) consider
- Additional needs emerge
 - Support for time-based content
 - More flexibility in Descriptive and Administrative metadata
- MOA2 revised
 - Starting in February 2001 concerned parties meet to review and revise MOA2
 - Outcome: mets.xsd



BENEFITS

19 October 2021 11



Main Provisions of METS schema

- 1. Identifying the files or parts of files that comprise the content of a digital entity, and expressing the structure or structures of this content
- 2. Linking Descriptive metadata with digital content
- 3. Linking Administrative metadata with digital content
- 4. Wrapping digital content, and associated descriptive and administrative metadata as binary data.
- 5. Wrapping digital content, and associated descriptive and administrative metadata as XML data.



Uses of METS

- Transfer syntax
 - standard for transmitting/ exchanging digital objects.
 - SIP (Open Archival Information Systems Reference Model)
 - DSpace SIP Toolkit uses a mandatory METS document
 - Fedora supports METS as a ingest package
- Functional syntax:
 - basis for providing end users with the ability to view and navigate digital content and its associated metadata
 - DIP
- Archiving syntax
 - standard for archiving digital objects.
 - combine with PREMIS (PREservation Metadata: Implementation Strategies)
 - AIP

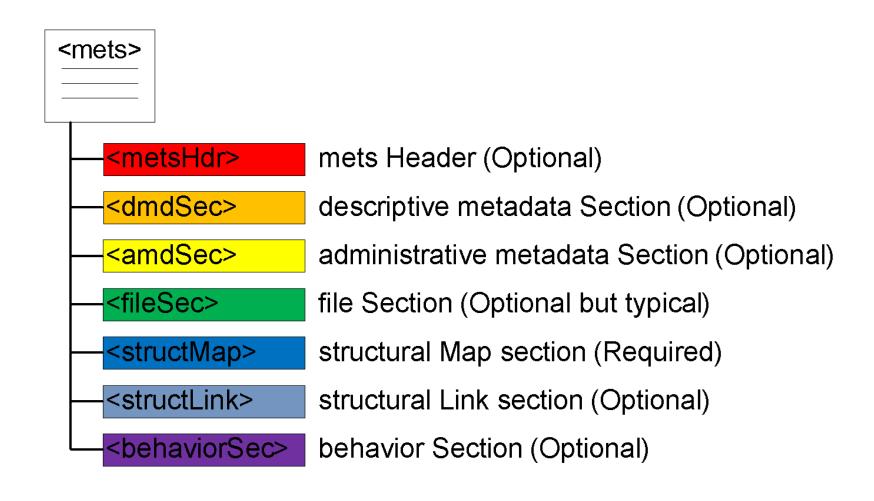


METS IN DETAIL CORE ELEMENTS

19 October 2021 14

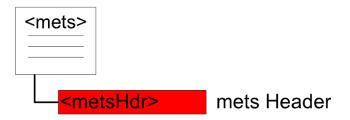


METS first level elements





METS Header



- Records administrative metadata about the METS document itself such as
 - Author/agent and role
 - Alternative identifiers for the METS document
 - Creation and update date and times
 - Status

<mets></mets>	
<dmdsec></dmdsec>	descriptive metadata Section

- Can record all of the units of descriptive metadata pertaining to the digital entity represented by METS document
 - Descriptive metadata could take any form including MARC record, Finding Aid, Dublin Core record
 - Descriptive Metadata may be
 - External to the METS document
 - Internal to the METS document
 - Both external and internal

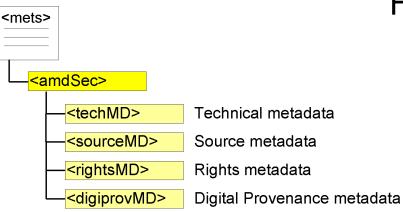
MEDS Administrative metadata



 Can record all of the units of administrative metadata pertinent to the METS object or its parts



Administrative metadata Flavors



- Administrative metadata elements come in 4 flavors
 - 1.Technical metadata
 - 2. Source Metadata
 - 3. Rights Metadata
 - 4. Digital Provenance Metadata
- You choose which to use
 - All
 - Just one
 - Any self-chosen number
- There are some recommendations of which flavor to use to which type of administrative metadata



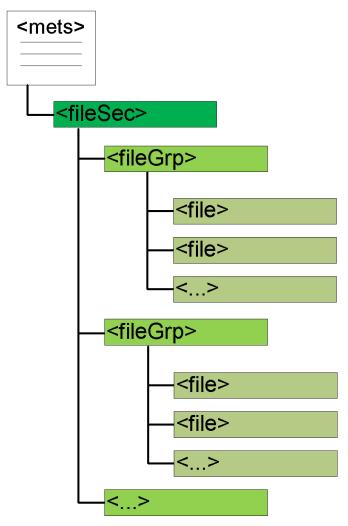
File Section



 Records all of the files that together comprise the content of the digital entity represented by the METS document



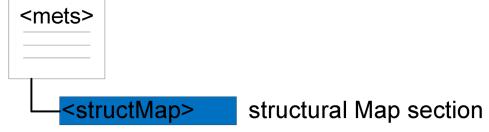
File Section Filegroups



Files are organized into File
Groups based on the grouping
you would like to do. One way
is to group by format (tiff, hires jpeg, med-res jpeg, gif,
etc)



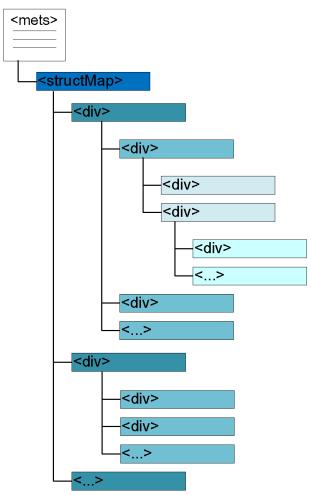
Structural Map Section



- Specifies the (hierarchical) structure of the digital entity represented by the METS document.
- Specifies how the content files (the files listed in the Files Section) fit into this structure.
- More than one structure may be specified. For example: a logical structure and a physical structure, a Webpage structure



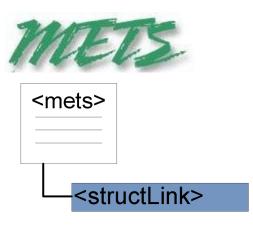
Expressing the Structure



The structural map analyzes a digital object into a hierarchy of Division (div) elements

Division (type="photoalbum")

Division (type="photo")



Structural Link Section

structural Link section

- Specification for hyperlinks between the different components of a METS structure that are delineated in a structural map.
- Used to note the existence of hypertext links between web pages, if you wished to record those links within METS.

MEIS

Behavior Section



- Can record all of the dissemination behaviors that pertain to a digital entity or its parts. A behavior unit may contain
 - A reference to an external interface definition that defines a set of related behaviors
 - A reference to an external executable that implements these behaviors
 - A reference to the Division or Divisions of the object structure to which the behaviors apply.

Internal or external



- Its possible to include descriptive and administrative metadata and files or point to them.
- Its possible to refer to an external file, or itself contain the file, or both.
 - External file. The element may point to an external file via a URI.
 - Internal file. The element may itself contain the file as binary data or XML data.



xsd:anyAttribute

- Add your own attributes to some elements
 - -mets
 - metsHdr
 - agent/note
 - dmdSec
 - amdSec
 - techMD
 - rightsMD
 - sourceMD
 - digiprovMD

- fileSec
 - fileGrp
 - file
- structMap
 - div/fptr
 - par
 - seq
 - area
- structLink
- behaviorSec

xsd:anyAttribute (cont'd)

Declare a custom namespace

 Ideally custom attributes are documented in a METS Profile



METS PROFILES

19 October 2021 29



Profiles

- An XML document
- Defines the rules of how METS is used
- One profile can extend another profile



Profiles

- Profiles schema version 2.1
- General information about who to contact regarding the profile and so on
- Element showing if its registered or not
- All the rules regarding the use of METS

METS

Profiles Registration

- Registration at the METS Editorial Board
- 46 registered profiles (October 2021)
- Review period at/by the METS-list
- Can be used by others
- Others can make an new profile extending an existing profile with their use



Validating

- The profile doesn't give validation rules
- The original schema has just the rules originally in METS
- For validating your own rules you must do one of these two options
 - 1. Edit the schema and add your restrictions
 - 2. Use Schematron to implement your restrictions

In addition you need to create the schema for the xsd:anyAttribute



Redundancy

- Using METS and other standards sometimes causes redundancy
- Same elements in several standards
- Same elements mandatory in several standards
- Own descisions regarding redundancy in cases when not mandatory in both standards.
- Regarding PREMIS and METS see the document mentioned earlier



CASE STUDY - FINNISH NATIONAL SPECIFICATIONS

19 October 2021

Digital Preservation Services in Finland

- CSC produces centralized digital preservation services for the Ministry of Education and Culture
 - CSC IT Center for Science:
 A non-profit state organization with special tasks

 Digital Preservation Service for Cultural Heritage (in production since 2015)

 Digital Preservation Service for Research Data (in production since 2019)

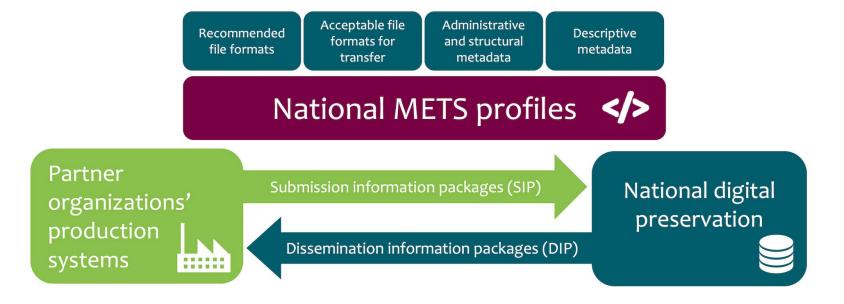
Large scale requires fully automated processes

Digital assets are stored on three different media types (disk and two tapes) Currently, we preserve 1.5 PB in 1.8 million archival information packages

reasing.

Digital Preservation Services in Finland

- Preparing and packaging digital assets is the most visible part of the digital preservation for the partner organizations
- Common specifications guides the packaging process
- Specifications are maintained in collaboration with partner organizations
 - Archives, Libraries, Museums, Universities, Research institutes...



Specifications and METS

- National specifications must be followed when creating SIPs or DIPs
- A SIP or DIP contains a METS document, a digital signature file, and digital objects
- The specification of METS document describe:
 - Mandatory and conditional metadata
 - Also in other formats than METS (e.g. PREMIS)
 - Conditional metadata is mandatory in certain conditions
 - Forbidden METS elements and attributes
 - Some nationally defined METS extensions
- Additionally, a large set of attributes are either "optional" or "not recommended"



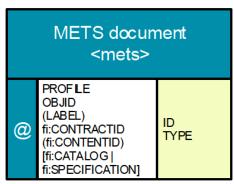
Main Sections

- Mandatory and optional sections:
 - metsHdr: METS Header [1..1]
 - dmdSec: Descriptive metadata [1..n]
 - amdSec: Administrative metadata [1..1]
 - techMD: Technical metadata (mandatory for each digital object)
 - sourceMD: Source metadata [0..n]
 - rightsMD: Rights metadata for DP service [0..n]
 - digiprovMD: Provenance metadata [1..n]
 - fileSec: File section [1..1]
 - structMap: Structural map [1..n]
- Forbidden sections:
 - behaviorSec and structLink



METS Root

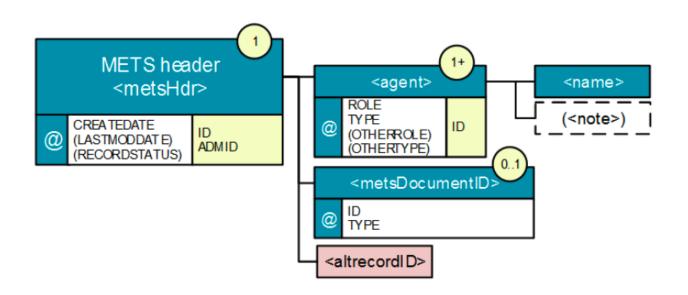
- The METS Root element <mets>:
 - OBJID: ID of the SIP/DIP
 - PROFILE:
 - Different profiles for different services
 - However, the specifications are currently same for all
 - National attribute extensions:
 - fi:CONTRACTID: ID of the contract between the producer and the service
 - fi:SPECIFICATION or fi:CATALOG: Version number of the specification or the corresponding XML Schema Catalog
 - fi:CONTENTID: Possibility to split the content to multiple SIPs (opt.)





METS Header

- METS header <metsHdr>:
 - CREATEDATE: Creation timestamp of the SIP/DIP
 - At least one METS agent (creator of the SIP/DIP)



METS

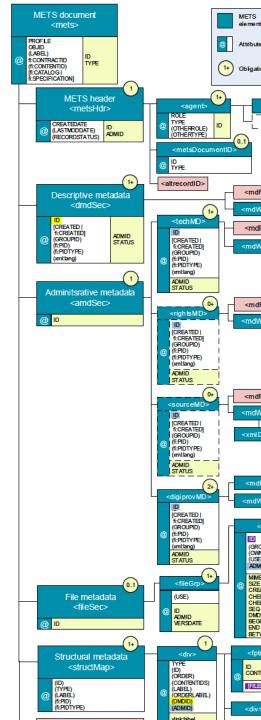
Example: METS Root and Header

```
<mets:mets
  PROFILE="http://digitalpreservation.fi/mets-profiles/cultural-heritage"
  OBJID="8ddf5aea-8bfc-4c29-9708-19c408c4fa63"
  fi:CATALOG="1.7.3"
  fi:CONTRACTID="urn:uuid:d7b44473-ff29-4e6a-86c9-9813292d730b">
    <mets:metsHdr CREATEDATE="2021-10-05T13:03:50">
       <mets:agent ROLE="CREATOR" TYPE="ORGANIZATION">
         <mets:name>CSC - IT Center for Science</mets:name>
       </mets:agent>
    </mets:metsHdr>
</mets:mets>
```



METS Sections

- Wrapping metadata inside METS file as XML format is mandatory.
 - Wrapping binary data inside METS is forbidden.
- In each metadata section:
 - ID: METS internal ID
 - fi:PIDTYPE and fi:PID for external ID (optional)
 - CREATED or fi: CREATED
 - Metadata creation time
 - CREATED requires accurate timestamp
 - Optionally fi:CREATED for imprecise time
 - Metadata wrapper <mdWrap>
 - MDTYPE: Metadata format
 - OTHERMDTYPE: Metadata format, used if not in dictionary of MDTYPE
 - MDTYPEVERSION: Version of the metadata format





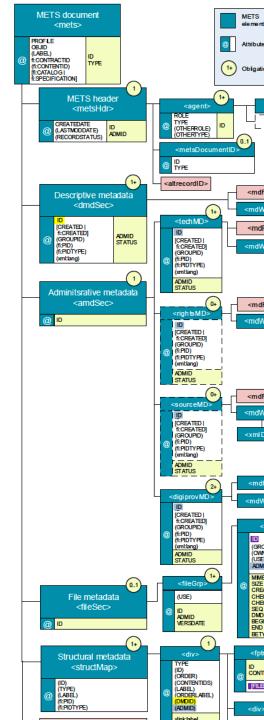
METS Sections: Example

```
<mets:dmdSec
  ID="dmdsec-001"
  CREATED="2021-10-05T13:03:49">
     <mets:mdWrap MDTYPE="DC" MDTYPEVERSION="1.1">
       <mets:xmlData>
         <dc:title>Test SIP</dc:title>
         <dc:description>Example for iPRES</dc:description>
         <dc:publisher>CSC</dc:publisher>
         <dc:identifier>test-id</dc:identifier>
       </mets:xmlData>
     </mets:mdWrap>
</mets:dmdSec>
```



METS Sections

- At least one <dmdSec> must include descriptive metadata in one of the formats listed in the specifications.
 - The list contains ~10 different descriptive metadata formats
- At least one <techMD> per digital object is required (PREMIS Object metadata).
 - Video containers and the included streams must be described separately as separate PREMIS Objects, each in their own <techMD> sections.
- Separate <techMD>:s required for
 - images: MIX metadata
 - audio streams: AudioMD metadata
 - video streams: VideoMD metadata
 - CSV files: ADDML metadata

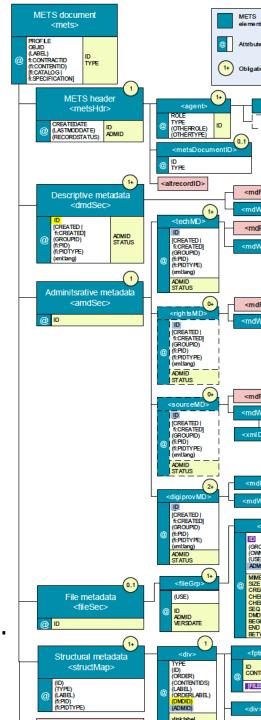




METS Sections

- Each PREMIS Event or Agent in its own <digiprovMD>
 - At least one PREMIS Event is mandatory
- It is allowed to use other metadata formats additionally, in their own sections with OTHERMDTYPE and MDTYPEVERSION attributes.

- Files (and streams) in the package are listed in a file section <fileSec>. Links between files and administrative metadata are in file section too (file@ADMID).
- At least one structural map <structMap> is mandatory. Links between an entity <div> and descriptive (or administrative) metadata are in structural map too (div@DMDID and div@ADMID).



PREMIS Object in METS

```
<mets:techMD
  ID= "tech-001"
  CREATED="2021-10-05T13:02:10">
    <mets:mdWrap
      MDTYPE="PREMIS:OBJECT"
      MDTYPEVERSION="2.3">
         <mets:xmlData>
            cobject> ... </premis:object>
         </mets:xmlData>
    </mets:mdWrap>
</mets:techMD>
```

PREMIS Event in METS

```
<mets:digiprovMD
  ID= "digiprov-001"
  CREATED="2021-10-05T13:02:10">
    <mets:mdWrap
      MDTYPE="PREMIS:EVENT"
      MDTYPEVERSION="2.3">
        <mets:xmlData>
            oremis:event> ... 
        </mets:xmlData>
    </mets:mdWrap>
</mets:digiprovMD>
```



AudioMD in METS

```
<mets:techMD
  ID= "tech-002"
  CREATED="2021-10-05T13:02:10">
    <mets:mdWrap
      MDTYPE="OTHER"
      OTHERMDTYPE="AudioMD"
      MDTYPEVERSION="2.0">
         <mets:xmlData>
            <audiomd:AUDIOMD> ... </audiomd:AUDIOMD>
         </mets:xmlData>
    </mets:mdWrap>
</mets:techMD>
```



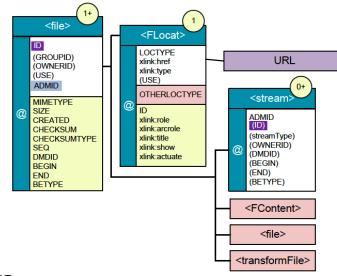
PREMIS Object in METS

```
▼<mets:techMD ID="tech-001" CREATED="2011-03-31T15:40:15">
 ▼<mets:mdWrap MDTYPE="PREMIS:OBJECT" MDTYPEVERSION="2.2">
  ▼<mets:xmlData>
   ▼mis:object xsi:type="premis:file">
    ▼objectIdentifier>
      objectIdentifierType>urn</premis:objectIdentifierType>
      objectIdentifierValue> URN:NBN:fi-fe201215081520 
     ▼▼objectCharacteristics>
      <premis:compositionLevel>0</premis:compositionLevel>
     ▼fixity>
       is:fixity>
     ▼ormat>
      ▼ormatDesignation>
        cpremis:formatName>image/png</premis:formatName>
        ormatVersion>1.2</premis:formatVersion>
       is:formatDesignation>
      ▼ormatRegistry>
        formatRegistryKey>fmt/13</premis:formatRegistryKey>
       </premis:formatRegistry>
      ▼▼on>
       <premis:dateCreatedByApplication> 2011-02-15T15:43:03 </premis:dateCreatedByApplication>
      </premis:creatingApplication>
     </premis:objectCharacteristics>
    </premis:object>
   </mets:xmlData>
 </mets:mdWrap>
</mets:techMD>
```

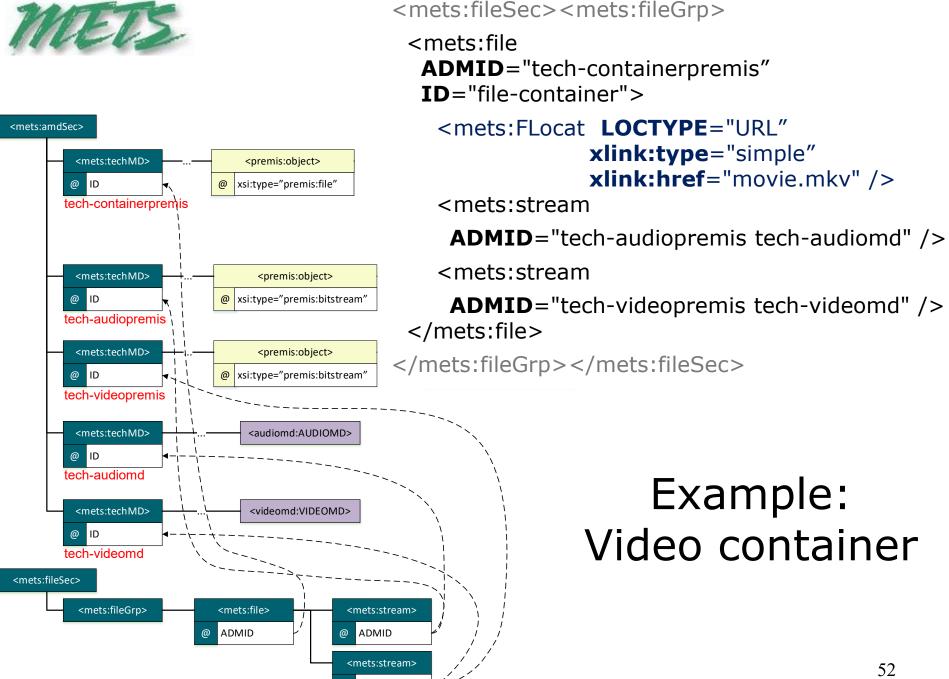


File Section

- File section <fileSec>:
 - one and only one
 - <fileGrp>, at least one
- File group <fileGrp>:
 - <file>, one for each file
 - ID: METS internal ID for file
 - ADMID: Links to administrative metadata
 - <FLocat>, one and only one
 - xlink:href: Relative path to digital object from the root of the package
 - Fixed values to LOCTYPE and xlink:type
 - <stream>, one for each stream (mandatory for a video container)
 - Internal ID for stream (if needed)
 - ADMID: Links to administrative metadata
 - Forbidden: <FContent>, <transformFile>, recursive use of <file>







ADMID

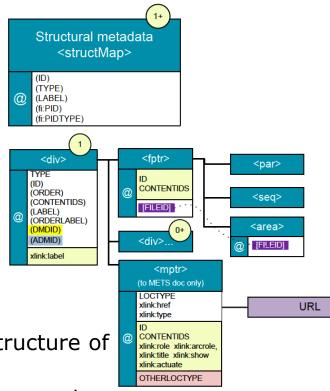


Structural Map

- Structural map <structMap>:
 - At least one is required
 - No predefined rules for the structure
 - For example,
 - a physical map describing the directory structure of the content
 - a logical map with following a given EAD3 metadata

Division <div>:

- TYPE is the only mandatory attribute
- DMDID for links to descriptive metadata
- ADMID for links to administrative metadata
- <div>: repeating and recursive use allowed
- <fptr>: File pointer
 - FILEID: METS internal ID to a file or a stream





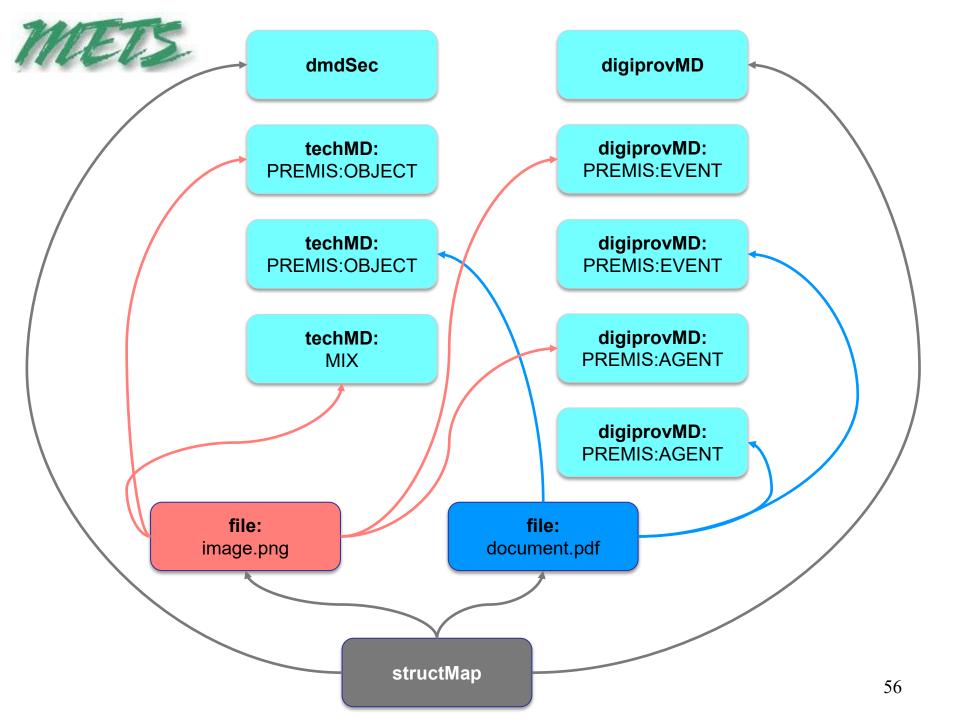
Structural Map

```
<mets:structMap TYPE="logical">
  <mets:div TYPE="audio-record" DMDID="descr-001">
     <mets:div TYPE="master">
       <mets:div TYPE="track" ORDER="1">
          <mets:fptr FILEID="file-001" />
       </mets:div>
       <mets:div TYPE="track" ORDER="2"> ... </mets:div>
     </mets:div>
     <mets:div TYPE="service-copy"> ... </mets:div>
     <mets:div TYPE="cover-images"> ... </mets:div>
  </mets:div>
</mets:structMap>
```



Links Inside METS

- The structural map...
 - contains links (<fptr>) to all <file> elements
 - contains links (DMDID) to all descriptive metadata sections
 - may contain links (<fptr>) to streams
 - may contain links (ADMID) to administrative metadata sections if the object is an entity described with <div>
- File section contains links to administrative metadata
 - Links (ADMID) to administrative metadata of a file or a stream in <file> and <stream> elements
 - At least one link per file and stream is mandatory: technical metadata for each file and stream
- Links to non-existing sections and unlinked sections are forbidden





ADMID Links and File Section

```
<mets:techMD ID="tech-2" ...> remis:object ...> ... </...>
 <mets:techMD ID="tech-3" ...> <mix:mix ...> ... </...>
 <mets:fileSec><mets:fileGrp>
 <mets:file ID="file-1" ADMID="tech-1 tech-3 dp-1 dp-3">
    <mets:FLocat ... xlink:href="image.png" />
 </mets:file>
 <mets:file ID="file-2" ADMID="tech-2 dp-2 dp-4">
    <mets:FLocat ... xlink:href="document.pdf" />
 </mets:file>
</mets:fileGrp></mets:fileSec>
```

ADMID/DMDID Links and Strutural Map

- For metadata concerning larger entities than files, we set the links in structural map
- <mets:dmdSec **ID**="dmd-desc" ...> ... </mets:dmdSec>
- <mets:digiprovMD **ID**="event-1" ...> ... </mets:digiprovMD>

Schemas, Source Code and Exercises

https://github.com/Digital-Preservation-Finland

- Large set of Schematron rules to meet the national specifications
- METS (and PREMIS) Library for Python
- Pre-Ingest Tool for Python:
 - Github repository: dpres-siptools
 - Creates SIPs according to our specifications
 - For automated use, but can be used manually too
- Exercises for Pre-Ingest Tool
 - Github repository: siptools-workshop-2019
- LGPLv3



Finnish National Specifications

https://digitalpreservation.fi/en/specifications

- Metadata requirements
- File formats
- Interfaces

Digital Preservation Services in Finland

Website: digitalpreservation.fi

Github: github.com/Digital-Preservation-Finland

Email: pas-support@csc.fi

Twitter: twitter.com/dpres_fi



PREMIS in METS



The PREMIS standard

- International de-facto standard for metadata to support the preservation of digital objects and ensure their long-term usability.
 - Information you need to know for preserving digital objects

Preservation Metadata: Implementation Strategies

- Developed by an international team of experts.
- Implemented in digital preservation projects around the world.
- Incorporated into commercial and open-source digital preservation tools and systems.





PREservation Metadata Implementation Strategies

The PREMIS standard

- Data Dictionary (PREMIS 3.0)
 - http://www.loc.gov/standards/ premis/v3/premis-3-0-final.pdf
 - Version 3 major release
- XML schema v3.0
 - http://www.loc.gov/standards/ premis/premis.xsd
- OWL ontology
- Supporting documentation





Scope

What PREMIS DD is:

- Common data model for organizing/thinking about preservation metadata
- Standard for exchanging information packages between repositories
- Implementable
- Technically neutral
- Core metadata

What PREMIS DD is not:

- Out-of-the-box solution
- All needed metadata
- Lifecycle management of objects outside repository
 - increasing support for integration with outside
- Rights management standard
 - strong support for rights statements





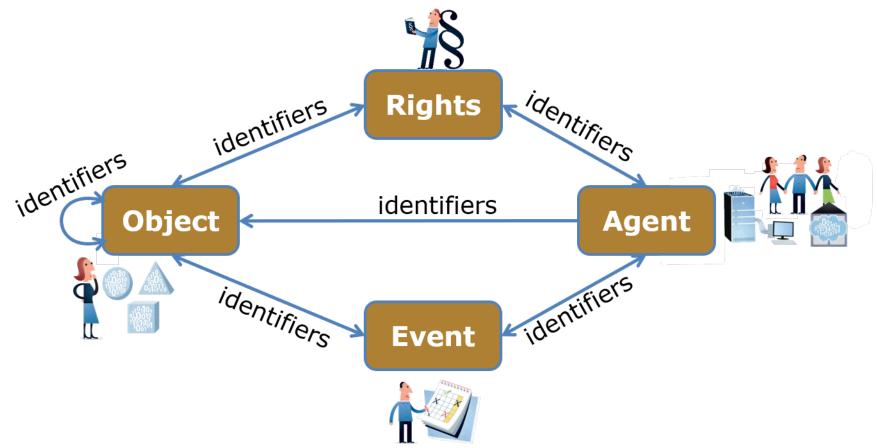


- Entities: "things" relevant to digital preservation that are described by preservation metadata (Intellectual Entities, Objects, Events, Rights, Agents)
- Properties of Entities (semantic units)
- Relationships between Entities
- Why have a data model?
 - Organizational convenience (for development and use)
 - Useful framework for distinguishing applicability of semantic units across different types of Entities and different types of Objects
 - But: not a formal entity-relationship model; not sufficient to design databases



PREservation Metadata Implementation Strategies

PREMIS 3 Entities





PREservation Metadata Implementation Strategies

PREMIS 3 Entities





object

Are what repository actually preserves Different types of objects: 'file', 'representation', 'bitstream' or 'intellectual entity'

event



Aggregates metadata about actions. Contains e.g. event identifier, event type (creation, migration, ...), and other event details.

agent



Can have different roles in relation to an event or objects.
Can be persons, organisations, programs etc.
Contains identifier and possibly other information like name and type

rights



Describes rights and permissions specifically related to digital preservation. Contains e.g. rights statements and possible extensions with other metadata



Issues in using PREMIS with METS

- Flexibility of METS requires implementation decisions:
 - Which METS sections to use
 - Use PREMIS container or separate packages?
 - Whether to record elements redundantly in PREMIS and METS
 - How to record elements that are also part of a format specific technical metadata schema (e.g. MIX)
 - Where to store structural relationships?
 - How to deal with locally controlled vocabularies



PREMIS and METS sections

- You can't put all PREMIS metadata directly under amdSec
- What sections to use for PREMIS metadata?
 - Alternative 1
 - Object (Representation, File and Bitstream) in techMD Object (Environment Intellectual Entity) in techMD Object (Intellectual Entity) in dmdSec
 - Event in digiProvMD
 - Rights in rightsMD
 - Agent with event or rights
 - Alternative 2
 - Everything in digiProvMD





Local decisions may vary depending on processing model

http://www.loc.gov/standards/premis/guidelines2017-premismets.pdf



SUPPORT AND THE METS COMMUNITY

Where to find support

- METS homepage: http://www.loc.gov/standards/mets/
- METS GitHub: https://github.com/mets
- METS listserv:

https://listserv.loc.gov/cgibin/wa?SUBED1=mets&A=1 and
METS@LISTSERV.LOC.GOV



Finding METS tools

COPTR Page:

https://coptr.digipres.org/index.php/ METS (Metadata Encoding and Transmission Standard)



CONFORMANCE



How to conform?

- Create a profile
- (Register profile)
- Remember to use the mandatory elements (StructMap)



SIMPLE EXAMPLE TO BUILD FAMILIARITY



Example

 There is an exercise available to create our own simple example and build familiarity

Digitized Postcard Example

- Each side digitized as separate hi-res TIFF images along with derived PNG thumbnail images
- A TEI transcription of the text written on the back
- MODS descriptive metadata record for the postcard
- Basic technical metadata for all files is available: format, size, checksum





Example: Folder and Files

- Postcard_Hemingway_001
 - front.tif
 - back.tif
 - front.png
 - back.png
 - back.tei
 - mods.xml



WRAP UP



Today we have:

- Gotten to know the elements of METS
- Seen 1 implementation
- Gotten a simple exercise to test at home
- In short we have started our journey into understanding and using METS



Thank you for today!

Karin Bredenberg
Juha Lehtonen
Tobias Steinke