



Metadata Encoding & Transmission Standard

METS Tutorial

iPRES 2019 – September 16th 2019



Welcome!

Presenters:

Karin Bredenberg

Juha Lehtonen

Sean Mosely



Content

- Introduction to METS
 - Short background (brief history and rationale of METS)
 - Benefits of implementing METS
 - METS in detail
 - Core elements
 - METS profile
- Implementation and tools
 - Case studies; Finland and New Zealand
- Simple example to build familiarity
- Wrap up
 - Support and the METS community
 - Conformance



INTRODUCTION TO METS



BACKGROUND



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



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



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

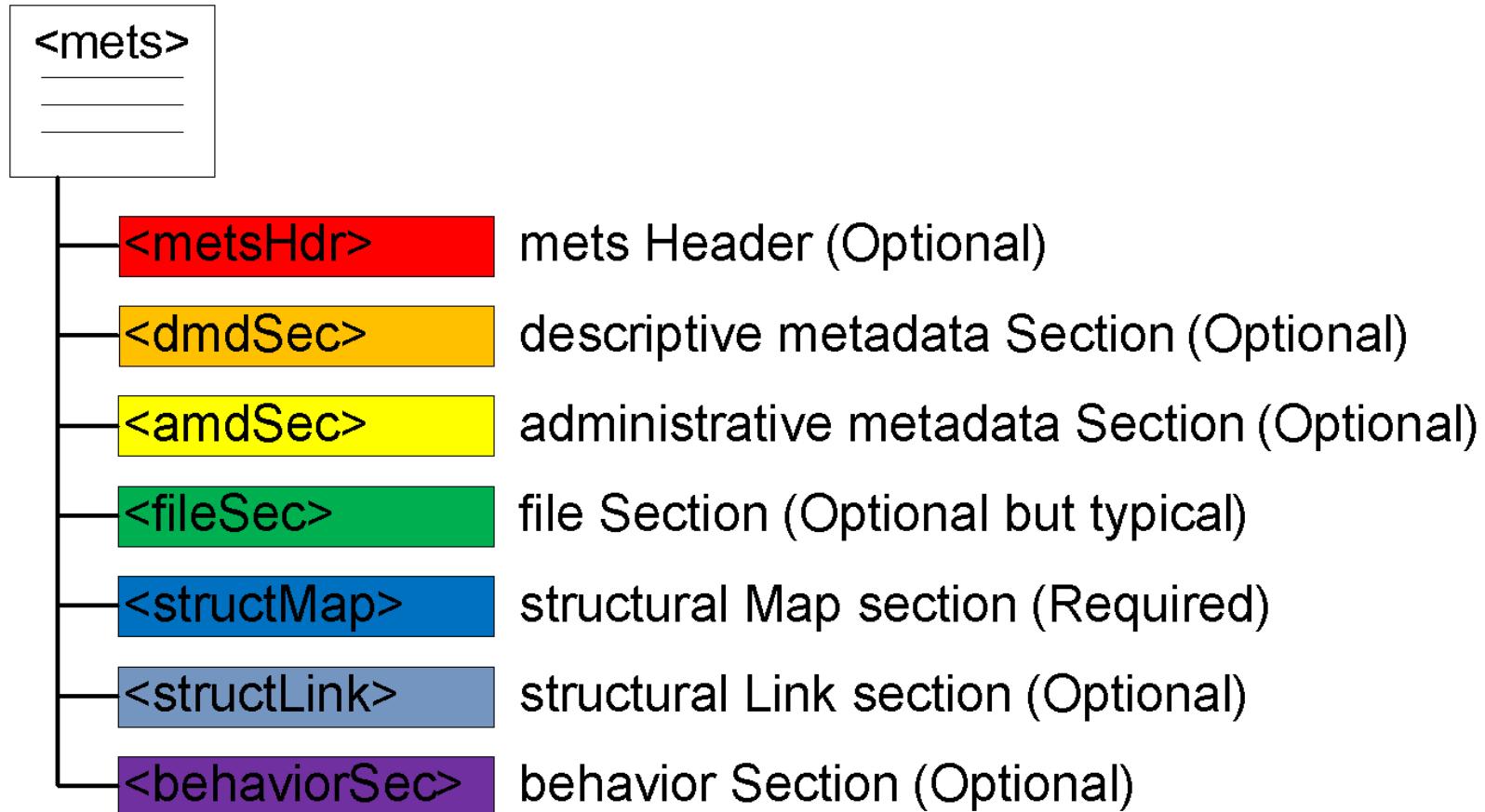


METS IN DETAIL

CORE ELEMENTS



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



Descriptive Metadata

```
<mets>
```

```
  <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



Administrative metadata

```
<mets>
```

```
  <!-- some lines -->
```

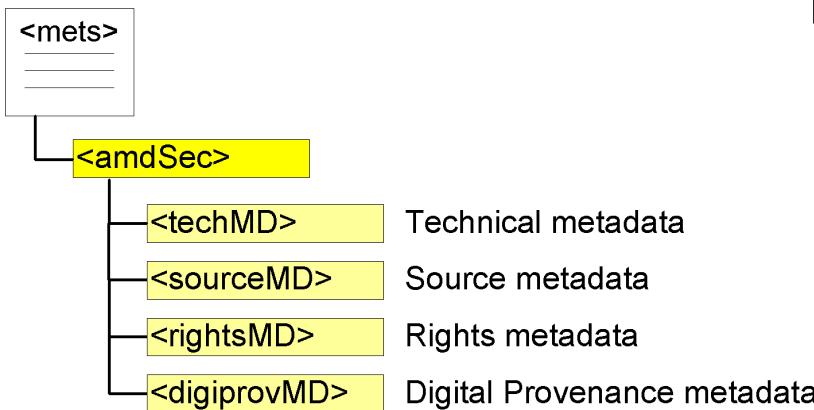
```
  <amdSec>
```

administrative metadata Section

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



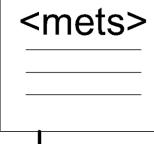
Administrative metadata Flavors



- 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



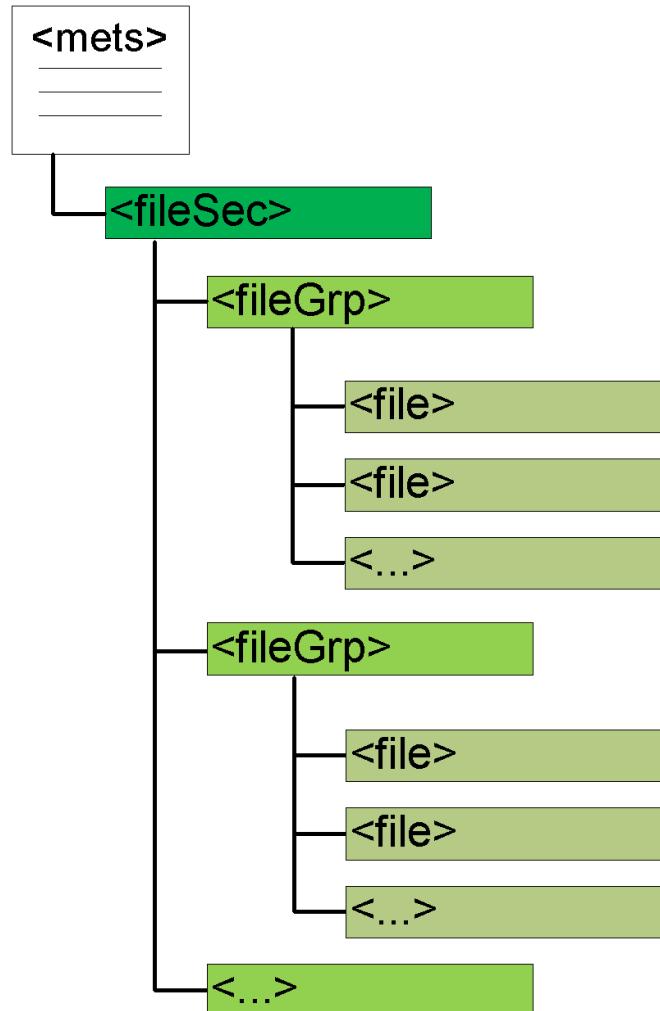
<fileSec>

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, hi-res jpeg, med-res jpeg, gif, etc)



Structural Map Section

```
<mets>
```

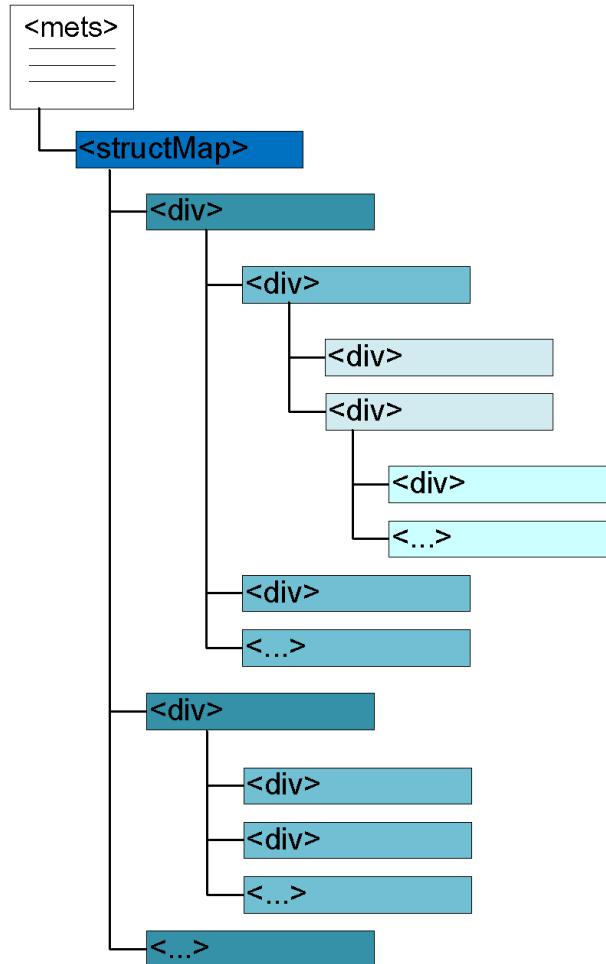
```
  <structMap>
```

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="page")
 - Division (type="photo")
 - Division (type="photo")
 - Division (type="photo")
 - Division (type="page")
 - Division (type="photo")
 - Division (type="photo")



Structural Link Section

```
<mets>
```

```
  <!--  
  <!--  
  <!--
```

```
  <structLink>
```

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.



Behavior Section

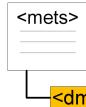
<mets>

<behaviorSec> 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.



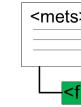
Internal or external



descriptive metadata Section



administrative metadata Section



file Section

- It's possible to include descriptive and administrative metadata and files or point to them.
- It's 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 namespace="##other" processContents="lax"/>
```



xsd:anyAttribute con't

- Declare a custom namespace

```
<mets xmlns="http://www.loc.gov/METS/
  xmlns:local="http://mets.library.illinois.edu/v1/">
  <metsHdr>
    <agent ROLE="CREATOR" >
      <name>Habing, Tytia</name>
      <note local:noteType="name-authority">
        http://viaf.org/viaf/7146331796818690920
      </note>
    </agent>
```

- Ideally custom attributes are documented in a METS Profile



METS PROFILES



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



Profiles Registration

- Registration at the METS Editorial Board
- August 2018 46 registered profiles
- Review period at/by the METS-list
- Can be used by others
- Others can make a new profile extending an existing profile with their use



Validating

- The profile dont 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 decisions regarding redundancy in cases when not mandatory in both standards.
- Regarding PREMIS and METS see the document mentioned earlier



IMPLEMENTATION AND TOOLS

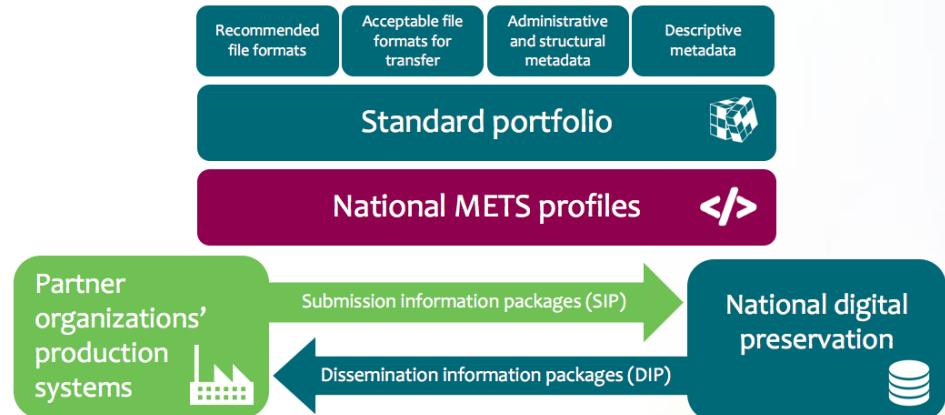
CASE STUDIES



CASE STUDY - FINNISH NATIONAL SPECIFICATIONS

Digital Preservation Services in Finland

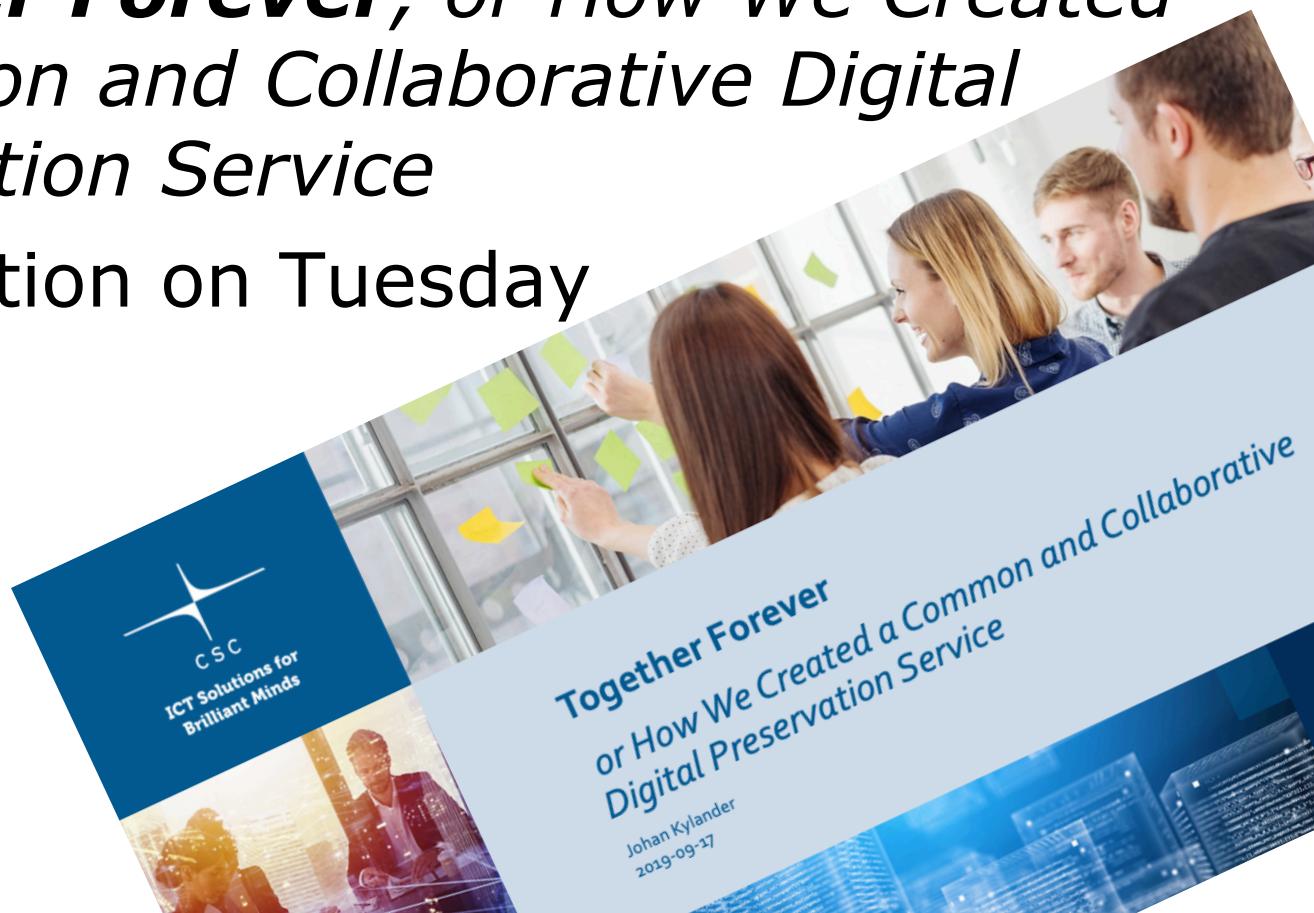
- CSC produces centralized digital preservation services for the Ministry of Education and Culture
 - Digital Preservation Service for Cultural Heritage (in production since 2015)
 - Digital Preservation Service for Research Data (in production 2019)
- Preparing and packaging digital assets is the most visible part of the digital preservation for the partner organizations
- Common specifications guides the packaging process
- Specifications are maintained in collaboration with partner organizations





Digital Preservation Services in Finland

- See iPRES 2019 short paper:
 - ***Together Forever, or How We Created a Common and Collaborative Digital Preservation Service***
 - Presentation on Tuesday





Information Packages

- An information package (SIP, DIP) contains a METS document, a digital signature file, and digital objects
- The METS document must follow the national specifications, describing:
 - 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”



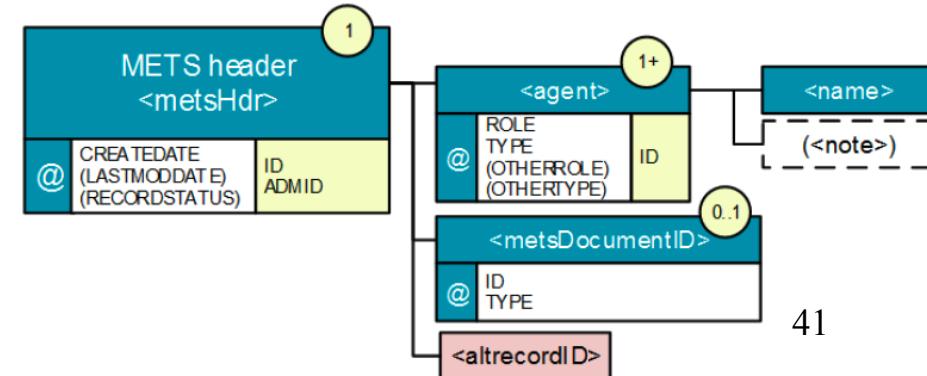
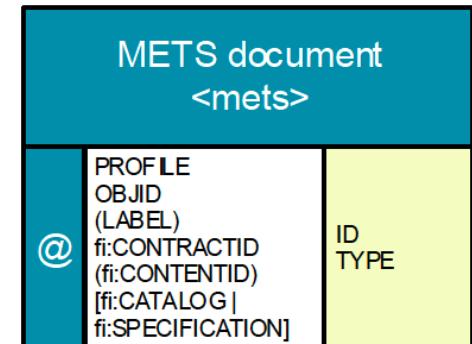
Main Sections

- 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 [2..n]
 - fileSec: File section [1..1]
 - structMap: Structural map [1..n]
- Forbidden sections:
 - behaviorSec and structLink

METS Root and Header

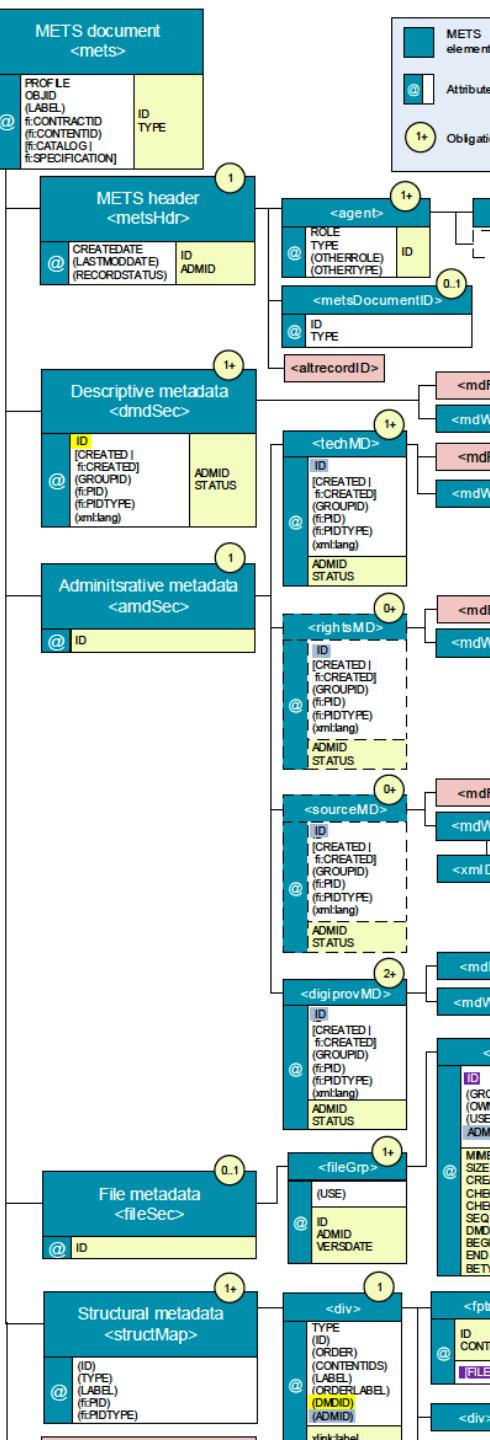
- The METS Root element `<mets>`:
 - OBJID: ID of the SIP/DIP
 - PROFILE:
 - Different profiles for different national services
 - However, the specifications are currently same for all
 - National attribute extensions:
 - `fi:CONTRACTID`: ID of the contract which justifies the submission
 - `fi:SPECIFICATION` or `fi:CATALOG`: Version number of the national specification or the used corresponding XML catalog
 - `fi:CONTENTID`: Possibility to split the content to multiple SIPs (opt.)
- METS header `<metsHdr>`:
 - CREATEDATE:

Creation timestamp of the SIP/DIP
 - At least one METS agent (creator of the SIP/DIP)



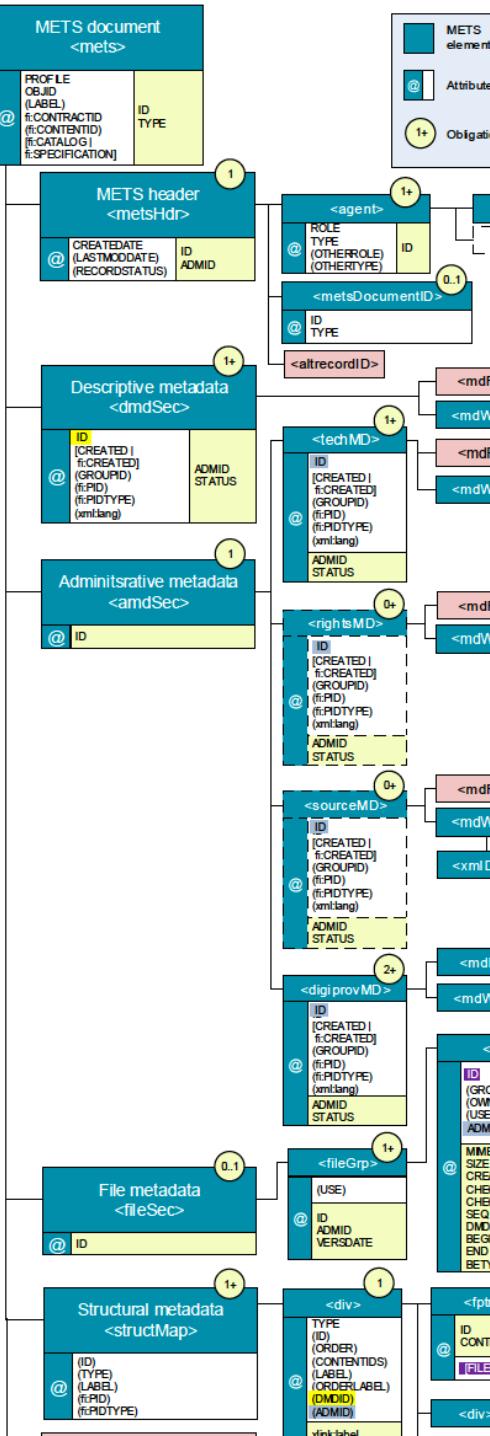
METS Sections

- 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
 - CREATED or fi:CREATED:
 - Metadata creation time
 - CREATED requires accurate timestamp, but optionally inaccurate presentation is possible with national extension fi:CREATED
 - Metadata wrapper <mdWrap>:
 - MDTYPE: Metadata format
 - OTHERMDTYPE: Metadata format, used if not in dictionary of MDTYPE
 - MDTYPEVERSION: Version of the metadata format



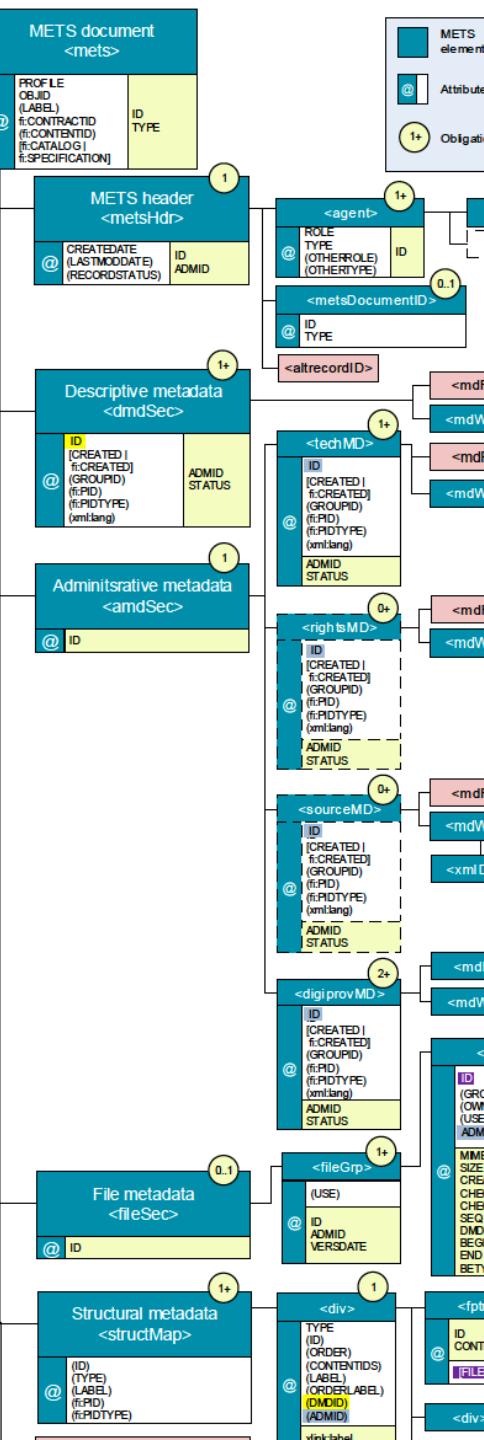
METS Sections

- At least one `<dmdSec>` must include descriptive metadata in one of the formats listed in the (national) Standard Portfolio.
- At least one `<techMD>` per digital object is required with 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

- Provenance metadata sections:
 - each PREMIS Event in its own <digiprovMD>. At least one Event for content creation is mandatory.
 - each PREMIS Agent in its own <digiprovMD>
 - Link to a preservation plan (digiprovMD/mdRef, TBD).
- Files (and streams) in the package are listed in a file section <fileSec>. Linkings between files and administrative metadata are in file section too (file@ADMID).
- At least one structural map <structMap> is mandatory. Linkings 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="2011-03-31T15:40:15">
  <mets:mdWrap MDTYPE="PREMIS:OBJECT" MDTYPEVERSION="2.2">
    <mets:xmlData>
      <premis:object xsi:type="premis:file">
        <premis:objectIdentifier>
          <premis:objectIdentifierType>urn</premis:objectIdentifierType>
          <premis:objectIdentifierValue>
            URN:NBN:fi-fe201215081520
          </premis:objectIdentifierValue>
        </premis:objectIdentifier>
        <premis:objectCharacteristics>
          <premis:compositionLevel>0</premis:compositionLevel>
          <premis:fixity>
            <premis:messageDigestAlgorithm>MD5</premis:messageDigestAlgorithm>
            <premis:messageDigest>
              aa4bddaacf5ed1ca92b30826af257a1b
            </premis:messageDigest>
          </premis:fixity>
          <premis:format>
            <premis:formatDesignation>
              <premis:formatName>image/png</premis:formatName>
              <premis:formatVersion>1.2</premis:formatVersion>
            </premis:formatDesignation>
            <premis:formatRegistry>
              <premis:formatRegistryName>PRONOM</premis:formatRegistryName>
              <premis:formatRegistryKey>fmt/13</premis:formatRegistryKey>
            </premis:formatRegistry>
          </premis:format>
          <premis:creatingApplication>
            <premis:dateCreatedByApplication>
              2011-02-15T15:43:03
            </premis:dateCreatedByApplication>
          </premis:creatingApplication>
        </premis:objectCharacteristics>
      </premis:object>
    </mets:xmlData>
  </mets:mdWrap>
</mets:techMD>
```



PREMIS Event in METS

```
<mets:digiprovMD ID="dp-001" CREATED="2011-03-15T12:13:14">
  <mets:mdWrap MDTYPE="PREMIS:EVENT" MDTYPEVERSION="2.2">
    <mets:xmlData>
      <premis:event>
        <premis:eventIdentifier>
          <premis:eventIdentifierType>uuid</premis:eventIdentifierType>
          <premis:eventIdentifierValue>
            1c002c38-837d-437f-8af2-de8c8864b5b1
          </premis:eventIdentifierValue>
        </premis:eventIdentifier>
        <premis:eventType>creation</premis:eventType>
        <premis:eventDateTime>2011-03-15T11:12:13</premis:eventDateTime>
        <premis:eventOutcomeInformation>
          <premis:eventOutcome>success</premis:eventOutcome>
        </premis:eventOutcomeInformation>
        <premis:linkingAgentIdentifier>
          <premis:linkingAgentIdentifierType>
            local
          </premis:linkingAgentIdentifierType>
          <premis:linkingAgentIdentifierValue>
            gimp-1
          </premis:linkingAgentIdentifierValue>
        </premis:linkingAgentIdentifier>
      </premis:event>
    </mets:xmlData>
  </mets:mdWrap>
</mets:digiprovMD>
```

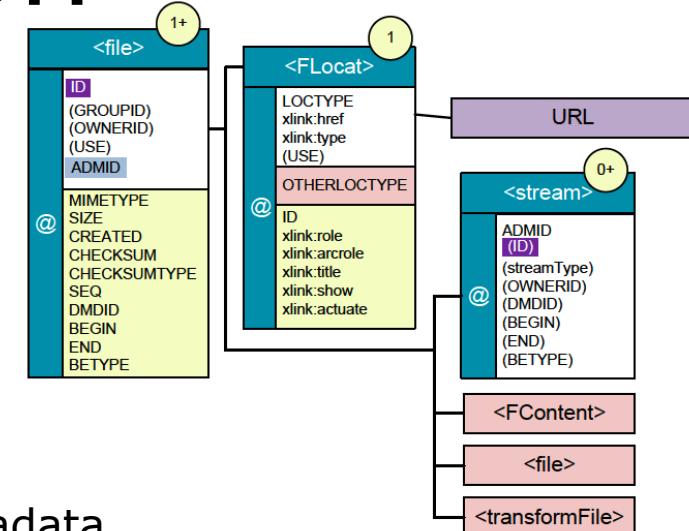


MIX metadata in METS

```
<mets:techMD ID="tech-002" CREATED="2011-03-31T15:40:15">
  <mets:mdWrap MDTYPE="NISOIMG" MDTYPEVERSION="2.0">
    <mets:xmlData>
      <mix:mix>
        <mix:BasicDigitalObjectInformation>
          <mix:Compression>
            <mix:compressionScheme>Deflate/zlib</mix:compressionScheme>
          </mix:Compression>
        </mix:BasicDigitalObjectInformation>
        <mix:BasicImageInformation>
          <mix:BasicImageCharacteristics>
            <mix:imageWidth>1920</mix:imageWidth>
            <mix:imageHeight>1080</mix:imageHeight>
            <mix:PhotometricInterpretation>
              <mix:colorSpace>sRGB</mix:colorSpace>
            </mix:PhotometricInterpretation>
          </mix:BasicImageCharacteristics>
        </mix:BasicImageInformation>
        <mix:ImageAssessmentMetadata>
          <mix:ImageColorEncoding>
            <mix:BitsPerSample>
              <mix:bitsPerSampleValue>8</mix:bitsPerSampleValue>
              <mix:bitsPerSampleUnit>integer</mix:bitsPerSampleUnit>
            </mix:BitsPerSample>
            <mix:samplesPerPixel>3</mix:samplesPerPixel>
          </mix:ImageColorEncoding>
        </mix:ImageAssessmentMetadata>
      </mix:mix>
    </mets:xmlData>
  </mets:mdWrap>
</mets:techMD>
```

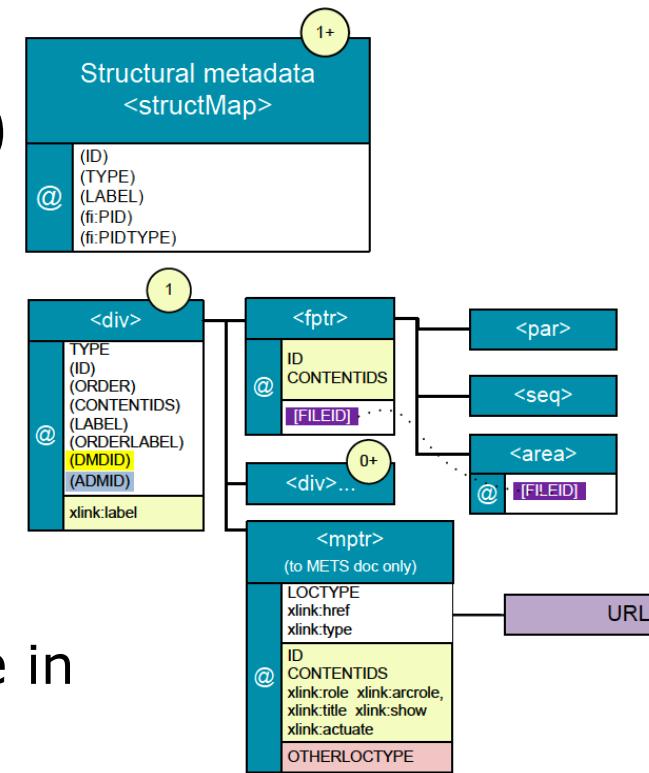
File Section

- File section `<fileSec>`:
 - one and only one
 - `<fileGrp>`, at least one
- File group `<fileGrp>`:
 - `<file>`, one for each file
 - ID: Internal ID for file
 - ADMID: Linkings 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 in a video container)
 - ID: Internal ID for stream
 - ADMID: Linkings to administrative metadata
 - Forbidden: `<FContent>`, `<transformFile>`, recursive use of `<file>`



StructMap

- Structural map `<structMap>`:
 - At least one
 - One and only one root `<div>`
- Division `<div>`:
 - TYPE is the only mandatory attribute in all `<div>`:s
 - DMDID for linkings to descriptive metadata
 - ADMID for linkings to administrative metadata
 - `<div>`: multiple elements allowed, recursive use allowed
 - `<fptr>`: File pointer, multiple elements allowed
 - FILEID: Pointer to a file or a stream





ADMID Linkings and File Section

- <mets:techMD ID="**tech-1**" ...> <premis:object ...> ... </...>
- <mets:techMD ID="**tech-2**" ...> <premis:object ...> ... </...>
- <mets:techMD ID="tech-3" ...> <mix:mix ...> ... </...>
- <mets:digiprovMD ID="dp-1" ...> <premis:event ...> ... </...>
- <mets:digiprovMD ID="dp-2" ...> <premis:event ...> ... </...>
- <mets:digiprovMD ID="dp-3" ...> <premis:agent ...> ... </...>
- <mets:digiprovMD ID="dp-4" ...> <premis:agent ...> ... </...>

```
<mets:fileSec><mets:fileGrp>
  <mets:file ID="file-1" ADMID="tech-1 tech-3 dp-1 dp-3">
    <mets:FLocat ... xlink:href="img.png" />
  </mets:file>
  <mets:file ID="file-2" ADMID="tech-2 dp-2 dp-4">
    <mets:FLocat ... xlink:href="doc.pdf" />
  </mets:file>
</mets:fileGrp></mets:fileSec>
```



ADMID/DMDID

Linkings and Structural Map

- For metadata concerning larger entities than files, use structural maps

- <mets:dmdSec ID="**dmd-desc**" ...> ... </mets:dmdSec>
- <mets:digiprovMD ID="**event-1**" ...> ... </mets:digiprovMD>

```
<mets:structMap>
  <mets:div TYPE="..." DMDID="dmd-desc" ADMID="event-1">
    <mets:ptr FILEID="file-1" />
    <mets:ptr FILEID="file-2" />
  </mets:div>
</mets:structMap>
```



Links to the Finnish Specifications and Code

- National specifications:
 - <http://digitalpreservation.fi/en/specifications>
 - Metadata requirements
 - File formats
 - Interfaces



Links to the Finnish Specifications and Code

- Schemas, source code and exercises:
 - <https://github.com/Digital-Preservation-Finland>
 - GPLv3
 - Large set of Schematron rules to meet the national specifications
 - METS and PREMIS Libraries for Python
 - Pre-Ingest Tool for Python:
 - Github repository: dpres-siptools
 - Creates SIPs according to our specifications
 - Exercises for Pre-Ingest Tool
 - Github repository: siptools-workshop-2019



CASE STUDY - NZ SPECIFICATIONS

NLNZ Digital Preservation

- The National Library of New Zealand is responsible for preserving NZ published works in perpetuity
- In addition, the Alexander Turnbull Library is responsible for acquiring and preserving special collections, including Māori and Pacific Islands cultural works
- NLNZ shares a digital preservation platform with Archives New Zealand



NLNZ Digital Preservation: Rosetta DP Platform

- Rosetta is a commercial digital preservation application, developed by Ex Libris (a ProQuest Company)
- It uses METS to structure the descriptive and technical metadata for SIPs, AIPs and DIPs
- Uses dmdSec, amdSec, fileSec and structMap sections
- Implements a flavour of PREMIS called DNX



NLNZ SIP/Ingest Tools

- Indigo (Java desktop application)
- Web Curator Tool
- Papers Past Bulk Ingest Tool

These handle 95% of use cases...



The other 5%...

- Very large web harvests
- One-off donations of a unique file structure
- Projects to scrape publications from a website
- Other “weird and wonderful” edge cases



Rosetta SIP Factory

- A Python library for creating Rosetta-friendly SIPs
- Allows tech-savvy staff members to build their own scripts for ordering their files and metadata
- A means of experimenting with new workflows, without the need to invest in developers



Rosetta SIP Factory

- Three callable functions:
 - `build_sip()`
 - `build_single_file_sip()`
 - `build_sip_from_json()`
- (relatively) simple interface
- Designed to handle 95% of the “5%” edge cases
- Anything more complicated can be finessed after running it through the SIP Factory



Rosetta SIP Factory

- Example:

```
from rosetta_sip_factory.sip_builder import build_sip
build_sip(
    ie_dmd_dict=[{"dc:title": "Title of IE"}],
    pres_master_dir="/path/to/ie/pm",
    general_ie_chars=[{"IEEntityType": "OneOffIE"}],
    accessRightsPolicy=[{"policyId": "100"}],
    input_dir=[{/path/to/ie}],
    digital_original=True,
    output_dir="/path/to/output_sip"
)
```



Rosetta SIP Factory

- The result:

```
/path/to/output_sip
  |
  |_content
  |
  |_mets.xml
  |
  |_streams
  |
  |_pm
  |_file001.tif
  |_file002.tif
```



Rosetta SIP Factory

- Github repository:

https://github.com/NLNZDigitalPreservation/rosetta_sip_factory

- Python Package Index:

pip install rosetta_sip_factory



**SIMPLE EXAMPLE TO BUILD
FAMILIARITY**



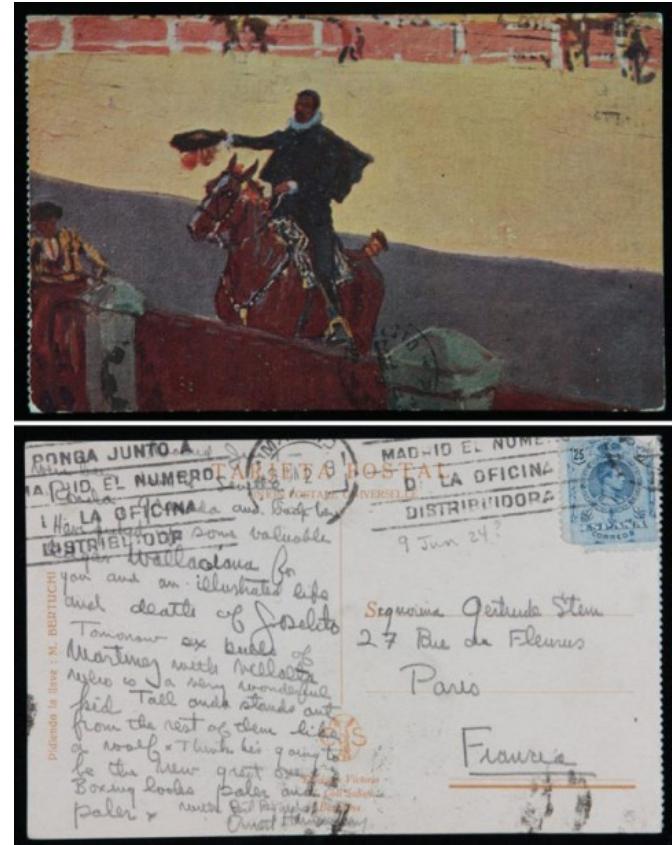
Example

- We use an exercise 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



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/cgi-bin/wa?SUBED1=mets&A=1> and
METS@LISTSERV.LOC.GOV



CONFORMANCE



How to conform?

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



Today we have:

- Gotten to know the elements of METS
- Seen 2 implementations with different tools used
- Seen a simple example
- In short we have started our journey into understanding and using METS



This presentation compiled with the help of:

- Nancy J. Hoebelheinrich, Knowledge Motifs LLC
- Rick Beaubien, Library Systems Office, U.C. Berkeley
- Joachim Bauer, Senior System Engineer, CCS Content Conversion Specialists GmbH
- Jerome McDonough, Asst. Professor, Graduate School of Library & Information Science, University of Illinois at Urbana-Champaign



Thank you for today!

Karin Bredenberg

Juha Lehtonen

Sean Mosely