The Metadata Encoding and Transmission Standard (METS)

From
Presentation
to Preservation

Markus Enders, GDZ
Göttingen State and University Library,
Germany

What is METS?

Ruleset expressed as XML Schema to describe a document

Keeps all objects of a document together:

      content files
      metadata
      structural data

makes no prescriptions for metadata schemas or format of content files

flexible container format

What is METS?

sections of METS:

```
                    ┌─────────────────┐
                    │  METS document  │
                    └─────────────────┘
        ┌──────────┬─────────┼─────────┬──────────┐
   ┌────────┐ ┌─────────┐ ┌──────────┐ ┌────────┐ ┌────────────┐
   │ Header │ │Descr.MD │ │ Adm. MD  │ │File List│ │ Behaviours │
   └────────┘ └─────────┘ └──────────┘ └────────┘ └────────────┘
      ┌────────────┐   ┌───────────┐
      │ Link Struct│   │ Struct Map│
      └────────────┘   └───────────┘
```
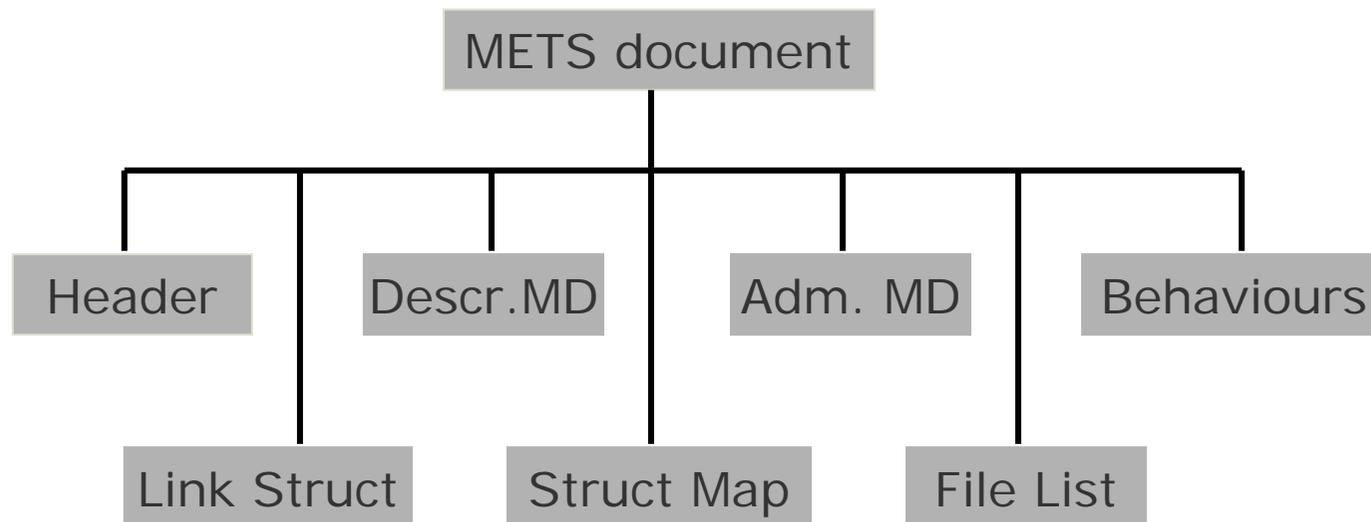
Example:
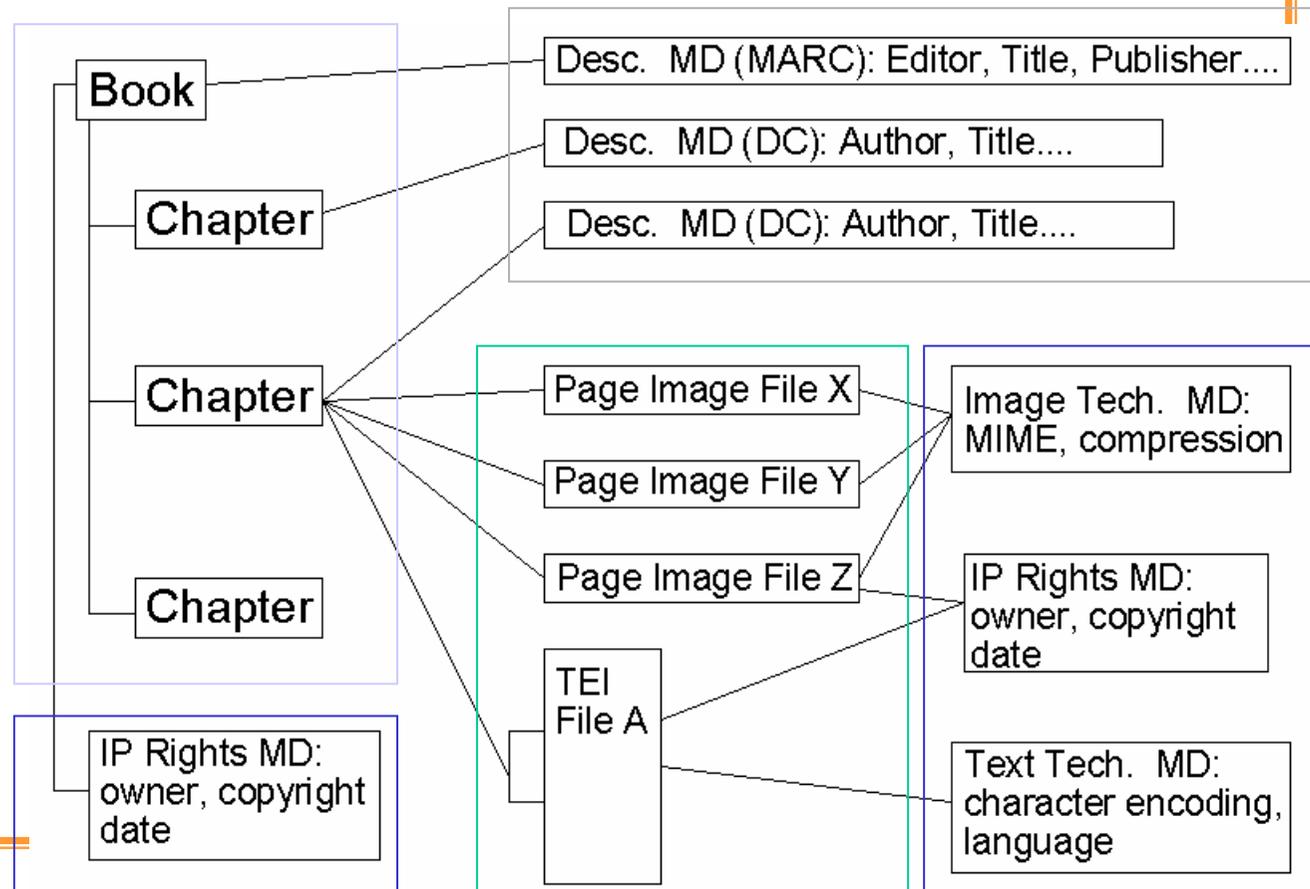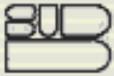
presentation

XSLT processing of METS file e.g. to create a TOC of struct map of web display.

script based processing based on METS files e.g. for page turner

```
<div type="monograph" label="title">
```

presentation

repository systems support more
elaborated document models.

different structures are supported

```
<structMap type="logical">
        <div ID="L1">....</div>
</structMap>
<structMap type="physical">
        <div ID="P1">....</div>
</structMap>
<structLink>
        <smLink FROM="L1" TO="P1"/>
</structLink>
```

presentation

repository systems support more elaborated document models.

different structures are supported

extensive use of more complex descriptive metadata models (DC, MODS...)

```
<dmdSec ID="dmd1>
        <dc:title>title</dc:title>
</dmdSec>
.....
<div dmdid="dmd1">
```

preservation

Why to use METS for preservation?

standarized way to describe a
document

flexible container format for SIP, DIP and AIP

Easy creation of SIPs as many delivery
systems are already supporting METS

METS profiles allows you to describe your
METS format in a human readable way
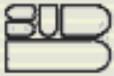
preservation

Why to use METS for preservation?

can be used to transfer data
(metadata and content) to keep data
redundant in different places.

What is needed
for Preservation?

preservation metadata

technical details on the format
structure
use of the digital content,
the history of all actions performed on the
resource
the authenticity information
responsibilities and rights information

What is needed for Preservation?

preservation metadata

technical details on the format

structMap structure

use of the digital content,
the history of all actions performed on the
resource
the authenticity information
responsibilities and rights information

What is needed
for Preservation?

preservation metadata

technical details on the format
structure
use of the digital content,
the history of all actions performed on the
resource
the authenticity information
responsibilities and rights information

fileSec

authenticity

available as attributes in METS file-section:

size in bytes, checksum, checksum-type, mimetype

```
<fileSec>
    <fileGrp>
        <file SIZE=.. CHECKSUM=.. MIMETYPE=..>
                <FLocat />
        </file>
    </ fileGrp >
</fileSec>
```
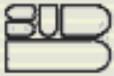
What is needed for Preservation?

preservation metadata

technical details on the format
structure
use of the digital content,
the history of all actions performed on the resource
the authenticity information
responsibilities and rights information

What is needed for Preservation?

preservation metadata

technical details on the format
structure
use of the digital content,
the history of all actions performed on the resource
the authenticity information
responsibilities and rights information

behaviour section

technical Metadata

metadata for files depending on media type as e.g resolution, color-depth…

use external metadata schema e.g. MIX for still-images in adm-section
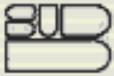
```
<admSec>
   <techMD>
       <MIX:mix>
               ....
       </MIX:mix>
   </techMD>
</admSec>
```

What is needed
for Preservation?

preservation metadata

technical details on the format
structure
use of the digital content,
the history of all actions performed on the
resource
the authenticity information
responsibilities and rights information

rightsMD
as administrative
metadata
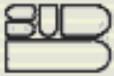
What is needed
for Preservation?

preservation metadata

technical details on the format
structure
use of the digital content,
the history of all actions performed on the resource
the authenticity information
responsibilities and rights information

digiprovMD
as administrative metadata

PREMIS and METS

PREMIS:

xml schemas to describe a ruleset to store
"… information a repository uses to support the digital preservation process…"
from Premis Data Dictionary

PREMIS and METS

PREMIS:

4 schemas:
  agents
  events
  objects
  rights

METS rightsMD-section:

```
<admSec>
   <rightsMD>
       <PREMIS:right>
              ....
       </PREIMIS:right>
   </rightsMD>
</admSec>
```

PREMIS and METS

PREMIS:

4 schemas:
      agents
      events
      objects
      rights

Information about an action that involes an object entity

METS digiprovMD-section:

```
<admSec>
   <digiprovMD>
       <PREMIS:event>
             ....
       </PREIMIS:event>
   </digiprovMD>
</admSec>
```
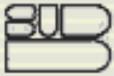
PREMIS and METS

example of an event

```
<admSec>
   <provMD>
      <PREMIS:event>
         <PREMIS:objectIdentifier>
            <PREMIS:eventIdentifierType>own</PREMIS:eventIdentifierType>
            <PREMIS:eventIdentifierValue>123</PREMIS:eventIdentifierValue>
         </PREMIS:objectIdentifier>
         <PREMIS:eventType>deletion</PREMIS:eventType>
         <PREMIS:eventDateTime> 2005-09-13T07:50:34+1:00 </PREMIS:eventDateTime>
         <PREMIS:eventDetail>deletion upon request from ...</PREMIS:eventDetail>
         ....
      </PREMIS:event>
   </provMD>
</admSec>
```

PREMIS and METS

PREMIS:

4 schemas:
      agents
      events
      objects   ???
      rights
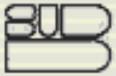
PREMIS and METS

PREMIS object can be:
representation, file or bitstream

what kind of information is stored for an object?

PREMIS and METS

what kind of information is stored for an object? – examples:

objectIdentifier

creating Application

enviroment: software, hardware

fixity – information (hashes)

format

storage information

relationships

PREMIS and METS

what kind of information is stored for an object?

objectIdentifier	— in dmdSec
creating Application	— in provMD
enviroment: software, hardware	— in provMD
fixity – information (hashes)	— in fileSec
format	— in techMD
storage information	— in fileSec
relationships	— in structMap

PREMIS and METS

what kind of information is stored for an object?

objectIdentifier	— in dmdSec
creating Application	— in provMD
enviroment: software, hardware	— in provMD
fixity – information (hashes)	— in fileSec
format	— in techMD
storage information	— in fileSec
relationships	— in structMap

PREMIS and METS

what kind of information is stored for an object?

objectIdentifier

creating Application

enviroment: software, hardware

fixity – information (hashes)

format

storage information

relationships

} in
digiprovMD

PREMIS and METS

Redundancy of metadata:

hashes:

in METS
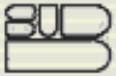
in MIX

in Premis

PREMIS and METS

Redundancy of metadata:

hashes:

in METS

```
<FileSec>
   <FileGrp>
       <File ID="FILE01" CHECKSUM="" CHECKSUMTYPE="" SIZE="">
               <FLocat .... />
       </File>
   </FileGrp>
</FileSec>
```

PREMIS and METS

Redundancy of metadata:

hashes:

    in METS

    in MIX

```
<File>
   <ImageIdentifier imageIdentifierLocation="system of identifier">
       unique persistent identifier
   </ImageIdentifier>
   <FileSize>1001000</FileSize>
   <Checksum>
       <ChecksumMethod>checksum</ChecksumMethod>
       <ChecksumValue>2224446888</ChecksumValue>
   </Checksum>
   ...
```

PREMIS and METS

Redundancy of metadata:

hashes:

    in METS

    in MIX

    in Premis

```
<METS:admSec>
<METS:digiProvMD>
<object>
    <objectIdentifier>.....</objectIdentifier>
    <objectCharacteristics>
        <fixitiy>
            <messageDigestAlgorithm>
            </messageDigestAlgorithm>
            <messageDigest></messageDigest>
            <messageDigestOriginator>
             </messageDigestOriginator>
        </fixity>
    </objectCharacteristics>
    .....
</object>
</METS:digiProvMD>
</METS:admSec>
```

PREMIS and METS

Redundancy of metadata:

format:

in METS	mimetype in fileSec
in MIX	mimetype in Format / MIMEType
in PREMIS	formatDesignation and format Name

METS and metadata schema

Best practices are needed

Current practice at SUB:

metadata which can be stored in container format, is only stored there.

If metadata schema regards redundant metadata as mandatory, it is stored in appropriate metadata section as well.
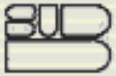
bundled files in METS

possibility to store information about nested files / bitstreams (zip or tar-archives)

bundled files

Proposed solution:

```
<file>
  <Flocat /> <!-- location of *tar.gz file -->
  <transformFile /> <!-- Instructions on reversing gzip -->
  <transformFile /> <! Instructions on reversing tar -->
  <file /> <!-- first embedded file -->
    <file > <!-- second embedded file -->
      <stream /> <!-- first embedded stream -->
      <stream /> <!-- second embedded stream -->
  </file>
</file>
```
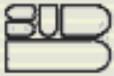
bundled files

possibility to store information about nested files / bitstreams (zip or tar-archives)

backward compatibility

currently under discussion for METS schema 1.5
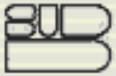
conclusions

use standarized container format (METS)
for SIP and DIP

ongoing collaboration necessary:

to enchance / extend METS

to create best practises

share tools to create SIPs / DIPs

the end