

# **METS und die digitale Bestandserhaltung**

**Die Verknüpfungsmöglichkeiten von METS mit  
speziellen Metadaten für technische Angaben  
und die digitale Bestandserhaltung**

**Markus Enders, SUB Göttingen**

**enders@sub.uni-goettingen.de**

# METS as a container format:

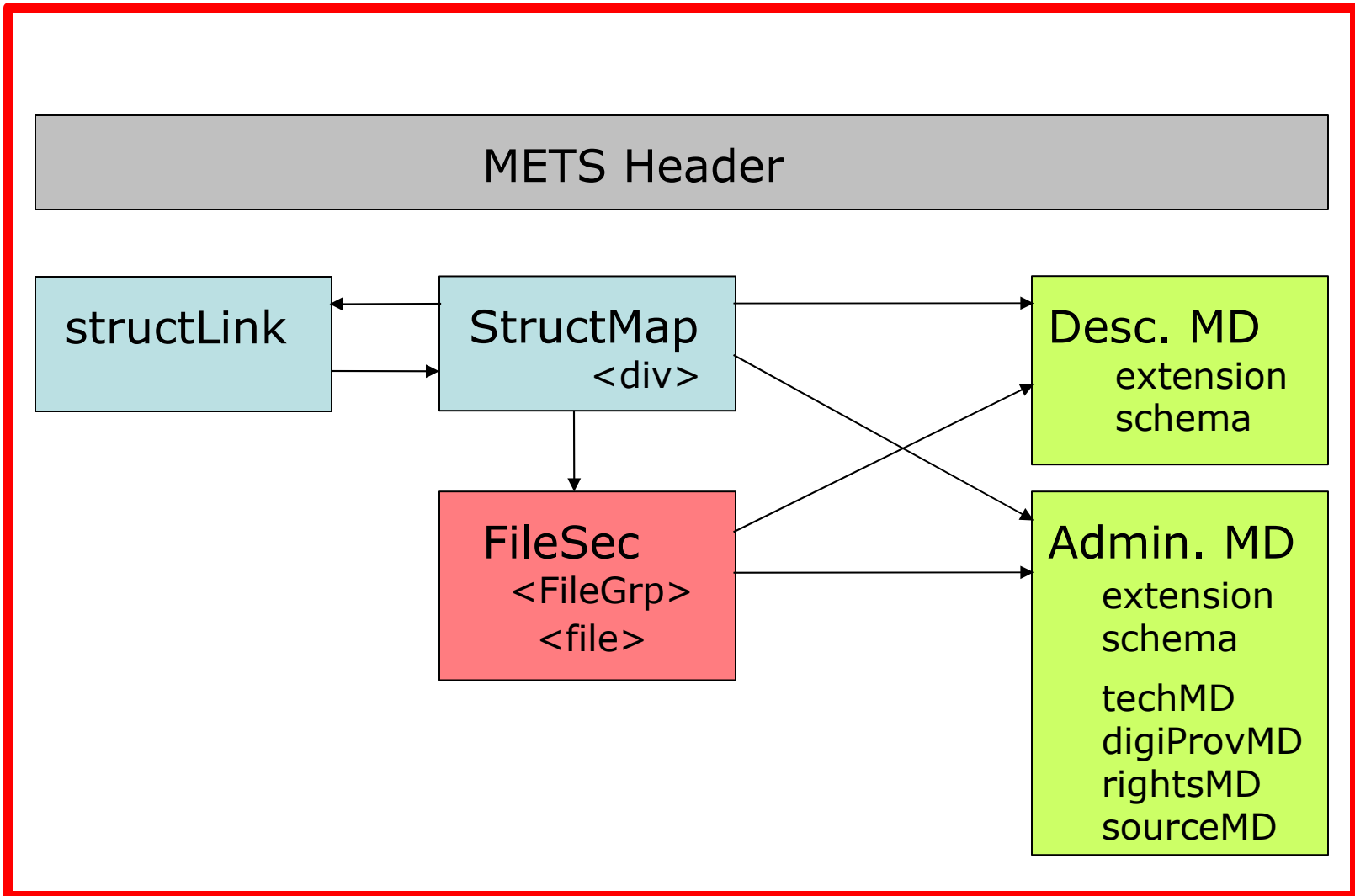
Extendable:  
usage of “extension schema” for  
descriptive and administrative metadata

Several ways to „serialize“ the model:  
xml is just one form;  
database structure,  
object-structure (java-objects)

Container as an infrastructure for your data

# METS abstract model

(simplified)



# Descriptive metadata

Metadata to describe an object:

division  
single file  
...

Any object can have one or more descriptive metadata sections.

Each section includes metadata according to one schema: DublinCore, MODS, MARC

# Administrative Metadata

Digital Preservation metadata

Technical metadata

Rights metadata (intellectual property)

Source metadata

# Preservation Metadata

e.g. Premis metadata schema

Object identification

Detailed information about an object:  
e.g. format of a file (e.g. PRONOM identifier)

Information about object creation: who? when?

Information about object usage:  
hardware/software environment

Information about object's integrity  
checksums

# Preservation Metadata

Event driven model to store history of object

Link to predecessor of an object

Type of event: e.g.  
migration event, deletion event

Can store detailed information about an agent  
which was responsible for the event  
(person or software)

# Preservation Metadata

Why is digital preservation metadata important?

For preservation planning:

selection process (which data needs to be migrated)

risk analysis: which files are affected, if a certain hardware / software environment is not available anymore

possibility to go back to earlier version, if data had been migrated to unsupported formats / lossy formats

# Preservation Metadata

Why is preservation metadata important?

To carry out preservation actions:

Hardware / software information are used to provide an emulation environment for accessing the content

Migration is based on detailed format information

# Technical Metadata

## (on file level)

Technical metadata:

metadata schema depends on type of file /  
media

e.g. MIX metadata schema for still images  
textMD for fulltext data

Format specific metadata:

e.g. Resolution, color depth, used compression  
schema as in `<mix:BasicImageCharacteristics>`

# Technical Metadata

Capture metadata:  
appliances used in analog-to-digital conversion  
process (e.g. Camera, camera settings, scanner  
model...)

```
<mix:ScannerCapture>...
```

.... also format independent metadata:

Detailed information about the source of a  
file: size of source (physical dimensions)

Fixity information (checksum)

Date and time of creation

## Why are technical metadata important?

Scenario:

failures due to software bugs:

e.g.

PDF creation with buggy software; might create PDF which is readable by most standard readers, but not conform to standard.

Might cause trouble with future migration / emulation toolkits

➡ Preservation staff needs to be able to find all affected assets in the preservation repository (based on technical metadata)

## **Administrative metadata can be stored in several places:**

Different metadata schema might provide similar elements:

```
<premis:fixity> ... </premis:fixity>  
<mix:fixity> ... </mix:fixity>
```

Basic technical metadata can be stored in METS itself: e.g. MIME-type, checksum

```
<mets:file checksum="xxx"/>
```

## Reasons for redundant metadata:

Some elements are regarded as mandatory by extension schema.

Different purpose / level of detail:

- access systems:

e.g. MIME-type in `<mets:file>` might be sufficient

- preservation systems:

very detailed level is needed, e.g. Used compression schema of a TIFF file is important

# Administrative Metadata

- Be careful when updating metadata / content
- Document your implementation of METS (METS profiles)

# Case study: GDZ

## Digitization Centre

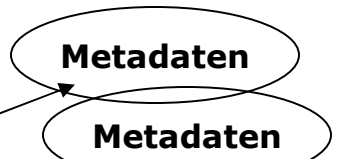
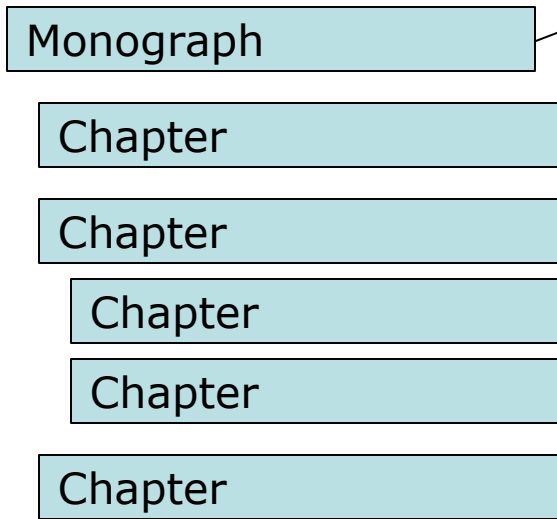
How to create this data in practice???

# METS example (1)

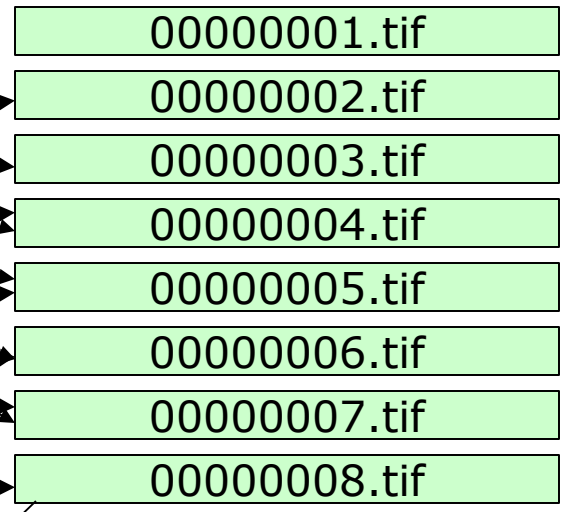
## Digitization Centre

### Simple logical document model

Logical structure  
<structMap>



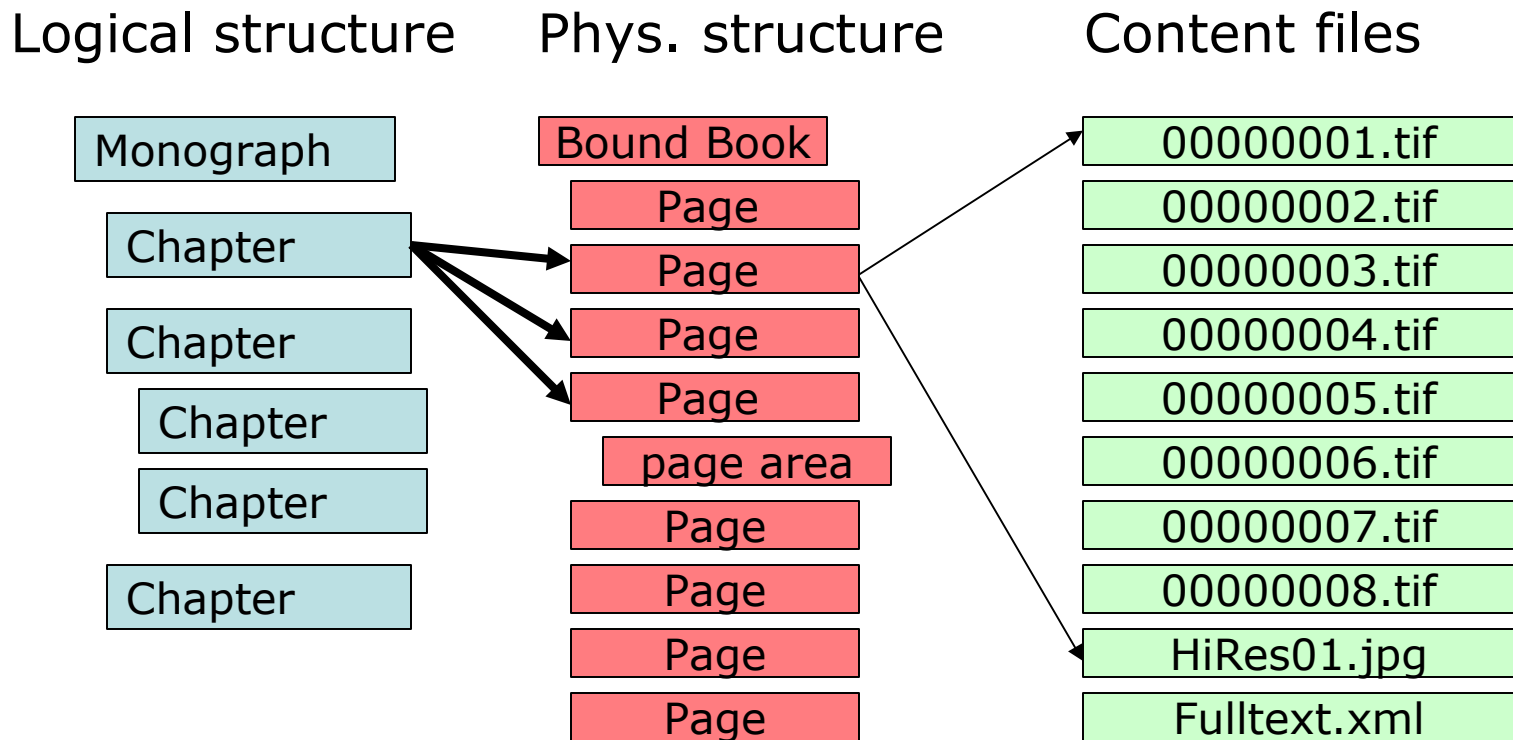
Content files  
<fileSec>



# METS example (2)

## Digitization Centre

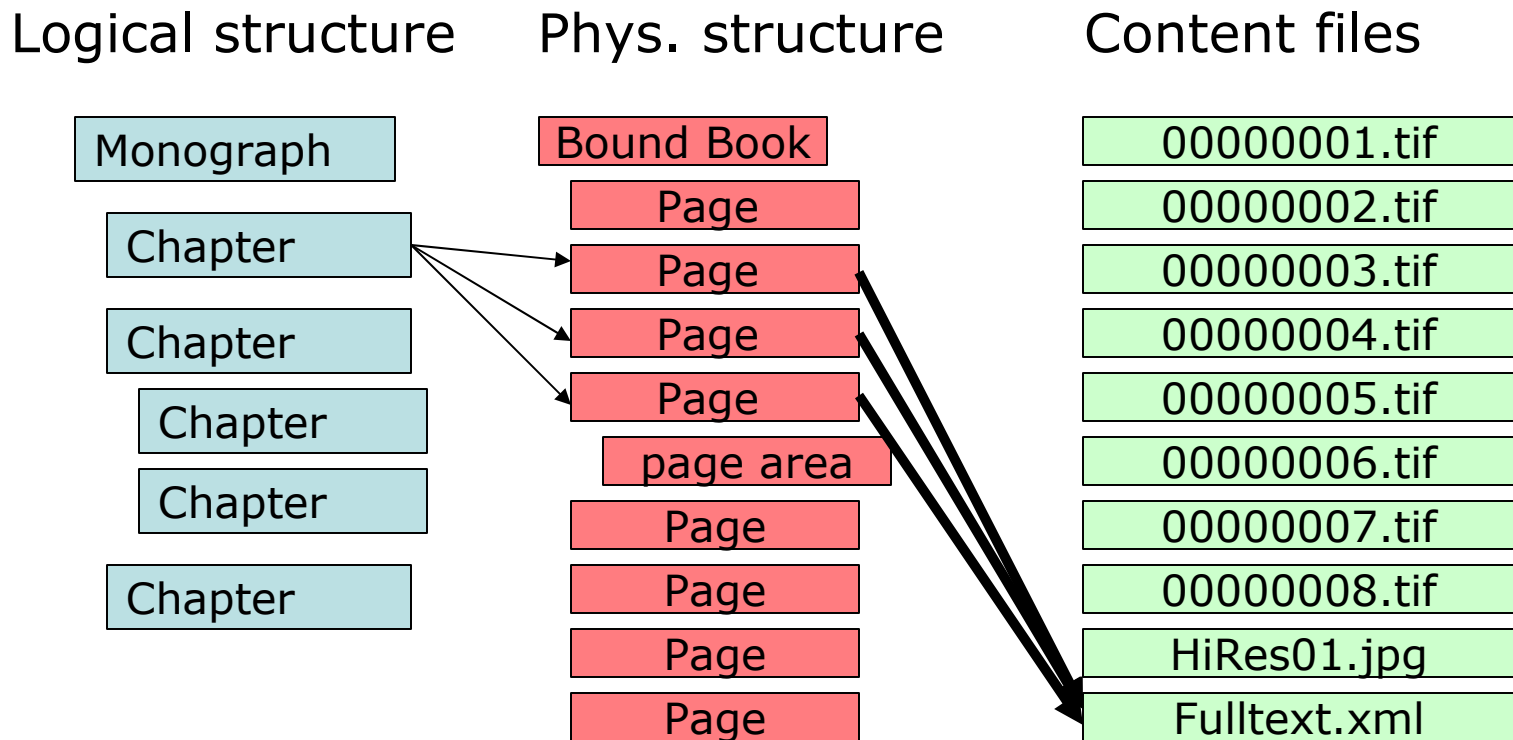
### Document model with two structures



# METS example (2)

## Digitization Centre

### Document model with two structures



# METS example (2)

## Digitization Centre

### Creating METS:

GOOBI – web-based workflow management tool

Create uppermost `<div>` (logical)  
with descriptive metadata (MODS) from OPAC

Read files and create `<fileSec>` and  
physical structure (pages)

Capture metadata for pages (page numbers)

Capture chapters and their metadata (logical)

Relationships between chapters and pages

# METS example (2)

## Digitization Centre

### Creating METS:

No technical metadata is captured;  
not needed for access system.

After data is ingested to the presentation system  
it is submitted to preservation team.

At the preservation team:

Run JHOVE to validate images, create techMD  
and digiProvMD.

Ingest data into preservation system

# METS workflow

## Distributed workflow:

METS may contain data

- which is captured at different locations
- for different purposes

Future usage of digitized files may simply add additional sections (e.g. if new metadata format becomes available, migrated files need to be stored etc.)



Interoperability is increased

METS Opening Day: May 7<sup>th</sup> 2007, Göttingen

[http://nestor.sub.uni-goettingen.de/mets\\_2007/](http://nestor.sub.uni-goettingen.de/mets_2007/)

<http://www.loc.gov/mets>