

PREMIS in METS: Die beiden Datenmodelle Matterhorn METS und Archivematica

Nestor-Praktikertag 2018: Metadaten für die digitale Langzeitarchivierung und
der Metadatenstandard PREMIS

Frankfurt am Main, 6. Juni 2018

Dr. Tobias Wildi, t.wildi@docuteam.ch

METS

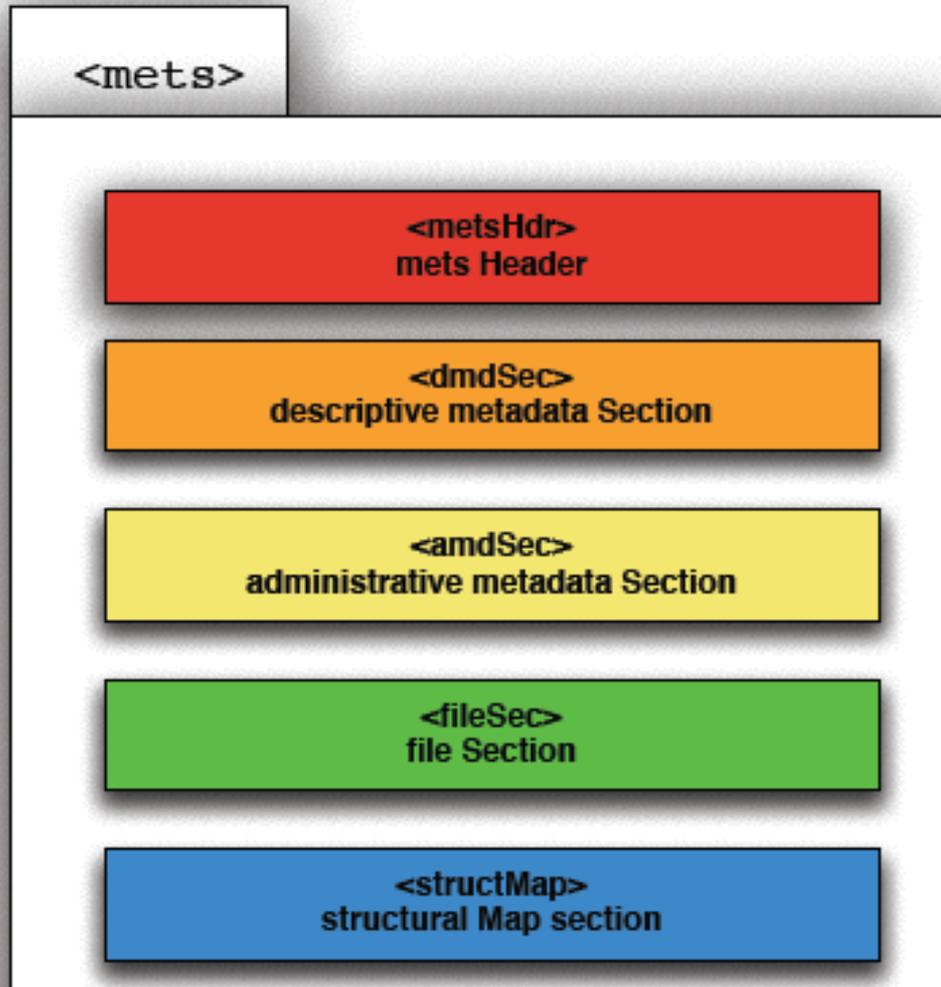
Was ist METS? Und was ist ein METS-Profil?

- **METS** = Metadata Encoding and Transmission Standard
Verwaltet durch die Library of Congress
<http://www.loc.gov/standards/mets/>
- METS ist ein Containerformat, welcher geeignet ist,
 - um das Informationsmodell von OAIS abzubilden
 - um Informationspakete zu serialisieren (systemunabhängig als eigenständige Paket weiterzugeben)
- Die konkrete Umsetzung von METS lässt grosse Flexibilität zu. In einem «Profil» wird im Detail beschrieben,
 - wie METS verwendet wird,
 - welche weiteren Standards eingebunden werden
 - wie diese Standards aufeinander abgestimmt werden.
- <http://www.loc.gov/standards/mets/mets-profiles.html>

METS Profiles

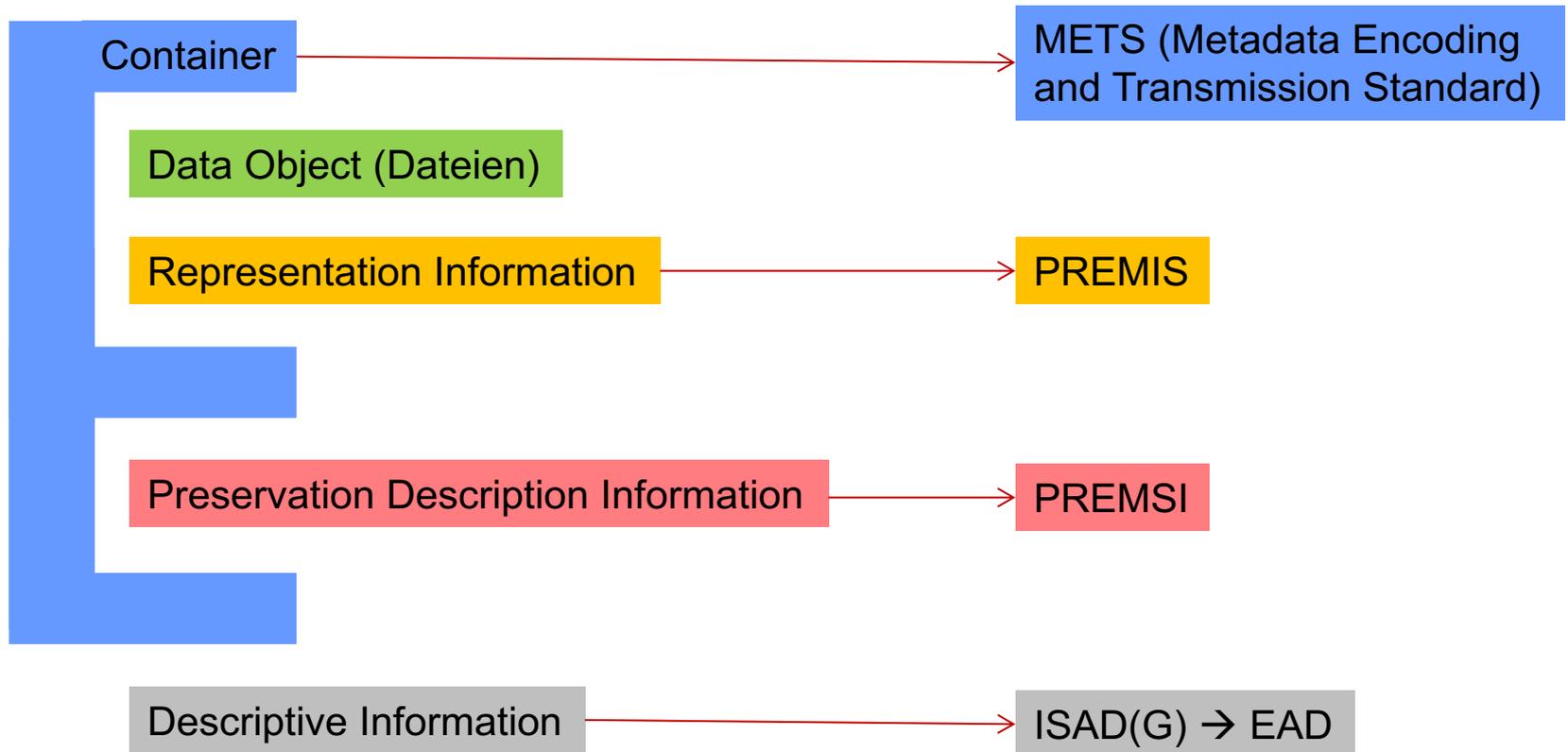
METS Profiles are intended to describe a class of METS documents in sufficient detail to provide both document authors and programmers the guidance they require to create and process METS documents conforming with a particular profile.

METS – Metadata Transmission and Encoding Standard



- Ist ein METS-Profil
- Basiert auf METS, EAD und PREMIS 2.2
- Ko-Entwicklung Staatsarchiv Wallis (CH) und docuteam.
- 2012 bei der Library of Congress registriert
<http://www.loc.gov/standards/mets/profiles/00000041.xml>
- Wird unterstützt durch die Open Source-Werkzeuge die docuteam für die digitale Archivierung entwickelt (docuteam packer, docuteam feeder)
- Wird zunehmend auch als Ablieferungsformat genutzt und nicht durch docuteam-Werkzeuge gebildet.
- Wird gegenwärtig weiterentwickelt zu «Matterhorn RDF», respektive «IP new generation», was in der Schweiz eCH-0160 ablösen soll.

Matterhorn METS-Profil bildet das OAIS-Informationsmodell ab



- Artefactual Systems, Inc., Toronto.
- Artefactual ist mit Evelyn McLellan selber im PREMIS-Board vertreten
- Archivemata bezeichnet sich als «Core Preservation System»: Bewertung, Ingest, Speicherung, Preservation Planning / Preservation Actions.
- Open Source (GPL3).
- Datenmodell basiert auf METS und Premis 2.2.
Dokumentation: <https://wiki.archivemata.org/METS>

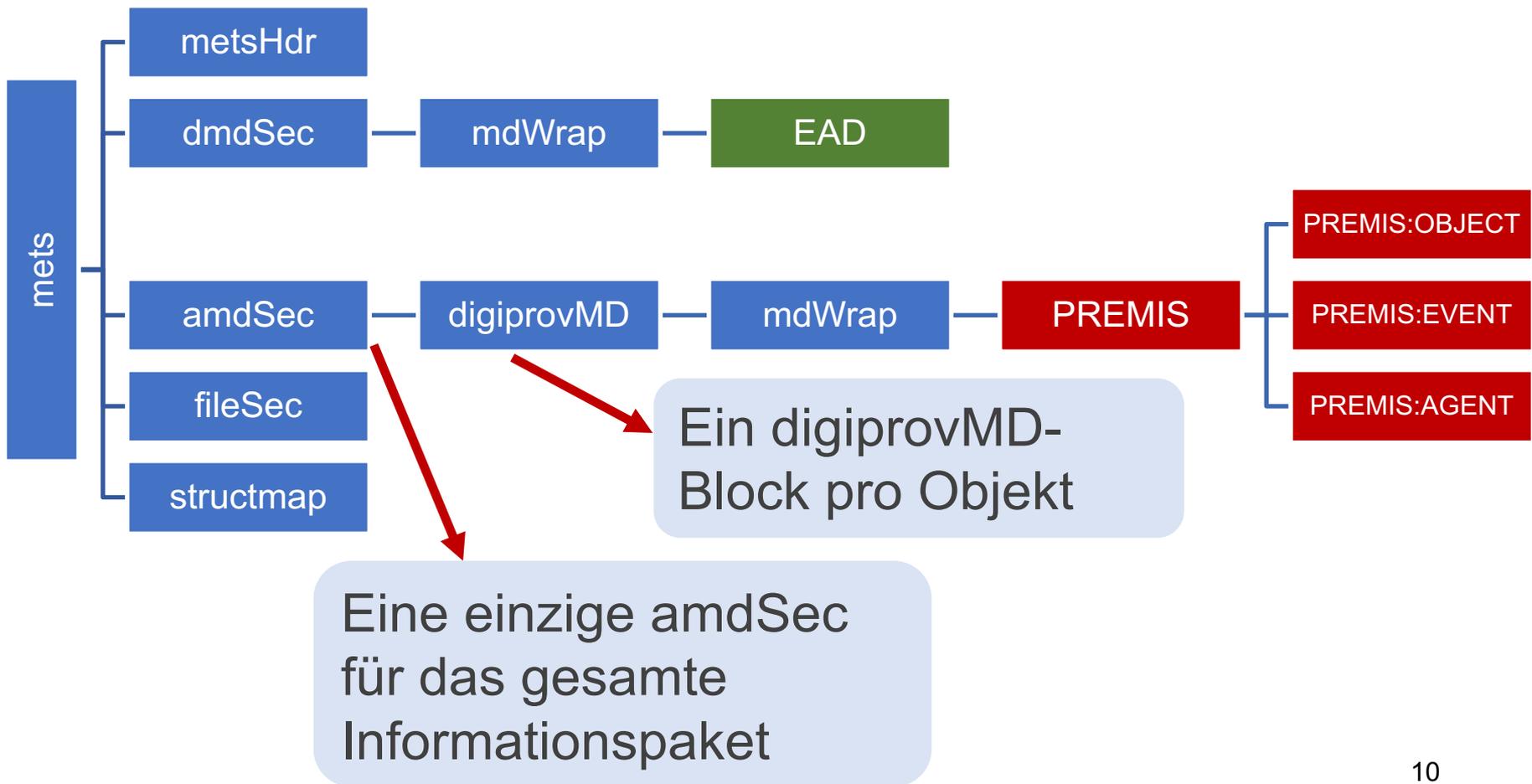
PREMIS in METS

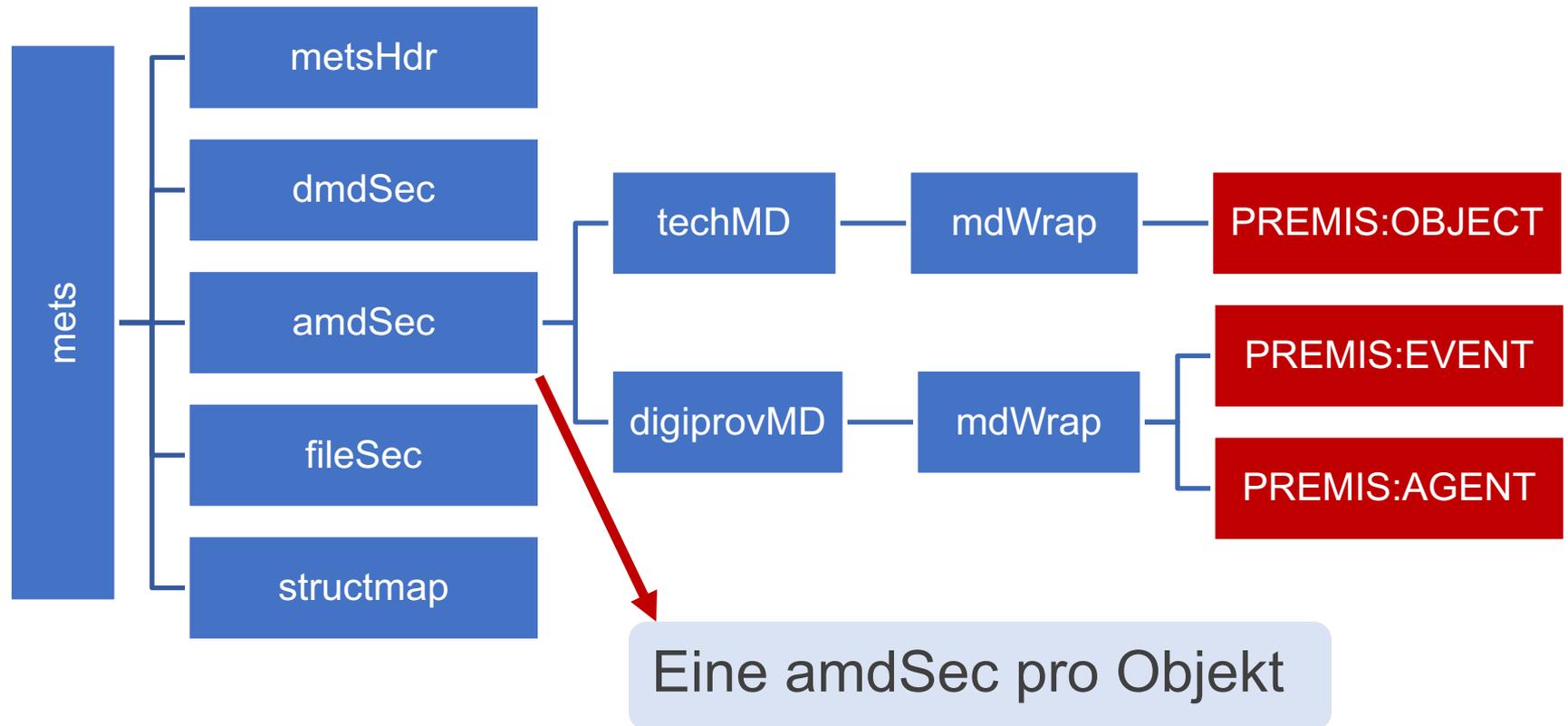
Library of Congress: Using PREMIS with METS

- Library of Congress: «Guidelines for using PREMIS with METS for exchange Revised September 17, 2008»
<https://www.loc.gov/standards/premis/guidelines-premismets.pdf>
- Library of Congress «Guidelines for using PREMIS with METS for exchange Revised January 2017»
<https://www.loc.gov/standards/premis/guidelines2017-premismets.pdf>

Folgende Punkte werden geregelt:

1. Using PREMIS in METS sections
2. Use of PREMIS container
«Use of one amdSec with repeating subelements (techMD, etc.) or repeating amdSec for each METS subelement is an implementer's preference. These are semantically equivalent»
3. Redundancies between PREMIS and METS
4. METS structMap and PREMIS structural relationship elements
5. METS ID/IDREF and PREMIS identifier elements
6. Use of METS profiles





Matterhorn METS: Eine amdSec für alle Objekte

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <METS:mets xmlns:METS="http://www.loc.gov/METS/" xmlns:EAD="urn:isbn:1-931666-22-9"
3   xmlns:PREMIS="info:lc/xmlns/premis-v2"
4   xmlns:OAI_DC="http://www.openarchives.org/OAI/2.0/oai_dc/"
5   xmlns:DC="http://purl.org/dc/elements/1.1/" xmlns:xlink="http://www.w3.org/1999/xlink"
6   xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
7   xsi:schemaLocation=" http://www.loc.gov/METS/ http://www.loc.gov/standards/mets/me
8   PROFILE="http://www.docuteam.ch/xmlns/sip-profile.xml" TYPE="sa_all-formats-01_dss-01">
9   <METS:metsHdr CREATEDATE="2018-06-03T08:16:53" LASTMODDATE="2018-06-03T08:17:03" [5 lines]
15  <METS:dmdSec ID="_20180603081653856"> [27 lines]
43  <METS:dmdSec ID="_20180603081653919"> [27 lines]
71  <METS:dmdSec ID="_20180603081653928"> [28 lines]
100 <METS:dmdSec ID="_20180603081653951"> [27 lines]
128 <METS:dmdSec ID="_20180603081653968"> [27 lines]
156 <METS:amdSec> [248 lines]
405 <METS:fileSec> [17 lines]
423 <METS:structMap> [30 lines]
454 </METS:mets>
---
```

Ein digiprovMD-Block pro Objekt

- PREMIS-Informationen zu einem Ordner oder einer Datei werden je in einer digiprovMD-Sektion zusammengefasst

```
156 <METS:amdSec>
157   <METS:digiprovMD ID="_20180603081653797"> [19 lines]
177   <METS:digiprovMD ID="_20180603081653843"> [30 lines]
208   <METS:digiprovMD ID="_20180603081653884">
209     <METS:mdWrap MDTYPE="PREMIS">
210       <METS:xmlData>
211         <PREMIS:premis version="2.2">
212           <PREMIS:object xsi:type="PREMIS:file"> [24 lines]
237           <PREMIS:event> [15 lines]
253         </PREMIS:premis>
254       </METS:xmlData>
255     </METS:mdWrap>
256   </METS:digiprovMD>
257   <METS:digiprovMD ID="_20180603081653924"> [48 lines]
306   <METS:digiprovMD ID="_20180603081653944"> [48 lines]
355   <METS:digiprovMD ID="_20180603081653954"> [48 lines]
404 </METS:amdSec>
```

Archivemata:

Eine amdSec pro Objekt

```
1 <?xml version='1.0' encoding='UTF-8'?>
2 <mets:mets xmlns:mets="http://www.loc.gov/METS/" xmlns:xlink="http://www.w3.org/1999/xlink" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://www.loc.gov/METS/ mets.xsd" >
3   <mets:metsHdr CREATEDATE="2018-04-12T12:03:16"/>
4   <mets:amdSec ID="amdSec_1"> [625 lines]
630   <mets:amdSec ID="amdSec_2"> [546 lines]
1177   <mets:amdSec ID="amdSec_3"> [445 lines]
1623   <mets:amdSec ID="amdSec_4"> [533 lines]
2157   <mets:amdSec ID="amdSec_5"> [470 lines]
2628   <mets:amdSec ID="amdSec_6"> [382 lines]
3011   <mets:amdSec ID="amdSec_7"> [470 lines]
3482   <mets:amdSec ID="amdSec_8"> [383 lines]
3866   <mets:amdSec ID="amdSec_9"> [186 lines]
4053   <mets:amdSec ID="amdSec_10"> [595 lines]
4649   <mets:amdSec ID="amdSec_11"> [445 lines]
5095   <mets:amdSec ID="amdSec_12"> [533 lines]
5629   <mets:amdSec ID="amdSec_13"> [497 lines]
6127   <mets:amdSec ID="amdSec_14"> [527 lines]
6655   <mets:amdSec ID="amdSec_15"> [546 lines]
7202   <mets:amdSec ID="amdSec_16"> [463 lines]
7666   <mets:amdSec ID="amdSec_17"> [464 lines]
8131   <mets:amdSec ID="amdSec_18"> [186 lines]
8318   <mets:amdSec ID="amdSec_19"> [560 lines]
8879   <mets:fileSec> [64 lines]
8944   <mets:structMap ID="structMap_1" LABEL="Archivemata default" TYPE="physical"> [76 lines]
9021   <mets:structMap ID="structMap_7a0f0769-86ee-4b20-b85e-13d955137004" LABEL="Hierarchical" TYPE="logical"> [58 lines]
9080 </mets:mets>
9081
```

Archivemataca: techMD und digiprovmD

- techMD für PREMIS:OBJECT
- digiprovmD für PREMIS:EVENT, PREMIS:AGENT

```
8131 ▾ <mets:amdSec ID="amdSec_18">
8132 ▶   <mets:techMD ID="techMD_18"> [43 lines]
8176 ▶   <mets:digiprovmD ID="digiprovmD_159"> [32 lines]
8209 ▶   <mets:digiprovmD ID="digiprovmD_160"> [32 lines]
8242 ▶   <mets:digiprovmD ID="digiprovmD_161"> [32 lines]
8275 ▶   <mets:digiprovmD ID="digiprovmD_162"> [13 lines]
8289 ▶   <mets:digiprovmD ID="digiprovmD_163"> [13 lines]
8303 ▶   <mets:digiprovmD ID="digiprovmD_164"> [13 lines]
8317 </mets:amdSec>
```

Redundanzen

LoC Implementation Guidelines:

«An implementer may decide whether it is easier to include the information redundantly, based on how the data will be used and/or supplied. Implementers should consider the use of the metadata (e.g. display or preservation) and whether the **METS (for display)** or **PREMIS (for preservation)** is primary when deciding which to use and whether to record redundantly.»

- Beispiel **Fixity Information** (Checksumme):
 - <METS:file CHECKSUM="[WERT]" CHECKSUMTYPE="[Wert]">
 - <PREMIS:fixity>
 - <messageDigestAlgorithm>
 - <messageDigest>
- Beispiel **Dateiformat**:
 - <METS:file MIMETYPE="[Wert]">
 - <PREMIS:formatDesignation>
 - <formatName>
 - <formatVersion>
- Entscheid bei **Matterhorn METS**:
 - Fixity in PREMIS
 - MIME-Type in METS (display), weiterführende Beschreibung von Dateiformat und Version in PREMIS (preservation).

Events

- Pro Event wird geschrieben:
 - eventIdentifier
 - eventType
 - eventDateTime
 - eventDetail
 - eventOutcomeInformation
 - eventOutcome eventOutcomeDetail
 - linkingAgentIdentifier
 - linkingAgentIdentifierType
(*eg. preservation system, repository code, Archivemata user*)
 - linkingAgentIdentifierValue
 - kein linkingObjectIdentifier, der Event bezieht sich jeweils auf das Object in der gleichen amdSec

- Pro Event wird geschrieben:
 - eventIdentifier
 - eventType (*standardisiertes Vokabular, abgestimmt auf LoC-Vokabular*)
 - eventDateTime (*Timestamp*)
 - eventDetail
(*hier auch textliche Beschreibung des Akteurs*)
 - eventOutcomeInformation
 - eventOutcome (success/failure)
 - eventOutcomeDetail
 - linkingObjectIdentifier
 - linkingObjectIdentifierType
 - linkingObjectIdentifierValue

- Die Event-Typen sind nicht Teil des METS-Profiles. Verwendet werden folgende Typen:
 - Creation
 - Rename
 - Fixity Check
 - Deletion
 - Path Modification
 - Migration
 - Replace
- Ausser “Replace” handelt es sich um Event Types aus dem Standardvokabular der Library of Congress:
<http://id.loc.gov/vocabulary/preservation/eventType.html>

```
<METS:mdWrap MDTYPE="PREMIS">
  <METS:xmlData>
    <PREMIS:premis version="2.2">
      <PREMIS:object xsi:type="PREMIS:file"> [24 lines]
        <PREMIS:event>
          <PREMIS:eventIdentifier> [3 lines]
          <PREMIS:eventType>Creation</PREMIS:eventType>
          <PREMIS:eventDateTime>2018-06-03T08:16:53</PREMIS:eventDateTime>
          <PREMIS:eventDetail>Performed by: 'tobias'</PREMIS:eventDetail>
          <PREMIS:eventOutcomeInformation>
            <PREMIS:eventOutcome>Success</PREMIS:eventOutcome>
          </PREMIS:eventOutcomeInformation>
          <PREMIS:linkingObjectIdentifier xlink:type="simple">
            <PREMIS:linkingObjectIdentifierType>Docu team</PREMIS:linkingObjectIdentifierType>
            <PREMIS:linkingObjectIdentifierValue>_20180603081653885</PREMIS:linkingObjectIdentifierValue>
          </PREMIS:linkingObjectIdentifier>
        </PREMIS:event>
      </PREMIS:premis>
    </METS:xmlData>
```

Vergleich und Fazit

Archivemata	Matterhorn METS
PREMIS 2.2	
PREMIS in METS eingebettet und nicht nur verlinkt	
dezentralisiert: <ul style="list-style-type: none">- Eine amdSec pro Objekt- Ein techMD-Block pro Objekt- Ein digiprovMD-Block pro Event, Agent	zentralisiert: <ul style="list-style-type: none">- Eine einzige amdSec pro Paket- Ein digiprovMD-Block pro Objekt
kein registriertes METS-Profil	METS-Profil: http://www.loc.gov/standards/mets/profiles/00000041.xml
Loglevel in Archivemata: Detailliert, möglichst umfassende technische Metadaten	Loglevel in docuteam feeder: Soviel wie nötig als Grundlage für Preservation Planning

- In den beiden Modellen werden praktisch identische Informationen kodiert.
- Matterhorn METS zentral, Archivemata dezentralisiert.
- LoC: «Use of one amdSec with repeating subelements (techMD, etc.) or repeating amdSec for each METS subelement is an implementer's preference. These are semantically equivalent»
- Dass unabhängig voneinander zwei Datenmodelle mit dieser hohen Ähnlichkeit entwickelt wurde zeigt, dass
 - die Standardisierungsdokumente hinreichend detailliert verfasst sind
 - die Implementation Guidelines der LoC praxisorientiert, umsetzbar und klar formuliert sind
- Am Vergleich von Archivemata und Matterhorn METS zeigt sich, dass ein Produktwechsel zwar syntaktische Änderungen an den API nach sich ziehen würde, die Semantik sich aber gleich bleibt.