

Paketering i METS: Metadata för digitaliserade monografier

Projektoberoende manual
KB - DIT Riksarkivet

Version 1.0
2020-06-17
(DOS-anpassad 2025-04-07)
KB 2019-986

Innehåll

Introduktion	4
Submission Information Package (SIP)	4
Datamodell	5
METS	7
Element och attribut i METS	7
<mets>	7
<metsHdr>	7
Descriptive Metadata Section <dmdSec>	8
Administrative Metadata Section <amdSec>	8
Technical Metadata <techMD>	9
File Section <fileSec>	9
Structure Map <structMap>	10
Exempel 1. Digitaliserade sidor av en volym + fil med ”performance data”. Bildfilerna otypade	11
MODS	13
Monografier och ”relaterade enheter”	13
<mods>	13
Metadata för monografin:	13
identifier	13
typeOfResource	13
genre	13
titleInfo	13
name	14
originInfo	14
physicalDescription	14
<relatedItem> = tryckt monografi som förlaga	14
identifier	15
originInfo	15
physicalDescription	15
location	15
<relatedItem> = projekt	16
genre	16
titleInfo	16
<relatedItem> = bilaga	16
genre	16
titleInfo	16
Lokala uppgifter om utgivare och leverantör	17
<mods>	17
name	17
role	17
PREMIS	18
PREMIS:OBJECT	18
1. object xsi:file=”representation”	18
1.1. objectIdentifier	18
1. object xsi:file=”file”	18
1.1. objectIdentifier	18
1.5. objectCharacteristics	19
1.5.1. compositionLevel	19
1.5.2. fixity	19

1.5.3. size.....	19
1.5.4. format	19
1.5.4.1. formatDesignation.....	19
1.5.4.2. formatRegistry.....	19
1.5.7. objectCharacteristicsExtension	19
MIX	20
BasicDigitalObjectInformation	20
BasicImageInformation	20
BasicImageCharacteristics	20
SpecialFormatCharacteristics.....	21
JPEG2000/CodecCompliance	21
JPEG2000/EncodingOptions.....	21
Image Capture Metadata	21
SourceInformation.....	21
GeneralCaptureInformation	22
ScannerCapture	22
DigitalCameraCapture.....	23
ImageAssessmentMetadata	23
SpatialMetrics.....	24
ImageColorEncoding	24
TargetData	24
Change History.....	24
ImageProcessing.....	25
Länkar/Referenser	26

Introduktion

Detta dokument innehåller en övergripande och generell redogörelse för hur digitala filer i digitalisering genomförd av KB och DIT (f.d. MKC) paketeras och beskrivs med hjälp av METS med flera standarder. Dokumentet är en anpassning för monografidigitalisering utifrån dokumentet för periodikadigitaliseringen.

Submission Information Package (SIP)

Resultatet av digitaliseringen paketeras i flera "leveranspaket" som Submission Information Package (SIP). Varje SIP kommer att motsvara en monografi, dvs. en SIP kommer att innehålla samtliga filer som utgör en digital avbildning av en monografisk resurs. En monografi definieras som en resurs bestående av en eller flera delar avsedd att utgöra en avslutad helhet. Exempel: Barabbas.

Varje SIP skapas enligt de specifikationer för monografidigitalisering som har tagits fram i samarbete mellan KB och DIT Riksarkivet. Specifikationerna utgår från KB:s metadataspecifikation för digitalisering av tidningar.

Den metadatastandard som används för paketering är [METS](#). Inom METS används för närvarande standarderna [PREMIS](#), [MODS](#) och [MIX](#). Standarden ALTO används också men inkluderas inte i METS utan registreras endast direkt i de ocr-lästa textfilerna (se dokument om användning av [ALTO](#)).

Datamodell

I nedanstående beskrivning av datamodellen utgås från ett exempel på digitalisering av en tryckt bok.

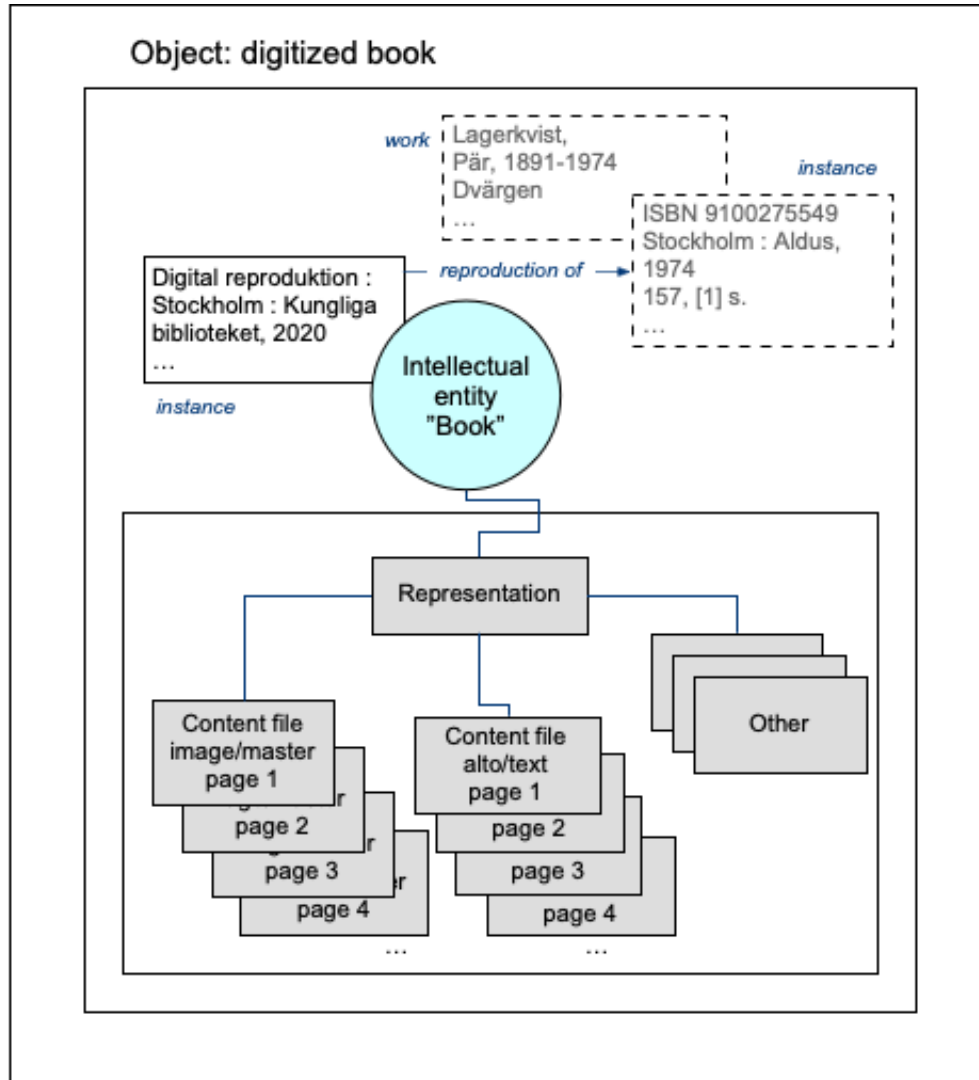


Bild 1

Bild 1 visar en digitaliserad tryckt bok, åskådliggjord med utgångspunkt från PREMIS datamodell.

Illustrationen exemplifierar förhållandet mellan PREMIS entiteterna *Intellectual entity* och *Objects* (files och representation).

En intellektuell entitet definieras i PREMIS som en samling av innehåll som kan behandlas och beskrivas som en intellektuell enhet. I det här fallet har vi den tryckta boken och dess digitala representation som båda enligt Bibframe-modellen är *instanser* av ett *verk*.

Metadata om dessa intellektuella entiteter, bibliografiska (beskrivande) metadata, registreras i METS-dokumentet i en eller flera [<dmdSec>](#).

Paketering i METS

Boken representeras av en uppsättning bildfiler (objekt). Varje sida i boken avbildas var för sig och sparas i olika format för olika syften, t.ex masterfiler för långtidslagring, ALTO-taggade filer för vidare textbearbetning, etc. Andra typer av objekt kan förekomma, t.ex. hela bokens innehåll sparad i en pdf-fil.

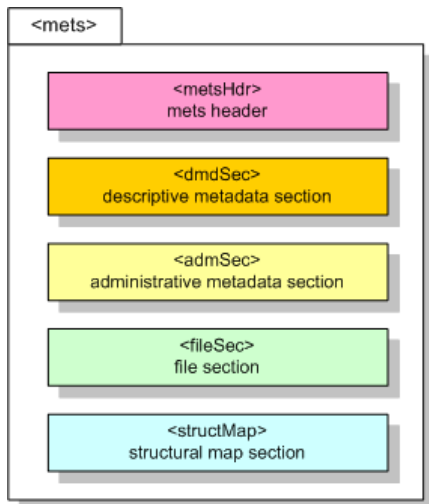
Administrativa (strukturella) metadata i METS [<admSec>](#) och [<structMap>](#) håller reda på vilka filerna är, deras egenskaper och relationer till varandra.

En "representation" i PREMIS är en annan typ av objekt. En representation utgör den samling filer och administrativa metadata som behövs för en komplett och begriplig återgivning av en intellektuell entitet.

METS

METS står för *Metadata Encoding and Transmission Standard* och används för att koda och packa ihop metadata (och data) för digitala objekt. METS är en flexibel standard, kodningen görs i XML och kan även inkludera andra metadatastandarder.

För varje resurs som paketeras i ett paket skrivs minst ett METS-dokument (en xml-fil med METS-koder). Ett METS-dokument kan delas in i upp till 7 sektioner, men i digitaliseringen kommer endast följande 5 sektioner att användas:



Element och attribut i METS

<mets>

Ett METS-dokument skapas för varje monografi. En monografisk resurs identifieras utifrån dess titel.

- För uppgifter om "xmlns" och "xsi:schemaLocation" Se även bilagorna [Tabell över element och attribut i METS](#) och [Tabell över ordlistor i METS](#).
- Ska innehålla attributen OBJID, PROFILE, TYPE och LABEL.
- OBJID är detsamma som paketets identifikator och måste vara unik i sitt sammanhang, dvs. inom DIT och i kontakten mellan DIT och KB.
- I PROFILE namnges den profil som mets-filen följer. Profilen för monografidigitalisering namnges i formen av en uri:
http://www.kb.se/namespace/mets/kbse_mets_profile_001.xml
LABEL ska innehålla monografins titel (title) från
dmdSec/@LABEL=Primary/.../mods:titleInfo/mods:title (plus ev. mods:subTitle).
Exempel på monografi: Dvärgen.

<metsHdr>

Innehåller information om METS-dokumentet i sig självt.

Ska innehålla: attributet CREATEDATE samt elementen metsDocumentID och agent.

- CREATEDATE ska innehålla datum och klockslag då METS-dokumentet skapades.
- RECORDSTATUS - här skrivs i förekommande fall något av värdena REPLACEMENT eller SUPPLEMENT om paketet ska antingen ersätta eller komplettera ett tidigare levererat paket.
- <metsDocumentID> innehåller METS-filens namn. Se även bilagan [Filnamngivning i digitalisering av monografier](#).
- Elementet <agent> upprepas för de institutioner som skapar respektive tar emot METS-dokumentet:
<agent ROLE="creator" TYPE="organization">
<name>Riksarkivet/DIT</name>
</agent>
<agent ROLE="archivist" TYPE="organization">
<name>Kungliga biblioteket</name>
</agent>

Descriptive Metadata Section <dmdSec>

Elementet utgör en behållare för beskrivande metadata om den intellektuella entitet som hör samman med de objekt som ingår i paketet. I det här fallet metadata om både den digitala reproduktionen och den tryckta förlagan av monografin. Beskrivande metadata kommer att hämtas från flera håll: Libris databas samt från digitaliseringsprocessen. Detaljer om vilka metadata som ska in här, se avsnittet om MODS.

- Det ska alltid förekomma en dmdSec med metadata i MODS för monografin. Denna dmdSec ska innehålla elementet mdWrap med attributet LABEL="Primary" enligt ordlistan vcDmdSec_LABEL_kbse.
- Det ska alltid förekomma en dmdSec med metadata i MODS för uppgifter om leverantör och utgivare. Denna dmdSec ska innehålla elementet mdWrap med attributet LABEL="Local", enligt ordlistan vcDmdSec_LABEL_kbse.
- Om tillhörande bilagor av monografin ska beskrivas i MODS ska dessa ha egna dmdSec.
- Varje dmdSec numreras löpande från dmdSec001, dmdSec002, etc. Dessa nummer registreras som värden i attributet ID. Dessa ID:n används som referenser inom METS-filen, t.ex. från <structMap>.
- Metadata i xml kapslas in i elementet <xmlData> under <mdWrap>.
- Attributet MDTYPE i <mdWrap> med värdet "MODS" är obligatoriskt för alla dmdSec.

Administrative Metadata Section <amdSec>

Elementet utgör en behållare för administrativa metadata om objekten (Fil eller Representation) och de händelser, agenter och rättigheter som förknippas med objekten. Med administrativa metadata menas t.ex. tekniska metadata från bildfångst, strukturella metadata som berättar om relationer mellan datafiler, vad som hänt med filerna i arbetsprocessen, restriktioner i användandet, etc.

- Varje METS-dokument innehåller en, och endast en, <amdSec>.
- Attributet ID är obligatoriskt med värdet "amdSec001".
- Innehåller flera element <techMD>.

Technical Metadata <techMD>

Elementet utgör en behållare för tekniska metadata om de datafiler som ingår i paketet. Metadata hämtas från processerna med bildfångst, OCR-läsning, ALTO-tagging m.m. Metadatastandard är PREMIS med tillägg av MIX. Detaljer om vilka metadata finns under avsnitten PREMIS respektive MIX.

- <techMD> upprepas för varje objekt (datafil).
- En <techMD> skapas också för representationen.
- I attributet ID numreras elementen enligt mönstret techMD001, techMD002, etc. Dessa ID används som referenser från <fileSec> och <structMap>.
- Metadata i xml kapslas in i elementet <xmlData> under <mdWrap>.
- Attributet MDTYPE i <mdWrap> med värdet "PREMIS:OBJECT" är obligatoriskt i alla techMD.

File Section <fileSec>

I fil-sektionen (fileSec) listas alla datafiler som ingår i paketet, med filnamn samt uppgifter om storlek, checksumma m.m.

En viss dubbling av information sker således här eftersom några uppgifter även förekommer som PREMIS-metadata i <techMD>.

Här refereras även till metadata i <techMD>.

- METS-dokumentet ska innehålla endast ett, <fileSec>-element.
- Attributet ID med värdet fileSec001 är obligatoriskt.
- Ska innehålla minst ett <fileGrp>-element, ett för varje filtyp som ingår i paketet. Dessa numreras enligt mönstret fileGrp001, fileGrp002, etc.

Varje <fileGrp> ska innehålla attributet USE som namnger gruppen enligt ordlista "vcUSE_kbse". Se bilagorna [Tabell över element och attribut i METS](#) och [Tabell över ordlistor i METS](#).

Exempel: värdet "image/master" anger att elementet innehåller referenser till bildfiler avsedda för arkivering.

- Elementet <file>, som underordnas <fileGrp>, upprepas för varje fil som ingår i paketet. Ett antal attribut är här obligatoriska:
 - ID - innehåller ett numeriskt värde, löpande från 1, föregånget av prefixet "file".
Exempel "file1", "file2", "file3", etc.
 - USE - upprepa det värde som anges i tillhörande <fileGrp>.
 - MIMETYPE - ange värde enligt <http://www.iana.org/assignments/media-types/>.
 - SIZE - ange samma värde som i premis <size> i motsvarande mets <techMD>.
 - CREATED - ange när filen skapades. Upprepa tidsangivelse t.ex. från mix <dateTimeCreated>. Skrivs som YYYY-MM-DD'TH:mm:ss+01:00.
 - ADMID - ange ID-värdet för den <techMD> som innehåller metadata om filen, t.ex. "techMD002"
 - CHECKSUM - ange samma värde som i premis <messageDigest> i motsvarande mets <techMD>. Om flera checksummor förekommer i premis, ange den senaste.
 - CHECKSUMTYPE - ange värdet "MD5", dvs. samma värde som i premis <messageDigestAlgorithm> i motsvarande mets <techMD>.
- Under <file> ingår elementet <Flocat> med en referens till filens namn. Obligatoriska attribut är:

- LOCTYPE - alltid med värdet "URL"
- xlink:type - alltid med värdet "simple"
- xlink:href - ange filens namn föregånget av prefixet "file:", exempel "file:Bib3j9bxjbg5k43n879_dig2020_0008.jp2". Se även bilagan [Filnamngivning i digitalisering av monografier](#).

Structure Map <structMap>

En "struktur-karta" beskriver hur datafilerna förhåller sig till varandra så att ett system kan rekonstruera och tillhandahålla de digitala objekten.

- METS-dokumentet ska innehålla en <structMap> som beskriver den "fysiska" strukturen.
- Attributet ID med värdet "structMap001" är obligatoriskt.
- Attributet TYPE ska vara "physical" (värde från vcStructMap_TYPE_kbse).
- <structMap> delas in i ett antal <div> som i sin tur kan delas in i en eller flera <div>. På så vis bildas en hierarkisk struktur som visar hur datafilerna relateras till varandra.
- Obligatoriska attribut i <div> är:
 - ID - löpande värden enligt mönstret div001, div002, etc.
 - TYPE – Obligatoriskt. Varje div ska namnges i TYPE med ett värde från ordlistan vcDiv_TYPE_kbse. Se exempel och förslag på struktur nedan.
 - LABEL - Kan användas om t.ex. en sida saknas och har ersatts av en "place holder" med förklarande text. Värden från vcDiv_LABEL_kbse.
 - ORDER - den ordning som filen ska visas i. Värden: 1, 2, 3, etc.
 - ORDERLABEL - Frivilligt. Textfält för sidnumrering som den förekommer i resursen.
 - DMDID - ange ID-värdet för den <dmdSec> som innehåller bibliografiska metadata om det som div-elementet refererar till. Exempel: För <div TYPE="monograph"> ange ID för den dmdSec som har LABEL="Primary". För div TYPE="volume" ange vid behov ID för den dmdSec som beskriver volymen (t ex en bilaga), etc.
- En <fptr> under <div> upprepas för varje datafil. I attributet FILEID samma värde som i motsvarande file@ID. Exempel "file1", "file2", "file3", etc.

Strukturmappen delas in så här med hjälp av <div> och attributet TYPE:

Exempel 1. Digitaliserade sidor av en volym + fil med "performance data". Bildfilerna otypade:

```
<structMap TYPE="physical">
  <div TYPE="files">Fungerar som en "wrapper" runt de andra
    <div TYPE="monograph">Högsta nivån för monografin
      <div TYPE="volume">Nivå för fysisk volym
        <div ORDER="1" TYPE="undefined">
          <fptr FILEID="">Referens till jpg-fil för bild 1
          <fptr FILEID="">Referens till alto-fil för bild 1
        </div>
        <div ORDER="2" TYPE="undefined">
          <fptr FILEID="">Referens till jpg-fil för bild 2
          <fptr FILEID="">Referens till alto-fil för bild 2
        </div>          etc.
        <div ORDER="3" TYPE="undefined">Här listas referenser till filerna för bild 3
          <fptr FILEID="">Referens till jpg-fil för bild 3
          <fptr FILEID="">Referens till alto-fil för bild 3
        </div>
        <div ORDER="4" TYPE="undefined">Här listas referenser till filerna för bild 4
          <fptr FILEID="">Referens till jpg-fil för bild 4
          <fptr FILEID="">Referens till alto-fil för bild 4
        </div>          etc.
      </div>
    </div>
    <div TYPE="performance">På samma nivå som "monograph"
      <fptr FILEID="">
    </div>
  </div>
</structMap>
```

Exempel 2. Samma som exempel 1 men med typade bildfiler

```
<structMap TYPE="physical">
  <div TYPE="files">Fungerar som en "wrapper" runt de andra
    <div TYPE="monograph">Högsta nivån för monografin
      <div TYPE="volume">Nivå för fysisk volym
        <div TYPE="[term för pärmens framsida]" ORDER="1">
          <fptr FILEID="">Referens till jpg-fil för pärmens framsida
          <fptr FILEID="">Referens till alto-fil för pärmens framsida
        </div>
        <div TYPE="[term för pärmens baksida]" ORDER="2">
          <fptr FILEID="">Referens till jpg-fil för pärmens baksida
          <fptr FILEID="">Referens till alto-fil för pärmens baksida
        </div>          etc.
        <div TYPE="page" ORDER="3">Här listas referenser till filerna för sidan 1
          <fptr FILEID="">Referens till jpg-fil för sidan 1
          <fptr FILEID="">Referens till alto-fil för sidan 1
        </div>
        <div TYPE="page" ORDER="4">Här listas referenser till filerna för sidan 2
          <fptr FILEID="">Referens till jpg-fil för sidan 2
          <fptr FILEID="">Referens till alto-fil för sidan 2
        </div>          etc.
      </div>
    </div>
    <div TYPE="performance">På samma nivå som "monograph"
      <fptr FILEID="">
    </div>
  </div>
</structMap>
```

Exempel 3. En monografi i två volymer. Bildfilerna otypade. Genomgående ordningsföljd.

```

<structMap TYPE="physical">
  <div TYPE="files">Fungerar som en "wrapper" runt de andra
    <div TYPE="monograph">Högsta nivån för monografin
      <div TYPE="volume">Nivå för fysisk volym (den första av två)
        <div ORDER="1" TYPE="undefined">Här listas referenser till filerna för bild 1
          <fptr FILEID="">Referens till jpg-fil för bild 1
          <fptr FILEID="">Referens till alto-fil för bild 1
        </div>
        <div ORDER="2" TYPE="undefined">Här listas referenser till filerna för bild 2
          <fptr FILEID="">Referens till jpg-fil för bild 2
          <fptr FILEID="">Referens till alto-fil för bild 2
        </div>
        etc.
        <div ORDER="3" TYPE="undefined">Här listas referenser till filerna för bild 3
          <fptr FILEID="">Referens till jpg-fil för bild 3
          <fptr FILEID="">Referens till alto-fil för bild 3
        </div>
        <div ORDER="4" TYPE="undefined">Här listas referenser till filerna för bild 4
          <fptr FILEID="">Referens till jpg-fil för bild 4
          <fptr FILEID="">Referens till alto-fil för bild 4
        </div>
        etc.
      </div>
      <div TYPE="volume">Nivå för fysisk volym (den andra av två)
        <div ORDER="100" TYPE="undefined">Här referenser till filerna för bild 100
          <fptr FILEID="">Referens till jpg-fil för bild 100
          <fptr FILEID="">Referens till alto-fil för bild 100
        </div>
        <div ORDER="101" TYPE="undefined">Här referenser till filerna för bild 101
          <fptr FILEID="">Referens till jpg-fil för bild 101
          <fptr FILEID="">Referens till alto-fil för bild 101
        </div>
        etc.
      </div>
    </div>
  </div>
  <div TYPE="performance">På samma nivå som "monograph"
    <fptr FILEID="">
  </div>
</div>
</structMap>

```

MODS

MODS (Metadata Object Description Schema) är ett xml-schema för bibliografiska metadata som kan användas för att beskriva många olika typer av objekt. Det är framförallt inom biblioteken man träffar på MODS. Det används ofta som ett "utbytesformat" till och från katalogposter i det traditionella biblioteksformatet MARC21, men även för att skapa nya beskrivningar av objekt i bibliotekens samlingar. Ansvariga för innehåll och utveckling av standarden är Network Development and MARC Standards Office vid Library of Congress.

Monografier och "relaterade enheter"

Registreras i `dmdSec/@ID="dmdSec001"/mdWrap/@LABEL="Primary"`

MODS-avsnittet inleds med metadata om den digitala versionen av monografin tillsammans med information om att paketet innehåller en digital version av ett tryckt exemplar.

Till beskrivningen hör även att registrera ett antal "relaterade enheter". Detta görs i behållaren `<relatedItem>` som upprepas för varje typ av relation. Vissa `<relatedItem>` är obligatoriska, andra kan tillkomma vid behov.

Element och attribut i MODS

`<mods>`

`<mods>` registreras under `mets/dmdSec/mdWrap/@MDTYPE="MODS"/xmlData`.

Metadata för monografin:

identifier

- Här en identifikator som tilldelats resursen vid reproduktionstillfället
- med attribut `type` (värde "local" om lokalt id).

typeOfResource

- med värdet "text".

genre

- med värdet "book".
- med attribut `authority="marcgt"` (står för marc genre term)

titleInfo

Huvudtitel och undertitel för monografin - samma som i `<mets@LABEL>`, se `<mets>`
Inkluderar elementen:

- `title`
 - med titel för monografin
 - avser framför allt huvudtitel men om det är opraktiskt att ange undertitel i separat element kan även undertitel anges här med kolon som skiljetecken, t.ex. "Hemsöborna : skärgårdsberättelse"
- `subTitle`

- med undertitel för monografin

name

Namn på person(er) som är upphov till monografin.

- med attribut type="personal"

Inkluderar elementen:

- namePart
 - med attributet type="family" – Här skrivs efternamn
 - med attributet type="given" – Här skrivs förnamn
- role
 - Inkluderar elementet:
 - roleTerm, som inkluderar:
 - attribut type="code"
 - authority="marcrelator"
 - med värdet "aut" enligt MARC Code List for Relators
<http://www.loc.gov/marc/relators/relacode.html>

originInfo

Innehåller information om när den tryckta monografin digitaliserades.

Inkluderar elementen:

- dateIssued - Här skrivs ett årtal enligt mönstret yyyy, t.ex. 2020.
 - med attributet encoding="w3cdtf".

physicalDescription

Inkluderar elementen:

- digitalOrigin:
 - "reformatted digital" om digitaliseringen är gjord från tryckt förlaga.
- note
 - med attributet type="reproduction".
Här i klartext att detta är en digital reproduktion, vem som är "utgivare" av denna, samt ett årtal då reproduktionen ägde rum.
Skrivs i formatet: Digital reproduktion: [Namn på ort där utgivaren har sin hemvist]: [Namn på utgivare], [årtal i formatet YYYY].
 - med attributet type="script".
Välj värde enligt lista (vcsScriptType_kbse).

<relatedItem> = tryckt monografi som förlaga

- med attribut TYPE="original"
Obligatoriskt om tillämpligt. Här ryms uppgifter om förlagan (tryckt monografi) till den digitala reproduktionen.

Inkluderar elementen:

identifier

Här anges den URI som identifierar den tryckta förlagan i Libris databas, t.ex. <https://libris.kb.se/0h98x84b3dk2gdx>. Om den tryckta monografin har ett ISBN-nummer anges detta i ett eget <identifier>.

- med attribut type="uri" för Libris-URI.
- <identifier> upprepas eventuellt med ISBN (attribut type="isbn")

originInfo

Inkluderar elementet:

- dateIssued
 - Värde: årtal i formen yyyy
 - med attributen encoding="w3cdtf"

physicalDescription

Inkluderar:

- form
 - med attribut authority="marcform" och värde: print.

location

Inkluderar:

- physicalLocation
För KB:s exemplar registreras här "S-SE".
- holdingSimple, inkluderar:
 - copyInformation, som inkluderar:
 - shelfLocator
 - note, utan attribut.
Här finns plats för noteringar om förlagan. Kan upprepas vid behov.
Uppgifter om skador på enskilda sidor sparas i ALTO under attributet QUALITY_DETAIL.
Alltid en note med värdet S-A om det är ett nationalexemplar.
 - note, med attribut type="condition".
Notering om förlagans fysiska kondition.

Exempel:

```
<mods:relatedItem type="original">
  <mods:location>
    <mods:physicalLocation>S-SE</mods:physicalLocation>
    <mods:holdingSimple>
      <mods:copyInformation>
        <mods:shelfLocator>h57 223 ( Ex. A)</mods:shelfLocator>
        <mods:note>S-A</mods:note>
      </mods:copyInformation>
    </mods:holdingSimple>
  </mods:location>
```

<relatedItem> = projekt

- med attributet type="host"

Inkluderar elementen:

genre

- med värde "project"

titleInfo

Inkluderar elementen:

- title, med projektets titel i klartext

Koden ser ut så här i samtliga fall:

```
<mods:relatedItem type="host">
  <mods:genre>project</mods:genre>
  <mods:titleInfo>
    <mods:title>[Namn på projekt]</mods:title>
  </mods:titleInfo>
</mods:relatedItem>
```

<relatedItem> = bilaga

- med attribut type="constituent">

<relatedItem TYPE="constituent"> registreras under en egen dmdSec:
dmdSec/mdWrap/@MDTYPE="MODS"/xmlData.

Här registreras information om delar som ingår i monografin och som behöver beskrivas för sig, t.ex. bilagor.

Inkluderar elementen:

genre

Värde i elementet väljs enligt genreterm_kbse: supplement

titleInfo

Inkluderar elementen:

- partName
 - Namn och eventuella delbeteckningar.

Lokala uppgifter om utgivare och leverantör

Hit bl.a. namn och identifikator i Mimers leverantörsregister för leverantör respektive utgivare av levererad resurs.

Registreras i `dmdSec/@ID="dmdSec002"/mdWrap/@LABEL="Local"`

Element och attribut i MODS

<mods>

`<mods>` registreras under `mets/dmdSec/mdWrap/@MDTYPE="mods"/xmlData`.

Inkluderar elementen `<name>` och `<role>` som upprepas för leverantör respektive utgivare::

name

- med attributen `TYPE="corporate"`, `AUTHORITY="local"` samt `valueURI="[KB:s id för organisationen]"`

Inkluderar:

- `namePart`

Värden för `namePart`:

- `"Kungliga biblioteket"`
- `"Riksarkivet/DIT"`

Värden för `valueURI`:

- `"http://id.kb.se/organisations/SE2021001710"`
- `"http://id.kb.se/organisations/SE2021001074-MKC"`

role

Inkluderar:

- `roleTerm` med attributen `type="text"` samt `authority=marcrelator` eller `local`
 - .. värden:
 - `"supplier"` (`authority=local`)
 - `"publisher"` (`authority=marcrelator`)

PREMIS

PREMIS står för Preservation Metadata: Implementation Strategies. Det är en standard framtagen för att effektivt hantera, finna och återskapa digital information. PREMIS innehåller bl.a. information om de digitala objektens tekniska egenskaper, proveniens, aktiviteter kring objekten, eventuella restriktioner i åtkomst och användande, etc.

Library of Congress tillhandahåller scheman i XML för PREMIS och det finns även en svensk version som Riksarkivet ansvarar för.

PREMIS-modellens entiteter Object, Event, Agent och Rights, motsvaras i schemat av metadatabehållare med samma namn.

I projekt kommer DIT att arbeta med <object> (se [PREMIS:OBJECT](#)) där varje digitalt objekts grunddata fångas.

PREMIS:OBJECT

Två olika typer av "object" registreras i paketen:

1. Ett <object> som avser representationen. Attributet xsi:file="representation"
2. Ett <object> upprepas för varje datafil som ingår i paketet. Attributet xsi:file="file"

<object> registreras under

mets/amdSec/techMD/mdWrap/@MDTYPE="PREMIS:OBJECT"/xmlData.

Upprepas för varje fil.

1. object xsi:file="representation"

Här samlas metadata som rör hela representationen och som man inte vill upprepa för varje datafil.

- Obligatoriskt attribut i <object> är xsi:file med värdet "representation".
- Obligatoriskt under <object> är:
 - 1.1. objectIdentifier

1.1. objectIdentifier

Element under <object> som identifierar representationen. Här finns två underelement:

- objectIdentifierType - Ange här typ av identifikator, värde: "local".
- objectIdentifierValue - Ange här samma identifikator som i mets/@OBJID.

1. object xsi:file="file"

Här samlas metadata som rör en fil.

- Obligatoriskt attribut är xsi:file med värdet "file".
- Obligatoriska element/behållare under <object> är:
 - 1.1. objectIdentifier
 - 1.5. objectCharacteristics

1.1. objectIdentifier

Identifierar objektet (filen). Innehåller:

- objectIdentifierType - Ange värde "filepath".

- objectIdentifierValue - Ange här filens namn. Se även bilagan [Filnamngivning i digitalisering av monografier](#).

1.5. objectCharacteristics

Innehåller:

- 1.5.1. compositionLevel
- 1.5.2. fixity
- 1.5.3. size
- 1.5.4. format
- 1.5.7. objectCharacteristicsExtension

1.5.1. compositionLevel

Värde alltid "0" (noll) för okomprimerad fil.

1.5.2. fixity

Information om checksummeberäkning. Uppgift om aktuell checksumma och algoritm upprepas i mets/fileSec/fileGrp/file/@CHECKSUMTYPE respektive /@CHECKSUM.

Innehåller:

- messageDigestAlgorithm MD5
- messageDigest - den beräknade checksumman
- messageDigestOriginator - t.ex. "DIT Riksarkivet".

1.5.3. size

Här anges storlek på filen uttryckt med ett tal i byte. Exempel: "3120". Uppgiften upprepas i mets/fileSec/fileGrp/file@SIZE.

1.5.4. format

Information om formatet. Innehåller:

- 1.5.4.1. formatDesignation
- 1.5.4.2. formatRegistry

1.5.4.1. formatDesignation

Information om filformat och version.

- formatName - Namn på filformat. Värden hämtas från PRONOM.
- formatVersion - Versionsnummer på filformatet enligt PRONOM.

1.5.4.2. formatRegistry

Information om formatet hämtat från PRONOM. Innehåller:

- formatRegistryName - alltid värdet "PRONOM".
- formatRegistryKey - ange identifikator/nyckel till format och formatversion i PRONOM.
- formatRegistryRole - alltid värde "specification".

1.5.7. objectCharacteristicsExtension

Behållare för att registrera objektkaraktäristiska uppgifter med hjälp av andra metadatastandarder. Se vidare under MIX.

MIX

MIX är en metadatastandard för att beskriva tekniska metadata för bildfiler. Den är en översättning till XML av en annan standard, ANSI/NISO Z39.87-2006.

Metadata som beskrivs med MIX lagras i ett PREMIS-objekt under taggen `<objectCharacteristicsExtension>`. Varje enskild bildfil ska ha sina metadata sparade under ett eget PREMIS-objekt.

Det finns en hel del överlapp mellan PREMIS och MIX, en del av de taggar som används inom MIX hämtas t.o.m. direkt från PREMIS. I de fall där något kan beskrivas både med PREMIS och med MIX kan vi valt att enbart använda PREMIS. Om vi istället valt att använda MIX hade vi normalt ändå varit tvungna att använda element från PREMIS, fast med skillnaden att elementen bäddats in i MIX-delen. Exempel på data som kan lagras i både MIX och PREMIS är grundläggande filinformation (filformat etc.) och checksummor.

BasicDigitalObjectInformation

Här lagras information av allmän karaktär som är applicerbar alla typer av digitala filer och inte enbart bildfiler. Mycket av informationen är redundant och spara på annan plats i schemat.

compression

Om bildfilen är komprimerad anges det i MIX eftersom PREMIS saknar stöd för detta.

Innehåller:

- `compressionScheme`. Värdet från ANSI/NISO Z39.87-2006
- `compressionSchemeLocalList`. Anges endast om egen ordlista skapas
- `compressionSchemeLocalValue`. Anges endast om egen ordlista skapas
- `compressionRatio`. Ska enligt standarden anges som ett positivt heltal men vi tillåter även flyttal då kompressionsgraden i JPEG2000 inte är begränsad till heltal.

BasicImageInformation

Denna sektion innehåller grundläggande metadata om bildfilen.

BasicImageCharacteristics

Metadata som inte relaterar till ett specifikt filformat.

Imagewidth/imageHeight

Bredd och höjd på bilden i pixlar

Alla metadata som relateras till färgrymd sparas i MIX. Det är möjligt att även lagra versionsnumret för en färgrymd i namnelementet (t.ex. AdobeRGB 1998) men detta gör det svårare att söka på versionsnummer. Om färgrymden inte kan anses som välkänd (t.ex. eciRGBv2) bör en URI lagras för att förenkla informationsinhämtning. MIX har även stöd för lokala färgprofiler men vi anser att sådana har en negativ påverkan på långtidsbeständigheten och att de därför inte ska användas.

PhotometricInterpretation

Innehåller:

- colorSpace

PhotometricInterpretation/colorProfile/iccProfile

Innehåller:

- iccProfileName
- iccProfileVersion
- iccProfileURI (Obs! notera att taggen slutar på URI, ej URL som det står i ANSI/NISO Z39.87-2006).

SpecialFormatCharacteristics

Metatdata som relaterar till ett specifikt filformat

MIX har stöd för JPEG2000. Fälten för codestreamProfile och complianceClass används inte då vi inte ser någon användning för dessa metatdata. Endast en mindre mängd av de parametrar som är av vikt för JPEG2000 kan anges i dedikerat element. Tyvärr saknas ett fritextfält där det är möjligt att lagra samtliga parametrar som används vid konvertering. Dessa ska därför lagras i de metadatafält som finns inuti bildfilen.

JPEG2000/CodecCompliance

Innehåller:

- codec
- codecVersion

JPEG2000/EncodingOptions

Innehåller:

- tiles
- qualityLayers
- resolutionLevels

Image Capture Metadata

Denna sektion innehåller tekniska metatdata om bildfångsten.

SourceInformation

Metatdata som härrör från det bildfångade (analog) objektet

Objektets orientering kan anges i MIX vilket gör det enklare att visa upp bilden på rätt ledd.

Innehåller:

- orientation

Det är möjligt att i MIX ange vilken typ av förlaga som digitaliserats. Denna information lagras normalt i MODS. Vi kan dock se ytterst sällsynta fall där majoriteten av bildfilerna har en viss fysisk förlaga och en enskild bildfil har en annan. I de fall då bildfilerna har skilda

förlagor ska sourceType användas för att notera förlagans typ för avvikande bildfiler. Fältet sourceID används för att tala om vilken fysisk förlaga som digitaliserats. sourceID formateras på samma sätt som motsvarande materialkategori i MODS. För materialtyper som ej täcks in kan även de kategorier som finns i ANSI/NISO Z39.87-2006 användas. Om detta blir aktuellt kommer vi att behöva skapa en ordlista för dessa fält.

SourceID

Innehåller:

- sourceType
- sourceID

Den fysiska förlagans storlek kan anges i MIX. Denna information är redundant om samplingsfrekvensen lagras eftersom objektets storlek då kan räknas fram utifrån bildens storlek och samplingsfrekvensen. Om utrustningen automatiskt kan mäta den fysiska storleken samtidigt som samplingsfrekvensen är okänd ska storleken anges (i millimeter). I alla övriga fall kan detta element lämnas tomt.

SourceSize/sourceX(Y)Dimension

Innehåller:

- sourceX(Y)DimensionValue
- sourceX(Y)DimensionUnit

GeneralCaptureInformation

Metadata om bildfångst som är oberoende av bildfångstutrustning.

Tid för bildfångst och den använda bildfångstutrustningen ska lagras. <imageProducer> lämnas normalt tomt då denna information finns i lagrad i MODS. På samma sätt som en enskild bild kan ha en annan fysisk förlaga än majoriteten av bilderna kan den dock även ha ett annat digitalt ursprung. I dessa fall ska bildens producent anges.

Innehåller:

- dateTimeCreated
- captureDevice
- imageProducer

ScannerCapture

Metadata om bildfångst och teknisk utrustning (skanner).

Vi ser ingen användning av fälten maximumOpticalResolution eller scannerSensor varför dessa inte inkluderas. Vi rekommenderar att skannerns serienummer inkluderas för att förenkla kvalitetssäkring om flera skannrar av samma modell används. För att underlätta sökningar ska skannermjukvara och mjukvaruversion anges i separata fält.

Innehåller:

- scannerManufacturer

ScannerModel

Innehåller:

- scannerModelName

- scannerModelNumber
- scannerModelSerialNo

ScanningSystemSoftware

Innehåller:

- scanningSystemSoftwareName
- scanningSystemSoftwareVersionNo

DigitalCameraCapture

Metadata om bildfångst och teknisk utrustning (digitalkamera).

Vi ser ingen användning av fältet cameraSensor som därför inte inkluderas. Vi rekommenderar att kamerans serienummer inkluderas för att förenkla kvalitetssäkring om flera kameror av samma modell används. Stöd för att ange mjukvara som använts för bildfångst saknas i MIX.

Innehåller:

- digitalCameraManufacturer

ScannerModel

Innehåller:

- digitalCameraModelName
- digitalCameraModelNumber
- digitalCameraModelSerialNo

Ett urval av de metadata som härrör från själva bildfångsten med kamera har gjorts. Metadata som kan vara användbar eller efterfrågas har inkluderats medan alla andra fält (t.ex. OECF och spectralSensitivity) inte tagits med i specifikationen. En del fält i standarden är dessutom duplicerade eftersom de beskriver samma sak men använder olika enheter och format (t.ex. ½ s och 0,5 s för slutartid).

CameraCaptureSettings/ImageData

Innehåller:

- fNumber
- exposureTime
- isoSpeedRatings
- exifVersion
- exposureBiasValue
- lightsource
- focalLength
- autoFocus

ImageAssessmentMetadata

Metadata som behandlar kvalitén av avbildningen.

SpatialMetrics

Om utrustningen kan mäta den praktiska samplingsfrekvensen ska denna anges och ha företräde framför förlagans fysiska storlek (se ovan). Det är viktigt att tänka på att samplingsfrekvensen förändras med avståndet till objektet.

samplingFrequencyUnit

Innehåller:

- xSamplingFrequency
- ySamplingFrequency

ImageColorEncoding

Bildens färgdjup ska anges både med det totala antalet bitar och med antalet bitar i varje färgkanal. Det är möjligt att ange mer avancerad färginformation men vi ser ingen nytta med detta.

Innehåller:

- samplesPerPixel

bitsPerSample

Innehåller:

- bitsPerSampleValue
- bitsPerSampleUnit

TargetData

Bildens kvalitet mäts med hjälp av mål och för att möjliggöra en efterkontroll bör information från kvalitetsmätning inkluderas i arkivpaketet. Detta görs genom att en resultatfil sparas i arkivpaketet. Det finns möjlighet att peka till resultatfilen från MIX men för att skapa enhetliga paket görs detta istället från fileSec/structMap. Information om det använda målet ska alltid sparas. Ett externt mål bildfångas ensamt medan ett internt mål syns på samma bild som objektet. Identifierare för enskilda mål bör inkluderas i metadatan. För mer information om kvalitetsmätning, se separat dokumentation

Innehåller:

- targetType ("external"/"internal", ej 0/1 som det står i ANSI/NISO Z39.87-2006)

TargetID

Innehåller:

- targetManufacturer
- targetName
- targetNo

Change History

Om bildfilen bearbetats ska detta anges under ChangeHistory. Det är viktigt att både beskriva vilken mjukvara som används, samtliga parametrar och syftet med efterbearbetningen. Syftet ska dock enbart anges om det avviker från normal bearbetning. Endast den senaste bearbetningen sparas under elementet <ImageProcessing>. Tidigare bearbetningar flyttas till

ett block under Previous Image Metadata. Om flera bearbetningar utförs ska de lagras kronologisk (den senaste överst i Previous Image Metadata och den första sista). Information som sparats under Previous Image Metadata får aldrig raderas.

ImageProcessing

Innehåller:

- dateTimeProcessed
- processingAgency
- processingRationale (förklarande text vid bildbearbetning som ej utförs regelmässig)
- processingActions

processingSoftware

Innehåller:

- processingSoftwareName
- processingSoftwareVersion

Länkar/Referenser

PREMIS, <http://www.loc.gov/standards/premis/>

MODS, <http://www.loc.gov/standards/mods/>

METS, <http://www.loc.gov/standards/mets/>

MIX, <http://www.loc.gov/standards/mix/>

ALTO, <http://www.loc.gov/standards/alto/>

KB:s profiler och scheman för paketering i METS, <http://www.kb.se/namespace/mets/>