

# Paketering i METS: Metadata för digitaliserade dagstidningar

Projektberoende manual  
KB - MKC Riksarkivet

Version 1.2  
2019-06-13  
(Med anvisningar för tidskrifter)  
KB 2019-515

## Innehåll:

Introduktion .....	3
Submission Information Package (SIP) .....	3
Datamodell .....	4
METS .....	6
Element och attribut i METS .....	6
<mets> .....	6
<metsHdr> .....	7
Descriptive Metadata Section <dmdSec> .....	7
Administrative Metadata Section <amdSec> .....	7
File Section <fileSec> .....	8
Structure Map <structMap> .....	9
MODS .....	12
Periodikanummer och “relaterade enheter” .....	12
Metadata för periodikanumret: .....	12
<relatedItem> = Värdepublikation .....	13
<relatedItem> = Projekt .....	15
<relatedItem> = mikrofilm som förlaga .....	15
<relatedItem> = tryckt periodika som förlaga .....	16
<relatedItem> sektion, bilaga, eller löpsedel .....	17
<relatedItem> edition .....	18
Lokala uppgifter om utgivare och leverantör .....	19
PREMIS .....	20
PREMIS:OBJECT .....	20
MIX .....	22
BasicDigitalObjectInformation .....	22
BasicImageInformation .....	22
Image Capture Metadata .....	23
ImageAssessmentMetadata .....	25
Change History .....	26
Länkar/Referenser .....	28

# Introduktion

Detta dokument innehåller en övergripande och generell redogörelse för hur digitala filer i digitalisering genomförd av KB och MKC paketeras och beskrivs med hjälp av METS med flera standarder. Dokumentet var ursprungligen framtaget för tidningsdigitaliseringen men har anpassats att gälla även tidskriftsdigitalisering. Hädanefter används benämningen periodika när tidningar och tidskrifter avses, till exempel periodikanummer i stället för tidnings- eller tidskriftsnummer.

## Submission Information Package (SIP)

Resultatet av digitaliseringen paketeras i flera "leveranspaket" som Submission Information Package (SIP). Varje SIP kommer att motsvara ett periodikanummer, dvs. en SIP kommer att innehålla samtliga filer som utgör en digital avbildning av ett periodikanummer.

Ett tidningsnummer definieras som en edition av en tidning så som den gavs ut en specifik dag. Exempel: *Stockholmseditionen av Aftonbladet 12 januari 1896*. Ett tidskriftsnummer definieras som en edition av en tidskrift så som den gavs ut som en numrerad del av en specifik årgång i ett specifikt år. Exempel: *Bonniers litterära magasin, årg. 1(1932): nr 1*.

Varje SIP skapas enligt de specifikationer för periodikadigitalisering som har tagits fram i samarbete mellan KB och MKC Riksarkivet. Specifikationerna utgår från KB:s grundprofil för METS som i sin tur bygger på SWEIPB - en basprofil framtagen av Riksarkivet m fl arkivinstitutioner och KB tillsammans.

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 tidningsdigitaliseringen men beskrivningen gäller även tidskriftsdigitaliseringen.

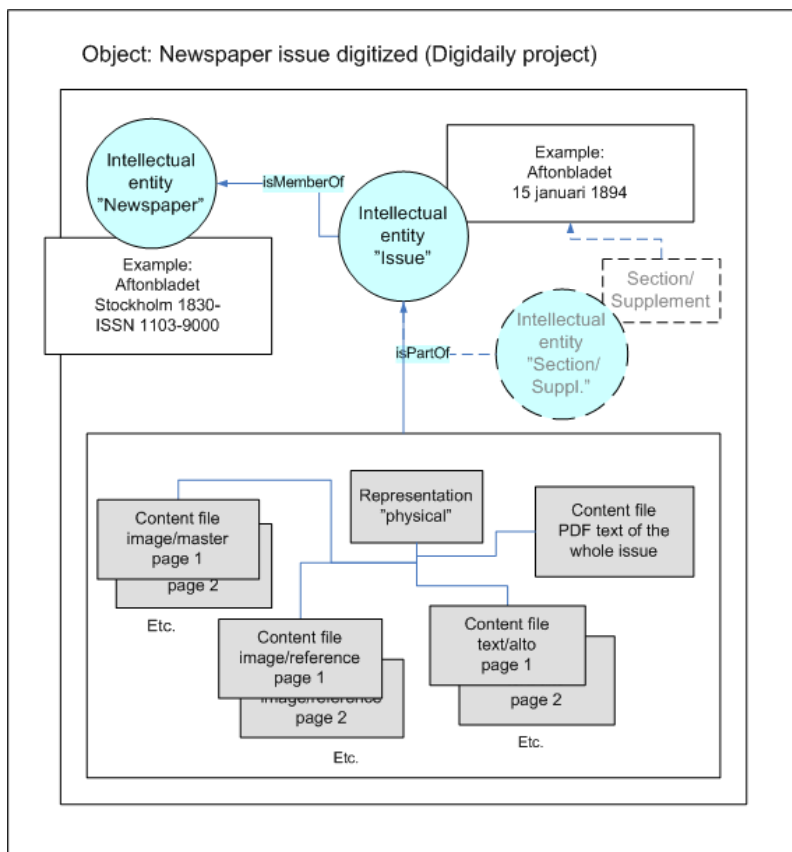


Bild 1

Bild 1 visar ett nummer av en digitaliserad tryckt dagstidning, åskådliggjord med utgångspunkt från PREMIS data model som är uppbyggd av fem entiteter: "Intellectual", "Object", "Event", "Agent" samt "Rights".

Illustrationen exemplifierar förhållandet mellan "Intellectual entites" (intellektuell entitet) och "Objects" (objekt).

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 flera enheter som kan beskrivas som egna intellektuella enheter - Först och främst själva tidningsnumret *Aftonbladet 1894-01-15* (tidningen som den gavs ut detta datum), som i sin tur är ett nummer av många i *Aftonbladet (1830-)*. Ett tidningsnummer kan dessutom delas in i flera underordnade enheter - *Sektion, Supplement, Artikel, Bild*, etc. som alla utgör intellektuella enheter. (De två senare kategorierna artikel och bild kommer däremot inte att behandlas).

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

Tidningsnumret representeras av en uppsättning bildfiler (en typ av objekt). Varje sida i tidningsnumret avbildas var för sig och sparas i olika format för olika syften, t.ex masterfiler

## Paketering i METS: Introduktion och Datamodell

för långtidslagring, ALTO-taggade filer för vidare textbearbetning, etc. Andra typer av objekt skulle kunna förekomma, t.ex. hela tidningsnumrets innehåll sparad i en pdf-fil.

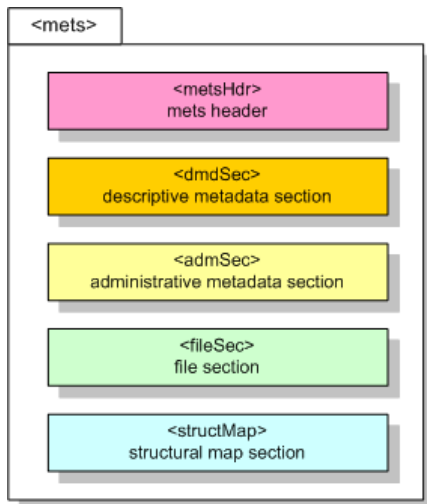
Administrativa (strukturella) metadata i METS [<admSec>](#) och [<structMap>](#) håller reda på vilka filerna är, deras egenskaper och relationer till varandra.

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 metadatatandarder.

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 periodikanummer. Ett tidningsnummer definieras utifrån det specifika datum då en tidningstitel publicerades. Ett tidskriftsnummer definieras utifrån det specifika numret som ofta ingår i en årgång samt utifrån det år då en tidskriftstitel publicerades. Ett periodikanummer kan vara indelat i sektioner, kan ha bilagor och löpsedlar etc.

- För uppgifter om "xmlns" och "xsi:schemaLocation" Se även bilagan [Tabeller över element, attribut och 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 MKC och i kontakten mellan MKC och KB.
- I PROFILE namnges den profil som mets-filen följer. Profilen för periodikadigitalisering namnges i formen av en uri:  
http://www.kb.se/namespace/mets/kbse\_mets\_profile\_001.xml
- LABEL ska innehålla tidningens eller tidskriftens titel (title) och datering/nr (date/number) från dmdSec/@LABEL=Primary/.../mods:titleInfo/mods:title. Exempel på tidningstitel med datum: *Aftonbladet 1851-12-04*. Exempel på tidskriftstitel med årgång, årtal och nummer: *Folket i bild/Kulturfront, årg. 1(1972):4*.

## <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](#).
- Elementet <agent> upprepas för de institutioner som skapar respektive tar emot METS-dokumentet:  
<agent ROLE="creator" TYPE="organization">  
<name>Riksarkivet/MKC</name>  
</agent>  
<agent ROLE="archivist" TYPE="organization">  
<name>Kungl. 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 periodika och det nummer som avbildas. Beskrivande metadata kommer att hämtas från flera håll: Libris databas, Signe, 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 periodikanumret. 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 eller sektioner av numret 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 bilagan [Tabeller över element, attribut och 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'THH: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:bib4112678\_18760203\_1\_24\_8\_m.jp2". Se även bilagan [Filnamngivning](#).

## 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 - Varje div bör 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, eller ett helt tidningsnummer, 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="issue"> ange ID för den dmdSec som har LABEL="Primary". För div TYPE="supplement" ange ID för den dmdSec som beskriver supplementet, 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:

## 1. Digitaliserade sidor + en pdf + fil med ”performance data”:

```
<structMap>
  <div TYPE="files">Fungerar som en ”wrapper” runt de andra
    <div TYPE="issue">Högsta nivån för periodikanumret
      <div TYPE="page" ORDER="1">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="2">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 TYPE="pdf">pdf-filen är på samma nivå som ”page”
        <fptr FILEID=""
      </div>
    </div>
    <div TYPE="performance">På samma nivå som ”issue”
      <fptr FILEID=""
    </div>
  </div>
</structMap>
```

## 2. Digitaliserade sidor indelade i sektioner + bilaga + löpsedlar + edition

```
<structMap>
  <div TYPE="files">
    <div TYPE="issue">Högsta nivån för periodikanumret
      <div TYPE="section">Indelning i sektioner, upprepas om flera.
        <div TYPE="page" ORDER="1">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="2">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 TYPE="supplement">
        <div TYPE="page" ORDER="1">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="2">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 TYPE="newsbill">
        <fptr FILEID=""
      </div>
      <div TYPE="pdf">pdf-filen
        <fptr FILEID=""
      </div>
    </div>
    Egen div för ”edition”:
      <div TYPE="edition">2 sidor ur annan edition
        <div TYPE="page" ORDER="5">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>
```

## Paketering i METS

```
<div TYPE="page" ORDER="6">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>
</div>
<div TYPE="performance">På samma nivå som "issue"
  <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.

## Periodikanummer och “relaterade enheter”

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

MODS-avsnittet inleds alltid med metadata om det fysiska periodikanumret men med tillägget att paketet innehåller en digital version, antingen av ett tryckt exemplar eller av en mikrofilmad kopia.

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 periodikanumret:

### `typeOfResource`

- med värdet ”text”.

### `genre`

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

### `titleInfo`

Inkluderar elementen:

- `title`
  - med en konstruerad titel för periodikanumret - samma som i `<mets@LABEL>`, se [<mets>](#)
  - konstrueras utifrån titel och datum/nr (date/number) för tidningar. Exempel: *Aftonbladet 1851-12-04*
  - konstrueras utifrån titel, årgång (volume/number), datering (date) och nummer (issue/number) för tidskrifter. Exempel: *Folket i bild/Kulturfront, årg. 1(1972):4*

## originInfo

Innehåller information om när numret gavs ut och om det tillhör en "edition". Uppgifterna här avser alltid den analoga förlagan.

Inkluderar elementen:

- dateIssued - Här skrivs ett datum enligt mönstret yyyy-mm-dd
  - med attributet encoding="w3cdtf".
  - Om datum saknas eller är felaktigt i förlagan alternativt finns endast angivet som årtal lägg till attributet qualifier, med värdet "inferred". Som värde i <dateIssued> skrivs ett antaget datum.
- edition - värde enligt kontrollerad lista. Se bilagan [Filnamngivning](#) för mer information.

## physicalDescription

Inkluderar elementen:

- digitalOrigin  
Något av följande värden:
  - "reformatted digital" om digitaliseringen är gjord från tryckt förlaga.
  - "digitized microfilm" om digitaliseringen är gjord från mikrofilm.
- 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 (vcScriptType\_kbse).

## <relatedItem> = Värddublikation

- med attributet TYPE="host"  
Obligatorisk uppgift för alla paket. Med "värddublikation" menas periodika som numret är en del av. Metadata om värddublikationen kommer att hämtas ur Libris databas. Detta görs av KB innan digitaliseringen påbörjas.

Här hittar man bl.a. titel, Libris-id, ISSN om sådan finns, samt den årgångs-, nummer- och/eller datumbeteckning som tillsammans med titeln identifierar periodikanumret. Registreras inom samma <dmdSec> som periodikanumret i övrigt.

Inkluderar elementen:

### titleInfo

Inkluderar elementen:

- title
  - med tidningens eller tidskriftens titel i klartext

### genre

- med värde "newspaper" för tidningar eller "journal" för tidskrifter

- med attribut `authority="marcgt"`

## originInfo

Inkluderar elementet:

- `dateIssued`
  - Värde: ett datum i formen yyyy-mm-dd för tidningar och yyyy för tidskrifter
  - med attributen `encoding="w3cdtf"`, och `point="start"` för periodikas startdatum
  - `dateIssued` upprepas om slutdatum finns, attribut `point="end"`

## language

Kan upprepas om värdpublikationen innehåller flera språk som alla är lika "dominerande" för innehållet. Inkluderar elementet:

- `languageTerm`.
  - Värde: bokstavskod enligt iso639-2b, t.ex. "swe"
  - med attributen `type="code"` och `encoding="iso639-2b"`

## identifier

Här anges den URI som identifierar den periodiska resursen i Libris databas. Om tidningen eller tidskriften har ett ISSN-nummer anges detta i ett eget `<identifier>`.

- med attribut `type="uri"` för Libris-id.
- `<identifier>` upprepas eventuellt med ISSN (attribut `type="issn"`).

## part

- Här anges periodikanumrets numrering och/eller datering.  
I de flesta fall kommer datumangivelsen dubbleras och se exakt likadan ut som i `mods/@TYPE=issue/originInfo/dateIssued`.  
Inkluderar elementen:
  - `detail`
    - med attributet `type="volume"`  
Inkluderar elementet:
      - `number`  
Här anges tidskriftens årgångsnumrering om det finns en sådan, dvs tryckt numrering av årgången eller annan beteckning på tidskriftsnumret.
  - `detail`
    - med attributet `type="issue"`  
Inkluderar elementet:
      - `number`  
Här anges periodikas numrering om det finns en sådan, dvs tryckt numrering eller annan beteckning (t.ex. Provenummer) på dagstidningsnumret. (Obs. nummer på mikrofilm i `relatedItem/@TYPE="original"`).
  - `date`  
Datum för numret.
    - Skriv datum enligt mönster yyyy-mm-dd, exempel: 1885-12-10.
    - Alltid attributet `encoding="w3cdtf"`.
    - Om datum saknas i förlagan eller anges endast som årtal lägg även till attributet `qualifier`, med värdet "inferred". Som värde i `<dateIssued>` skrivs ett antaget datum.

## note

- med attributet type="date"  
Hit eventuell anmärkning om saknat eller felaktigt datum.
- med attributet type="numbering"  
Hit eventuell anmärkning om saknad, felaktig eller dubbelnumrering.
- med attributet type="date/sequential designation"  
Hit eventuell anmärkning om saknad, felaktig eller dubbelvolym/årgång.

## <relatedItem> = Projekt

- med attributet type="host"  
Obligatorisk uppgift för alla paket. Metadata om projektet kommer att hämtas ur Libris databas vilket görs av KB innan digitaliseringen påbörjas.

Inkluderar elementen:

### genre

- med värde "project"

### identifier

Här anges den URI som identifierar projektet i Libris databas.

- med attribut type="uri" för Libris-id.

### 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:identifier type="uri">http://libris.kb.se/resource/bib/[libris-id]</mods:identifier>  
</mods:relatedItem>
```

## <relatedItem> = mikrofilm som förlaga

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

Inkluderar elementen:

### identifier

- Värde: mikrofilmens nummer.
- attribut type="reel number".

## physicalDescription

Inkluderar elementen:

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

## location

Inkluderar:

- holdingSimple
  - Inkluderar elementet:
    - copyInformation, som inkluderar:
      - note, utan attribut.  
Här finns plats för noteringar om förlagan. Är under utredning av TE.  
Uppgifter om skador på enstaka sidor sparas i ALTO under attributet QUALITY\_DETAIL.
      - note, med attribut type="condition".  
Här uppgift om förlagens fysiska kondition, enligt lista med kodade värden. Utredning pågår av TE.

## <relatedItem> = tryckt periodika som förlaga

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

Inkluderar elementen:

## identifier

- Värde: Kod som identifierar vilket exemplar som används som förlaga. Värden i position 1-2 i preparationskoden från Signe. Exempel: "S-A"
- attribut type="local".

## physicalDescription

Inkluderar:

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

## location

Inkluderar:

- holdingSimple
  - Inkluderar elementet:
    - copyInformation, som inkluderar:
      - note, utan attribut.  
Här finns plats för noteringar om förlagan t ex att den har slängts efter digitalisering. Kan upprepas vid behov.  
Alltid en note med kod som identifierar exemplarets härkomst, kondition och om det är ett nationalexemplar (hämtas från Signe, t.ex. S-A-1).



Uppgifter om skador på enstaka sidor sparas i ALTO under attributet QUALITY\_DETAIL.

- note, med attribut type="condition".  
Notering om förlagans fysiska kondition. Hit förs det värde som utgör position 4 i preparationskoden från Signe. Exempel: "1"

## <relatedItem> sektion, bilaga, eller löpsedel

- 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 periodikanumret och som behöver beskrivas för sig: sektioner, bilagor, löpsedlar...

Inkluderar elementen:

### genre

Värde i elementet väljs enligt genreterm\_kbse: supplement, section, newsbill

### titleInfo

Inkluderar elementen:

- partName
  - Namn och eventuella delbeteckningar. KB kommer att förse MKC med mer information för varje periodika.

### subject

Inkluderar elementet:

- Subject
  - Topic.  
Värde: välj från kontrollerad ordlista: "bilagetyp\_kbse".

### originInfo

Inkluderar:

- dateIssued
  - Skriv här datum om det skiljer sig från periodikanumret. Skriv enligt mönstret yyyy-mm-dd.
  - med attributet encoding="w3cdtf".
  - Om datering saknas i förlagan lägg även till attributet qualifier, med värdet "inferred". Som värde i <dateIssued> skrivs ett antaget datum.

### note

- med attributet type="date"  
Hit eventuell anmärkning om saknat eller felaktigt datum.
- utan attribut  
T.ex. för uppgift om tillagd titel.

## <relatedItem> edition

<relatedItem> med attribut type="otherVersion">

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

Här registreras information om annan edition av periodika för samma datum som följer med i paketet.

Inkluderar elementen:

### genre

Värde i elementet: edition (väljs enligt genreterm\_kbse)

### titleInfo

Inkluderar elementen:

title

- Här registreras varianttitel på ett enstaka nummer av editionen.

### originInfo

Inkluderar:

edition

- Skriv här editions-beteckning enligt uppgift från KB

dateIssued

- Skriv här datum om det skiljer sig från huvudeditionen. Skriv enligt mönstret yyyy-mm-dd.
- med attributet encoding="w3cdtf".
- Om datering saknas i förlagan lägg även till attributet qualifier, med värdet "inferred". Som värde i <dateIssued> skrivs ett antaget datum.

### note

note med attribut type=date: Anmärkningstext som gäller dateIssued.

För note utan attribut type: Här finns möjlighet att lägga till en anmärkning som inte gäller specifikt för dateIssue.

## 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="Lcal"`

### 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`:

- "Kungl. biblioteket"
- "Riksarkivet/MKC"

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

## PREMIS:OBJECT

Två olika typer av ”object” registreras i paketet:

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](#).

## 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. "MKC 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. Värdet 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/imageHeigth

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 dedikerad 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 (t.ex. en saknad tidningssida som i

efterhand digitaliseras från mikrofilm). 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/>