

# Skapa informationspaket i Projekt Digidaily Specifikationer

## Innehåll:

Paketering i METS: Metadata för digitaliserade dagstidningar i Projekt DigiDaily (manual)

Tabeller: Metadatelement och attribut i METS, MODS, PREMIS och MIX; ALTO - element och attribut; Ordlistor

Filnamngivning

Instruktion för saknad sida/saknat nummer

# Paketering i METS: Metadata för digitaliserade dagstidningar i Projekt DigiDaily

Manual  
KB - MKC Riksarkivet

Version 2.0  
2012-07-10 (corr. 2013-02-04)

**Innehåll:**

Introduktion .....	3
Submission Information Package (SIP) .....	3
Datamodell .....	4
METS .....	6
Element och attribut i METS .....	6
<mets> .....	6
<metsHdr> .....	6
Descriptive Metadata Section <dmdSec> .....	7
Administrative Metadata Section <amdSec> .....	7
File Section <fileSec> .....	8
Structure Map <structMap> .....	9
MODS .....	11
Tidningsnummer och “relaterade enheter” .....	11
Metadata för Tidningsnumret: .....	11
<relatedItem> = Värdepublikation .....	12
<relatedItem> = Projekt .....	13
<relatedItem> = mikrofilm som förlaga .....	14
<relatedItem> = tryckt tidning som förlaga .....	15
<relatedItem> sektion, bilaga, eller löpsedel .....	15
Lokala uppgifter om utgivare och leverantör .....	16
PREMIS .....	18
PREMIS:OBJECT .....	18
MIX .....	20
BasicDigitalObjectInformation .....	20
BasicImageInformation .....	20
Image Capture Metadata .....	21
ImageAssessmentMetadata .....	23
Change History .....	24
ALTO .....	26
Description .....	26
Layout .....	28
Länkar/Referenser .....	31
Bilagor .....	32

# Introduktion

Detta dokument innehåller en övergripande och generell redogörelse för hur digitala filer i projektet Digidaily tidningsdigitalisering paketeras och beskrivs med hjälp av METS med flera standarder.

## Submission Information Package (SIP)

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

Ett tidningsnummer definieras som en edition av en tidning så som den gavs ut en specifik dag. Exempel: *Stockholmseditionen av Aftonbladet 12 januari 1896*

Varje SIP skapas enligt de specifikationer för tidningsdigitalisering i projektet Digidaily 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.

De metadatastandarder som används i paketeringen är: [METS](#), [PREMIS](#), [MODS](#), [MIX](#) och [ALTO](#). Fler standarder kan komma att bli aktuella.

# Datamodell

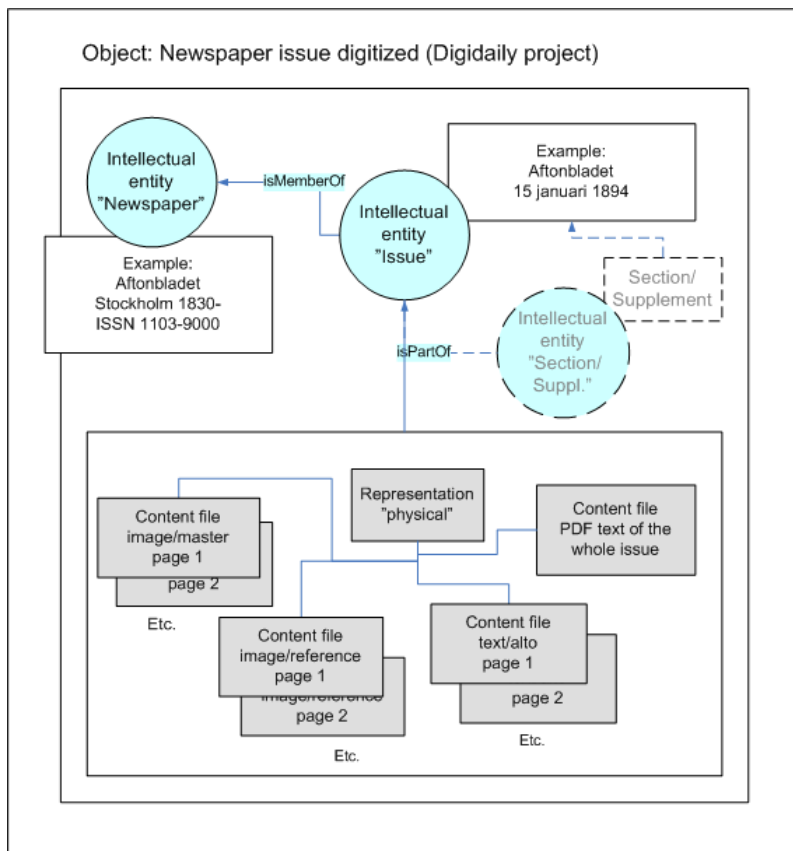


Bild 1

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 i detta projekt).

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 för långtidslagring, ALTO-taggade filer för vidare textbearbetning, etc. Dessutom kommer hela tidningsnumrets innehåll sparas ner i en pdf-fil. Administrativa (strukturella) metadata i

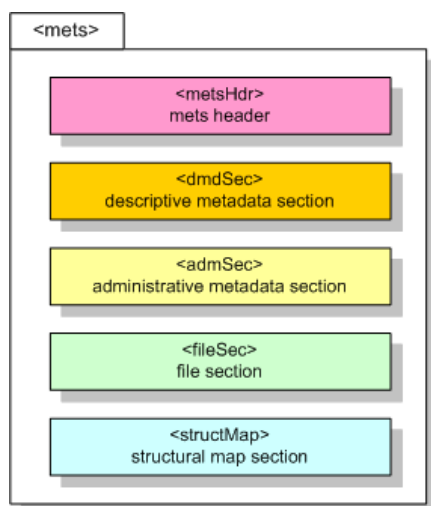
METS [<amdSec>](#) och [<structMap>](#) håller reda på vilka filerna är, deras egenskaper och relationer till varandra.

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 projektet kommer endast följande 5 sektioner att användas:



## Element och attribut i METS

### <mets>

Ett METS-dokument skapas för varje tidningsnummer. Ett tidningsnummer definieras utifrån det specifika datum då en tidningstitel publicerades. Ett tidningsnummer kan vara indelat i sektioner, kan ha bilagor och löpsedlar etc.

- För uppgifter om ”xmlns” och ”xsi:schemaLocation” Se även [Bilagor](#). Metadataelement och attribut 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:  
[http://www.kb.se/namespace/mets/kbse\\_mets\\_profile\\_001.xml](http://www.kb.se/namespace/mets/kbse_mets_profile_001.xml)
- LABEL ska innehålla tidningens titel (title) och datum/nr (date/number) från dmdSec/@LABEL=Primary/.../mods:titleInfo/mods:title. Exempel: *Aftonbladet 1851-12-04*

### <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 [Bilagor](#): 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>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 tidningen 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 tidningsnumret. 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-taggning 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 dubblering 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 ETT, och 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 [Bilagor](#) Metadataelement och attribut i METS... (tabell).

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 [Bilagor](#): 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 tidningsnumret
      <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 etc.

```
<structMap>
  <div TYPE="files">
    <div TYPE="issue">Högsta nivån för tidningsnumret
      <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>
    <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.

## Tidningsnummer och "relaterade enheter"

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

MODS-avsnittet inleds alltid med metadata om det fysiska tidningsnumret 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 Tidningsnumret:

### `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 tidningsnumret - samma som i `<mets@LABEL>`, se [<mets>](#)

### `originInfo`

Innehåller information om när tidningsnumret 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 datering saknas eller är felaktigt i förlagan lägg till attributet `qualifier`, med värdet `"inferred"`. Som värde i `<dateIssued>` skrivs ett antaget datum.
- `edition` - värde enligt kontrollerad lista. Se [Bilagor](#) 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"`.  
Värde: Digital reproduktion: Stockholm : Projekt Digidaily, yyyy (det årtal som digitaliseringen gjordes).
  - med attributet `type="script"`.  
Välj värde enligt lista (`vcScriptType_kbse`).

## <relatedItem> = Värdepublikation

- med attributet `TYPE="host"`  
Obligatorisk uppgift för alla paket. Med `"värdepublikation"` menas tidningen som tidningsnumret är en del av. Metadata om värdepublikationen kommer att hämtas ur LIBRIS databas. Detta görs av KB innan digitaliseringen påbörjas.

Här hittar man bl.a. tidningens titel, LIBRIS-id, ISSN om sådan finns, samt det nummer- och/eller datumbeteckning som tillsammans med titeln identifierar tidningsnumret. Registreras inom samma `<dmdSec>` som tidningsnumret i övrigt.

Inkluderar elementen:

### titleInfo

Inkluderar elementen:

- `title`
  - med tidningens titel i klartext

### genre

- med värde `"newspaper"`
- med attribut `authority="marcgt"`

### originInfo

Inkluderar elementet:

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

## language

Kan upprepas om tidningen 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 tidningen i LIBRIS databas. Om tidningen 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 tidningsnumrets 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="issue"  
Inkluderar elementet:
      - number  
Här anges tidningens numrering om det finns en sådan, dvs tryckt numrering eller annan beteckning (t.ex. Provnnummer) 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 datering saknas i förlagan lägg även till attributet qualifier, med värdet "inferred". Som värde i <dateIssued> skrivs ett antaget datum. 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 eller felaktig numrering.

## <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>Digidaily</mods:title>
  </mods:titleInfo>
  <mods:identifier type="uri">http://libris.kb.se/resource/bib/12101075</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 enskilda sidor sparas i ALTO under attributet QUALITY\_DETAIL.

- note, med attribut type="condition".  
Här uppgift om förlagans fysiska kondition, enligt lista med kodade värden. Utredning pågår av TE.

## <relatedItem> = tryckt tidning som förlaga

- med attribut TYPE="original"  
Obligatoriskt om tillämpligt. Här ryms uppgifter om förlagan (tryckt tidning) 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 tidningsnumret 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 tidning.

### **subject**

Inkluderar elementet:

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

### **originInfo**

Inkluderar:

- dateIssued
  - Skriv här datum om det skiljer sig från tidningsnumret. 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. 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.

## **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:

- "Kungliga 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 [Bilagor](#): Filnamnsgivning.

## 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 teknisk 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 andra pplats 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

All 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 teknisk 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 den 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

# ALTO

ALTO är en metadatastandard som används för att spara information om både textinnehållet och tidningens layout. Hela ALTO-standarden ingår i metadata-specifikationen för Digidaily. Däremot förväntas inte samtliga element och attribut som förekommer i standarden att användas. De element och attribut som anses behövas eller är önskvärda för att en bra metadatanivå har dock specificerats nedan. Majoriteten av dessa element och attribut är krav medan användningen av en mindre del är beroende på om mjukvaran automatiskt kan leverera dem. Det är möjligt att mjukvaran som används för OCR-tolkning och segmentering på ett bra sätt dessutom kan leverera element och attribut som ej explicit inkluderats i specifikationen nedan. När så är fallet kan dessa element och attribut komma att inkluderas i specifikationen efter en utvärdering. För att möjliggöra detta krävs ett nära samarbete med MKC.

Det måste alltid finnas en ALTO-fil för varje enskild bildfil. Detta gäller även för sidor som saknas i den fysiska förlagan eftersom dessa ersätts med tomma sidor i arkivpaketet.

Data i ALTO specificeras oftast som attribut istället för som element vilket är vanligare i de andra metadatastandarderna som ingår i specifikationen. Attribut har i punktlistorna indenterats en extra nivå.

## Description

Denna sektion innehåller metadata av allmän karaktär.

Som måttenheten används 1/10 mm eftersom övriga enheter både är svårare att visualisera och att översätta till SI-enheter.

Innehåller:

- MeasurementUnit (mm10)

Filnamnet för den fil som användes för att generera informationen i ALTO-filen ska alltid anges. Det är visserligen möjligt att koppla ihop ALTO-filen med originalfilen genom information i METS-filen men genom att inkludera filnamnet skapas en försäkring mot dataförlust eftersom filnamnet innehåller Librisnumret för posten för den digitaliserade publikationen. Detta gör det även enklare att identifiera innehållet i en fristående ALTO-fil.

### sourceImageInformation

Innehåller:

- fileName

All bildbehandling som sker innan OCR-tolkning redovisas i METS-filen (under MIX), varför dessa metadata ej behöver upprepas inuti ALTO-filen. Övriga inställningar anges i en instans av elementet OCRprocessing. I en sådan instans tillåts endast en beskrivning av de parametrar som användes vid teckentolkningen. Eftersom ALTO-filen beskriver en enda bildfil så är detta inget problem. Om ALTO-filen istället beskriver flera sidor och sidorna teckentolkats med olika inställningar hade flera instanser av detta element behövts. Om resultatet från teckentolkningen efterbearbetats så måste detta anges inom ramen för detta element. Notera

att flera olika inställningar för efterbearbetning får anges. Det är viktigt att parametrar och inställningar, både för teckentolkning och för efterbearbetning, redovisas utförligt för att möjliggöra omtolkning och utvärdering av resultatet.

### **OCRProcessing/ocrProcessingStep**

Innehåller:

- processingDateTime
- processingAgency
- processingStepSettings
- processingSoftware

### **OCRProcessing/ocrProcessingStep/processingSoftware**

Innehåller:

- softwareCreator
- softwareName
- softwareVersion
- applicationDescription (Beskrivning av viktiga egenskaper hos mjukvaran. Används vid t.ex. icke-kommersiell mjukvara)

### **OCRProcessing/postProcessingStep**

Innehåller:

- processingDateTime
- processingAgency
- processingStepSettings
- processingSoftware

### **OCRProcessing/postProcessingStep/processingSoftware**

Innehåller:

- softwareCreator
- softwareName
- softwareVersion
- applicationDescription (Beskrivning av viktiga egenskaper hos mjukvaran. Används vid t.ex. icke-kommersiell mjukvara)

För att minska ALTO-filens omfång definieras ett antal fontstorlekar. I texten anges sedan referenser till dessa storlekar. Granulariteten på storlekarna ska initialt vara begränsad till 0,5 punkter och endast fontstorlekar som förekommer på tidningssidan ska inkluderas i listan. Granulariteten bör utvärderas för att avgöra om det räcker med en skillnad på 1 punkt.

### **Styles**

Innehåller:

- TextStyle
  - ID
  - FONTSIZE

# Layout

I denna sektion sparas all metadata som härrör från objektets struktur, såsom mått, textinnehåll, bilder, rubriker etc. I ALTO används hierarkin PAGE->PrintSpace->Textblock->TextLine->String för att beskriva textinnehållet på en sida

Viss metadata anges på sidnivå. I attributet QUALITY anges ett grovt mått på sidan kvalit  (OK, Damaged, Missing). H r  r det framf rallt viktigt att ange om sidan saknas. Det  r  ven m jligt att komplettera m ttet med en beskrivande text om det beh vs, t.ex. ”sidan inneh ller ett h l”. Om f rlagans paginering automatiskt kan identifieras ska denna inkluderas. Sidans position ska ocks  anges (h ger, v nster etc.) om detta kan g ras automatiskt. Den uppskattade noggrannheten vid teckentolkningen p  sidniv  ska anges om programvaran kan leverera denna. Noggrannheten p  sidniv  kan anges p  tv  platser i standarden (<Accuracy> och Page Confidence, <PC>) och det  r sv rt att utl sa anledningen till detta. Det  r m jligt att det  r s  enkelt att de b da attributen s rskiljs genom att de anv nder olika enheter (<Accuracy> anges i procent och <PC> som ett tal mellan 0 och 1). B da m tten b r anges om mjukvaran har st d f r detta. Om s  inte  r fallet ska enbart det attribut som st ds anges. Sidans marginaler anges ej explicit d  de enkelt kan ber knas utifr n sidans utstr ckning och printSpace (textens utstr ckning, ej inkluderat sidnummer och liknande text).

## Page

Inneh ller:

- PAGECLASS
  - HEIGHT
  - WIDTH
  - PHYSICAL\_IMG\_NR (sidnummer inom publikationen, normalt l pnumret p  bildfilen)
  - PRINTED\_IMG\_NR (det tryckta sidnumret,  nskv rt om det kan identifieras)
  - PROCESSING (L nk till OCRProcessing-element)
  - QUALITY (Kvalit n p  sidan i den fysiska f rlagan)
  - QUALITY\_DETAIL (fritext)
  - POSITION
  - ACCURACY
  - PC
- PrintSpace
  - HEIGHT
  - WIDTH
  - HPOS
  - VPOS

## PrintSpace

En fysisk tidningssida ska delas upp i hierarkin TextBlock/TextLine/String (d r String kan ers ttas med SP f r blanksteg och HYP f r bindestreck). F r samtliga element anges storlek, position och ID. ID-numret kan komma att f renkla framtida textkorrigeringar.

Innehåller:

- TextBlock
  - ID
  - HEIGHT
  - WIDTH
  - HPOS
  - VPOS
  - ROTATION (Textblockets lutning. Värdet anges som grader motsols)

Det är möjligt att ange språket för ett textblock. Denna möjlighet bör endast utnyttjas om språket inte är svenska. Om mjukvaran automatiskt kan känna igen språket får det igenkända språket användas i fältet.

### **PrintSpace/TextBlock**

Innehåller:

- TextLine
  - ID
  - HEIGHT
  - WIDTH
  - HPOS
  - VPOS
- language

För varje ord kan alternativa stavningar anges. Antalet alternativa stavningar bör minimeras till två eftersom för många felaktiga alternativa stavningar kan resultera i en försämrad sökbarhet (vi räknar med att de alternativa stavningarna kommer att vara sökbara).

Det är möjligt att ange den uppskattade noggrannheten för teckentolkningen både på tecken- och på ordnivå. Vi har dock inte funnit någon användning av noggrannhet på teckennivå som motiverar dess lagring. Vi har däremot sett ett fåtal fall där ordnoggrannheten uppskattats till 100% men där alternativa och felaktiga stavningar ändå angetts. I dessa fall kan ordnoggrannheten kunna användas för att exkludera de alternativa stavningarna.

Bindestreck (Hyp) får endast anges som sista tecken på en rad (bindestreck inkluderas i/som en vanlig String om de förekommer någon annanstans på raden). Bindestrecket talar om att den sista strängen fortsätter på nästa rad. Eftersom bindestreck kan ha olika utformning (långt, kort etc.) så ska även dess utseende anges i attributet Content. I åtminstone USA verkar man inte använda Hyp för att explicit markera avstavningar vilket kan tyda på att inte all mjukvara har stöd för detta. Det är viktigt att använda Subs\_type (för de båda String-elementen) vid avstavning för att ange hela det avstavade ordet så att det blir sökbart i texten.

Om en förkortning påträffas i texten ska om möjligt subs\_type även användas för att inkludera det oförkortade ordet. Detta medför att sökningar som utförs på oförkortade ordet även ger träff på förkortningen.

## **PrintSpace/TextBlock/TextLine**

Innehåller:

- String (Ett ord, separeras av blanksteg eller bindestreck)
  - ID
  - HEIGHT
  - WIDTH
  - HPOS
  - VPOS
  - CONTENT (Innehållet (ordet) i strängen)
  - STYLEREFS
  - SUBS\_TYPE (Innehåller hela ordet vid avstavning eller förkortning (HypPart1, HypPart2 samt Abbreviation))
  - WC (Word confidence)
- ALTERNATIVE (Alternativstavningar som OCR-tolken kan generera)
- SP (blanksteg)
  - ID
  - HEIGHT
  - WIDTH
  - HPOS
  - VPOS
- HYP (bindestreck)
  - ID
  - HEIGHT
  - WIDTH
  - HPOS
  - VPOS
  - CONTENT (bindestreckets utformning)

# 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/>



# Bilagor

Metadataelement och attribut i METS, MODS, PREMIS och MIX  
(tabell)

ALTO - element och attribut (tabell)

Ordlistor (tabell)

Filnamngivning

Instruktion för saknad sida/saknat nummer

Projekt Digidaily: Metadataelement och attribut i METS, MODS, PREMIS och MIX				Uppdaterad 2012-07-10			
Schema	Elementnamn	Attribut	Värde i Attribut	Värde i element	M/O	Upp-repas?	Anvisningar/Anmärkning
							M: mandatory MA: mandatory if applicable O=Optional
							Blå text representerar metadata som levereras från KB och/eller KB:s system
							Röd text = uppdateringar, kommentarer m m som behöver kontrolleras
METS	mets	<b>obligatoriska namespace och attribut:</b>			M		Innehåller: metsHdr, dmdSec, amdSec, fileSec, StructMap
		xmlns:mets	http://www.loc.gov/METS/				
		xmlns:mods	http://www.loc.gov/mods/v3				
		xmlns:premis	info:lc/xmlns/premis-v2				
		xmlns:mix	http://www.loc.gov/mix/v20				
		<i>xmlns:alto</i>	<i>http://www.loc.gov/standards/alto/ns-v2#</i>				<i>Behövs inte när alto ej inkl i METS</i>
		xmlns:xlink	http://www.w3.org/1999/xlink				
		xmlns:xsi	http://www.w3.org/2001/XMLSchema-instance				
		xsi:schemaLocation	http://www.kb.se/namespace/mets/kbse_mets_001.xsd				
			http://www.kb.se/namespace/mods/kbse_mods_001.xsd				
			http://www.kb.se/namespace/premis/kbse_premis_001.xsd				
			http://www.kb.se/namespace/mix/kbse_mix20_001.xsd				
			http://www.kb.se/namespace/alto/kbse_alto-v2.0_001.xsd				<i>Behövs inte när alto ej inkl i METS</i>
		PROFILE	http://www.kb.se/namespace/mets/kbse_mets_profile_001.xml				
		ID			M		Namnet på mets-filen. Se separat anvisning för filnamsgivning. Upprepas i metsHdr/@metsDocumentID
		OBJID			M		Unik identifikator för paketet som bygger på filnamnet.
		LABEL			M		Hämta tidningens titel från dmdSec/mdWrap/@LABEL=Primary/.../mods:title. Exempel: Aftonbladet 1851-12-04
		TYPE	SIP		M	NR	Värde ska här alltid vara "SIP"

	metsHdr				M	NR	Under: mets. Innehåller: agent, metsDocumentID	
		CREATEDATE			M	NR	Datum och klockslag då mets-filen skapades. Skrivs YYYY-MM-DD'T'HH:mm:ss+01:00	
		RECORDSTATUS			MA	NR	Används för att indikera om ett paket är en ersättning eller komplettering av ett tidigare levererat paket. RECORDSTATUS skrivs bara om något av värdena SUPPLEMENT eller REPLACEMENT ska åsättas paketet	
			REPLACEMENT				ersättning - Hela paketet, med all data, byts ut	
			SUPPLEMENT				komplettering - Delar av paketet byts ut/och eller kompletteras med mer data	
	agent				M		Under: metsHdr. Innehåller: name. Agent registreras 2 ggr för MKC Riksarkivet (role=creator) respektive Kungliga biblioteket (role=archivist).	
		ROLE			M			
			CREATOR					
			ARCHIVIST					
	name				M		Under: agent. Obligatoriskt attribut type med värde organization	
		TYPE	ORGANIZATION		M			
				Riksarkivet/MKC				
				Kungliga biblioteket				
	metsDocumentID				M		Under: metsHdr. Samma värde som i mets/@ID	
				<b>Tidningsnumret</b>				
	dmdSec				M	R	Under: mets. Innehåller: mdWrap	
		ID	dmdSec001, dmdSec002, etc.		M	NR	Löpande numrering av dmdSec enligt mönster "dmdSec001", "dmdSec002", "dmdSec003", etc.	
	mdWrap				M	NR	Under: demdSec. Innehåller xmlData	
		LABEL	Primary					
		MDTYPE	MODS		M	NR		
	xmlData				M	NR	Under: mdWrap. Innehåller: mods	
MODS	mods				M	NR	Innehåller: identifier, typeOfResource, genre, originInfo, physicalDescription, relatedItem	
	identifier	type	local	[mkc:s resursid]	M	NR	Under: mods.	
	typeOfResource			text	M	NR	Under: mods.	
	genre	authority	marcgt	issue	M	NR	Under: mods.	

	titleInfo				M	NR	Under: mods. Innehåller: title
	title				M	NR	Under: titleInfo. Tidningens titel och datum/nr (date/number). Samma värde som i mets:mets@LABEL. Exempel: Aftonbladet 1851-12-04
	originInfo				M	NR	Under: mods. Innehåller: dateIssued, note, edition. Uppgifterna om originInfo gäller den analoga (tryckta) förlagan.
	dateIssued				M	NR	Under: originInfo. Exempel: 1834-08-11.
		encoding	w3cdtf		M	NR	
		qualifier	inferred		MA	NR	Om datering saknas, eller är felaktigt, i förlagan anges ett antaget datum och attributet QUALIFIER med värde "inferred".
	<i>note</i>	<i>type</i>	<i>date</i>		<i>O</i>	<i>R</i>	<i>Under: originInfo. Diskuteras: I en "note" ange om datumet i förlagan saknas/är felaktigt. Upprepning. Använd istället note i relatedItem type=host</i>
	edition			Värden enligt kontrollerad lista	MA	NR	Under: originInfo. Värde är samma som bibliografiskt ID i Signe
	physicalDescription				M	NR	Under: mods. Innehåller: digitalOrigin, note
	digitalOrigin				M	NR	Under: physicalDescription
				reformatted digital			digitalisering från tryckt förlaga
				digitized microfilm			digitalisering från mikrofilm
	note	type	reproduction	Digital reproduktion: Stockholm : Riksarkivet/MKC i samarbete med Kungliga biblioteket, [yyyy]	M	NR	Under: physicalDescription.
	note	type	script	Enligt vcScriptType_kbse	M	NR	Under: physicalDescription.
<b>Microfilm som förlaga</b>							
	relatedItem	type	original		MA	R	Under: mods. Innehåller: identifier, location, physicalDescription
	identifier	type	reel number	[sträng]	M	NR	Under: relatedItem. Mikrofilmens nummer
	location				O	NR	under: relatedItem. Innehåller holdingSimple

	holdingSimple				O	NR	Under: location. Innehåller copyInformation
	copyInformation				O	NR	Under: holdingSimple. Innehåller note
	note				O	R	Under: copyInformation. Noteringar om förlagorna. Kan upprepas
	note	type	condition		O	NR	Under: copyInformation. Notering om förlagans fysiska kondition.
	physicalDescription				M	NR	Under: relatedItem. Innehåller: form
	form	authority	marcform	microfilm			Under: physicalDescription
	<b>Tryckt exemplar som förlaga</b>						
	relatedItem	type	original		MA	R	Under: mods. Innehåller: identifier, location, physicalDescription. Om fler än ett exemplar av tidningen har levererats upprepas relatedItem
	identifier	type	local	[sträng]	M	NR	Under: relatedItem. Kod som identifierar vilket exemplar som används som förlaga. Värdet i position 1-2 i preparationskoden från Signe. Exempel: "S-A"
	location				O	NR	Under: relatedItem. Innehåller holdingSimple.
	holdingSimple				O	NR	Under: location. Innehåller copyInformation
	copyInformation						Under: holdingSimple. Innehåller note
	note				O	R	Under: copyInformation. Noteringar om förlagorna. Kan upprepas
	note	type	condition		O	NR	Under: copyInformation. Notering om förlagans fysiska kondition. Hit förs det värde som utgör position 4 i preparationskoden från Signe. Exempel: "1"
	physicalDescription				M	NR	Under: relatedItem. Innehåller: form
	form	authority	marcform	print	M	NR	Under: physicalDescription

Värdepublikationen (dagstidningen) inklusive numrering och datum på tidningsnumret:							
relatedItem	type	host			M	NR	Under: mods. Innehåller: genre, titleInfo, originInfo, language, identifier, part, note
genre	authority	marcgt	newspaper		M	NR	Under: relatedItem
titleInfo					M	NR	Under: relatedItem. Innehåller: title
title					M	NR	Under: titleInfo. Värdepublikationens titel. Metadata från KB/LIBRIS
originInfo					M	R	Under: relatedItem. Innehåller: dateIssued
dateIssued	encoding	w3cdtf			M	NR	Under originInfo. Ska vara datum enl mönster YYYY-MM-DD
	point	start			M	NR	Datum för första utg nummer (startdatum)
	point	end			M	NR	Datum för ev sista utg nummer (slutdatum)
language					M	R	Under: relatedItem. Innehåller: languageTerm
languageTerm	type	code			M	NR	Under: language
	encoding	iso639-2b			M	NR	En treställig kod enl iso639-2b
identifier	type				M	R	Under: relatedItem. Identifierare till tidningen i sin helhet (värdepublikationen)
		uri	http://libris.kb.se/resource/bib/[librisid]		M	NR	URI till libris-posten över den digitala tidningen (värdepublikation)
		issn			MA	NR	ISSN till tidningen online. Metadata från KB/LIBRIS. Skrivs utan bindestreck
part					MA	NR	Under: relatedItem. Innehåller: detail, date
detail	type	issue			MA	NR	Under: part. Innehåller: number
number					MA	NR	Under: detail. Tidningens numrering hit, om det finns en sådan. Dvs tryckt numrering eller annan beteckning (t.ex. Provnnummer) på dagstidningsnumret. Nummer på mikrofilm läggs i relatedItem/TYPE="original"/identifier/TYPE="reel number"
date	encoding	w3cdtf			MA	NR	Under: part. Datum för numret. Exempel: 1885-12-10
	qualifier	inferred			MA	NR	Om datering saknas, eller är felaktigt, i förlagan anges ett antaget datum och attributet QUALIFIER med värde "inferred".
note	type	date			O	R	Under: relatedItem. Diskuteras: I en "note" ange om datumet i förlagan saknas/är felaktigt. Standardfras: "Datering saknas eller är felaktigt"

	note	type	numbering		O	R	Under: relatedItem. Diskuteras: I en "note" ange om numrering i förlagan saknas/är felaktig. Standardfras: "Numrering saknas eller är felaktig"
<b>Projektet:</b>							
	relatedItem	type	host		M	NR	Under: mods. Referens till information i LIBRIS om projektet. Innehåller: genre, titleInfo, identifier
	genre			project	M	NR	Under: relatedItem. Värde från lista genreterm_kbse
	titleInfo				M	NR	Under: relatedItem. Innehåller: title
	title			Digidaily	M	NR	Under: titleInfo. Projektets titel. Alltid samma värde
	identifier	type	uri	<a href="http://libris.kb.se/resource/bib/12101075">http://libris.kb.se/resource/bib/12101075</a>	M	NR	Under: relatedItem. URI till LIBRIS-post över projektet. Kommer alltid att vara samma id
<b>Bilaga, sektion, löpsedel:</b>							
							Löpsedel skulle kunna uteslutas här och istället registreras enbart i structMap
METS	dmdSec				MA	R	Under: mets. Innehåller: mdWrap
		ID	dmdSec001, dmdSec002, etc.		M	NR	Löpande numrering av alla dmdSec: dmdSec001, dmdSec002 etc.
	mdWrap	MDTYPE	MODS		M	NR	Under: dmdSec. Innehåller:xmlData
	xmlData				M	NR	Under: mdWrap. Innehåller: mods
MODS	mods				M	NR	Under: xmlData. Innehåller: relatedItem
	relatedItem	type	constituent		M	NR	Under: mods. Innehåller: titleInfo, subject, genre, originInfo, note
	titleInfo						Under: relatedItem. Innehåller: partName
	partName				MA	NR	Under: titleInfo. Namn på bilagan eller sektionen. Här även eventuell bokstavs- eller nummerbeteckning.
	<i>PartNumber</i>				MA	NR	<i>Används ej</i>
	subject				MA	R	Under: relatedItem. Ämnesord som beskriver innehållet i bilaga. Innehåller: topic
	topic	authority	bilagetyb_kbse	Värde enligt kontrollerad lista	MA	NR	Under: subject. KB tillhandahåller termer
	genre			supplement	MA	NR	Under: relatedItem. Värde enligt genreterm_kbse
				section	MA	NR	
				newsbill	MA	NR	
	originInfo				O	NR	Under: relatedItem. Innehåller: datelssued
	datelssued	encoding	w3cdf		O	NR	Under: datelssued. Uppgifterna om originInfo gäller den analoga (tryckta) förlagan. Anges om datum skiljer sig från issue.

		qualifier	inferred		O	NR	Om datering saknas, eller är felaktigt, i förlagan anges ett antaget datum och attributet QUALIFIER med värde "inferred".
	note	type	date		O	R	Under: relatedItem. I en "note" ange om datumet i förlagan saknas/är felaktigt
	note				O	R	Under: relatedItem. Standardfras: "Titel tillagd för hand, sannolikt tillfällig bilaga"
<b>Lokala uppgifter om leverans:</b>							
METS	dmdSec				M	R	Under: mets. En lokal dmdSec, för att hantera namn + id för leverantör. Innehåller: mdWrap
		ID	dmdSec001, dmdSec002, etc		M	NR	
		LABEL	Local		M	NR	
	mdWrap	MDTYPE	MODS		M	NR	Under: dmdSec. Innehåller: xmlData
	xmlData				M	NR	Under: mdWrap. Innehåller: mods
MODS	mods				M	NR	Under: xmlData. Innehåller: name
	name	type	corporate		M	R	Under: mods. Innehåller: namePart, role
		authority	local		M	NR	
		valueURI	http://id.kb.se/orga nisations/SE20210 01710		M	NR	Formen för organisationsid: SE[organisationsnr]-[suffix]
	namePart			Kungliga biblioteket	M	NR	Under: name. Utgivarens namn, här alltid Kungliga biblioteket
	role				M	NR	Under: name. Innehåller: roleTerm
	roleTerm	type	text		M	NR	Värde alltid publisher
		authority	marcrelator	publisher			
	name	type	corporate		M	R	Under: mods. Innehåller: namePart, role
		authority	marcrelator		M	NR	
		valueURI	http://id.kb.se/orga nisations/SE20210 01074-MKC		M	NR	Formen för organisationsid: SE[organisationsnr]-[suffix]
	namePart			Riksarkivet/MKC	M	NR	
	role				M	NR	Under: name. Innehåller: roleTerm
	roleTerm	type	text		M	NR	Under: role. Värde alltid supplier
		authority	marcrelator	supplier			



METS PREMIS:OBJECT							
METS	amdSec				M	NR	Under: mets. Innehåller: techMD
		ID	amdSec001		M	NR	
	techMD				M	R	TechMD upprepas för varje objekt som ingår i paketet. Innehåller: mdWrap
		ID	techMD001, techMD002 etc.		M	NR	
	mdWrap	MDTYPE	PREMIS:OBJECT		M	NR	Innehåller: xmlData
	xmlData				M	NR	Innehåller: premis
PREMIS	premis				M	NR	Innehåller: object
	object				M	NR	Innehåller: objectIdentifier, objectCharacteristics, relation
		xsi:type			M	NR	
			file		MA	NR	Upprepa techMD/..object=file för varje datafil som ingår i paketet.
			representation		MA	NR	Alltid en techMD/..object=representation för varje paket.
	objectIdentifier				M	NR	Innehåller: objectIdentifierType, objectIdentifierValue
	objectIdentifierType				M	NR	
				filepath			object=file
				uri			
				local			object=representation
	objectIdentifierValue				M	NR	När object=file blir värdet samma som filens namn (filepath). När object=representation blir värdet samma som i mets/@OBJID
	objectCharacteristics				MA	NR	Innehåller: compositionLevel, fixity, size, format, objectCharacteristicsExtension
	compositionLevel		0		M	NR	0(noll) = Okomprimerad fil
	fixity				M	R	Innehåller: messageDigestAlgorithm, messageDigest, messageDigestOriginator
	messageDigestAlgorithm		MD5		M	NR	KB:s metsprofil stödjer MD5 och SHA-1
	messageDigest				M	NR	checksumman
	messageDigestOriginator		MKC Riksarkivet		M	NR	
	size				M	NR	ett tal i enheten byte
	format				M		Innehåller: formatDesignation, formatRegistry
	formatDesignation				M	NR	Innehåller: formatName, formatVersion
	formatName				M	NR	ange formatets namn (Name:) enligt PRONOM, <a href="http://http://www.nationalarchives.gov.uk/PRONOM/">http://http://www.nationalarchives.gov.uk/PRONOM/</a>
			JPEG2000				

				Extensible Markup Language				
	formatVersion				M	NR	ange formatversion enligt PRONOM, eller lämna tomt om uppgiften saknas	
				1.0			Formatversion för xml	
				[Ingen uppgift]			Formatversion lämnas tomt för jpg2000	
	formatRegistry				M	NR	Innehåller: formatRegistryName, formatRegistryKey, formatRegistryRole	
	formatRegistryName			PRONOM	M	NR	Värde hämtas alltid från PRONOM	
	formatRegistryKey				M	NR	Formatets ID i PRONOM	
				x-fmt/392			Formatnyckel för jpg2000	
				fmt/101			Formatnyckel för xml	
	formatRegistryRole			specification	M	NR	Alltid "specification"	
	objectCharacteristicsExtension				M		Behållare för metadata enligt andra standarder. Innehåller: mix	
	<b>objectCharacteristicsExtension: MIX</b>							Tekniska metadata för bildfiler
MIX	mix							
	BasicDigitalObjectInformation							
	compression				M	R	(För varje fil)	
	compressionScheme			string/enumerated in local list	M	NR	Initialt används värdena i ANSI/NISO. Om dessa inte räcker till skapas en egen lista	
	compressionSchemeLocalList			reference/URL	MA	NR	location of the file containing the local enumerated list	
	compressionSchemeLocalValue				MA	NR	Värden ut den lokala ordlistan	
	compressionRatio			float (egentligen int)	MA	NR	6, 8, 10 etc. 6 motvarar 6:1. JPEG2000 kan ha t.ex. 1:13.3 så jag föreslår att vi använder float. O o standarden, MA hos oss	
	BasicImageInformation							
	BasicImageCharacteristics				M	NR		
	imageWidth			int	M	NR	Bredd i pixlar. (Specifies the width of the image in horizontal or X dimension, should be registrered in pixels)	
	imageHeigth			int	M	NR	Höjd i pixlar (Specifies the length of the image in vertical or or Y dimension, should be registrered in pixels)	
	BasicImageInformation							
	BasicImageCharacteristics							
	PhotometricInterpretation							
	colorSpace			string	M	NR	RGB, CMYK, Lab etc. Bör vara från lista även om det inte står i standarden	
	BasicImageInformation							

	BasicImageCharacteristics						
	PhotometricInterpretation						
	ColorProfile				MA	NR	
	iccProfile				MA	NR	
	iccProfileName		string		MA	NR	sRGB, adobeRGB etc
	iccProfileVersion		string		MA	NR	T.ex 1998
	iccProfileURI		string/URL		MA	NR	Om profilen inte är välkänd måste en URI anges.
	SpecialFormatCharacteristics				MA	NR	Endast för JPEG2000
	JPEG2000						
	CodecCompliance				MA (O)	NR	(Några O i standarden, MA hos oss)
	codecVersion		string		MA (O)	NR	Kakadu
	codecVersion		string		MA (O)	NR	3.1
	SpecialFormatCharacteristics						
	JPEG2000						
	EncodingOptions				MA	NR	
	tiles		string		MA	NR	1024x1024
	qualityLayers		int		MA	NR	14
	resolutionLevels		int		MA	NR	6
							Obs: De parametrar som används vid kodningen ska sparas inuti filen i ett XML-fält.
	ImageCaptureMetadata						
	SourceInformation						
	SourceID						
	sourceType		string		MA	NR	Används endast om den aktuella bilden har annan ursprungskälla än majoriteten av filerna i paketet (t.ex. tidigare saknad sida som skannats från mikrofil). Normala värden är digitized microfilm, born digital och reformatted digital.
	sourceID		string		MA	NR	Identifierare till det fysiska objektet, t.e.x vilken mikrofilmsrulle som använts. Formateras på samma sätt som motsvarande materialkategori i MODS
	ImageCaptureMetadata						
	SourceInformation						
	SourceSize						
	sourceXDimension						
	sourceXDimensionValue		real		O	NR	Skall anges om utrustningen automatiskt kan leverera värdet och om inte samplingsfrekvens kan anges. Kan användas för skalning.
	sourceXDimensionUnit		enumerated		MA	NR	millimeter
	ImageCaptureMetadata						

	SourceInformation						
	SourceSize						
	sourceYDimension						
	sourceYDimensionValue		real	O	NR	Skall anges om utrustningen automatiskt kan leverera värdet och om inte samplingsfrekvens kan anges. Kan användas för skalning.	
	sourceYDimensionUnit		enumerated	MA	NR	millimeter	
	ImageCaptureMetadata						
	orientation		enumerated	M	NR	Lägg till ett värde enligt <a href="http://www.loc.gov/standards/mix/mix20/mix20.xsd">http://www.loc.gov/standards/mix/mix20/mix20.xsd</a> , typeOfOrientationType. Se Ordlistor, rad 137ff	
	GeneralCaptureInformation			M	NR		
	dateTimeCreated		w3cdtf	M	NR	Datum och tid för bildfångsten. Skrivs enligt <a href="#">w3cdtf</a> som YYYY-MM-DD'T'HH:mm:ss+01:00.	
	captureDevice		enumerated	M	NR	Lista på tillåtna värden tas fram i samarbete med MKC, t.ex. Scanner, digital camera etc.	
	imageProducer		string	MA	NR	Används endast om bilden har en annan ursprungskälla än majoriteten av bilderna, dvs. analogt med sourceType/sourceID	
	ScannerCapture			MA	NR	Samtliga är R i standarden	
	scannerManufacturer		string	MA	NR	Tillverkare	
	ScannerCapture						
	ScannerModel						
	scannerModelName		string	MA	NR	Namn på modell.	
	scannerModelNumber		string	MA	NR	Modellnummer	
	scannerModelSerialNo		string	R	NR	Serienummer. Kan användas för QA om flera exemplar av samma skanner används.	
	ScannerCapture						
	ScanningSystemSoftware						
	scanningSoftwareName		string	MA	NR	Namn på scannermjukvara.	
	scanningSoftwareVersionNo		string	MA	NR	Namn på scannermjukvaruversion	
	DigitalCameraCapture			MA	NR	Samtliga är R i standarden	
	digitalCameraManufacturer		string	MA	NR		
	DigitalCameraCapture						
	DigitalCameraModel						

	digitalCameraModelName			string	MA	NR	Namn på modell.
	digitalCameraModelNumber			string	MA	NR	Modellnummer
	digitalCameraModelSerialNo			string	R	NR	Serienummer. Kan användas för QA om flera exemplar av samma kamera används.
	DigitalCameraCapture						
	CameraCaptureSettings						
	ImageData				MA		O i standarden
	fNumber			real	MA	NR	8.0
	exposureTime			real	MA	NR	0.5
	isoSpeedRatings			int	MA	NR	T.ex. 100. Värden från ISO12232
	exifVersion			enumerated	MA	NR	T.ex. 0220 (2.2). Värden hämtas från EXIF 2.2 och formateras lite annorlunda
	exposureBiasValue			rational	MA	NR	1.0
	lightsource			enumerated	MA	NR	T.ex. fluorescent. Värden från EXIF 2.2
	focalLength			real	MA	NR	T.ex. 0.08. Anges i meter
	autoFocus			enumerated	MA	NR	T.ex. Manual. Värden från DIG35, B3.2.5
	ImageAssessmentMetadata						
	SpatialMetrics						
	samplingFrequencyUnit	enumerated	1,2,3		MA	NR	1 = no unit, 2 = inch, 3 = centimeter (Är detta ett attribut?)
	xSamplingFrequency			rational	MA	NR	Anges endast om värdet kan beräknas automatiskt. Har företräde framför sourceDimension
	ySamplingFrequency			rational	MA	NR	Anges endast om värdet kan beräknas automatiskt. Har företräde framför sourceDimension
	ImageColorEncoding						
	bitsPerSample				M	NR	
	bitsPerSampleValue			string	M	NR	T.ex. 8,8,8. Anger antalet bitar per färgkanal
	bitsPerSampleUnit			int	M	NR	Anger enhet för bitsPerSample
	ImageColorEncoding						
	samplesPerPixel			int	M	NR	Anger antalet färgkanaler
	TargetData						(R i standarden)
	targetType			enumerated	MA	NR	MA. 0=external, 1=internal. Ett externt mål tar upp en egen bild, ett internt mål syns på samma bild som objektet
	TargetData						
	TargetID						
	targetManufacturer			string	MA	NR	GreytagMacbeth
	targetName			string	MA	NR	ColorChecker
	targetNo			string	MA	NR	Vilket exemplar som används

	ChangeHistory				MA	NR	Något komplicerad struktur. Se dokumentationen av MIX/NISO för detaljer
	ImageProcessing				MA	R	Sparas i kronologisk ordning om den inbördes ordningen har betydelse
	dateTimeProcessed		w3cdtf		MA	NR	Senaste tidpunkten för modifikation sparas. Skrivs enligt YYYY-MM-DD'THH:mm:ss+01:00
	processingAgency		string/MKC		MA	R	T.ex. MKC
	processingRationale		string		R	NR	Möjlighet att ange förklarande text vid bildbearbetning som ej utförs regelmässigt
	ImageProcessing						
	processingSoftware				R	R	
	processingSoftwareName		string		R	NR	
	processingSoftwareVersion		string		R	NR	
	ImageProcessing						
	processingActions		string		R	R	T.ex. deskew 1.6%. Bör om möjligt vara kommandot som skickats till programmet
<i>PREMIS:relationship</i>							<i>Borttaget. Se istället under metsHdr@RECORDSTATUS</i>
<b>METS: fileSec</b>							
METS							
	fileSec	ID	fileSec001		M	NR	Under: mets. En fileSec per METS. Innehåller: fileGrp
	fileGrp				M	R	Under: fileSec. fileGrp upprepas, en per USE. Innehåller: file
		ID	fileGrp001, fileGrp002 etc.		M	NR	
		USE	[Värde från vcUSE_kbse]		M	NR	Under: fileGrp. Plocka värde från ordlista ID="vcUSE_kbse"
	file				M	R	Gruppera filer med samma USE inom en fileGrp. Innehåller: Flocat
		ID	file1, file2, file3, etc.		M	NR	Numrera från 1, 2, 3 etc. Prefix "file"
		USE			M	NR	enl. vocabulary ID="vcUSE" (samma som fileGrp)
		MIMETYPE	text/xml		M	NR	Värde enligt MIME
			image/jp2				
		SIZE			M	NR	samma som i premis/.../size
		CREATED			M	NR	YYYY-MM-DD'THH:mm:ss+01:00
		ADMID			M	NR	ID för den techMD som innehåller metadata om filen
		CHECKSUM			M	NR	Samma checksumma som i premis/fixity/messageDigest
		CHECKSUMTYPE	MD5		M	NR	
	Flocat	LOCTYPE	URL		M	NR	Under: file.

		xlink:type	simple		M	NR	
		xlink:href	file:[filnamn]		M	NR	
<b>METS: structMap</b>							
	structMap				M	R	Under: mets. Är repeterbart men normalt bara en structMap som beskriver hur objekten förhåller sig till varandra "fysiskt". Delas in i en hierarkisk struktur med hjälp av elementen div. Se exempel i manualen. Innehåller: div
		ID	structMap001		M	NR	
		TYPE	physical		M	NR	värde "physical" från ordlistan vcStructMap_TYPE_kbse
	div				M	R	Under: structMap. Varje div kan innehålla ingen, en eller flera div. Hierarkier byggs upp med värden från ordlistan vcDiv_TYPE_kbse. Innehåller: fptr
		ID	div001, div002 etc.		M	NR	
		TYPE	Välj från vcStructMap_TYPE_kbse		M	NR	Varje div namnges i TYPE med ett värde från vcDiv_TYPE_kbse
		LABEL	Se Anvisningar		MA	NR	Om div har type-värdet page eller issue, och om sidan eller numret t.ex. saknas, kan värde väljas från vcLABEL_kbse
		ORDER	1,2,3, etc.		M	NR	Den ordning som filen ska visas i
		ORDERLABEL			M	NR	Textfält. Här kan man lägga sidnumrering (förmodl inte aktuellt för tidningarna)
		DMDID			M	NR	ID för den DMDID som innehåller metadata på den nivå som div-elementet representerar
		ADMID	techMDXXX		M	NR	ID för techMD som innehåller metadata om representationen.
	fptr	FILEID	file1, file2, file3, etc.		M	R	Under: div. Ange samma värde som i motsvarande file@ID

Projekt Digidaily: ALTO -element och attribut							
Container	Elementnamn	Attribut	Värde i Attribut	Värde i element	M/O	Upp-repas?	Anvisningar/Anmärkning
Description	MeasurementUnit			string	M	N	Innehåller information av allmän karaktär Ska vara "mm10". Enhet för mätningarna, tiodelsmilimeter
Description	sourceImageInformation						
	fileName			string	M	N	Filnamn för fil som OCR-tolkats
Description	OCRProcessing	ID			M	R	Kan förekomma i flera instanser men bör inte göra det eftersom varje fil endast ska innehålla resultatet från en sida
Description	OCRProcessing						
	ocrPreProcessingStep						Täcks av MIX och ImageProcessing. Inkluderas ej.
Description	OCRProcessing						
	ocrProcessingStep				M	N	Själva OCR-tolkningen får endast förekomma i en instans (inom varje OCRProcessing-instans)
	processingDateTime			dateTimeType	M	N	ISO8601. Lite osäker på formatet i standarden men det måste väl vara bäst att använda samma format genomgående?
	processingAgency			MKC Riksarkivet	M	N	
	processingStepSettings			string	M	N	Samtliga inställning för OCR programvaran, se exempel från LoC och Australien. Med dessa ska det vara möjligt att läsa tolka om texten och få ett identiskt resultat
	processingSoftware			processingSoftwareType	M	N	Länk, se nedan
Description	OCRProcessing						
	ocrProcessingStep						
	processingSoftware						
	softwareCreator			string	M	N	
	softwareName			string	M	N	
	softwareVersion			string	M	N	
	applicationDescription			string	MA		Beskrivning av viktiga egenskaper hos mjukvaran. Används vid t.ex. icke-kommersiell mjukvara
Description	OCRProcessing						



postProcessingStep					MA	R	Främst aktuellt om OCR-tolkning följs av ett manuellt steg för t.ex. segmentering. Kan upprepas, flera olika steg kan utföras
	processingDateTime			dateTimeType	M	N	w3cdtf.
	processingAgency			MKC Riksarkivet	M	N	
	processingStepSettings			string	M	N	
	processingSoftware			string	M	N	Referens. Används av PROCESSING
Description							
OCRProcessing							
postProcessingStep						R	
processingSoftware							
	softwareCreator			string	MA	N	
	softwareName			string	MA	N	
	softwareVersion			string	MA	N	
	applicationDescription			string	MA	N	Beskrivning av viktiga egenskaper hos mjukvaran. Används vid t.ex. icke-kommersiell mjukvara
Styles							
	TextStyle	ID	ID			R	
		FONTSIZE	float			R	Begränsa till hela och halva tal
Layout							
Page					M	R	Ett ALTO-fil kan innehålla flera sidor
	PAGECLASS	HEIGHT	int		M		Mäts i Measurement unit (1/10 mm)
		WIDTH	int		M		
		PHYSICAL_IMG_NR	int		M		Sidnumret inom publikationen
		PRINTED_IMG_NR	string		MA		Det tryckta sidnumret. Önskvärd
		PROCESSING	IDREF		MA		Länk till OCRProcessing-instans.
		QUALITY	lista i standarden		MA		Kvaliteten på sidan i den fysiska förlagan (obs, ej OCR)
		QUALITY_DETAIL	string		O		Fritext för utökad beskrivning av den fysiska förlagens kvalitet. Bör hållas så kontrollerad som möjligt, rutiner måste utarbetas
		POSITION	lista i standarden		MA		left, right etc.
		ACCURACY	float		MA		OCR, i %. Används ej av Aus eller USA men vi anser att det är önskvärdt.
		PC	float		MA		Page confidence. Används av USA och Australien (men vi förstår inte hur de använder attributen då det står under Page i standarden). Också möjligt att använda accuracy men det gör de inte. Värdet begränsat till intervallet 0-1. Bör anges.
							(Både ACCURACY och PC ska anges om mjukvaran har stöd för detta. Om så inte är fallet ska enbart det attribut som stöds anges)
Page							

	PrintSpace	HEIGHT	int		M		Endast en printspace per sida och fil
		WIDTH	int		M		
		HPOS	int		M		Horizontal position
		VPOS	int		M		Vertical position
Page							
PrintSpace							
	TextBlock				MA	R	Sammanhängande textblock, minst ett per sida (om sidan har text).
		ID	ID		M		
		HEIGHT	int		M		
		WIDTH	int		M		
		HPOS	int		M		
		VPOS	int		M		
		ROTATION	float		MA		Anges om textblocket lutar och inte rätats upp. Värdet anges som grader motsols
Page							
PrintSpace							
TextBlock							
	TextLine				MA	R	En textrad
		ID	ID		M		
		HEIGHT	int		M		
		WIDTH	int		M		
		HPOS	int		M		
		VPOS	int		M		
Page							
PrintSpace							
TextBlock							
	language			language	MA	N	Språket i ett textblock. Bör endast användas om det ej är svenska och om OCR-tolken kan identifiera det automatiskt

Page							
PrintSpace							
TextBlock							
TextLine							
	String				MA	R	Ett ord, separeras av blanksteg eller bindestreck
		ID	ID		M		
		HEIGHT	int		M		
		WIDTH	int		M		
		HPOS	int		M		
		VPOS	int		M		
		CONTENT	string		M		Innehållet (ordet) i strängen
		STYLEREFS	IDREF		MA		Refererar till TextStyle
		SUBS_TYPE	enumeration		MA		Innehåller hela ordet vid avstavning eller förkortning (HypPart1, HypPart2 samt Abbreviation)
		WC	float		MA		Word confidence
	ALTERNATIVE			string	MA	R	Innehåller alternativstavningar som OCR-tolken kan generera. Är ett element (inte ett attribut) under String
Page							
PrintSpace							
TextBlock							
TextLine							
	SP				MA	R	Blanksteg
		ID	ID		M		
		HEIGHT	int		M		
		WIDTH	int		M		
		HPOS	int		M		
		VPOS	int		M		
TextLine							
	HYP				MA	N	Bindestreck, kan endast förekomma vid radslut
		ID	ID		M		
		HEIGHT	int		M		
		WIDTH	int		M		
		HPOS	int		M		
		VPOS	int		M		
		CONTENT	string		M		Bindestreck kan se olika ut.

Projekt Digidaily: Ordlistor				Uppdaterad 2012-07-10		
Schema	Elementnamn	Attribut	Värde i Attribut	Godkänt värde i element	Referens	Kommentarer
<b>METS</b>						
	metsHdr	RECORDSTATUS	REPLACEMENT		vcRECORDSTATUS_kbse	
			SUPPLEMENT			
			VERSION			
	agent	ROLE	CREATOR		SWEIPB	
			ARCHIVIST		SWEIPB	
				Riksarkivet/MKC		
	name	ROLE			SWEIPB	
		TYPE	ORGANIZATION		SWEIPB	
	name			Kungliga biblioteket		
	dmdSec	LABEL	Primary		vcDmdSec_LABEL_kbse	
			Local		vcDmdSec_LABEL_kbse	
	fileGrp	USE				
	file		image/master		vcUSE_kbse	bildfil endast avsedd för långtidslagring
			image/reference		vcUSE_kbse	bildfil avsedd för visning (ska användas när masterfil även används för visning)
			image/dynamic		vcUSE_kbse	bildfil med dynamiska egenskaper avsedd för visning (ska användas när masterfil även används för visning)
			text/alto		vcUSE_kbse	ALTO
			text/performance		vcUSE_kbse	Fil som innehåller resultat av kvalitetsmätning
			text/pdf		vcUSE_kbse	textfil i formatet pdf
			<i>text/metadata</i>		vcUSE_kbse	<i>(tillagt 20110301:) fil som innehåller metadata, t.ex. originalmetadaa från leverantören. Internt för KB</i>
		MIMETYPE			MIME type, <a href="http://www.iana.org/assignments/media-types/">http://www.iana.org/assignments/media-types/</a>	
			image/jp2			
			text/xml			

	structMap	TYPE				
			physical		vcSTRUCTMAP_TYPE_kbse	
			logical		vcSTRUCTMAP_TYPE_kbse	
	div	TYPE				
			files		vcDiv_TYPE_kbse	
			issue		vcDiv_TYPE_kbse	
			pdf		vcDiv_TYPE_kbse	
			section		vcDiv_TYPE_kbse	
			page		vcDiv_TYPE_kbse	
			supplement		vcDiv_TYPE_kbse	
			newsbill		vcDiv_TYPE_kbse	
			performance		vcDiv_TYPE_kbse	
			<i>origmetadata</i>		vcDiv_TYPE_kbse	<i>tillagt 20110301. Internt för KB</i>
	div	LABEL	missingpage		vcDiv_LABEL_kbse	"tom sida" ersätter saknad sida
			missingissue		vcDiv_LABEL_kbse	"tom sida" ersätter ett saknat tidningsnummer
			damagedpage		vcDiv_LABEL_kbse	sida med "trasigt" innehåll
			misplaced		vcDiv_LABEL_kbse	sida som är felplacerad/okänd - sorteras in sist
<b>MODS</b>						
	edition			[editionsbeteckning]	Unik lista för varje värdpublikation	
	dateIssued	qualifier	approximate		MODS	a date that may not be exact, but is approximated, such as "ca. 1972".
			inferred		MODS	a date that has not been transcribed directly from a resource, such as "[not before 1852]".
			questionable		MODS	a questionable date for a resource, such as "1972?".
	digitalOrigin			reformatted digital	MODS	Digitalisering från tryckt förlaga
				digitized microfilm	MODS	Digitalisering från mikrofilm
				born digital	MODS	
	note	type	reproduction	[text]	Unik lista byggs upp efter behov	

			script	gothic	vcScriptType_kbse	Texten i förlagan är i fraktur-stil.
				roman	vcScriptType_kbse	Texten i förlagan är i antikva-stil.
				mixed	vcScriptType_kbse	Texten i förlagan innehåller både fraktur och antikva
			condition	[Position 4 i prepareringskod fr Signe]		Notering om förlagens fysiska kondition
	identifier	type	uri	[en uri]	MODS	URL, URN, etc.
			local	[ospecificerad]	MODS	
			issn		MODS	
			reel number	[microfilmens nr]	Signe	
	physicalLocation			[Bibliotekssigel]		Exempel: S
	topic			[Kategoriord som beskriver bilagans innehåll]	bilagetyyp_kbse	Lokal ämneslista
	genre			supplement	genreterm_kbse	
				section	genreterm_kbse	
				newsbill	genreterm_kbse	
				project	genreterm_kbse	
				issue	marcgt	
				newspaper	marcgt	
	roleTerm			supplier	roleterm_kbse	
				publisher	marcrelator	
<b>PREMIS</b>						
	formatName			JPEG2000	PRONOM, <a href="http://www.nationalarchives.gov.uk/PRONOM/">http://www.nationalarchives.gov.uk/PRONOM/</a>	
				Extensible Markup Language		
	formatRegistryName			PRONOM	PRONOM, <a href="http://www.nationalarchives.gov.uk/PRONOM/">http://www.nationalarchives.gov.uk/PRONOM/</a>	

	formatRegistryKey			x-fmt/392	PRONOM, <a href="http://http://www.nationalarchives.gov.uk/PRONOM/">http://http://www.nationalarchives.gov.uk/PRONOM/</a>	
				fmt/101		
<b>MIX</b>						
	compressionScheme			Uncompressed	NISO Z39.87-2006	(Urval). Om ytterligare värden måste användas krävs en egen lista
				LZW		
				JPEG Baseline sequential		
				JPEG 2000 lossy		
				JPEG 2000 lossless		
	colorSpace			RGB	NISO Z39.87-2006	(Urval). Om ytterligare värden måste användas krävs en egen lista
				sRGB		
				CIE Lab		
				Lab		
				CMYK		
				YCbCr		
				Other		(ersätt med lämpligt värde)
	iccProfileName				[Unik lista tillsammans med versionsnumret]	Beror på arbetsflöde hos MKC och restriktioner i JPEG 2000
	iccProfileVersion					
	codecVersion				[Unik lista som byggs ut efter behov]	
	codecVersion					
	sourceType			[Unik lista, samma värden som digitalOrigin]	MODS/MIX	reformatted digital, digitized microfilm, born digital
	orientation				<a href="http://www.loc.gov/standards/mix/mix20/mix20.xsd">http://www.loc.gov/standards/mix/mix20/mix20.xsd</a>	
				normal		
				normal, image flipped		

				normal, rotated 180		
				normal, image flipped, rotated 180		
				normal, image flipped, rotated cw 90		
				normal, rotated ccw 90		
				normal, image flipped, rotated ccw 90		
				normal, rotated cw 90		
				unknown		
						cw=clockwise; ccw=counterclockwise;
	imageProducer			MKC Riksarkivet		
	scannerManufacturer				[Unik lista som byggs ut efter behov]	
	scannerModelName				[Unik lista som byggs ut efter behov]	
	scannerModelNumber				[Unik lista som byggs ut efter behov]	
	scanningSystemSoftwareName				[Unik lista som byggs ut efter behov]	
	scanningSystemSoftwareVersionNo				[Unik lista som byggs ut efter behov]	
	digitalCameraManufacturer				[Unik lista som byggs ut efter behov]	
	digitalCameraModelName				[Unik lista som byggs ut efter behov]	
	digitalCameraModelNumber				[Unik lista som byggs ut efter behov]	
	isoSpeedRatings				ISO12232	
	exifVersion				EXIF 2.2	
	lightsource				EXIF 2.2	



	autofocus				EXIF 2.2	
	samplingFrequencyUnit			Siffrorna 1, 2, 3	NISO Z39.87-2006	1 = no unit, 2 = inch, 3 = centimeter
	targetType			Siffrorna 0 och 1	NISO Z39.87-2006	0=external, 1=internal
	targetManufacturer				[Unik lista som byggs ut efter behov]	
	targetName				[Unik lista som byggs ut efter behov]	
	processingAgency				MKC Riksarkivet	
	processingRationale				[Unik lista som byggs ut efter behov]	
	processingSoftwareName				[Unik lista som byggs ut efter behov]	
	processingSoftwareVersion				[Unik lista som byggs ut efter behov]	
	processingActions				[Unik lista som byggs ut efter behov]	
<b>ALTO</b>						
	processingAgency			MKC Riksarkivet		
	processingStepSettings				[Unik lista som byggs ut efter behov]	
	processingSoftware				[Unik lista som byggs ut efter behov]	
	softwareCreator				[Unik lista som byggs ut efter behov]	
	softwareName				[Unik lista som byggs ut efter behov]	
	softwareVersion				[Unik lista som byggs ut efter behov]	
	applicationDescription				[Unik lista som byggs ut efter behov]	
	processingAgency			MKC Riksarkivet		

	processingStepSettings				[Unik lista som byggs ut efter behov]	
	processingSoftware				[Unik lista som byggs ut efter behov]	
	QUALITY_DETAIL				[Unik lista som byggs ut efter behov]	
<b>JPEG2000</b>	XML-box					Ska innehålla de parametrar som användes när filen skapades.
<b>Noteringar</b>						
	ISO8601			yyyymmdd		Flera olika formatteringar används i standarden, aktuellt här är det som benämns "basic"
	w3cdf			yyyy-mm-dd		Profil inom ISO8601 som lägger till "-" mellan siffrorna, <a href="http://www.w3.org/TR/NOTE-datetime">http://www.w3.org/TR/NOTE-datetime</a>

## Preliminär filnamngivning för Digidaily v.1.0

### Ändringshistorik

Datum	Ändring	Ändrad av

Samtliga filnamn ska börja med prefixet bib (filnamn som börjar med en siffra är ej tillåtna).

*libris* = librisnummer för värdpublikationen (alltså inte posten för det digitala numret eller för den tryckta förlagan), föregången av prefixet "bib".

*ååååmmdd* = datum för det tryckta tidningsnumret, såsom specificerat i fältet date i METS-filen

*edition* = editionsbeteckning enligt unik. Listan konstrueras av tidningsenheten vid KB. Den ursprungliga editionsbeteckningen kan ej lagras i filnamnet då beteckningen ofta innehåller tecken som ej är tillåtna (t.ex. "\*""). För varje värdpublikation skapas därför en översättning mellan editionsbeteckning och ett löpnummer. En editionsbeteckning måste alltid anges i samtliga filnamn. I projekt Digidaily digitaliseras normalt huvudeditionen. Om editionsbeteckning saknas ska edition alltid anges med siffran 0 om inga andra instruktioner ges. Om fler än en edition digitaliseras kommer instruktioner alltid att åtfölja tidningarna om annan tidning än huvudeditionen saknar editionsbeteckning. Detta betyder att MKC ej behöver bokföra de fall då editionsbeteckning saknas eftersom detta arbete görs vid KB.

*nummer* = tidningens ordningsnummer, såsom specificerad i fältet part/number i METS-filen. Om nummerbeteckning saknas anges bokstaven s.

### **METS-fil**

I Mimer (KB:s lagringsplattform som just nu utvecklas) vill man ha en enhetlig namngivning för de filer som innehåller metadata så att systemet lätt ska känna igen dessa. Namnen ska vara utförda så att metadatafilen inte ska kunna förväxlas med andra filer. Man har därför beslutat att ersätta filändelsen ".xml", med filändelsen ".[namn på metadatastandard].metadata".

libris\_ååååmmdd\_edition\_nummer.mets.metadata

Exempel

bib4112678\_18760203\_1\_24.mets.metadata (normalfallet)

bib4112678\_18760203\_0\_24.mets.metadata (saknad editionsbeteckning på huvudedition)

bib4112678\_18760203\_1\_s.mets.metadata (saknad nummerbeteckning)

### **Bildfil**

libris\_ååååmmdd\_edition\_nummer\_löpnummer\_filversion.jp2

löpnummer = sekvensnummer för bilden i tidningsnumret  
filversion = m och r (för master och reference). Om master- och visningsfil kombineras så  
namnges bildfilen som masterfil.

Exempel

bib4112678\_18760203\_1\_24\_8\_m.jp2

### ***Pdf-fil***

libris\_ååååmdd\_edition\_nummer\_pdf.pdf

Exempel

bib4112678\_18760203\_1\_24\_pdf.pdf

### ***ALTO-fil***

libris\_ååååmdd\_edition\_nummer\_löpnummer\_alto.xml

Exempel

bib4112678\_18760203\_1\_24\_8\_alto.xml

### ***Performance-fil***

(Denna filtyp innehåller information om bildkvalitén)

libris\_ååååmdd\_edition\_nummer\_performance.xml

Exempel

bib4112678\_18760203\_1\_24\_performance.xml

### ***ObjektID***

Varje paket måste ha ett unikt objektID som registreras i attributet OBJID i fältet mets i Mets-  
filen. Detta id har samma struktur som filnamnet.

libris\_ååååmdd\_edition\_nummer

Exempel

bib4112678\_18760203\_1\_24

## Instruktion för saknad sida/saknat nummer

### Ändringshistorik

Datum	Ändring	Ändrad av

### **Saknad sida**

Om en sida saknas ska en bild infogas som innehåller tidningens titel, publiceringsdatum samt texten ”Saknad sida”. Om möjligt ska även sidnummer inkluderas i bilden. Bildfilen bör ha samma dimensioner som övriga bilder som härrör från tidningsnumret.

Den infogade bilden behandlas som en vanlig bildfil, både med avseende på metadata och med på filnamn. Viss teknisk metadata kommer naturligtvis att saknas. I structMap tilldelas sidan en LABEL med värdet ”missingpage”.

En ALTO-fil skapas för den saknade sidan. Filen behöver endast inkludera den data som lagras i taggen <sourceImageInformation>.

### **Saknat nummer**

Om ett helt nummer saknas ska ändå en SIP konstrueras enligt specifikationen. Paketet ska innehålla en enda bildfil som visar tidningens titel, publiceringsdatum samt texten ”Saknat nummer”. Bildfilen bör ha snarlika dimensioner som bildfiler som härrör från den utrustning där tidningen sannolikt skulle ha bildfångats. All bibliografisk metadata på nummernivå hanteras som vanligt. Eftersom den enda bildfilen inte är bildfångad på vanligt vis måste dock en del metadatafält att lämnas tomma (t.ex. teknisk metadata).

Den infogade bilden behandlas som en vanlig bildfil. Som edition i filnamnet anges huvudeditionen om inget annat meddelats. I structMap tilldelas bildfilen en LABEL med värdet ”missingissue”.