

ADDML

Archival Data Description Markup Language

Utvidete betingelser for bruk i Arkivverket, Norge

Versjon PA 1.1

Sist oppdatert: 10. juni 2020 av TPD

| | |
|--|----|
| Innledning | 4 |
| Forutsetninger..... | 4 |
| Den norske pakkemodellen | 4 |
| Hvordan benytte ADDML | 5 |
| Språkforvirring!? | 5 |
| Endringer i dokumentet | 5 |
| Arkivverkets bruk av ADDML | 6 |
| Ett og bare ett datasett..... | 6 |
| addml..... | 6 |
| Egenskaper for elementet flatFile..... | 6 |
| fileName | 7 |
| numberOfOccurrences | 8 |
| checksum..... | 8 |
| algorithm | 9 |
| value | 9 |
| flatFile | 9 |
| Tillatte verdier for elementet charset | 9 |
| Tillatte verdier for elementet dataType | 9 |
| string | 10 |
| integer | 10 |
| decimal | 11 |
| date..... | 11 |
| boolean | 12 |
| link | 12 |
| Tillatte verdier for elementet alignment | 13 |
| Tilleggselementer for elementet context..... | 13 |
| Tilleggselementer for elementet content..... | 16 |
| Tillatte verdier for elementet recordSeparator | 19 |
| Nummerering | 20 |
| Prosesser | 20 |
| Analyser | 20 |
| Kontroller | 20 |
| Konvertering..... | 21 |

| | |
|-----------------------------|----|
| Andre typer prosesser | 22 |
| Eksempel | 23 |
| Dataobjekter | 23 |

Innledning

Dette dokumentet beskriver det norske arkivverkets standard for teknisk metadata - Archival Data Description Markup Language (ADDML) versjon 8.3 og den profilen arkivverket benytter av standarden, navngitt 8.3.1.

Dokumentet fungerer som et supplement til den generelle beskrivelsen. Der hvor profilen angir et sterkere krav enn standarden generelt, vil dette bli angitt etter samme prinsipper som i den generelle beskrivelsen.

Forutsetninger

Dette dokumentet forklarer hvordan ADDML skal benyttes ved innlevering av materiale til Arkivverket. Denne beskrivelsen vil også kunne være veiledende ved innlevering av materiale til kommunale arkiver.

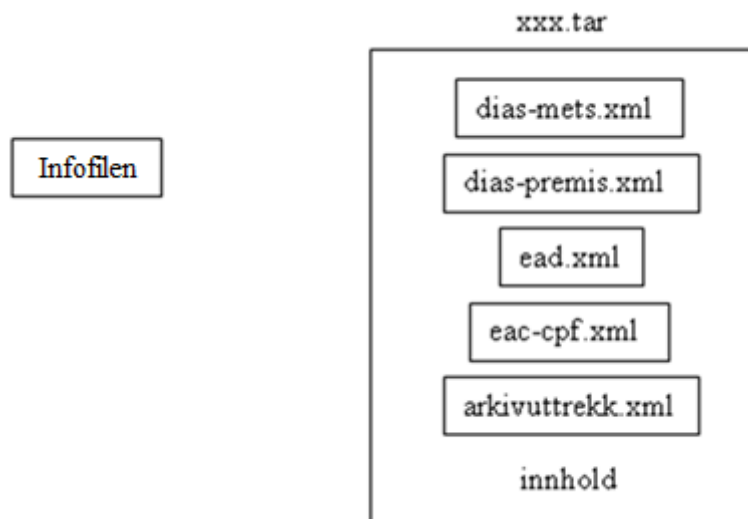
For å kunne forstå innholdet anbefales det å lese dokumentet om ADDML generelt først for å få et grunnlag for hva ADDML er og hvordan den er bygget opp.

Den norske pakkemodellen

Gjennom DIAS-prosjektet som var et samarbeid mellom Arkivverket og kommunale arkiver, ble det vedtatt en modell for arkivpakker (AIP). Denne modellen er siden av Arkivverket blitt benyttet til også å definere en modell for en innleveringspakke (SIP). De to modellene er dermed noenlunde like i grunnstruktur.

Selve innholdet og dets metadata pakkes inn i en tar-fil. I tillegg overføres også en fil som kalles infofilen som inneholder informasjon om tar-filen og informasjon knyttet til overføringen til depotet.

I selve pakken benyttes standardene dias-mets som er en profil av den internasjonale standarden mets og inneholder en oversikt over alle filer som inngår i pakken. Også infofilen er en profil av mets. I pakken følger dessuten dias-premis som er en profil av den internasjonale standarden og inneholder informasjon om hendelser og rettigheter til innholdet. Dessuten kan det forekomme filer som følger standardene ead og eac-cpf som inneholder arkiv- og aktørbeskrivelse. (Disse er valgfrie for en SIP.) Og endelig kommer den tekniske strukturbeskrivelsen i filen arkivuttrekk.xml som skal følge ADDML For ADDML må derfor hvert depot også definere sin profil av standarden og dette dokumentet beskriver en slik profil som da benyttes for overføring og i Arkivverket.



Figur 1. Innholdet i en pakke.

Hvordan benytte ADDML

ADDML benyttes for å beskrive datasett som avleveres eller deponeres i Arkivverket. Etter dagens regelverk er ADDML kun påkrevd for avlevering/deponering av uttrekk fra fagsystem (ikke Noark system) og for uttrekk som følger Noark. Det har imidlertid vært planer om at filen arkivuttrekk.xml skal medfølge for alle typer avleveringer/deponeringer til Arkivverket. Med endring av arkivloven kan det dog være at kravene blir helt annerledes.

I ADDML kan man legge inn prosesser som fungerer som instruksjoner for et testverktøy. I dagens versjon av Arkade (5.0) er det valgfritt om programmet bruker disse eller om de skal forholde seg til brukergrensesnittet. Alle prosessene er beskrevet i et eget kapittel i dette dokumentet, slik at det fortsatt kan være mulig å benytte seg av instruksjoner i ADDML-filen.

Språkforvirring!?

I dette dokumentet forekommer det både engelske og norske begreper. Det er imidlertid ikke så tilfeldig som det kan virke.

ADDML standarden er laget i engelsk språkdrakt for at det skal være mulig også for andre å kunne benytte standarden. Dette betyr at alle tagger som hører til ADDML er engelske termer. Dette inkluderer også tagger definert for norske formål, for å holde det konsistent. Derimot vil verdiene være på norsk, med unntak for gitte kodesett – de som blir ansett som en del av standarden.

Endringer i dokumentet

Fra versjon 1.0 til 1.1:

- Korrigert skrivefeil

Arkivverkets bruk av ADDML

Denne profilen vil sette en del begrensninger som ikke er i den generelle standarden. Disse begrensningene er beskrevet i dette kapitlet. Noen av begrensningene vil dog referere til andre kapitler, hvor begrensningene vil bli nærmere utdypet. Dette gjelder spesielt for de stedene hvor ADDML er generisk oppbygd, dvs properties, additionalElements, dataObjects og processes.

Begrensningene er som følger:

1. En beskrivelse skal beskrive ett og kun ett datasett
2. Elementet flatFile skal ha egenskaper (properties). (Hvilke egenskaper er beskrevet nedenfor.)
3. Elementet charset kan kun inneholde forhåndsdefinerte verdier.
4. Elementet dataType kan kun inneholde forhåndsdefinerte verdier.
5. Elementet alignment kan kun inneholde forhåndsdefinerte verdier.
6. Elementet context skal ha tilleggselementer (additionalElements). (Hvilke elementer er beskrevet nedenfor.)
7. Elementet content skal ha tilleggselementer (additionalElements). (Hvilke elementer er beskrevet nedenfor.)
8. Elementet recordSeparator kan kun inneholde forhåndsdefinerte verdier.
9. Nummerering

Noen av begrensningene vil også medføre begrensninger også i andre element.

Ett og bare ett datasett

I arkivverket ønskes det å ha rene linjer, slik at det blir en beskrivelse til hvert datasett. Dette medfører at vi ikke vil tillate at en ADDML-fil vil beskrive mer enn ett datasett. Samtidig skal det alltid være beskrivelse av ett datasett i en fil.

Dette kravet vil medføre endringer for elementet addml.

addml

Siden elementet addml alltid skal ha ett og kun ett element datasett under seg, vil dette medføre at forholdet mellom addml og datasett blir 1 til 1.

| Underliggende elementer | Kravtype | Forekomster |
|-------------------------|-----------|-------------|
| datasett | mandatory | 1 |

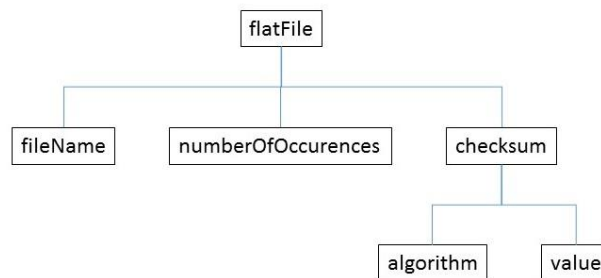
Egenskaper for elementet flatFile

Det er bestemt at elementet skal ha underliggende egenskaper (properties). Egenskapene er enten påkrevd eller ikke. Det er tillatt å ha med flere egenskaper enn de som her er definert.

Følgende egenskaper skal være med:

- **fileName**
Det fysiske filnavnet til filen.
- **numberOfOccurrences**
Angivelse av det totale antall poster i en fil.
- **checksum**
Sjekksuminformasjon (overordnet egenskap).
 - **algorithm**
Algoritme for beregning av sjekksum.
 - **value**
Verdien til sjekksummen

Strukturen for egenskapene skal følge strukturen i figur 2.



Figur 2. Egenskaper for flatFile.

fileName

I egenskapen **fileName** skal man angi filnavnet til den fysiske filen. Navnet skal angis med sti. Alternativt kan man angi to under egenskaper, en med selve filnavnet og en med stien.

Eksempel:

```
<properties>
  <property name="fileName">./ArkivA/dias-mets.xml</property>
  ...
</properties>
```

alternativt

```
<properties>
  <property name="fileName">
    <properties>
      <property name="path">./ArkivA/</property>
      <property name="name">dias-mets.xml</property>
    </properties>
  </property>
  ...
</properties>
```

numberOfOccurrences

I egenskapen numberOfOccurrences skal man angi antallet poster i filen man omtaler.

Hvis det er forskjellige posttyper i filen, kan man i tillegg ha tilsvarende egenskaper – en for hver posttype, hvor man angir posttypen og antallet poster for denne.

Eksempel 1:

```
<properties>
  ...
  <property name="numberOfOccurrences">1214</property>
  ...
</properties>
```

Eksempel 2:

```
<properties>
  ...
  <property name="numberOfOccurrences">1214</property>
  <property name="numberOnType">
    <properties>
      <property name="type">Alfa</property>
      <property name="numberOfOccurrences">608</property>
      <property name="type">Beta</property>
      <property name="numberOfOccurrences">606</property>
    </properties>
  </property>
  ...
</properties>
```

checksum

I egenskapen checksum skal man angi sjekksummen for filen man beskriver. Siden man trenger informasjon både om algoritmen og verdien, vil det være to under egenskaper for checksum.

Eksempel:

```
<properties>
  <property name="checksum">
    <properties>
      <property name="algorithm">SHA-286</property>
      <property name="value">[Selve sjekksummen]</property>
    </properties>
  </property>
</properties>
```

algorithm

I egenskapen algorithm skal man angi hvilken algoritme som er benyttet for å beregne sjekksummen. Til Arkivverket er det krav om at denne algoritmen skal være SHA-286.

value

I egenskapen value skal man angi selve verdien til sjekksummen.

flatFile

Kravet om egenskaper vil medføre endringer for elementet flatFile. I tabellen under er det justert for kravene i henhold til dette notatet.

| Underliggende elementer | Kravtype | Forekomster |
|-------------------------|----------|-------------|
| properties | optional | 1 |

Tillatte verdier for elementet charset

I elementet charset tillates kun følgende verdier for overføring av en arkivversjon til Arkivverket:

- ISO-8859-1
- ISO-8859-4
- UTF-8

Andre tegnsett må avtales med Arkivverket for å være tillat.

Til andre arkivinstitusjoner kan listen av gyldige verdier være annerledes.

Tillatte verdier for elementet dataType

I elementet dataType tillates kun følgende verdier for overføring av en arkivversjon til Arkivverket:

- string
- integer
- decimal
- date
- boolean
- link

Andre tegnsett må avtales med Arkivverket for å være tillat.

Til andre arkivinstitusjoner kan listen av gyldige verdier være annerledes.

For hver aksepterte datatype vil det kun være ett sett av formater som kan aksepteres. Hver av disse kombinasjonene kan så medføre at det blir behov for å definere både regler og prosesser for å kontrollere gyldigheten av kombinasjonen.

string

Det er ingen formater knyttet direkte til string som omfattes av dette punktet.

Eksempel for et vanlig tekstfelt:

```
<dataType>string</dataType>
```

Imidlertid er det likevel noen situasjoner som skal nevnes.

Noen ganger kan det være spesielle ”felt” som forefinnes. Noen av disse spesialfeltene vil det være mulig å legge på noen kontroller. Her skal nevnes:

- Fødselsnummer - fnr
- Organisasjonsnummer - org
- Kontonummer - knr

Alle disse tre har så mange tegn, at de i visse systemer ikke kan defineres som numeriske. I tillegg skal disse feltene alltid ha et gitt antall siffer og dermed kunne ha ledende nuller.

Angivelse av disse spesialvariantene gjøres ved å benytte taggen for format hvor kortversjonen (fnr, org eller knr) angis som verdier for de respektive.

Eksempel for et organisasjonsnr:

```
<dataType>string</dataType>
```

```
<fieldFormat>org</fieldFormat>
```

Det vil bli definert en prosess for hver av disse spesialvariantene. Se under prosesser – andre prosesser.

integer

Alle integer-felt skal være numeriske.

Unntak for dette er følgende:

- Eksponensial verdier (av 10) skal noteres på formen <tall>E+<eksponent>
- Tusenskilte (se nedenfor)

Eksponensial verdier skal benytte taggen fieldFormat for å angi formen som skal være nnEeks. Hvor nn betegner verdien i feltet, E benyttes som skille mellom tall og eksponent, og eks angir eksponentverdien. Det er ingen korrelasjon mellom lengdene på nn og eks og de faktiske lengdene på disse verdiene.

Eksempel for et eksponensialtall:

```
<dataType>integer</dataType>
```

```
<fieldFormat>nnE+exp</fieldFormat>
```

Og hvor selve verdien da kan være f.eks. 4E+5 som betyr 400000.

Fortegn skal kun angis for negative tall med – tegnet tett inntil før tallet. Det vil ikke være behov for å benytte fieldFormat i dette tilfellet.

Eksempel:

```
<dataType>integer</dataType>
```

Et tall kan ha ledende nuller eller ikke, dette er valgfritt. Det vil uansett ikke være behov for andre tagger enn dataType.

Tusenskilte bør ikke være benyttet, men kan aksepteres etter avtale. Tegnet for tusenskilte skal være punktum.

Eksempel for et tusenskilte med punktum:

```
<dataType>integer</dataType>  
<fieldFormat>n.nnn</fieldFormat>
```

Justering bør være høyrestilt, men andre varianter aksepteres. Eventuelle blanke tegn fjernes under behandlingen. Mulige varianter er høyre, venstre og midt. Dersom justeringen er høyrestilt, kan taggen sløyfes. Dersom justering ikke er med, forventes at feltet er høyrestilt.

Eksempel på høyrejustering:

```
<alignment>right</alignment>
```

decimal

Alle desimalfelt skal følge reglene for heltall.

Desimaltegnet bør være ”,” (komma). Selve tegnet som benyttes som desimaltegn angis ved hjelp av fieldFormat taggen.

Eksempel for desimaltall (komma som desimaltegn):

```
<dataType>float</dataType>  
<fieldFormat>nn,nn</fieldFormat>
```

Det skal ikke være behov for å tagge tusenskilte i fieldFormat i utgangspunktet. I de tilfellene dette benyttes likevel, skal tegnet benyttet som tusenskilte angis.

Eksempel for et tusenskilte for desimaltall (punktum som tusenskilte og komma som desimaltegn):

```
<dataType>float</dataType>  
<fieldFormat>n.nnn,nn</fieldFormat>
```

date

Datofelt skal i utgangspunktet være satt sammen av en datodel og en tidsdel. Imidlertid kan tidsdelen bortfalle slik at bare datodelen forefinnes.

Datodelen skal bestå av 3 elementer samt skilletegn. De tre elementene er dag, måned og år. Dag skal alltid beskrives med dd i formatet, måned med MM (siffer) eller MMM (bokstaver) og år med yy (2-sifret) eller yyyy (4-sifret). Rekkefølgen på elementene angis i formatet. Dersom skilletegn mellom elementene er benyttet, angis også dette i formatet.

Tidsdelen skal bestå av 3 elementer samt skilletegn. De tre elementene er time, minutt og sekund. Time skal alltid beskrives med HH i formatet, minutt med mm og sekund med ss, alle 2-sifret. Rekkefølgen skal være i nevnte orden og elementene skal skilles ad med kolon (:). I tillegg skal tidsdelen startes med T for å angi skille mellom datodel og tidsdel, og avsluttes med zzz for å angi Europeisk tid. Elementet sekund er valgfritt.

Eksempel for en dato med tidsdel og skilletegn:

```
<dataType>date</dataType>
```

```
<fieldFormat>dd.MM.yyyyTHH:mm:sszzz</fieldFormat>
```

Det vil dermed bli definert en prosess som kontrollerer at verdiene alltid er i henhold til det oppgitte formatet. Se under prosesser – andre prosesser.

Husk også at om man ønsker å dele opp et date element, kan man benytte seg av fieldParts for å skille ut de enkelte bestanddelene. De enkelte bestanddeler vil ikke være definert som date, men som heltall, dvs integer.

boolean

Boolske verdier angis med formatet T/F. Her angir T ”true”-verdien og F ”false”-verdien. De skal alltid skilles med / (skråstrek/slash).

Eksempel for en boolsk variabel:

```
<dataType>boolean</dataType>
```

```
<fieldFormat>J/N</fieldFormat>
```

Det vil dermed bli definert en prosess som kontrollerer at verdiene alltid er en av de 2 oppgitte. Se under prosesser – andre prosesser.

link

Link vil bli behandlet som en string.

Link er typen når feltet inneholder en eller annen form for referanse.

Link kan dermed være:

- En referanse til et annet element (som en fremmednøkkel) - forkey
- En sti til et annet dokument - doc
- En internett adresse - www
- En URL - url
- osv.

I utgangspunktet vil fieldFormat ikke være benyttet for link. Dette kan imidlertid endre seg senere. Tanken kan være å oppgi typen link i formatfeltet. Ved bruk i dag vil formatfeltet bli oversett. Men det kan likevel benyttes ved å angi typen link (forkey, doc, www, url, ...).

For referanser til dokumenter skal elementet inneholde relativ filsti i tillegg til dokumentnavnet.

Tillatte verdier for elementet alignment

I elementet alignment tillates kun følgende verdier for overføring av en arkivversjon til Arkivverket:

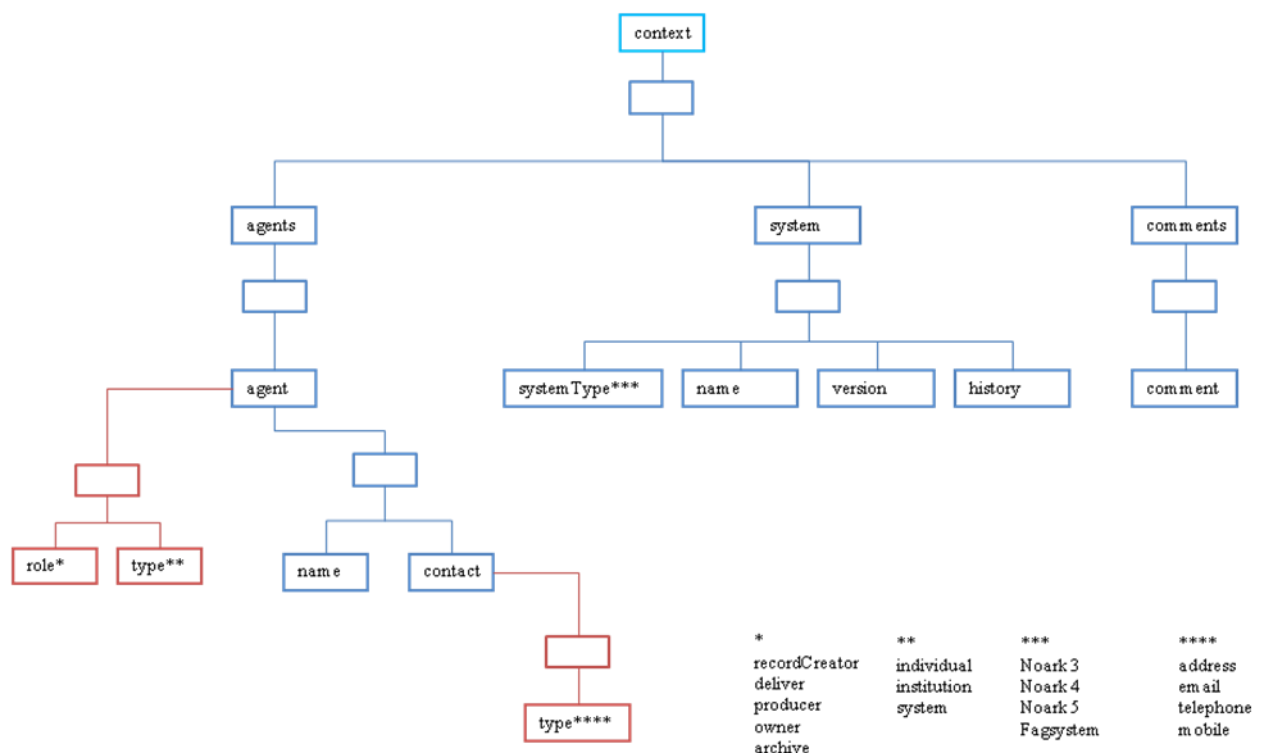
- left
- right
- center

Andre verdier for tekstjustering må avtales med Arkivverket for å være tillat.

Til andre arkivinstitusjoner kan listen av gyldige verdier være annerledes.

Tilleggselementer for elementet context

For elementet context er det krav om et sett av tilleggselementer. Disse tilleggselementene skal alle inneholde informasjon om materialets kontekst, altså informasjon knyttet til det opprinnelige system informasjonen er skap i, både hva systemet er og hvordan og av hvem det er benyttet. En oversikt over tilleggselementer som skal forefinnes i en arkivversjon er vist i figur 3. Deretter følger en nærmere beskrivelse av det enkelte tilleggselement.



Figur 3. En oversikt over foreslåtte elementer til context.

Under kontekstuelle elementer er det foreslått tre bolker med elementer. Flere av elementene kan også gjenfinnes som elementer i infofilen og i mets-filen.

- agents

1. agent

1. role (**property**)

Hvilken agent det er snakk om.

Mulige verdier er: recordCreator, deliver, producer, owner, archive.

(Verdiene kan endres over tid.)

2. type (**property**)

Hvilken type agent det er snakk om.

Mulige verdier er: individual, institution, system.

(Verdiene kan endres over tid.)

3. name

navn på agenten

4. contact

kontaktinformasjon til agenten

1. type (**property**)

Hva slags type informasjon som oppgis.

Mulige verdier er: address, email, telephone, mobile.

(Verdiene kan endres over tid.)

- system

Samleelement for opprinnelig system hos arkivskaper.

1. systemType (**property**)

Hvilken type system det er snakk om.

Mulige verdier er: Noark 3, Noark 4, Noark 5, Fagsystem.

(Verdiene kan endres over tid.)

2. name

Navn på systemet.

(Her må vi bli enige om en navnekonvensjon.)

3. version

Versjon av systemet.

(Se kommentar under name.)

4. history

Informasjonens historie i systemet og tidligere system.

- comments

Samleelement for kommentarer.

1. comment
Eventuelt kommentar.

Eksempel

```
<context>
  <additionalElements>
    <additionalElement name="agents">
      <additionalElements>
        <additionalElement name="agent">
          <properties>
            <property role="recordCreator" type="institution">
          </properties>
          <additionalElements>
            <additionalElement name="name">[navn]</additionalElement>
          </additionalElements>
        <additionalElement name="agent">
          <properties>
            <property role="deliver" type="institution">
          </properties>
          <additionalElements>
            <additionalElement name="name">[navn]</additionalElement>
            <additionalElement name="contact">
              <properties>
                <property type="email">
                  <value>[Avgivers epostadresse]</value>
                </property>
              </properties>
            </additionalElement>
          </additionalElements>
        <additionalElement name="agent">
          <properties>
            <property role="deliver" type="individual">
          </properties>
          <additionalElements>
            <additionalElement name="name">[navn]</additionalElement>
            <additionalElement name="contact">
              <properties>
                <property type="email">
                  <value>[kontakpersonens epostadresse]</value>
                </property>
                <property type="mobile">
                  <value>[kontakpersonens mobilnummer]</value>
                </property>
              </properties>
            </additionalElement>
          </additionalElements>
        </additionalElement>
      </additionalElements>
    </additionalElement>
  </additionalElements>
  [Her vil ytterligere agenter være definert!]
</additionalElements>
</additionalElement>
```

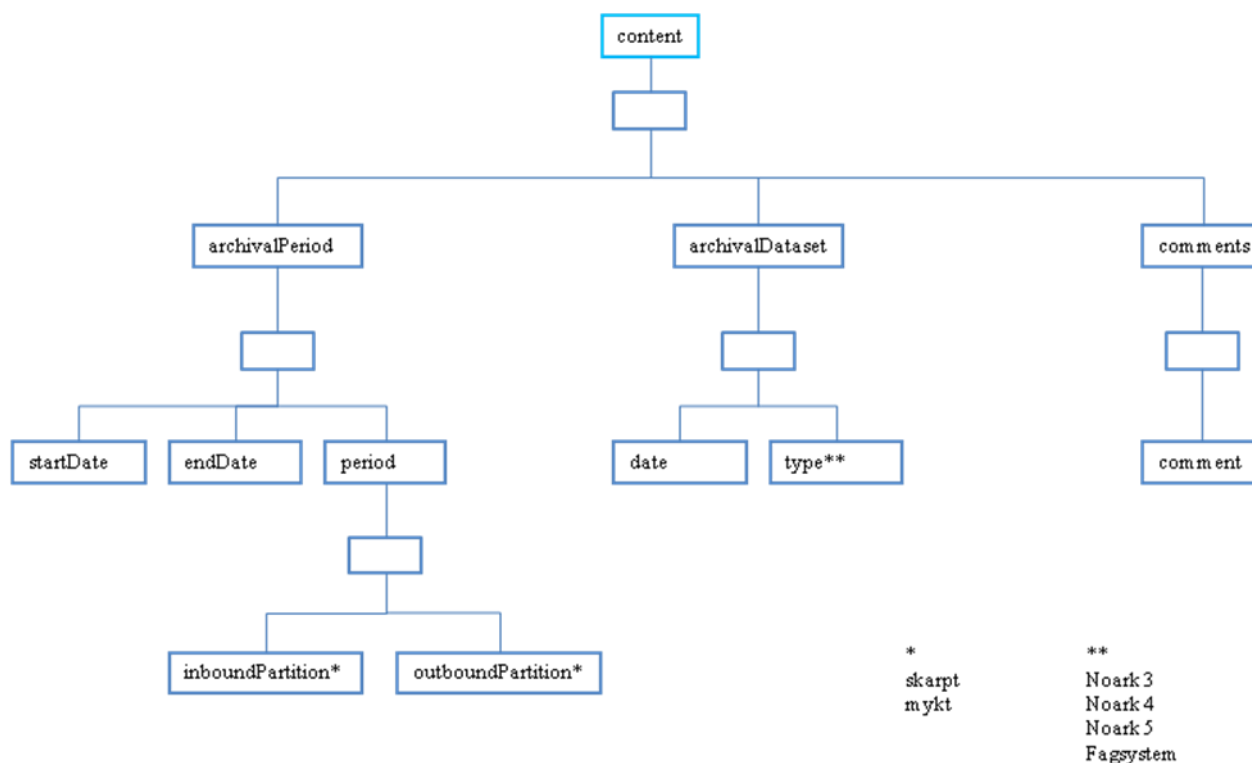
```

<additionalElement name="system">
  <additionalElements>
    <additionalElement name="systemType">[type]</additionalElement>
    <additionalElement name="name">[navn]</additionalElement>
    <additionalElement name="version">[versjon]</additionalElement>
    <additionalElement name="history">
      [Dette har hendt]
    </additionalElement>
  </additionalElements>
</additionalElement>
<additionalElement name="comments">
  <additionalElements>
    <additionalElement name="comment">
      [Kommentar]
    </additionalElement>
    <additionalElement name="comment">
      [Kommentar]
    </additionalElement>
  </additionalElements>
</additionalElement>
</additionalElements>
</context>

```

Tilleggselementer for elementet content

For elementet content er det også krav om et sett av tilleggselementer. Disse tilleggselementene skal alle inneholde informasjon om arkivuttrekket. En oversikt over tilleggselementer som skal forefinnes i en arkivversjon er vist i figur 4. Deretter følger en nærmere beskrivelse av det enkelte tilleggselement.




Figur 4. En oversikt over foreslåtte elementer i content.


I archivalPeriod finnes det elementer som angir start og sluttdato for arkivperioden som uttrekket omfatter. Dessuten informasjon om inngående og utgående skille for perioden, hvorvidt det er skarpt eller mykt. I archivalDataset finnes det to elementer som er uttrekksdato og uttrekksformatet, hvor valgene i det sistnevnte er Noark 3, Noark 4, Noark 5 og Fagsystem.

Dessuten vil det være mulig å legge til kommentarer i *comments*.

- archivalPeriod
Arkivperiode for det angjeldende uttrekket, Samleelement.

1. startDate
Startdato for informasjonen i uttrekket.
2. endDate
Sluttdato for informasjonen i uttrekket.
3. period
Periodeskiller, Samleelement.

 inboundPartition
Inngående skille for uttrekket
Mulige verdier er: Skarpt, Mykt.

 outboundPartition
Utgående skille for uttrekket
Mulige verdier er: Skarpt, Mykt.

- archivalDataset
Samleelement om uttrekket.
 1. date
Dato for når uttrekket er laget.
 2. type
Type av uttrekk.
Mulige verdier er: Noark 3, Noark 4, Noark 5, Fagsystem.
- comments
Samleelement for kommentarer.
 1. comment
Eventuell kommentar.

Eksempel

```

<content>
  <additionalElements>
    <additionalElement name = "archivalPeriod">
      <additionalElements>
        <additionalElement name="startDate">[dato]</additionalElement>
        <additionalElement name="endDate">[dato]</additionalElement>
        <additionalElement name="period">
          <additionalElements >
            <additionalElement name="inboundPartition">
              [Inngående skille]
            </additionalElement>
            <additionalElement name="outboundPartition">
              [Utgående skille]
            </additionalElement>
          </additionalElements>
        </additionalElement>
      </additionalElements>
    </additionalElement>
    <additionalElement name = "archivalDataset">
      <additionalElements>
        <additionalElement name="date">[dato]</additionalElement>
        <additionalElement name="type">[type]</additionalElement>
      </additionalElements>
    </additionalElement>
    <additionalElement name = "comments">
      <additionalElements>
        <additionalElement name="comment">
          [Kommentar]
        </additionalElement>
        <additionalElement name="comment">
          [Kommentar]
        </additionalElement>
      </additionalElements>
    </additionalElement>
  </additionalElements>
</content>

```

Tillatte verdier for elementet recordSeparator

I elementet recordSeparator tillates kun følgende verdier for overføring av en arkivversjon til Arkivverket:

- CRLF
- Uten postskilletegn
(Dette valget medfører at elementet utelates i xml-filen da dette er standardverdien.)

Andre postskiller som bare CR eller bare LF, må avtales med Arkivverket for å være tillatt.

Til andre arkivinstusjoner kan listen av gyldige verdier være annerledes.

Nummerering

All nummerering skal begynne på 1

Prosesser

For bruk i verktøyet Arkade 5 er det definert et sett av prosesser. For at en prosess skal bli aktivert i Arkade5, må det være angitt et flagg i addml-filen. I det etterfølgende er det satt opp en oversikt over de definert flaggene og hva slags prosess de vil initiere. Prosessene er delt inn i grupper avhengig av hva slags type prosess som skal utføres.

Analyser

Følgende prosessflagg er definert som analyse:

- Analyser på filnivå
 - Analyse_CountRecords
Teller opp antall poster i en fil.
 - Analyse_CountChars
Teller opp antall tegn i en fil.
- Analyser på postnivå
 - Analyse_FindExtremeRecords
Finner lengste og korteste post for hver posttype.
 - Analyse_CountRecordDefinitionOccurences
Teller opp antall poster for hver posttype.
 - Analyse_AllFrequenceList
Utfører en frekvensanalyse for samtlige felt hvor det er definert et kodesett.
 - Analyse_CrossTable
Utfører en kryss-sjekk mellom 2 felt i samme post.
- Analyser på feltnivå
 - Analyse_CountNULL
Teller opp antall forekomster av verdien null i feltet.
 - Analyse_FindExtremeValues
Finner lengste og korteste verdi i feltet.
 - Analyse_FindMinMaxValue
Finner laveste og høyeste verdi i feltet.
 - Analyse_FrequenceList
Teller opp antall forekomster for hver verdi i feltet.

Kontroller

Følgende prosessflagg er definert som kontroller:

- Kontroller på filnivå

- Control_AllFixedLength
Kontrollerer at oppgitt postlengde på samtlige posttyper er korrekt (kun fast format).
- Control_NumberOfRecords
Teller opp og kontrollerer at antall poster i filen er lik antallet som er oppgitt i egenskapen numberOfOccurrences i flatFile.
- Kontrollerer på postnivå
 - Control_FixedLength
Kontrollerer om oppgitt postlengde er korrekt (kun for fast format)
 - Control_NotUsedRecordDef
Kontrollerer om denne posttypen benyttes i datasettet.
 - Control_Key
Kontrollerer at nøkkelen (primary eller alternate) er unik.
 - Control_ForeignKey
Kontrollerer fremmednøkler for å sjekke at de går til en faktisk forekomst.
- Kontrollerer på feltnivå
 - Control_MinLength
Kontrollerer om oppgitt minste lengde faktisk er minste lengde i feltet.
 - Control_MaxLength
Kontrollerer om oppgitt største lengde faktisk er største lengde i feltet.
 - Control_DataFormat
Kontrollerer om oppgitt dataformat er korrekt.
 - Control_NotNull
Kontrollerer om det finnes null-verdier i feltet.
 - Control_Uniqueness
Kontrollerer om verdiene i feltet er unike.
 - Control_Codes
Kontrollerer om definerte koder benyttes og om det benyttes koder som ikke er definert i kodelisten.

Konvertering

Følgende prosessflagg er definert for konvertering:

- Kontrollerer på dataset og filnivå
 - Convert_Charset
Utfører konvertering fra ett tegnsett til et annet.
Som parameter angis det nye tegnsettet.
 - Convert_Format
Utfører konvertering fra ett filformat til et annet.
Som parameter angis det nye formatet.
 - Convert_RecordSeparator
Utfører konvertering fra ett postskille til et annet.
Som parameter angis det nye postskillet.

- Convert_PreserveStructure
Utfører konvertering ved å beholde filstrukturen som den var. Den vil altså beholde filen som en fil selv om det skulle være flere posttyper, hvor det vil være default å splitte disse. Den vil også beholde tegnsettet og feltene slik de er, selv om noen i utgangspunktet skulle vært pakket ut og konvertert til et gyldig tegnsett.
Prosessen har ingen parametre.
- Kontroller på postnivå:
 - Convert_AddKeyFields
Prosessen legger på nøkkelfelt ved splitting på flere posttyper.
Som parametre angis feltene som skal legges på med deres innbyrdes rekkefølge, samt hvor de skal plasseres i posttypen.
 - Convert_SplitRepeatingGroup
Splitter en posttype med repeterende gruppe opp i en fil med hoveddelen av posten og en fil med den repeterende gruppen med nøkkelfelt.
Som parametre angis nøkkelfeltene som skal legges på.
- Kontroller på feltnivå:
 - Convert_Dataformat
Konvertering av verdien i et felt fra en datatype til en annen.
Som parameter angis den nye datatypen.
 - Convert_PreservePacked
Konvertering hvor dette feltet beholdes i pakket form.
Prosessen har ingen parametre.

Andre typer prosesser

Følgende prosesser er definert som andre prosesser:

- Prosess på datasetnivå:
 - Create_Database
Utfører opprettelse av en database og lasting av dataene inn i denne.
Prosessen har ingen parametre.
- Prosesser på feltnivå:
 - Control_Birthno
Kontroll av kontrollsiffrerne i et fødselsnummer.
Prosessen har ingen parametre, men forutsetter at dataType er *string*.
 - Control_Organisationno
Kontroll av kontrollsiffrerne i et organisasjonsnummer.
Prosessen har ingen parametre, men forutsetter at dataType er *string*.
 - Control_Accountno
Kontroll av kontrollsiffrerne i et kontonummer.
Prosessen har ingen parametre, men forutsetter at dataType er *string*.
 - Control_Date_Value
Kontroll av verdiene i et dato-felt.
Prosessen har ingen parametre, men forutsetter at dataType er *date*.

- Control_Boolean_Value
Kontroll av verdiene i et boolsk felt.
Proessen har ingen parametre, men forutsetter at dataType er *boolean*.

Eksempel

Eksempel på prosesser som instruksjoner i ADDML-filen:

```
<flatFileProcesses flatFileReference="fileA">  
  <recordProcesses definitionReference="recordB">  
    <processes>  
      <process name="Analyse_FindMinMaxRecordLength">  
        <process name="Control_Key">  
      </processes>  
    </recordProcesses definitionReference="fieldC">  
      <processes>  
        <process name="Analyse_FindMinMaxValues">  
          <process name="Control_DataFormat">  
        </processes>  
      </fieldProcesses>  
    </recordProcesses>  
  </flatFileProcesses>
```

Dataobjekter

I addml kan man også angi dataobjekter for andre typer filer enn flate filer. Det er ikke satt noen krav til dataObjects eller dataObject i denne profilen.

Hvordan man kan bruke dataobjekter er beskrevet i et eget notat – *Bruk av komponenter i ADDML*. For nærmere informasjon om dataobjekter og bruk av disse henvises til dette notatet.