



ARKIVVERKET
RIKSARKIVET OG STATSARKIVENE

ADDML

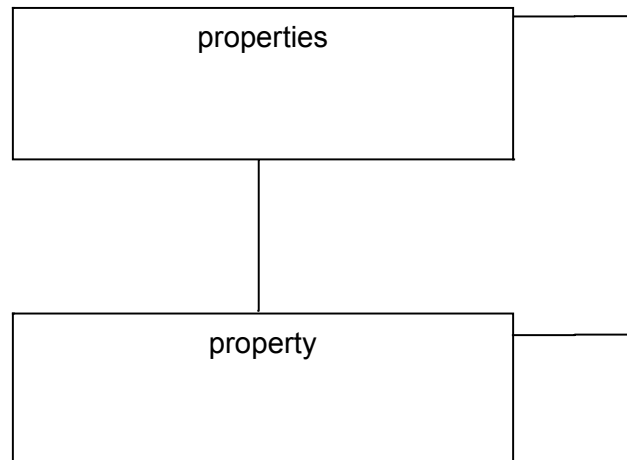
Status og videre arbeid

Terje Pettersen-Dahl
Seksjon for Digitalt Depot
Riksarkivet

KDRS konferansen
Trondheim 13. november 2014



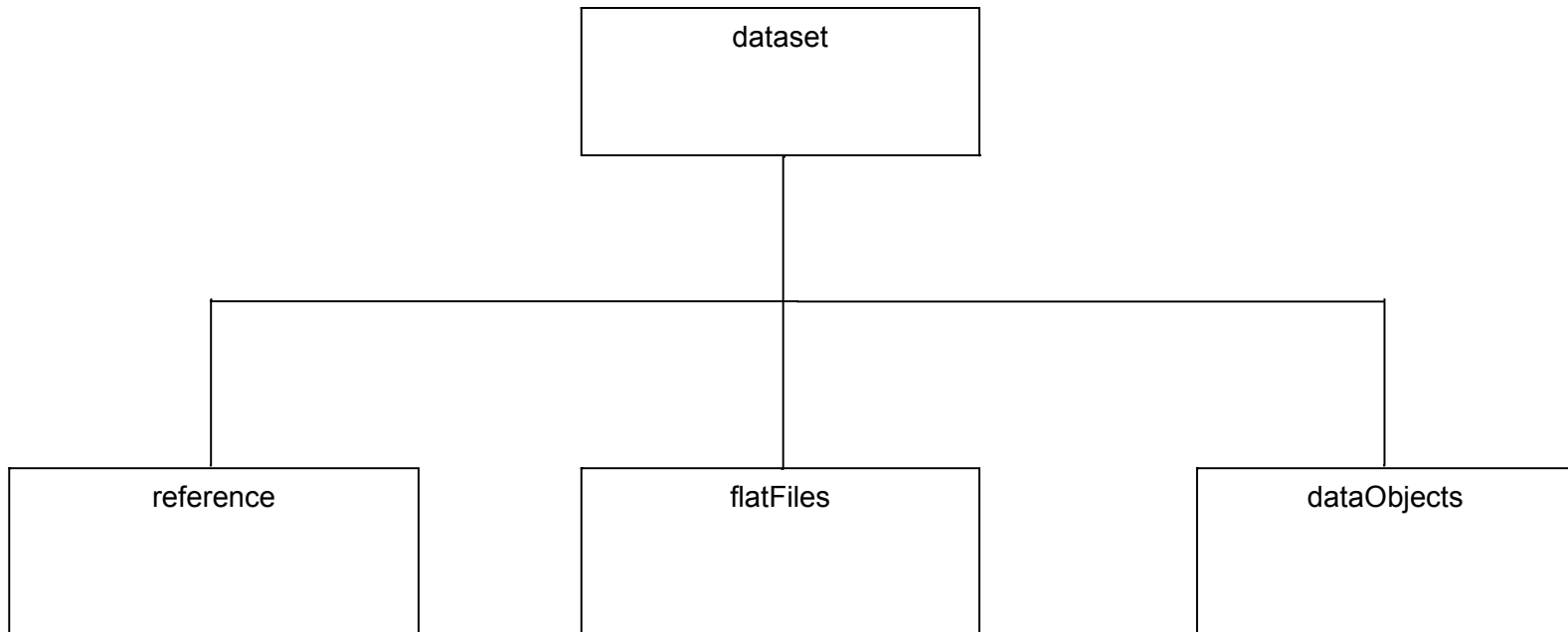
ADDML - Egenskaper



De elementene som har egenskaper er markert med en P.
Eksempel på en generisk struktur som er vanlig i ADDML.



ADDML - Overordnet struktur



reference – metadata om uttrekket

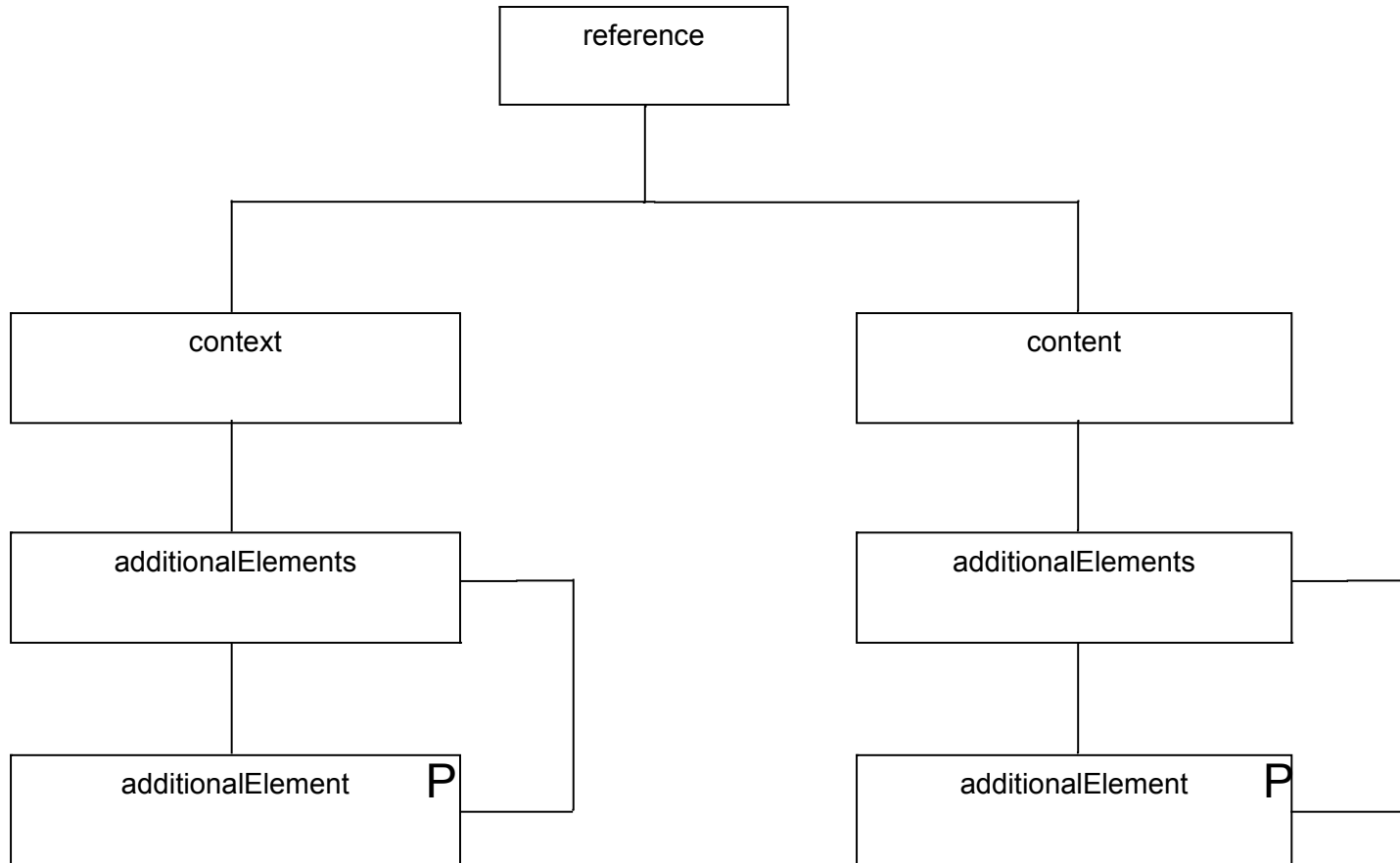
flatFiles – strukturbeskrivelse for flate filer

dataObjects – for å beskrive alle andre typer filer

ADDML godtar flere datasett, men vi har valgt å godta kun ett



ADDML - reference



Generisk oppbygning i reference

Hver enkelt bruker definerer sine sett av elementer/attributter



ADDML - reference

Kontekstuelle elementer:

recordCreators

recordCreator – navn på arkivskaper

system

systemType – tillatte verdier: Noark 3, Noark 4, Noark 5, Fagsystem

systemName – navn på systemet (*navnekonvensjon må vedtas*)

version – versjon av systemet

history – tekstslig informasjon om system og evt. tidligere system

archive

archive – navn på arkivet (*kan muligens fjernes?*)

comments

comment – kommentar



ADDML - reference

Innholdsrelaterte elementer:

archivalPeriod

startDate – startdato

endDate – sluttdato

period

inngåendeSkille

utgåendeSkille

archivalDataset

date – uttrekksdato (*må defineres nærmere hvilken dato*)

type – tillatte verdier: Noark 3, Noark 4, Noark 5, Fagsystem

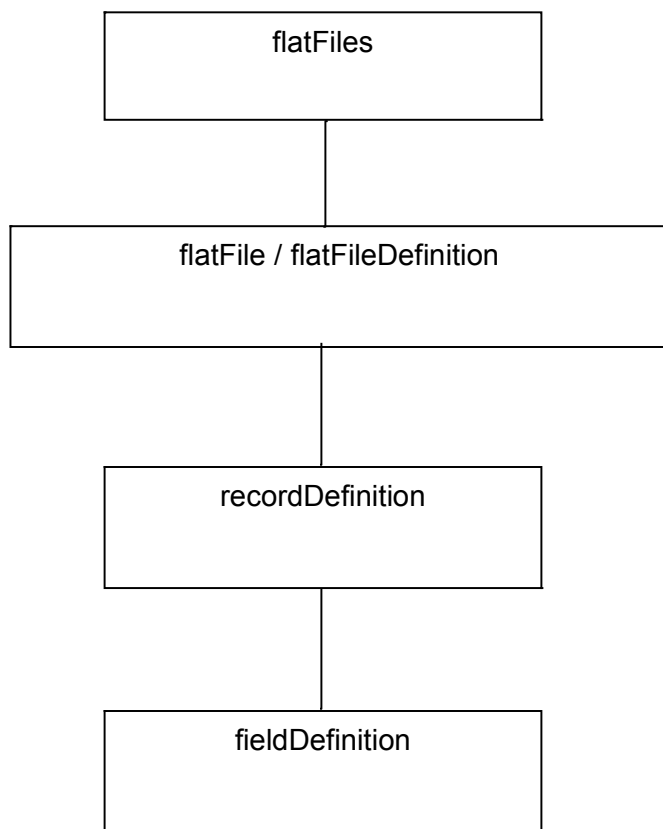
packageID – uttrekkets UUID (*korrespondere med info.xml*)

comments

comment – kommentar

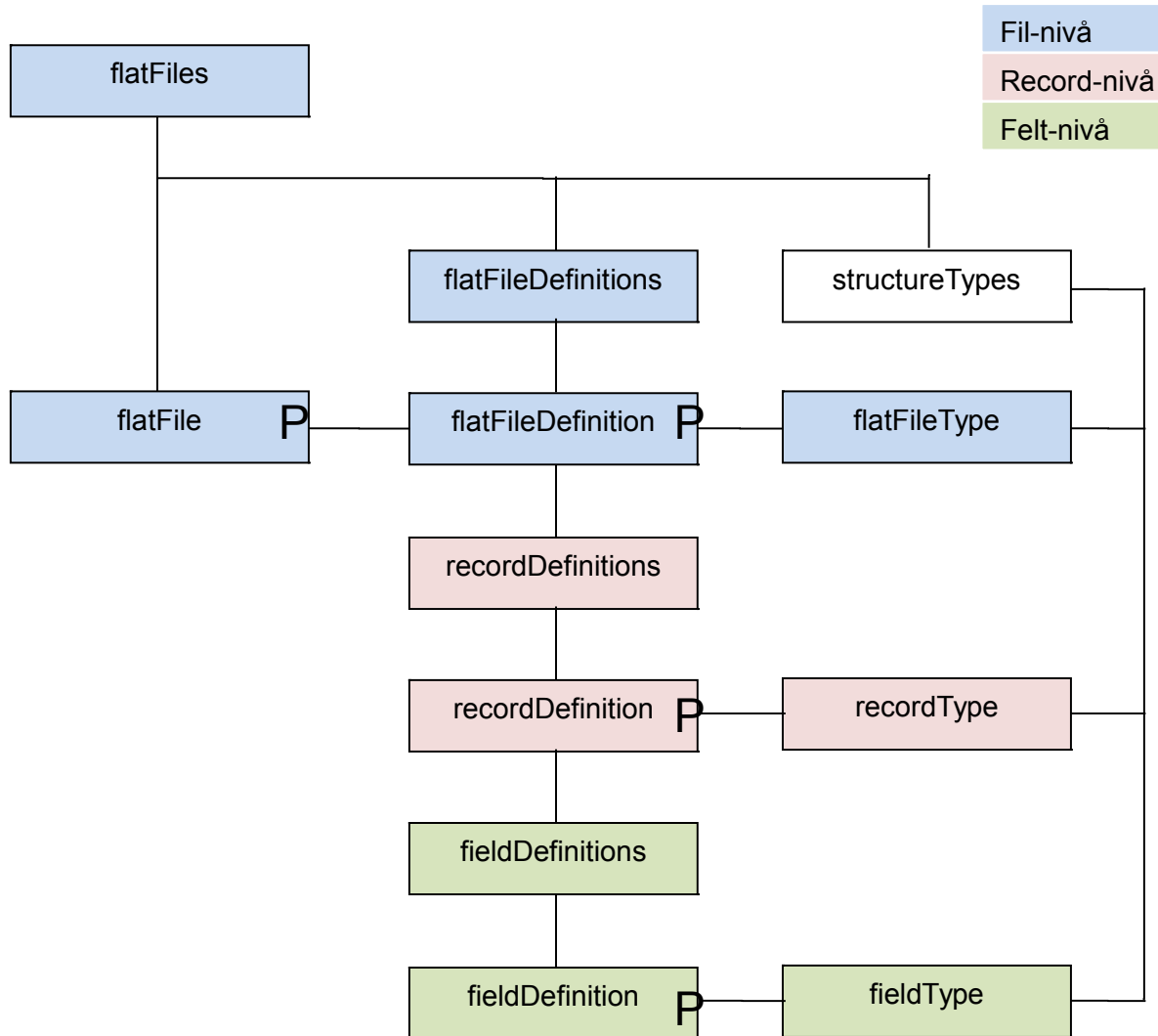


ADDML - flate filer (enkel)





ADDML - flate filer (komplett)





ADDML - definisjonsnivåene

Filnivå

external – angir at filen ikke er med i uttrekket

Postnivå

incomplete – ikke fullstendig beskrivelse av posten

fixedLength – fast postlengde

repeatingGroups – repeterende grupper

keys - nøkkeldefinisjoner

Feltnivå

startPos – startposisjon

endPos – sluttposisjon

fixedLength – fast feltlengde

minLength – minimums lengde

maxLength – maksimums lengde

unique – entydig (unik) verdi

notNull – feltet kan ikke inneholde NULL

fieldParts – feltet består av underliggende felt

codes – kodeverdier



ADDML - typenivåene

Filnivå

charset – tegnsett

charDefinitions – mulighet for å definere egne verdier for enkelttegn

fixedFileFormat/delimFileFormat – format i filen, med skilletegn

Postnivå

trimmed – om posten er trimmet (fjernet blanke tegn)

Feltnivå

dataType – feltets datatype

fieldFormat – feltets format

alignment – justering i feltet

padChar – om feltet er er paddet og hvilket tegn som er benyttet

packType – om feltet er pakket og hvordan

nullValues – om andre NULL verdier benyttes og hvilket



ADDML - flatFile

Egenskaper til flatFile:

fileName – fysisk filnavn inklusive sti (*relativ*)

numberOfRecords – antall poster i filen



ADDML - datatyper

String

Ikke noe spesielt format.

Spesialfelt:

fødselsnummer – FNR

organisasjonsnummer – ORG

kontonummer – KNR



ADDML - datatyper

Integer

Skal være numerisk.

Eksponensial verdier (av 10) – <tall>E+<eksponent>.

Negative tall – minustegn tett inntil før tallet.

Tusenskilte – bør ikke benyttes.

Justering – bør være høyrestilt.



ADDML - datatyper

Float

Følger reglene for Integer.

Desimaltegn – bør være komma (europeisk standard).



ADDML - datatyper

Date

Skal i utgangspunktet bestå av en datodel og en tidsdel.
Tidsdelen er ikke obligatorisk.

fieldFormat benyttes for å angi formen.

Anbefalt format – YYYYMMDDTHHMMSSZ.



ADDML - datatyper

Boolean

Skal alltid oppgi verdiene.

fieldFormat – T/F.



ADDML - datatyper

Link

Behandles som en String.

Et sett av verdier må defineres.

Eksempler:

Referanse som fremmednøkkel – FORKEY

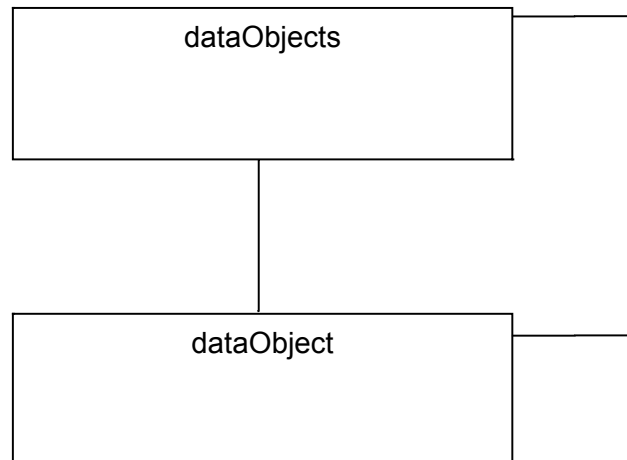
Sti til dokument – DOC

Internettadresse – WWW alt URI

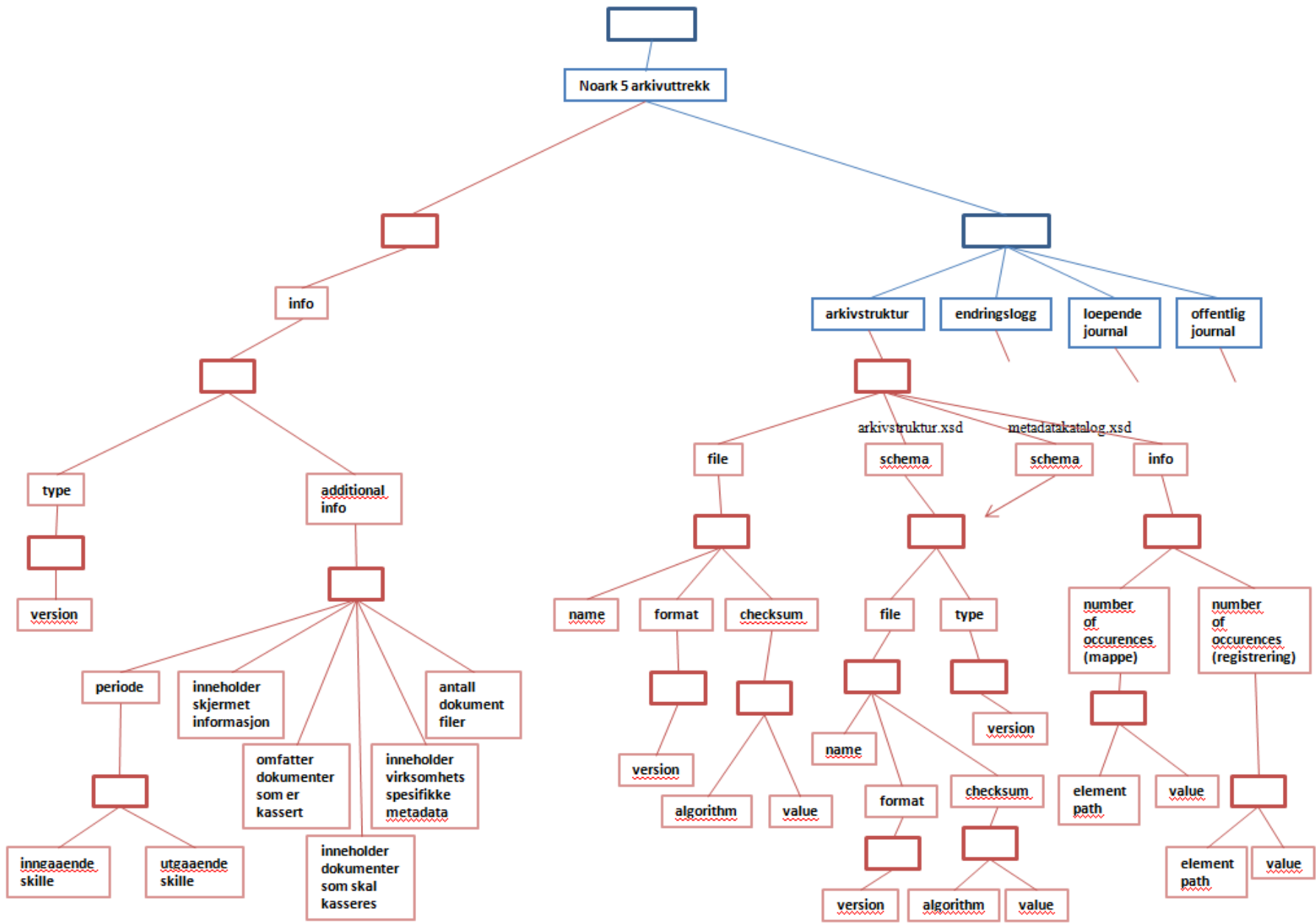
Osv.



ADDML - dataObjects

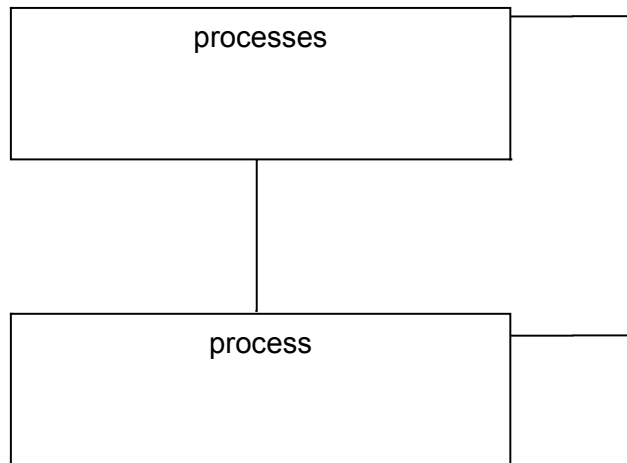


dataObjects benyttes for å beskrive filer i datasettet som ikke er flate filer. Et eksempel på dette kan være xml-filer. Også denne delen er generisk og bygges opp av brukerne etter behov.





ADDML - prosesser - analyser



Prosesser følger den samme velkjente generiske formen.



ADDML - prosesser - analyser

Filnivå

Analyse_CountRecords

Analyse_CountChars

Postnivå

Analyse_FindExtremeRecords – lengste og korteste post

Analyse_CountRecordDefinitionOccurrences

Analyse_AllFrequencyList

Analyse_CrossTable

Feltnivå

Analyse_CountNULL

Analyse_FindExtremeValues – lengste og korteste feltverdi

Analyse_MinMaxValue – laveste og høyeste verdi

Analyse_FrequencyList



ADDML - prosesser - kontroller

Filnivå

Control_AllFixedLength – kun for fast format
Control_NumberOfRecords

Postnivå

Control_FixedLength – kun for fast format
Control_NotUsedRecordDef
Control_Key
Control_ForeignKey

Feltnivå

Control_MinLength
Control_MaxLength
Control_DataFormat
Control_NotNull
Control_Uniqueness
Control_Codes



ADDML - prosesser - konvertering

Filnivå og datasetnivå

Convert_Charset

Convert_Format – fra ett filformat til et nytt

Convert_RecordSeparator

Convert_PreserveStructure – uten å splitte på posttyper

Postnivå

Convert_AddKeyFields – legger på nøkkelfelt ved splitting av posttyper

Convert_SplitRepeatingGroup

Feltnivå

Convert_Dataformat

Convert_PreservePacked



ADDML - prosesser - andre

Datasetnivå

Create_Database

Feltnivå

Control_Birthnr

Control_Organisationnr

Control_Accountnr

Control_DateValue

Control_BooleanValue



ADDML - nyheter

reference

Ikke lenger mandatory.

fieldReference

Entydig gjennom hele ADDML-filen.

header

Flate filer kan ha en eller flere header linjer.

Dokumentasjon

Endelig lage en dokumentasjon av standarden!

I første omgang kun på selve standarden.

Denne kommer også på engelsk.

Hvordan standarden skal brukes i Norge er under arbeid, men vil vil også påvirkes av Arkade 5.0-prosjektet.