

Arkivet i dokumentet
eller
XMP i PDF

tsodring@oslomet.no

Merk: Dette er et konsept som Jean-Philippe Caquet utviklet mens han jobbet hos Trondheim kommune og er implementert som demo i Documaster

Arkivet i dokumentet

- Kan vi bygge inn arkivmetadata i arkivdokumentet (PDF)
- Kan vi påføre dokumentet metadata som gjør at dokumentet kan leve utenfor arkivet og framstå som en autentisk arkivdokument
- På en måte kan dette gjøre Noark overflødig

PDF og metadata

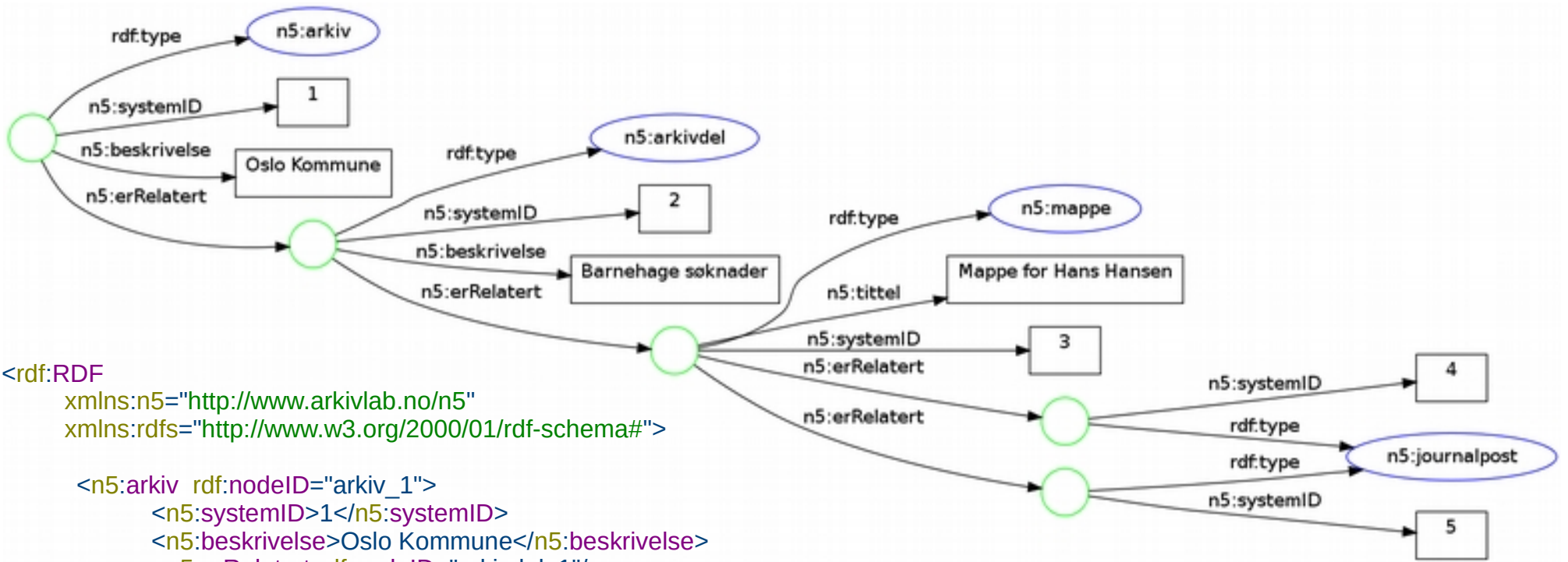
- PDF har et eget innebygd sett av metadata
 - **Document Information Dictionary**
 - forfatter, tittel, emne, nøkkelord, opprettetDato, sistEndretDato, opprettetProgramvare og konverteringsProgramvare
- Siden PDF 1.4 har PDF støtte for utvidet metadata beskrevet i en standard som heter Extensible Metadata Platform (XMP)

XMP og RDF

- XMP er basert på RDF/XML og er derfor utvidbar
- XMP er standardisert av ISO
 - ISO 16684-1:2012 part 1 og ISO 16684-2:2014 part 2

Arkivet i dokumentet

- Det er mulig at det lar seg gjennomføre med XMP
- Vi trenger en forståelse av Noark i RDF



```

<rdf:RDF
  xmlns:n5="http://www.arkivlab.no/n5"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"

  <n5:arkiv rdf:nodeID="arkiv_1">
    <n5:systemID>1</n5:systemID>
    <n5:beskrivelse>Oslo Kommune</n5:beskrivelse>
    <n5:erRelatert rdf:nodeID="arkivdel_1"/>
  </n5:arkiv>
  <n5:arkivdel rdf:nodeID="arkivdel_1">
    <n5:systemID>2</n5:systemID>
    <n5:beskrivelse>Barnehave søknader</n5:beskrivelse>
    <n5:erRelatert rdf:nodeID="mappe_1"/>
  </n5:arkivdel>
  <n5:mappe rdf:nodeID="mappe_1">
    <n5:tittel>Mappe for Hans Hansen</n5:tittel>
    <n5:systemID>3</n5:systemID>
    <n5:erRelatert rdf:nodeID="journalpost_1"/>
    <n5:erRelatert rdf:nodeID="journalpost_2"/>
  </n5:mappe>
  <n5:journalpost rdf:nodeID="journalpost_1">
    <n5:systemID>4</n5:systemID>
  </n5:journalpost>
  <n5:journalpost rdf:nodeID="journalpost_2">
    <n5:systemID>5</n5:systemID>
  </n5:journalpost>
</rdf:RDF>

```

RDF

```

<arkiv>
  <systemID>1</systemID>
  <beskrivelse>Oslo Kommune</beskrivelse>
  <arkivdel>
    <systemID>2</systemID>
    <beskrivelse>Barnehave søknader</beskrivelse>
    <mappe>
      <systemID>3</systemID>
      <tittel>Mappe for Hans Hansen</tittel>
      <journalpost>
        <systemID>4</systemID>
      </journalpost>
      <journalpost>
        <systemID>5</systemID>
      </journalpost>
    </mappe>
  </arkivdel>
</arkiv>

```

XML

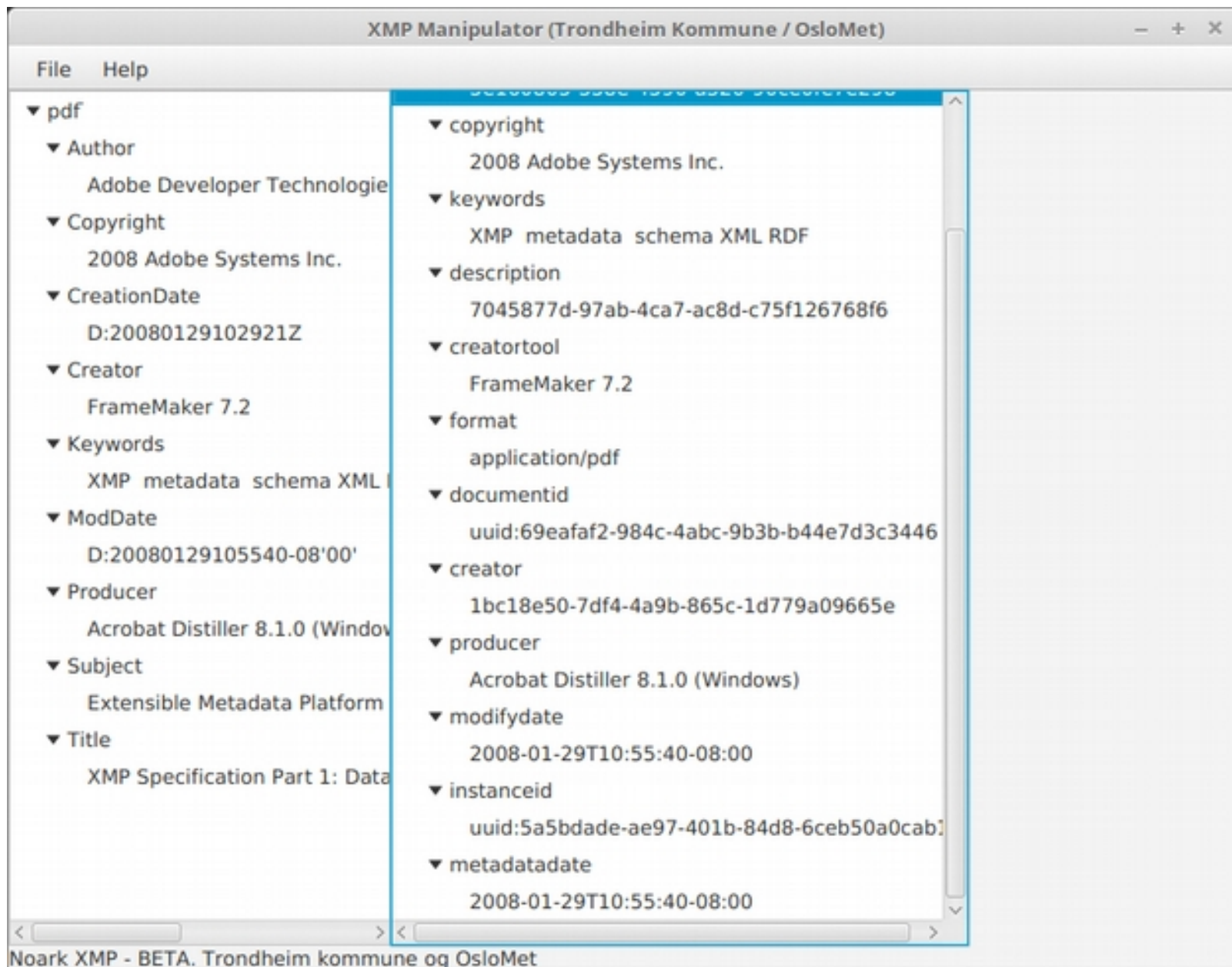
Bouvet og RDF

- Bouvet lagde en RDF-modell av Noark 5 arkivstruktur
- <https://github.com/SesamResearch/Records-Management-and-Archive-Systems-Research>
- Dette mener vi kan være utgangspunktet for en implementasjon

Så langt

- Laget et java program som kan ekstrahere DID og XMP metadata fra en PDF
 - Dette var enkelt, finnes gode biblioteker
- Vi kan også oppdatere XMP metadata

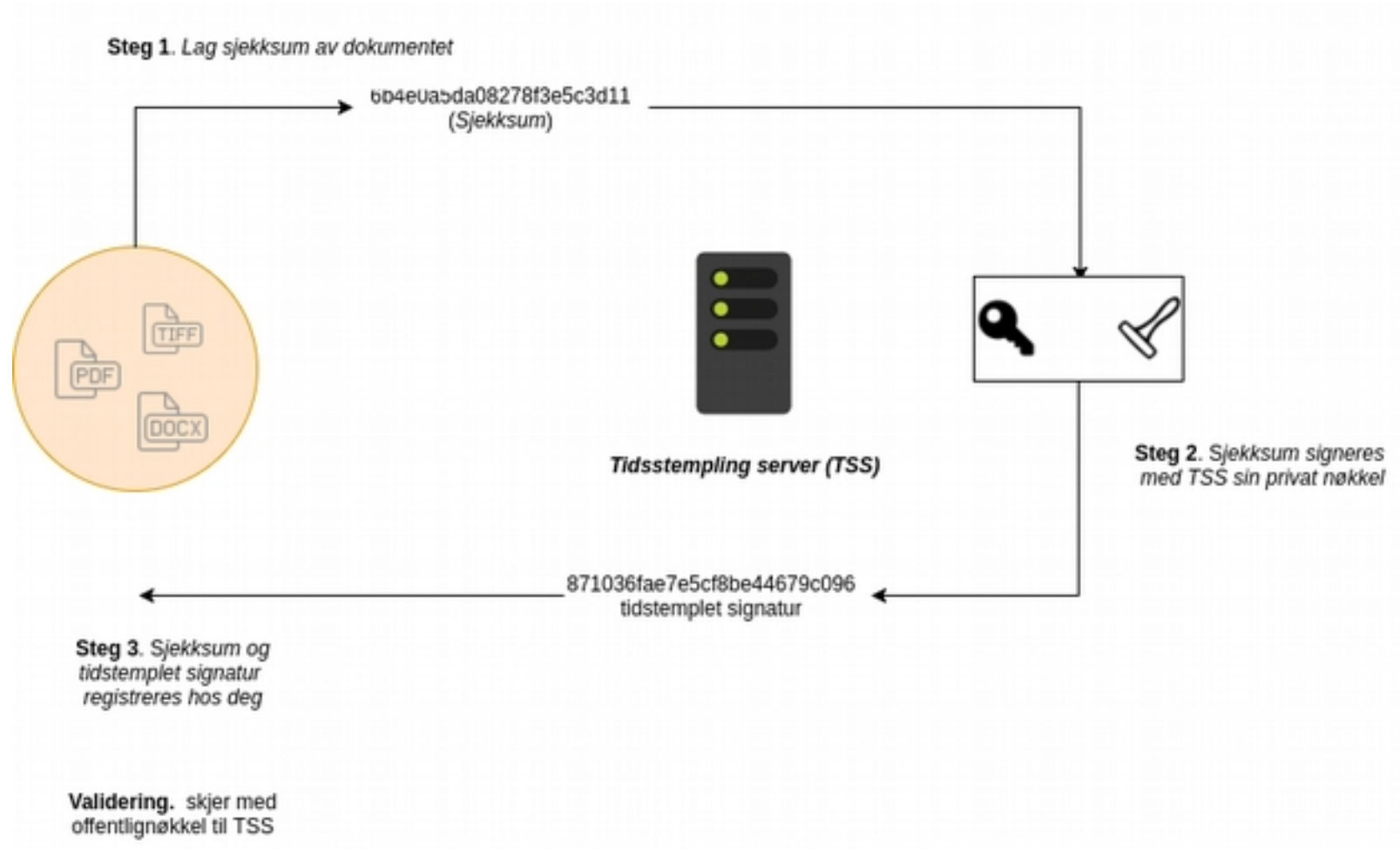
Demo



Autentisitet eller kontekst?

- Autentisiteten til dokumentet kan verifiseres med blokkjede eller tidsstempling
- Så dette handler også om kontekst
 - Klassifikasjon
 - Mappe
- Egentlig er nytteverdien her muligheten til å gjøre dokumentene maskinlesbar med arkiv informasjon uten å måtte lese innholdet

Tidsstempling



Gjenskape arkivet

- Kanskje litt av poenget her er egentlig er at det kan være mulig å gjenskape arkivets struktur utifra et dokumentsamling