



NOARK-5 innsyn

Digitalt to dager til ende | Ole Liabø | Piq AS

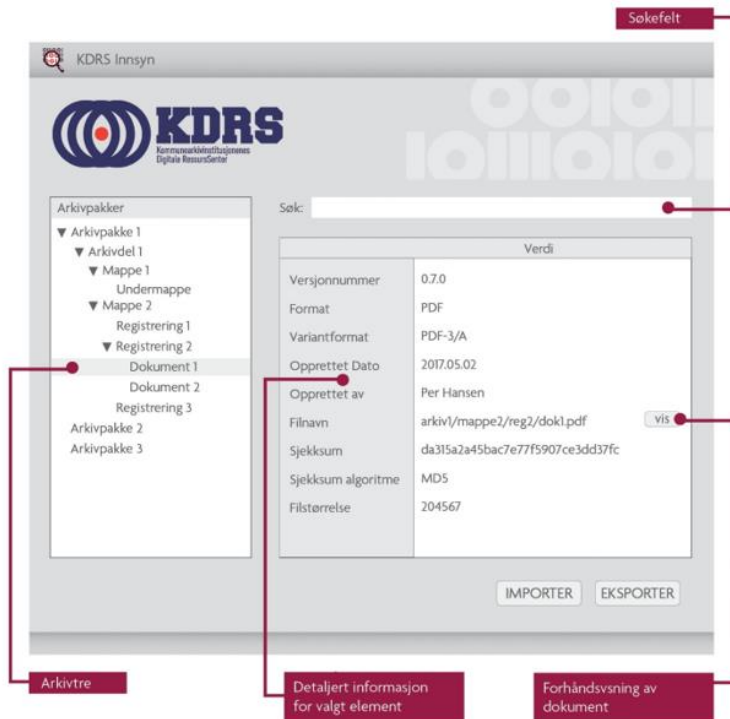


Om Piql

- Ole Liabø – Piql AS
- Forretningside: Bevare digitale data på lagringsmedium med lang levetid
 - Beskytter mot økonomiske/politiske svingninger
 - Autentisitet



KDRS innsyn



- Applikasjon for å se på NOARK-5 uttrekk
 - Oppdragsgiver: KDRS
 - Windows / Mac / Linux – enkel installering
 - Åpen kildekode
 - Viktig med raske søk
 - Generering av rapporter fra søkeresultat
 - Ferdig Q1-2018
- Hva det ikke er:
 - NOARK-5 validering
 - Flerbruker

Testdata

- Mini
 - 2 saksmapper
- Stor
 - XML 4GB
 - 1 000 000 PDF dokumenter
 - Totalt: 26 GB
- Ekte data: Hva er reelle størrelser?

```
<?xml version="1.0" encoding="UTF-8"?>
<arkiv xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns="http://www.arkivverket.no/standarder/noark5/arkivstruktur"
  xmlns:n5mdk="http://www.arkivverket.no/standarder/noark5/metadatakatalog">

  <systemID>e7bc7dc4-7d7e-4c8d-8a50-b4271f685199</systemID>
  <tittel>Oslo Kommune</tittel>
  <beskrivelse>arkiv</beskrivelse>
  <arkivstatus>Avsluttet</arkivstatus>
  <dokumentmedium>Elektronisk arkiv</dokumentmedium>
  <opprettetDato>2011-09-09T12:01:20</opprettetDato>
  <opprettetAv>Arkivleder</opprettetAv>
  <avsluttetDato>2014-09-09T12:01:20</avsluttetDato>
  <avsluttetAv>Arkivleder</avsluttetAv>
  <arkivskaper>...</arkivskaper>
  <arkivdel>
    <systemID>6f25c136-9abc-450a-8394-30a13e6bf1c8</systemID>
    <tittel>Tekniske tjenester</tittel>
    <beskrivelse>Søknader/Vedtak</beskrivelse>
    <arkivdelstatus>Avsluttet periode</arkivdelstatus>
    <dokumentmedium>Elektronisk arkiv</dokumentmedium>
    <opprettetDato>2011-09-09T12:02:40</opprettetDato>
    <opprettetAv>Arkivleder</opprettetAv>
    <avsluttetDato>2013-09-09T12:00:00</avsluttetDato>
    <avsluttetAv>Arkivleder</avsluttetAv>
    <mappe xsi:type="saksmappe">...</mappe>
    <mappe xsi:type="saksmappe">...</mappe>
  </arkivdel>
</arkiv>
```

Hastigheter

- Uttrekk kan være store (XML på >4GB)
- Lesing fra disk: 0.05-0.1 GB/s
- Lesing fra minne: >20 GB/s
- Teoretisk bestetid for 4GB XML ($4 / 0.05$): 80 sekunder

Søk

- Søk skal være interaktivt
- Indeksering av PDFer valgfritt
- Fulltekstsøk
 - logiske uttrykk (OG, ELLER)
- Nøkkel/verdi
- XML spørre språk (x-query / x-path)

Generell løsning

- Ingen innebygde referanser til NOARK-5
 - Alt styres gjennom enkel konfigurasjonsfil
 - Kan tilpasses andre XML baserte AIP format (kanskje NOARK-4?)
 - Vil se på mulighet for å bruke det til å se på innholdet av AIPer (METS) generert av Archivematica
- Språk:
 - Bokmål / nynorsk / samisk?