



ARKIVVERKET
RIKSARKIVET OG STATSARKIVENE

Testverktøy Status og videre tanker

Terje Pettersen-Dahl
Seksjon for Digitalt Depot
Riksarkivet

KDRS konferansen
Trondheim 13. november 2014



Status

- Mange verktøy
 - Proteus for Noark-3
 - ArkN4 for Noark-4
 - Arkade 4.0 for Noark-5
 - Arkade 4.0 for Fagsystem samt Arkade 2.0 for å danne strukturbeskrivelse
 - ESSArch Tools for å generere pakker (SIP)
- Forskjellige formater
 - Hver type system har sitt spesifikke



Hva har skjedd dette året

- Krav fra KUD om max 4 uker behandlingstid (pr. 31.12)
- Vi deler teknologi med uttrekksprodusenter
- Vi har effektivisert viruskontrollen
- Vi har rustet opp maskinparken
- Vi kjører "dobbelt" /(test / andre oppgaver)
- Vi endrer "småfeil" (og dokumenterer dette)



Hva kan andre gjøre?

- Validitet/Velformethet av XML filer
 - Også i forhold til maler (.xsd filer) Det burde ikke være nødvendig å levere inkonsistente filer.
- Absolutt/relativ filsti
 - En gjenganger. Burde være lett å unngå absolutt filsti.
- Rare tegn eller kontrolltegn i XML filene
 - Arkn4 går nedennom hvis den støter på veldig uventede tegn
- TIF vs. PDF (med mer)
 - En annen gjenganger er at filer konvertert fra TIF til PDF fortsatt står flagget som TIF i DOKVER.XML.
- Andre generelle feil i Noark4 rapport

Veriliggende



Vi må gjøre mer!

- Prosjektene vi jobber med har ulike utfordringer:
 - Små prosjekter med kun metadata (Noark 3)
 - Middels prosjekter med dokumenter (Noark 4)
 - Store prosjekter med mye metadata (Seismikk)
 - Store prosjekter med mye dokumentdata (pantebøker 14 Terrabyte)



Arkivloven

- Avlevert materiale skal være ordnet – ansvaret er arkivskapers!

Loven

§ 10. Avlevering.

- Kostnadene ved avlevering skal dekkjast av det avleverande organet.

§ 6. Arkivansvaret.

- Offentlege organ pliktar å ha arkiv, og desse skal vera ordna og innretta slik at dokumenta er tryggja som informasjonsskjelder for samtidig og ettertid

Forskriften

§ 5-4. Generelle krav til materiale som skal avleverast

- Arkivmateriale som blir avlevert til arkivdepot, skal vere ordna etter opphav (proveniens). Dette inneber at arkiv frå ulike arkivskapande einingar skal haldast for seg, og at **den opphavlege orden og indre samanhengen** i kvart enkelt arkiv så langt som mogleg skal haldast ved lag



Prosjekt Arkade 5.0

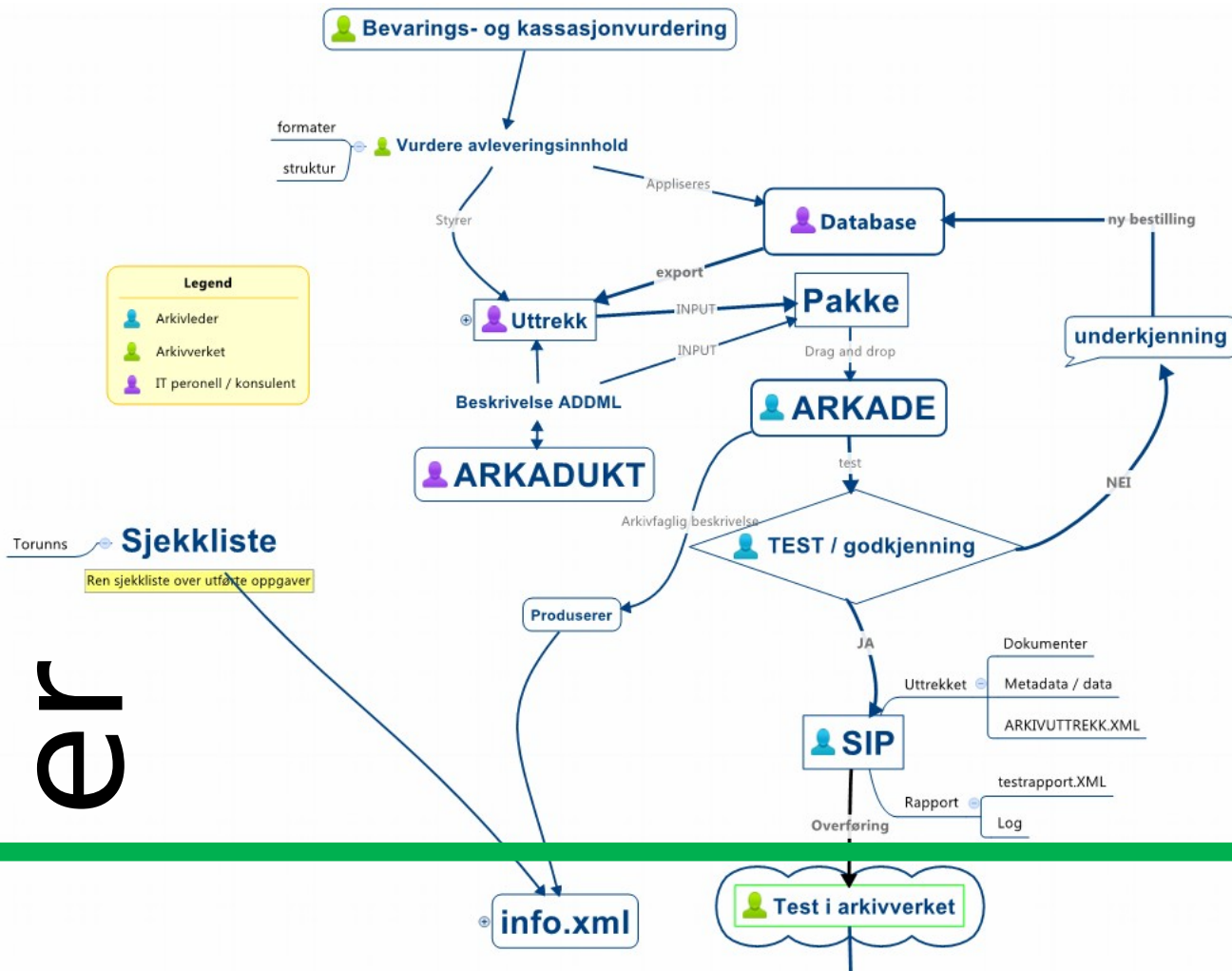
- Arkivskaper må ta på seg ansvaret for testløpet (tiden er inne)
- All risiko knyttet til tid og kostnader må fjernes fra Riksarkivet
- Arkivverkets oppgaver:
 - ... å kvalitetssikre innleveringene
 - ... samt å langtidsbevare
 - ... samt å tilrettelegge for (historisk) innsyn
- For å legge til rette for dette ...



Prosjekt Arkade 5

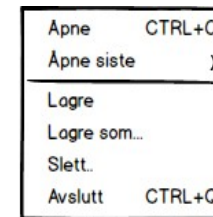
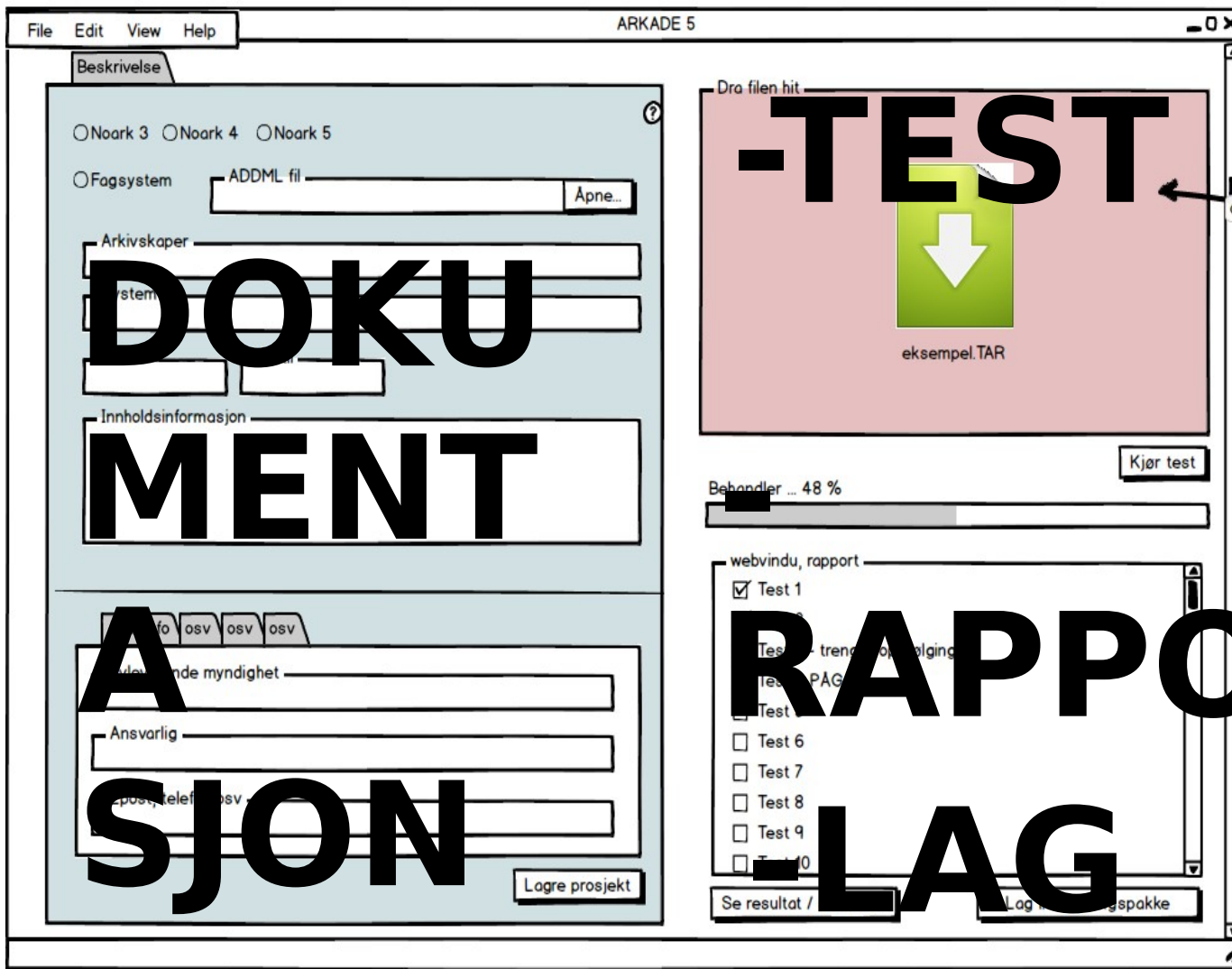
Arkivskap

er





Skjermbilde, Arkade 5.0



drag and drop



RAKKE



Forventet status

- Proteus – utgår!
 - Evt Noark 3-uttrekk behandles som Fagsystem.
- ArkN4 – består i sin nåværende form, men med noe videreutvikling av programvaren.
- Arkadukt 2.0 – består, men utvikles videre.
- Arkade 4.0 – erstattes av Arkade 5.0.



På lengre sikt

- SAMDOK prosjektet
 - piloter:

Arkivmatisk metode



SAMDOK-prosjektet

http://samdok.com/



SAMDOK / Samla samfunnsdokumentasjon

Arkivmeldinga: Arkivutvikling i eit samla perspektiv

FØLG OSS VIA EPOST

Skriv inn epost-adresse i feltet og klikk på 'Følg med', så får du nye innlegg som epost.

Følg med

HVA LETER DU ETTER?

EMNEORD

arbeidsgrupper
arkivdanning Arkiv i

Reisebrev fra ARMA 2014: Åpningshelg og zombie data

28/10/2014 · by Kari Frodesen | Under bevaring og kassasjon, big data, information governance, konferanse, skytjenester | Kommenter innlegget

Vi har mottatt reisebrev fra Torunn Tveit Gaasemyr og Gro Hendriksen, Riksarkivets deltakere på ARMA-konferansen 2014 i San Diego: "ARMA-hva-for-noe?", sa du? ARMA er en organisasjon for personer som jobber innen information governance stiftet i 1955. Organisasjonen har over 27000 medlemmer fra over 30 land. Hvis du er enda mer nysgjerrig kan du jo se informasjonsfilmen deres [...]

De nye encyklopedistene – NordLod2014

27/10/2014 · by Kari Frodesen | Under big data, DIFI, e-forvaltning, formidling, geocoding, konferanse, lenka data, LoCloud, nordisk, NordLod, opne data, RDF,

HEIM

KONTAKT OSS

OM SAMDOK-PROGRAMMET OG DENNE BLOGGEN

SAMDOK-KONFERANSEN

OM PROSJEKTENE

ORGANISERING

PRIORITERTE OPPGAVER 2014

MOTE, REFERAT O.A.

FELLESdokUMENT

FAGDAGAR, KONFERANSAR O.A.

KVA SKJER? (KALENDER)

PÅMELDING

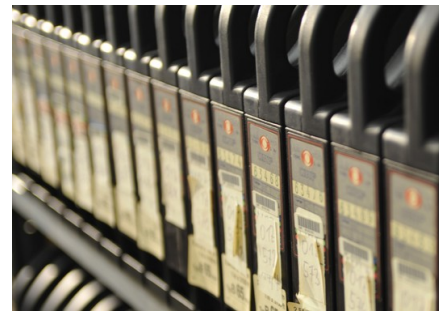
- Påmelding og program NORDLOD

VI MINNER OM

NORDLOD-konferansen
oktober 23rd, 2014

Presentasjoner fra konferansen blir lagt ut på kulturognaturreise.no

Bakgrunn



- Arkivverket mottar tildels svært store arkivuttrekk
- Store uttrekk kan medføre overførings- og ytelses problemer
- Store uttrekk kan ha stor indre variasjon
- Store uttrekk har ofte mange feil, og mange kategorier av feil
- Endringer i struktur over tid medfører problemer i

STORE KOSTNADER



Ved mindre steg av gangen

- ... kan man minske krav til datakapasitet
 - ... kan man i større grad sikre "like barn" i overføringene
 - ... vil man ha mindre variasjon i typer feil
 - ... kan man få mye raskere tilbakemelding på feil
 - ... og dermed mye raskere kunne rette feil
 - ... og det er mindre sjans for tap av data over tid
- SUM: Raskere, sikrere og ressursbesparende**



3 prosjekter relevant for

OSS Kongsberg

NOARK leveranser fra kommuner



- Riksantikvaren / Mattilsynet

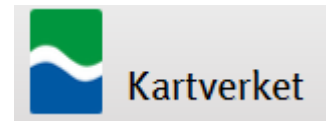
Inspirasjonen er hentet fra "semantisk web",

Det vil si overføring basert på "RDF tripler"



- Kartverket

Overføring av dokumenter fra
"Panteboka"





- IKA Kongsberg dekker Buskerud, Vestfold og



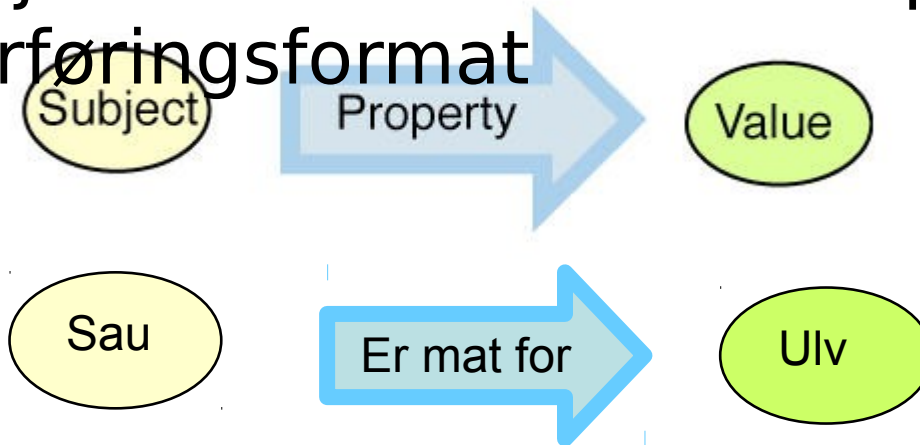
Kommunesektorens organisasjon

- KS deltar som partner
- Fase1: Kravspek klar for tilbudsinnbydelse (31/12)
- Fase2: Løsningsutvikling - vår 2015
- Fase3: Pilot - høst 2015 (barnehageopptak, NOARK)
- Et par kommuner vil være testpiloter



Mattilsynet

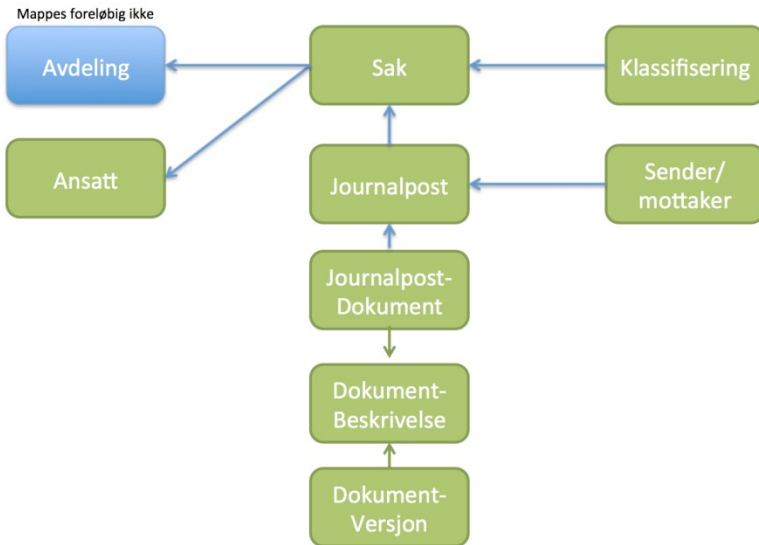
- Data fra ePhorte og P360
- Kontinuerlig avlevering -eksport fra Noark 5
- Prosjektet utforsker "RDF tripler" som overføringsformat



- Ingen relasjonsdatabase skjemaer eller tabeller

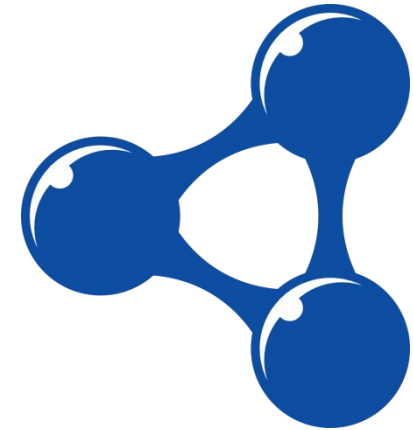


Pilotforsøk hos



transformering

N5 “triplestore”



Løpende eksport av **RDF tripler/relasjoner** ved **SDShare**

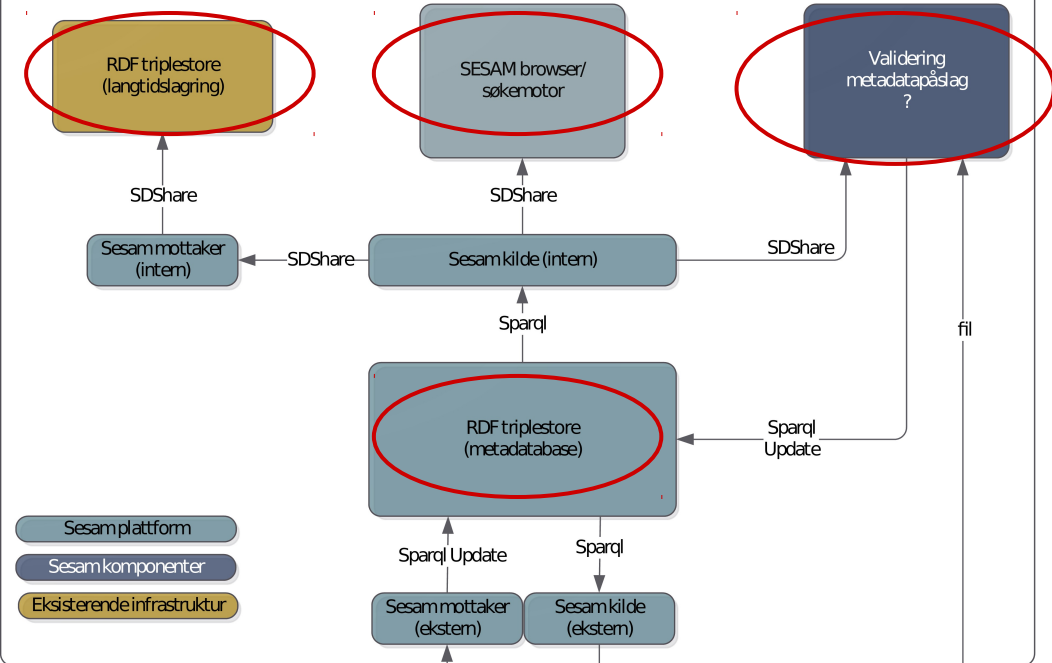
A Protocol for the Syndication of Resource

Riksarkivet

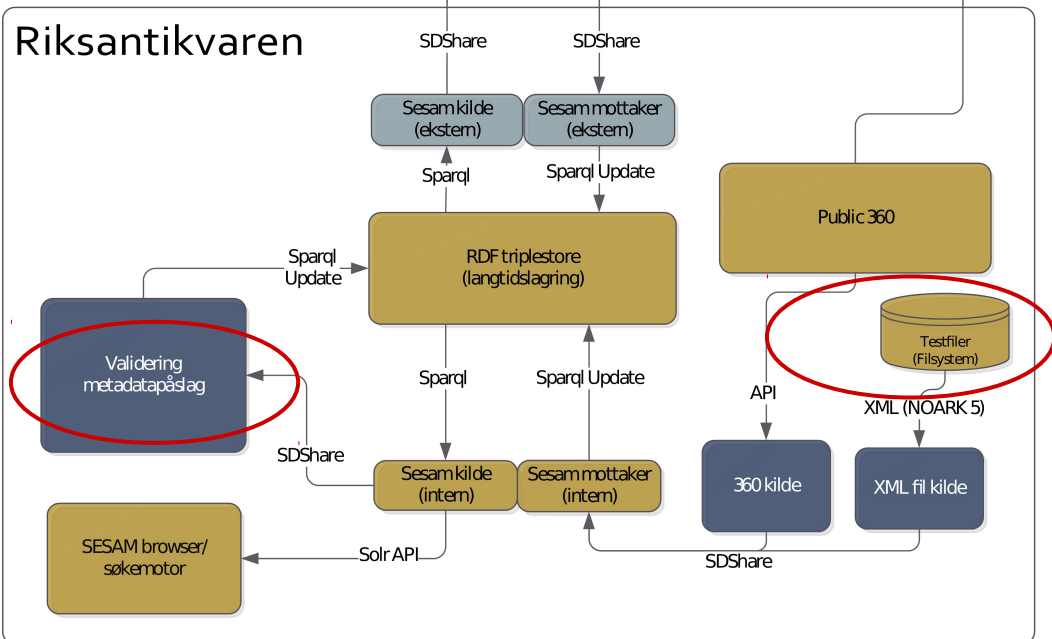


Prosjektskis se lånt fra Bouvet v/ Øystein Isaksen / Graham Moore

Riksarkivet

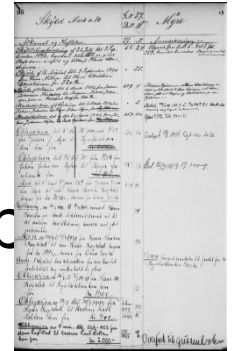


Riksantikvaren





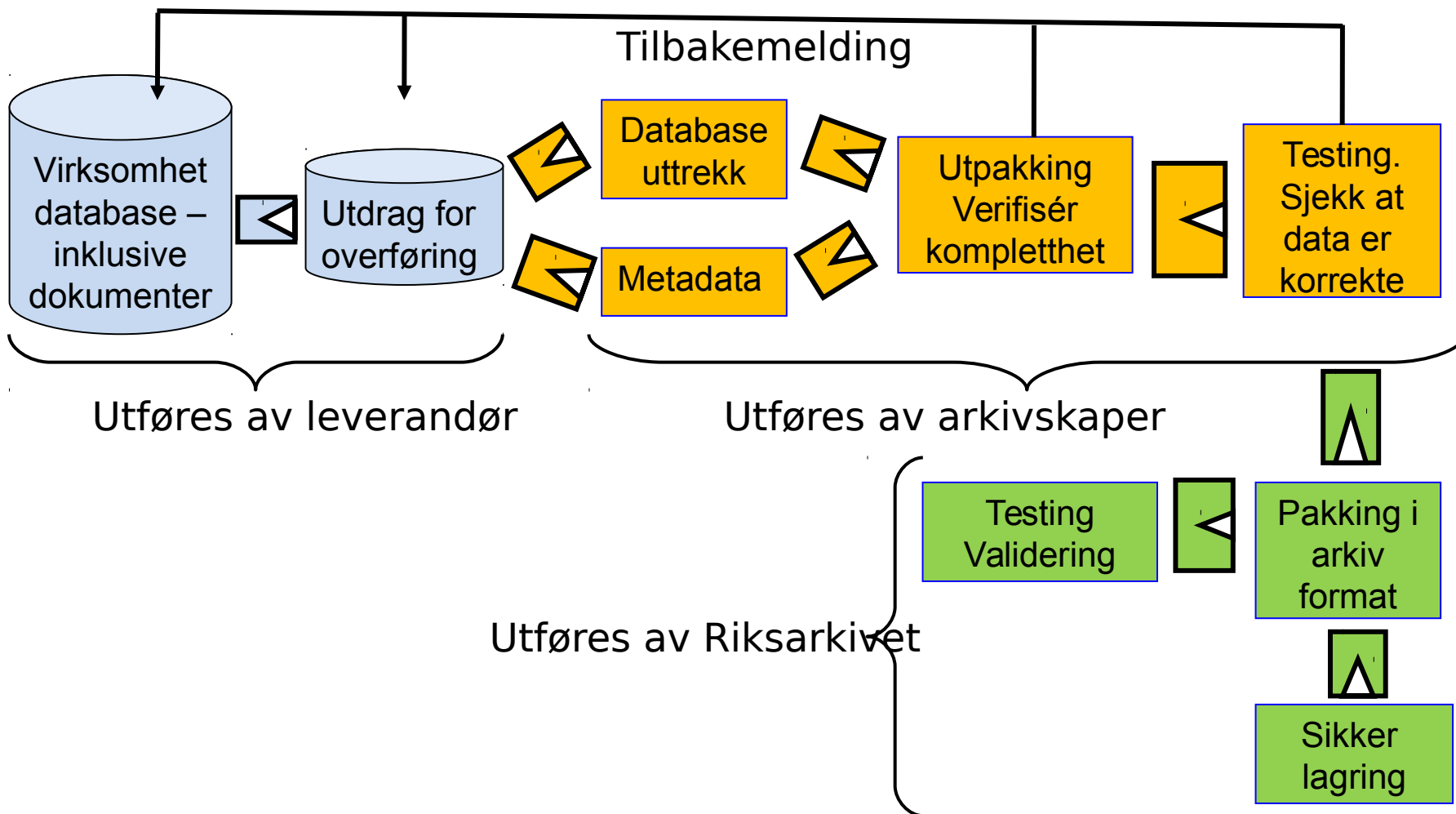
Deponering av digital Pantek



- Lang historikk – ikke vellykket så langt
- Men i år har vi under etablering ny lovende metodikk
- Etterslep på 14.5 Tb deponeres stegvis via linje
- Dette er et **fagsystem** – ikke NOARK basert
- Pilot utvikles høst 2014
- **Ambisjon - å være ajour til julaften 2015**



Systemskisse





ProsjektMål (Piloter)

- Utforske ulike strategier
- Ulike typer datakilder
- Samarbeid kommune / stat
- Beste praksis
 - Store mengder metadata
 - Store mengder dokumentdata



Men igjen...

- Automagi er framtiden
- Vi må automatisere mest mulig av prosessene
- OEP har vist at dette er mulig

