



ARKIVVERKET
RIKSARKIVET

Samdok

samla samfunnsdokumentasjon

RAPPORT 2014



PRIORITERT OPPGAVE

Arkiv i e-forvaltning

(3a) "Synkron" avlevering (STAT)

/Riksantikvaren (RDF)

Utarbeidet av
Øystein Isaksen, Gunnar Urtegaard, Kari Frodesen

Rapportdato
14.01.2015



SAMDOK delprosjekt: eForvaltning	Prioritert oppgave: 3a)	SAK (ePhorte): 2014/15569	Dato: 14.01.2015	Side: 1 av 8
Forfatter: ØI/gunurt/karfro	Tittel: Rapport Samdok 2014 – Synkron avlevering (STAT) Riksantikvaren (RDF)			

OPPGAVE	(3a) “Synkron” overføring av arkivmateriale til depot (STAT) / Riksantikvaren (RDF)
Ansvarlig strategigruppe	Arkiv i e-forvaltning
Arbeidsgruppens leder	Gunnar Urtegaard
Arbeidsgruppens medlemmer	<u>Bouvet</u> : Graham Moore, Øystein Isaksen, Tom Bech <u>Riksantikvaren</u> : Marte Brekke, Lise Storfjord <u>Riksarkivet</u> : Kari Frodesen, Gunnar Urtegaard, Ole Myhre-Hansen, Jon Atle Haugen, Tor Anton Gaarder, Hans Fredrik Berg, Arne-Kristian Groven, Espen Tønnesen
Målformulering	Etablere en pilot for avlevering av NOARK 5 data fra en statlig etat til Riksarkivet, der overføring og lagring skal benytte RDF basert teknologi.
Sammendrag	<p>Samdok-prosjektet ville prøve ut metodikk for løpende eller synkron overføring av data fra arkivskaper til Riksarkivet. Arbeidet startet opp i november 2014 og skal være avsluttet i februar 2015.</p> <ul style="list-style-type: none">• Riksantikvaren er pilotetat og data fra deres sak/arkiv-system (P360) skal være testdata i prosjektet.• Prosjektet henter data løpende fram Riksantikvaren, og overfører disse til et mellomarkiv hos Riksantikvaren som er bygget opp basert på RDF som datamodell. Sommeren 2014 ble det utviklet en RDF-basert versjon Noark5. Den ble langt til grunn for datamodellen i prosjektet.• I overføringen bli data validert hos Riksantikvaren og deretter hos Riksarkivet. Ved feilvalidering sendes det løpende meldinger til arkivtjenesten hos Riksantikvaren.• Det er etablert en enkel innsynsløsning både hos Riksantikvaren og hos Riksarkivet.• Alle endringer i metadata hos Riksantikvaren fører til ny validering og automatisk oppdatering av data i mellomarkiv.• Ved avsluttede saker i P360 blir saken sendt til Riksarkivet. Dersom den valideres der blir dokumenter med tilhørende vedlegg hentet fra Riksantikvaren.• I tillegg til å henta data fra P360, kan og XML-uttrekk i Noark5-format plasseres i filkatalog hos Riksantikvaren og sak for sak valideres og overføres til Riksarkivet• Prosjektet avsluttes i slutten av februar. Data fram Mattilsynet er og planlagt som testdata i piloten.• Prosjektet er finansiert av Riksarkivet/Samdok. Kostnadsramme: 500.000



SAMDOK delprosjekt: eForvaltning	Prioritert oppgave: 3a)	SAK (ePhorte): 2014/15569	Dato: 14.01.2015	Side: 2 av 8
Forfatter: ØI/gunurt/karfro	Tittel: Rapport Samdok 2014 – Synkron avlevering (STAT) Riksantikvaren (RDF)			

Vedlegg	<p>PDF – “On the use of RDF in records management and archiving”</p> <p><i>Lenker:</i></p> <p><u>Sesam:</u> Informasjon om Sesam konsept og teknologi: www.sesam.no Prosjektrom benyttet i prosjektet: https://pingvinen.bouvet.no/kunder/Riksarkivet/Prosjekter/SynkronavleveringRiksarkivet/</p> <p><u>Github:</u> https://github.com/SesamResearch https://github.com/SesamResearch/Records-Management-and-Archive-Systems-Research</p>
Oppfølging 2015	<p>Riksarkivet vil oppsummere prosjektet på en fagdag i februar/mars.</p> <p>Der vil en og drøfte videre arbeidet med RDF som datamodell og overføringsformat.</p> <p>Flere statlige etater har signalisert at de kan være interessert i å prøve ut RDF som overføringsformat.</p>
Status	Ferdig



SAMDOK delprosjekt: eForvaltning	Prioritert oppgave: 3a)	SAK (ePhorte): 2014/15569	Dato: 14.01.2015	Side: 3 av 8
Forfatter: ØI/gunurt/karfro	Tittel: Rapport Samdok 2014 – Synkron avlevering (STAT) Riksantikvaren (RDF)			

Innledning

Riksarkivet ønsker å prøve ut nye måter å få avlevert data fra forvaltningen til langtidslagring i Riksarkivet sitt digitale sikringsmagasin. I flere år har man arbeidet med uttrekk fra eldre og avsluttede systemer/databaser. Dette har av ulike grunner vist seg vanskelig og kostbart.

Det er grunn til å tro at en mer løpende eller synkron overføring etter hvert som saker eller journalposter blir avsluttet, kan visa seg mer effektivt, billigere og kan trolig gi bedre kvalitet på data som blir overført. Trolig vil en slik metode være bedre for både arkivskaper/avleverende organ og for Riksarkivet.

Utgangspunktet i dette prosjektet er at et statlig organ har etablert en server som lagrer data i RDF og at Riksarkivet får tilgang til å hente data etter nærmere definerte regler fra denne serveren.

Spesifisering av oppgaven

Det ble i starten av prosjektet gjort en del spesifiseringer i forhold til omfanget og fortolkning av oppgaven:

- Riksantikvaren ble forespurt og takket ja til å være arkivskaper/avleverende organ.
- Et sett med NOARK 5 xml eksempelfiler ble bestemt laget for å teste flyt mellom Riksantikvaren og Riksarkivet på et så tidlig tidspunkt som mulig.
- Public 360 er NOARK 5 system det skal arkiveres fra.
- Det skal leveres data fra Public 360 hos Riksantikvaren som er i produksjon. (Ikke testserver)
- NOARK 5 dataene skal valideres først hos Riksantikvaren og kun godkjente data skal synkroniseres videre til Riksarkivet.
- Data som ligger hos både Riksantikvaren og Riksarkivet skal kunne vises i et grensesnitt. (Søk)
- Det skal sendes en epost med tilbakemelding om en validering hos Riksantikvaren feiler.
- Langtidslagring skal leveres til en RDF database.
- Bouvet sitt konsept/tjeneste Sesam benyttes.

Sammensetning av arbeidsgruppen

Arbeidsgruppen er sammensatt av representanter fra Riksantikvaren, Riksarkivet og Bouvet.

Hos Bouvet er følgende personer vært mest involvert:

- Graham Moore
- Øystein Isaksen
- Tom Bech

Hos Riksantikvaren er følgende personer vært mest involvert:

- Marte Brekke
- Lise Storfjord

Fra Riksarkivet er følgende personer mest involvert:

- Kari Frodesen
- Gunnar Urtegaard



SAMDOK delprosjekt: eForvaltning	Prioritert oppgave: 3a)	SAK (ePhorte): 2014/15569	Dato: 14.01.2015	Side: 4 av 8
Forfatter: ØI/gunurt/karfro	Tittel: Rapport Samdok 2014 – Synkron avlevering (STAT) Riksantikvaren (RDF)			

- Ole Myhre-Hansen
- Jon Atle Haugen
- Tor Anton Gaarder
- Hans Fredrik Berg
- Arnt-Kristian Groven
- Espen Tønnesen

Avhengigheter

Prosjektet er i hovedsak avhengig av at IT drift hos både Riksantikvaren og Riksarkivet klarer å behandle forespørsler innen rimelig tid, da det er en del nettverkskommunikasjon organisasjonene som må på plass for at prosjektet blir vellykket.

Prosjektet er også avhengig av at arbeidsgruppe klarer å besvare spørsmål angående validering av NOARK 5 data, samt andre punkter hvor det kreves domenekunnskap og inngående arkivkunnskap.

Funn

Selv om prosjektet først avsluttes i slutten av februar, ser en alt nå at denne tilnærmingen fungerer både teknisk og metodisk. Validering av data fungerer godt og tallet sjekkpunktet for Noark5 kan reduseres dramatisk. På samme tid sikrer en bedre kvalitet. Løsningen vil bli testet live og med større datamengder i februar

Drøfting

Det er ulike elementer med løsningen som gjør at prosjektet mener framgangsmåten som er valgt egner seg godt til problemstilling den søker å løse, de mest fremtredende egenskapene er diskutert i de kommende avsnitt.

Konsept og teknologi

Det er i dette prosjektet benyttet Sesam som konsept/tjeneste for å tilfredsstillere Riksarkivets krav til synkron avlevering. Nedenfor beskrives de mest relevante prinsippene i konseptet og hvorfor dette er nyttig i nettopp dette scenarioet.

Konseptuelt prinsipp	Verdi for Riksarkivet
Alle Sesam komponenter er løst koblet. noe som oppnås ved bruk av «publish-subscribe» modellen, alle kildekomponenter publiserer sine data uten å ha noe forhold til mottakerkomponenten(e) som abonnerer på dataene.	Dette er viktig for å ikke få noen direkte avhengigheter mellom arkivskaper og Riksarkivet. Hvis en lignende modell som denne skal benyttes mot flere hundre statlig etater(arkivskaper) i fremtiden, er det ekstremt viktig at hverken statlig etat er avhengig av at Riksarkivet til enhver tid er tilgjengelig (100% opptid på mottaksservere), eller at Riksarkivet er avhengig av at alle statlige etater er tilgjengelige til enhver tid for å utveksle data. 100% tilgjengelighet er svært kostnadsdrivende og heller ikke nødvendig i denne sammenheng.
Datadrevet arkitektur. Data flyter mellom komponenter ved bruk av	Data er stadig i endring, behov er stadig i endring og det er aldri mulig å definere et ferdig modell for hvilke data som skal overføres



SAMDOK delprosjekt: eForvaltning	Prioritert oppgave: 3a)	SAK (ePhorte): 2014/15569	Dato: 14.01.2015	Side: 5 av 8
Forfatter: ØI/gunurt/karfro	Tittel: Rapport Samdok 2014 – Synkron avlevering (STAT) Riksantikvaren (RDF)			

åpne standarder, det defineres ingen faste grensesnitt mellom kilde og mottaker, slik som i en tjenesteorientert arkitektur.	for all fremtid. Riksarkivet har svært langt perspektiv på data og det vil også være en stor fordel om en tjeneste for synkron avlevering har et svært fleksibelt forhold til hvilke data de ulike statlige etatene oversender. Her er det snakk om mange ulike systemer over svært mange år og både kravene til avleverte data samt måten de avleverte dataene lages på vil endres hele tiden. Hvis Riksarkivet skulle tilbudt et sett med statiske grensesnitt for avlevering (web services for eks), måtte disse i tilfelle til stadighet måtte endres i tiden fremover. Da måtte Riksarkivet enten støttet svært mange ulike versjoner eller krevd at statlige etater endret sine systemer til å avlevere mot siste oppdatert versjon, ingen av delene er spesielt gunstig.
Fleksibel skjematøs datamodell (RDF)	Bruken av en skjematøs database betyr at i motsetning til mange datavarehusprosjekter, trenger man ikke å bruke tiden på å bestemme, lage og vedlikeholde en felles datamodell for hele virksomheten. Sesam gjør det mulig å lagre og sammenkoble data i fra forskjellige kilder uten å måtte endre på deres "form".
Publiser endringer fortløpende	Istedenfor å kjøre store jobber svært sjelden, oppdateres alle dataene med jevne mellomrom med mye hyppigere intervall. Det er kun endringer siden forrige synkronisering som blir overført, dermed slipper løsningen å overføre voldsomme mengder med data hver gang. Det blir dermed mye enklere å rette opp i eventuelle feilsituasjoner, mindre sjanse for brudd på forbindelser mellom servere i løpet av en overføring, samt mindre belastning av arkivet hos statlig etat, da det kun blir bedt om endringer siden forrige synkronisering og ikke store jobber som henter ut alt eller svært store deler av arkivet hver gang en synkronisering kjøres.
Lagre alle metadata i et sentralt Datanav	Alle metadata som kommer igjennom systemet blir lagret i et sentralt Datanav (RDF database), dette betyr at dataene enkelt kan eksponeres i en søkemotor. Dette gir svært gode muligheter for gjenfinning, både av arkivert materiale i seg selv, men også om data rundt arkiveringsprosessen, som valideringsfeil eller andre operasjonelle data. Dataene kan også brukes til rapportering, eksport til ulike formater, eksponering mot statlig etat som har avlevert dataene eller integrasjon mot andre systemer hos Riksarkivet hvor alle eller deler av dataene er relevante.



SAMDOK delprosjekt: eForvaltning	Prioritert oppgave: 3a)	SAK (ePhorte): 2014/15569	Dato: 14.01.2015	Side: 6 av 8
Forfatter: ØI/gunurt/karfro	Tittel: Rapport Samdok 2014 – Synkron avlevering (STAT) Riksantikvaren (RDF)			

Valideringsfunksjon

For dette prosjektet er det laget en egen valideringskomponent, som har som oppgave å validere RDF baserte NOARK 5 data opp mot NOARK 5 sine valideringsregler. For å oppnå dette er følgende gjort:

- Det er laget en syntaks for å beskrive NOARK 5 modellen i RDF, inkludert betingelser (constraints).
- Syntaksen er basert på vedlagte dokument «On the use of RDF in records management and archiving», skrevet av arbeidsgruppe deltaker Graham Moore fra Bouvet.
- Både modell og regler (constraints) er beskrevet i data og lagres i RDF databasen sammen med resten av dataene. Dette gjør det enkelt å synkronisere modell/regler fra Riksarkivet til arkivskaper, slik at validering av NOARK 5 RDF hos arkivskaper skjer med best mulig forutsetninger.
- Det er valgt ut et begrenset antall regler (constraints) i dette prosjektet.
- Valideringskomponenten skriver tilbake resultatet av valideringen til RDF database.
- Hvis resultatet av validering hos arkivskaper er vellykket, sendes dataene videre til Riksarkivet der dataene valideres på nytt, er resultatet igjen vellykket sendes dataene til langtidslagring.
- Valideringskomponent støtter også enkel validering av binær fil som skal arkiveres. (PDF-A)

Public 360 integrasjon

For å hente ut data fra et eksisterende NOARK 5 system benyttes det en Sesam Public 360 komponent mot Riksantikvarens Public 360 server. Denne fungerer på følgende måte:

- Henter automatisk operasjonelle data fortløpende fra Public 360 etter hvert som det skjer endringer i databasen. Bruker av systemet trenger ikke å eksportere manuelt eller vite om at dataene hentes.
- Henter data via standard Public 360 API.
- I pilot hentes et begrenset sett av metadataene, hovedsakelig de dataene som skal til for å tilfredsstille de reglene (constraints) som er definert i valideringskomponenten.
- Henter ikke data fra personalmapper.
- Data blir først transformert til RDF, deretter til NOARK 5 RDF formatet, slik at valideringskomponent kan gjøre valideringen.



SAMDOK delprosjekt: eForvaltning	Prioritert oppgave: 3a)	SAK (ePhorte): 2014/15569	Dato: 14.01.2015	Side: 8 av 8
Forfatter: ØI/gunurt/karfro	Tittel: Rapport Samdok 2014 – Synkron avlevering (STAT) Riksantikvaren (RDF)			

Konklusjoner

Prosjektet viser lovende resultater selv om det ikke blir sluttført for i februar. Komponenter utviklet i prosjektet vil bli tilgjengelig som åpen kildekode.

Videre arbeid

Riksarkivet vil oppsummerer prosjektet på en fagdag i februar/mars. Der vil en og drøfte videre arbeidet med RDF som datamodell og overføringsformat.

Flere statlige etater har signalisert at de kan være interessert i å prøve ut RDF som overføringsformat.