



ARKIVVERKET  
RIKSARKIVET

# Samdok

samla samfunnsdokumentasjon

RAPPORT 2016



## PRIORITERT OPPGAVE

7 Nasjonal publiseringsplattform for skanna arkivdokumenter

Leveranse 1a

Utarbeidet av  
**Anette S. Clausen**

Rapportdato  
**16.06.16**



SAMDOK delprosjekt: <b>Kommunale arkiv</b>	Prioritert oppgave: <b>7</b>	SAK (ePhorte):	Dato: <b>16.06.16</b>	Side: <b>1 av 5</b>
Forfatter: <b>Anette S. Clausen</b>	Tittel: <b>Rapport 2016 – Nasjonal publiseringsplattform for skanna arkivdokumenter, leveranse 1a</b>			

<b>OPPGAVE</b>	<i>7 Nasjonal publiseringsplattform for skanna arkivdokumenter</i>
<b>Ansvarlig delprosjekt</b>	<i>Kommunale arkiv Anna Malmø-Lund</i>
<b>Arbeidsgruppens leder</b>	<i>Anette Skogseth Clausen (RA)</i>
<b>Arbeidsgruppens medlemmer</b>	<i>Espen Tønnessen (RA), Kristian Hunskaar (RA), Ingrid Jørgensen (Arbark), Aasta Karlsen (Arkiv i Nordland), Espen Sæterbø (Fylkesarkivet SF), Øystein Eike (Oslo byarkiv), Ottar Anderson (IK AMR), Egil Blomsø (AA Museum og arkiv), Tove Wefald Pedersen (Norsk folkemuseum)</i>
<b>Målformulering</b>	<i>Det er behov for å etablere en nasjonal publiseringsplattform for skanna arkivdokumenter i Norge, som kan dekke arkivinstitusjonenes behov for å publisere digitaliserte (skanna) papirbaserte arkiver og samlinger på Internett. Løsningen skal være både en stor-skala publiseringstjeneste for papirbasert arkiv fra alle samfunnssektorer, og en leverandør av digitale tjenester til innbyggerne, basert på innhold fra arkivinstitusjonene.</i>
<b>Sammendrag</b>	<i>Rapporten er en beskrivelse av tjenestene og funksjonene som er ønskelig å ha i en nasjonal publiseringsplattform for skanna arkivdokumenter. Hovedfunksjonen er at innholdsleverandørene (depotinstitusjonene) skal kunne laste inn skannede bilder av arkivdokumenter, knytte bildene til arkivinformasjon i ASTA hvis det fins, administrere tilganger til materialet (selv om alt i utgangspunktet skal være åpent og fritt tilgjengelig) og kunne benytte seg av API for å dele eller gjenbruke materialet i andre sammenhenger. Brukere skal ha fri tilgang til bildene og det skal være enkelt å navigere i plattformen for å finne materialet etter geografi, emne, arkivskaper, oppbevaringssted, periode osv.</i>
<b>Oppfølging 2016</b>	<i>– Dette er første delrapport. Det skal komme flere delrapporter utover 2016 som beskriver mer teknisk løsning, organisasjon og drift.</i>
<b>Vedlegg</b>	



SAMDOK delprosjekt: <b>Kommunale arkiv</b>	Prioritert oppgave: <b>7</b>	SAK (ePhorte):	Dato: <b>16.06.16</b>	Side: <b>2 av 5</b>
Forfatter: <b>Anette S. Clausen</b>	Tittel: <b>Rapport 2016 – Nasjonal publiseringsplattform for skanna arkivdokumenter, leveranse 1a</b>			

## Del 1a: Funksjons- og tjenestebeskrivelse

### Generelt

- NAPSA skal være en gratis, alternativt veldig rimelig, tjeneste for innholdsleverandørene. Følgende må finansieres:
  - Utvikling av NAPSA versjon 1.0.
  - Datalagring.
  - Drift og vedlikehold av NAPSA.
  - Eventuelt videreutvikling av NAPSA.
- NAPSA skal være en skalerbar tjeneste mtp mengde og innholdstyper.
  - Tjenesten skal være bygd for store datamengder og massepublisering i industriell skala.
  - Tjenesten skal ikke være låst for videreutvikling.
  - Nye innholdstyper skal kunne komme til.
- NAPSA skal bygge på åpne standarder.
- NAPSA's innholdsdata skal være tilgjengelige gjennom et åpent API. APIet kan forholde seg til tilgangsstyringa i NAPSA, og det er derfor ikke nødvendig å holde deler av innholdsdataene utenfor APIet.
- Den enkelte innholdsleverandøren skal ha fullt eierskap til sitt innhold.
  - Publisert innhold skal kunne trekkes tilbake, helt eller delvis.
  - Innhold som er publisert fritt tilgjengelig, kan siden bli regulert med adgangskontroll.
- Den enkelte innholdsleverandøren er alene ansvarlig for at sitt fritt tilgjengelig publiserte innhold kan publiseres på Internett.
  - Taushetsbelagt innhold skal ikke publiseres fritt tilgjengelig.
  - Innholdsleverandører (organ) som er underlagt Offentleglova, skal ikke publisere opplysninger som nevnt i Forskrift til offentliglova § 7 (<http://lovdata.no/forskrift/2008-10-17-1119/§7>), fritt tilgjengelig.
  - Det må framgå tydelig hvem som er innholdsleverandør, m.a.o. hvem som er ansvarlig.

### Innholdsleverandørsiden ("back end")

- NAPSA skal være organisert med ett sentralt fillager, én sentral database med metadata på arkivenhetsnivå (metadata/kataloginformasjon som er felles for alle bilder knyttet til en arkivenhet) og metadata på filnivå (metadata som gjelder ett eller noen av bildene knyttet til en arkivenhet), og ett sentralt brukerregister.
- Innholdsleverandørene skal kunne utføre operasjoner mot fillager og database via et administrasjonsgrensesnitt, som de har tilgang til gjennom innlogging med brukernavn og passord.
  - En innholdsleverandør kan ha flere brukerprofiler (brukernavn/passord), for eksempel én brukerprofil for hver medarbeider.
  - Hver brukerprofil skal kunne gis spesifikke og avgrensede rettigheter etter behov. Typisk vil en brukerprofil ha rettigheter til å administrere filer og metadata som tilhører vedkommendes arbeidsgiver (innholdsleverandør).
  - Et begrenset antall brukerprofiler skal ha admin-rettigheter for bl.a. å kunne tildele rettigheter til andre brukere.
- Det skal være mulig å registrere metadata på arkivenhetsnivå i databasen via administrasjonsgrensesnittet.
  - Dersom en arkivenhet er registrert i Arkivportalen, skal registreringen starte med å opprette ei kobling til oppføringa i Arkivportalen og høste relevante data derfra. Databasen skal jevnlig synkroniseres mot Arkivportalen, slik at endringer i arkivenhetens kataloginformasjon tilflyter databasen (forutsatt at arkivenhetens Asta-ID er uendret).
  - Det skal være mulig å masseregistrere flere arkivenheter, f.eks. alle stykker i en serie, i én operasjon.
  - Dersom en arkivenhet ikke er registrert i Arkivportalen, skal det være mulig å registrere relevante data manuelt, eller man kan registrere den overordnede arkivenheten med merknad om at bare en del av denne er digitalisert.
    - Dersom en arkivenhet som er registrert i NAPSA *avpubliseres* i Arkivportalen, må den registreres manuelt i stedet, med mindre den ønskes avpublisert også i NAPSA.
  - Det skal være mulig å registrere ytterligere informasjon som ikke finnes i Arkivportalen, men som kan forbedre brukernes søkemuligheter, for eksempel emneord og geografiske tilknytninger. (Digitalarkivet har felt for bl.a. emneknagger, 1947-kommuner og dagens kommuner, men sistnevnte felt er i lite praktisk bruk.)



SAMDOK delprosjekt: <b>Kommunale arkiv</b>	Prioritert oppgave: <b>7</b>	SAK (ePhorte):	Dato: <b>16.06.16</b>	Side: <b>3 av 5</b>
Forfatter: <b>Anette S. Clausen</b>	Tittel: <b>Rapport 2016 – Nasjonal publiseringsplattform for skanna arkivdokumenter, leveranse 1a</b>			

- Det skal være mulig å definere én eller flere "profiler" som legger den samme tilleggsinformasjonen på en serie av arkivenheter.
  - Det skal være mulig å registrere at en arkivenhet skal være omfattet av tilgangsstyring og definere tilgangsgrupper som skal ha tilgang.
- Det skal være mulig å laste opp et filsett til fillageret via Internett.
  - Den enkelte innholdsleverandør velger og står inne for bildekvaliteten for sine filer.
  - Følgende filformater skal støttes:
    - TIFF
    - JPEG2000
    - JPEG
    - PNG
    - PDF (for uthenting av bildefiler i PDF-filene)
  - Opplasting skal skje ved å velge en forhåndsregistrert arkivenhet fra databasen som filsettet skal kobles til.
  - Ved opplasting skal filsettet valideres (virussjekk etc.). Originalfilene konverteres til visningsfiler i JPG-format med tilhørende UUIDer, men de originale filnavnene lagres også.
    - Av hensyn til hurtig nedlasting for brukerne, bør ikke visningsfiler være særlig større enn 1 MB. Der det er behov for større visningsfiler (detaljerte tegninger og kart etc.) må en tiles-funksjon være tilgjengelig.
    - De originale filene slettes, slik at NAPSA kun lagrer visningsfilene. Innholdsleverandørene må selv ta ansvaret for å lagre de originale filene i de originale formatene, dersom de ønsker å lagre dem for ettertida (anbefalt).
      - Opsjon om at originalfilene kan lagres i en NOARK5-kjerne (vil medføre kostnader for innholdsleverandører som benytter opsjonen).
    - Innholdsleverandørene skal kunne laste ned en "mapping" mellom originale filnavn og permanente UUIDer i NAPSA.
  - En "default" visningsrekkefølge for filene etableres ved at filene sorteres alfabetisk-numerisk etter originale filnavn.
- Det skal være mulig (men ikke nødvendig) å registrere ytterligere metadata på filnivå (for enkeltfiler eller grupper av filer). Dette inkluderer:
  - Beskrivelse av hva hvert bilde (hver fil) viser, så som hva (f.eks. "Formannskapsmøte"), når (f.eks. "25.01.1933"), sidetall i kilden osv.
  - Endring av visningsrekkefølgen, herunder ekskludering av bilder fra visningsrekkefølgen.
  - Klausulering / tilgangsstyring av enkeltfiler.
  - Automatisk OCR-gjenkjenning av maskinskrevet tekst skal kunne utnyttes til registrering av metadata på filnivå.
  - Registrering av ytterligere metadata skal tilrettelegges for crowdsourcing, slik at innholdsleverandørene kan velge å åpne for crowdsourcing på hele eller deler av sitt materiale (taushetsbelagt materiale kan f.eks. ikke crowdsources).
- Det skal være mulig å laste opp nye filsett til en arkivenhet som allerede har filsett knyttet til seg, erstatte filer etc.
- Publisering / avpublisering skal skje med et tastetrykk, enten for en enkelt arkivenhet eller for ei gruppe arkivenheter.
- Statistikkverktøy skal gjøre det mulig å hente ut:
  - Antall besøk og sidevisninger totalt, for den enkelte innholdsleverandør, for enkeltkilder og for grupper av kilder (også på tvers av innholdsleverandører, for eksempel likeartede kilder).
  - Antall publiserte bilder og kilder, fordelt på innholdsleverandør og/eller fordelt kronologisk (minimum måned).
  - Antall opplastede bilder og kilder, fordelt på innholdsleverandør og/eller fordelt kronologisk (minimum måned). Dette inkluderer bilder og kilder som er lastet opp, men som ikke er publiserte.
  - Antall registrerte kilder, fordelt på innholdsleverandør og/eller fordelt kronologisk (minimum måned). Dette inkluderer kilder som er registrert, men som ikke har fått bilder koblet til seg.



SAMDOK delprosjekt: <b>Kommunale arkiv</b>	Prioritert oppgave: <b>7</b>	SAK (ePhorte):	Dato: <b>16.06.16</b>	Side: <b>4 av 5</b>
Forfatter: <b>Anette S. Clausen</b>	Tittel: <b>Rapport 2016 – Nasjonal publiseringsplattform for skanna arkivdokumenter, leveranse 1a</b>			

## Brukersiden ("front end")

- NAPSA og alt innhold skal være gratis å bruke for brukere (publikum).
- Det skal være færrest mulig begrensninger på gjenbruk av innhold, men f.eks. åndsverksrettigheter må respekteres.
  - Den enkelte innholdsleverandør er ansvarlig for å sette lisens på sitt materiale; fri lisens er standard og idealet.
  - Den enkelte innholdsleverandør må ha oppmerksomhet mot bestemmelser om nettpublisering i forbindelse med avtaler med arkivskapere.
- NAPSA's brukergrensesnitt skal være generøst, enkelt og moderne og tilgjengelig på alle digitale plattformer (PC, Mac, tablet, mobiltelefon).
  - Søk etter publiserte arkivenheter skal være raske og effektive. Det skal være mulig å søke etter:
    - oppbevaringssted
    - arkiv
    - administrativ tilknytning / arkivskaper (ikke alltid identisk med arkiv)
    - emneord
    - geografisk tilknytning
    - tidsintervall
    - fritekst
    - OCR-lest tekst (innholdstekst)
  - Blading og fremvisning skal være raskt og effektivt, og det skal være god lesefunksjonalitet:
    - bilder skal kunne skaleres i størrelse
    - bilder skal kunne roteres
  - Det skal være mulig å dele innhold på e-post og i sosiale medier.
  - Det skal være mulig å kommunisere med NAPSA og/eller innholdsleverandørene via tilbakemeldingsfunksjoner. Funksjonene skal legge til rette for toveis-kommunikasjon:
    - Tilbakemeldingsfunksjon for tekniske spørsmål og meldinger om driftsfeil betjenes av et sentralt ledd på vegne av NAPSA.
    - Tilbakemeldingsfunksjon for spørsmål om arkivinnhold rutes til den enkelte innholdsleverandøren, basert på metadata om oppbevaringssted, for eksempel med e-postskjema eller ved at e-postadresse oppgis. Den enkelte innholdsleverandør er ansvarlig for å svare, men det er viktig at brukerne blir oppmerksomme på at ikke alle innholdsleverandører vil ha like stor kapasitet til å svare, og at det kan ta tid før de får svar.
- En bruker skal ikke være nødt til å logge inn med brukerprofil for å bruke NAPSA, men det skal være mulig å logge inn for å få tilgang til særskilte brukervalg og arkivenheter med tilgangsstyring, samt personlig toveis-kommunikasjon med innholdsleverandørene.
- Publisert materiale skal vises med korrekt og detaljert kildereferanse, dvs. oppbevaringssted, arkivnavn og eventuelt ytterligere arkivbeskrivelse fra arkivkatalog.
  - Dersom det publiserte materialet er registrert i Arkivportalen, skal det etableres toveis lenke mellom NAPSA og Arkivportalen.
- Innholdet i NAPSA skal kunne indekseres av markedsledende søketjenester som Google.



SAMDOK delprosjekt: <b>Kommunale arkiv</b>	Prioritert oppgave: <b>7</b>	SAK (ePhorte):	Dato: <b>16.06.16</b>	Side: <b>5 av 5</b>
Forfatter: <b>Anette S. Clausen</b>	Tittel: <b>Rapport 2016 – Nasjonal publiseringsplattform for skanna arkivdokumenter, leveranse 1a</b>			

## Integrasjon og samarbeid med andre relevante innholdstjenester

- Det skal være god integrasjon mot andre relevante innholdstjenester som:
  - Arkivportalen
  - Norvegiana
  - Digitalt Museum
  - bokhylla.no (nb.no)
- Det skal være mulig å lenke mellom arkivenheter i NAPSA og relevant innhold i andre innholdstjenester.
- Intelligente søk skal gi relevante treff i NAPSA og andre innholdstjenester samlet.

/eof