



# Metadata for arkiv

---

Videre arbeid

# Bakgrunn for metadata for arkiv

---

- › Samfunnsendringer
  - › Behov for modernisering - nye typer arkiver og endrede forventninger til bruk
  - › Tettere sammenheng mellom informasjonsforvaltning, dokumentasjonsforvaltning, langtidsbevaring og tilgjengeliggjøring
  - › Dagens metadatakrav er utfordrende for forvaltningen
- › Arkivverkets satsing på Digitalarkivet
  - › Både bevaring og tilgjengeliggjøring er avhengig av gode metadata
  - › Forutsetning for arbeid med katalog
- › Bygger på tidligere arbeid og erfaringer

# Arbeidet med metadata for arkiv

---

- › Internt arbeid så langt
- › Ønskelig å se orden i eget hus og arkiv i sammenheng for å unngå dobbeltarbeid
- › Hovedantakelser:
  - › Endrede behov og teknologi, «kun-en-gang» gjør at vi kan forenkle kravene vi stiller til virksomheter som skal overføre arkiv til langtidsbevaring
  - › Arkivene mer anvendbare og prosessene blir smidigere dersom vi får strukturerte metadata på mottakstidspunktet
- › Fokus på
  - › Hvilke arkivfaglige hensyn skal metadata understøtte?
  - › Hvor i verdikjeden er det behov for metadata?

## Litt mer detaljert om arbeidet

---

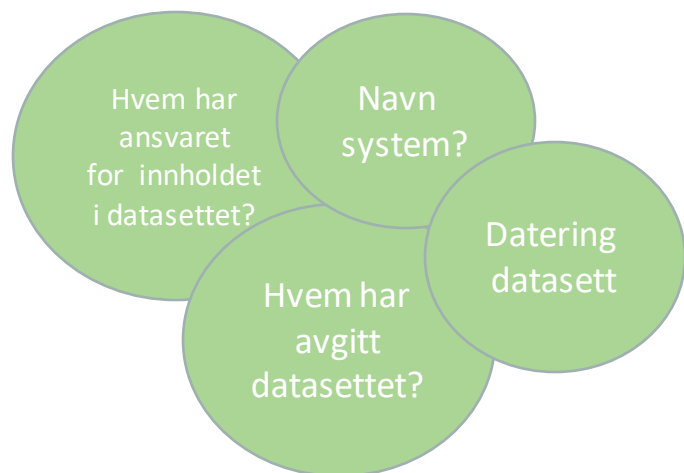
- › Skaffe oversikt over metadata som er i bruk i dag. Det har vært identifisert over 200 metadatabehov.
- › Tok utgangspunkt i metadataene vi bruker i dag. På datasettnivå hentet fra:
  - › DIAS-standardene, særlig DIAS-METS, DIAS-PREMIS og EAD
  - › ADDML
- › Vurdert opp mot metadatabehovene i dagens verdikjede. Metadata som ikke kan knyttes til et konkret behov ble tatt bort og metadata som er nødvendig for å dekke nye behov ble lagt til.

# Hovedfunn og resultater så langt

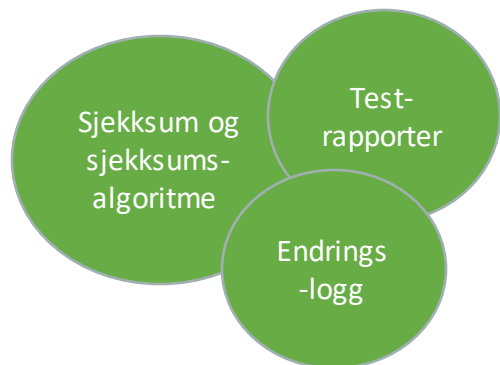
---

- Identifisert et utvalg metadata som dekker behovene til langtidslagring og tilgjengeliggjøring
- Oversikt over metadatabehov på datasett- og hovedobjektnivå
- Testes ut i Proof of Concept med Felles datakatalog
- Behov for bedre beskrivelser og tydeligere kobling til arkivfaglige hensyn og hvor de oppstår i arbeidet med langtidsbevaring og tilgjengeliggjøring

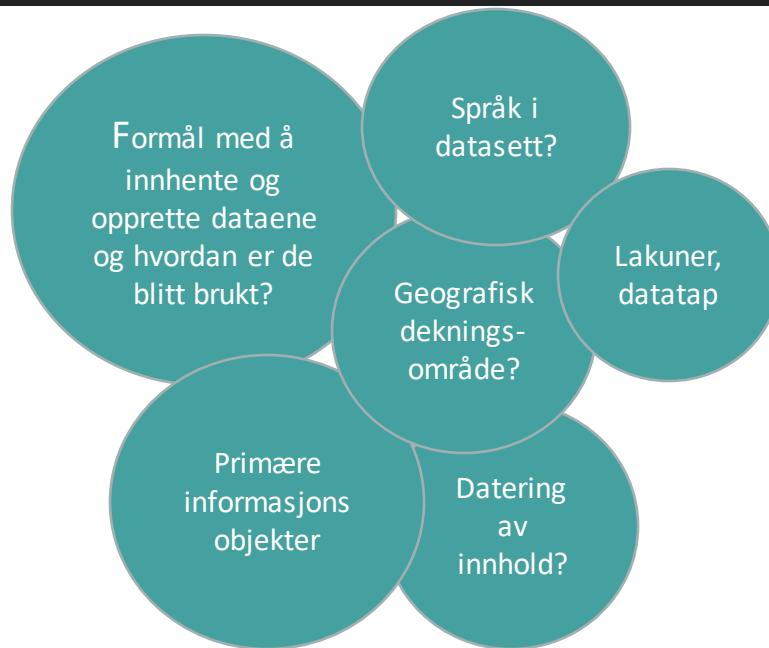
# Behov for metadata...



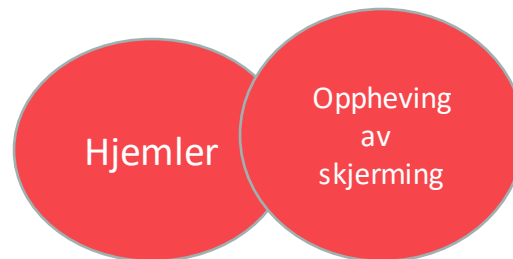
## Opphavet til dataene...



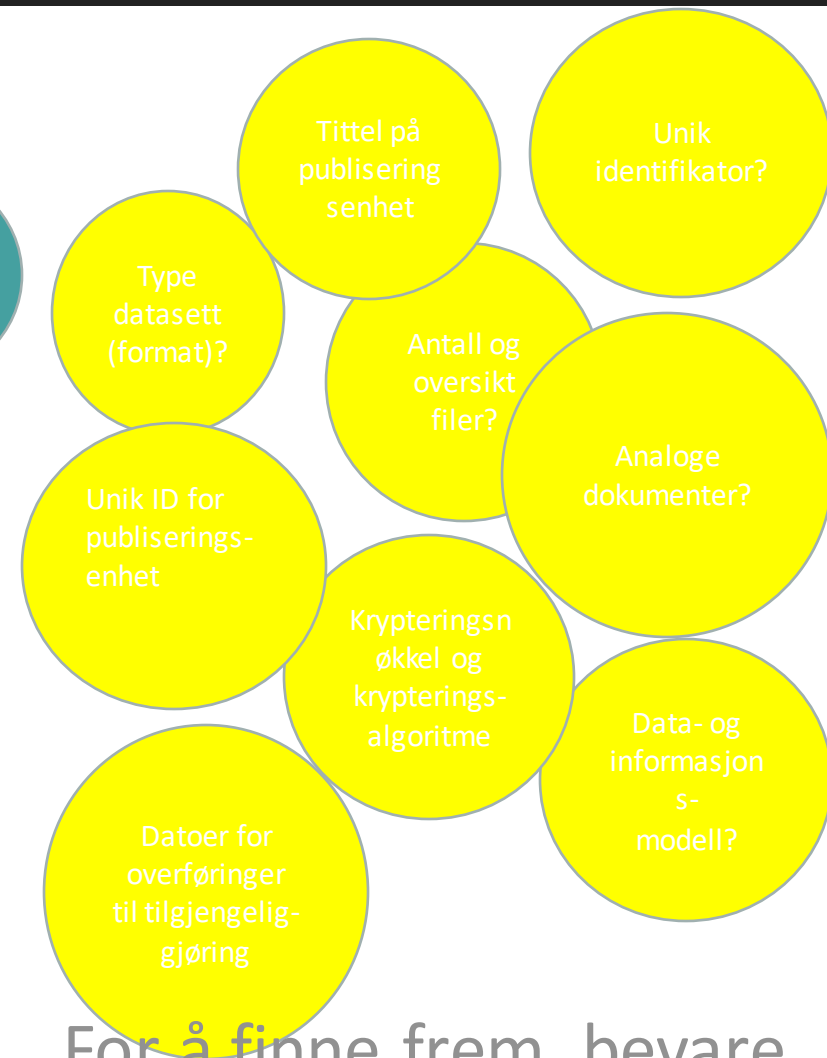
At de ikke har vært endret uautorisert...



## Innholdet i datasettet...

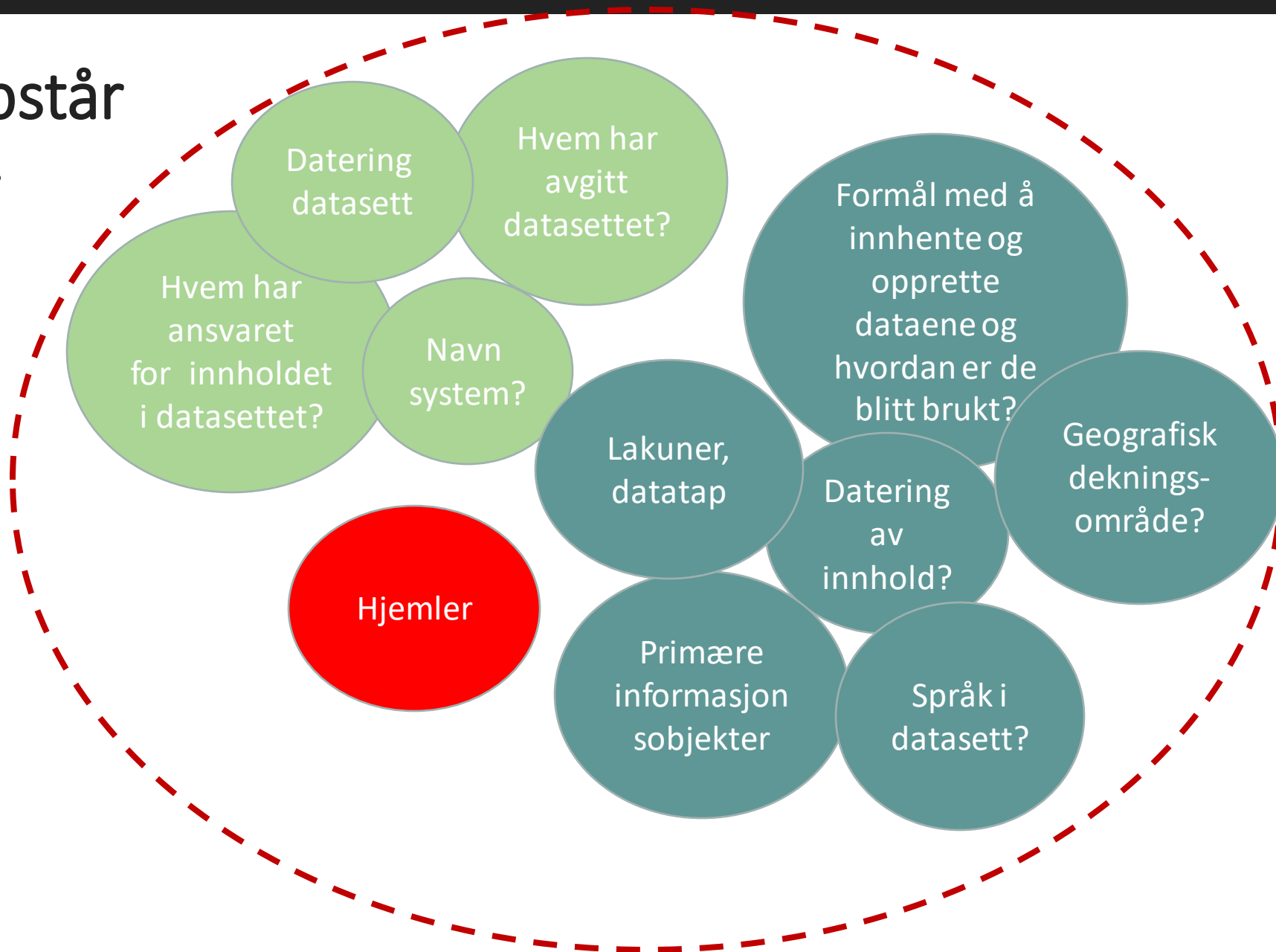


Hvem som skal ha tilgang...

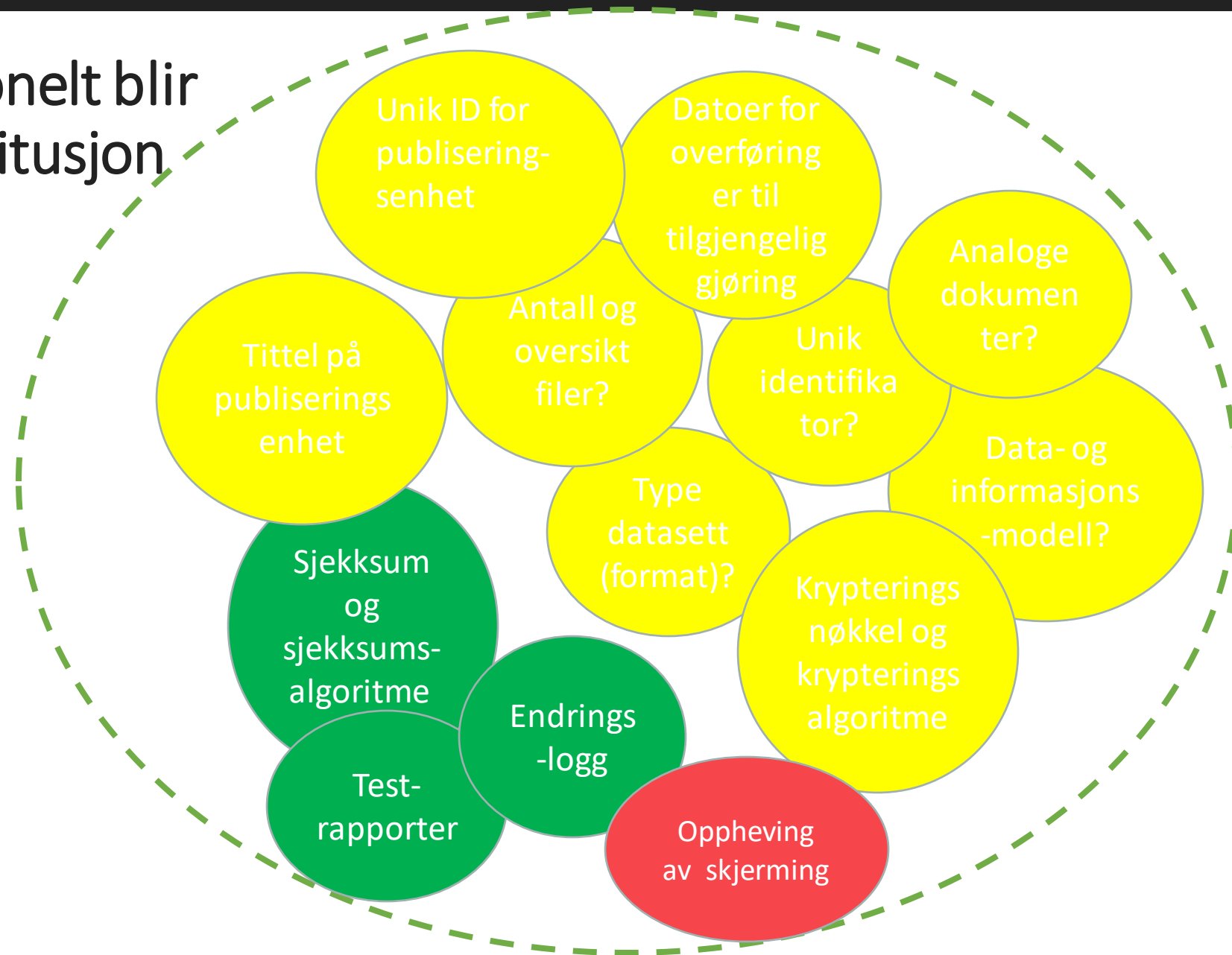


For å finne frem, bevare trygt og tilgjengeliggjøre...

# Metadata som oppstår i dokumentasjons- forvaltnings- prosesser



# Metadata som tradisjonelt blir tilført av bevaringsinstitusjon





# Proof of Concept med Felles datakatalog

---

- › Samarbeid med Felles datakatalog (Digdir)
- › Dialog informasjonsforvaltningsmiljø:
  - › Arbeidsgruppa (ca. 20 deltakere, statlig og kommunal)
  - › Innføringsmøter i felles datakatalog (ca 80-100 deltakere fra forvaltning)
- › Basert på oversikten over metadatabehov fra metadata for arkiv
- › Diskusjoner rundt NOU 2019:9. Ny lov om samfunnsdokumentasjon og arkiver

# Proof of Concept - Spørsmål

---

- › Utfordring: Vi beskriver datasett i mange omganger i ulike sammenhenger
- › Er det mulig å se disse beskrivelsene i større sammenheng?
  - › Som økosystem for deling av data, nasjonal samhandling og tjenesteutvikling?
- › Avsjekke hypoteser:
  - › Er det konseptuelt match mellom metadatabehov og Felles datakatalog? Hva er gapet?
  - › Hvilke standarder kan vi bruke for å dekke dette gapet?

# POC - funn

---

- Metadatabehovene på datasett nivå er bra dekket av Felles datakatalog
- På nivå av enkeltobjekt er metadatabehovene i svært liten grad dekket
- Det er ikke tydelig hva *hjemmel er*, og hvordan den skal refereres
- Begrepene *prosess* og *behandling* trenger forklaring

# Neste skritt

---

1. Utvide arbeidsgruppen med eksterne
2. **Identifisere det minimum av metadata** som er nødvendig for å ivareta arkivfaglige hensyn for mottak, langtidsbevaring og tilgjengeliggjøring
  - Utgangspunkt i arkivfaglige behov (tillit og anvendbarhet)
3. **Gi eksempler på hvordan metadata kan beskrives.**
4. Lage et utkast til en **oppdatert oversikt og beskrivelse over metadatabehov**, og eksempler på hvordan de kan beskrives.
5. Teste (enten i form av skrivebordsøvelser eller reelle tester) for å **verifisere om metadatabehovene er dekkende for de fleste typer arkiver.**

# Hovedantakelse for dette arbeidet

---

- "Vi tror at det langt på vei er samme metadatabehov for statlig, kommunale og private arkiv selv om kravene som stilles til ulike typer arkiv og ulike virksomheter vil kunne variere..."
  
- Teste:
  - Et statlig, et kommunalt og to privatarkiv (bedrift/organisasjon og personarkiv)
  - Et arkiv fra et Noark-system
  - Et fotoarkiv
  - Et SIARD databaseuttrekk
  - Et fagsystem (fellesløsning) som benyttes -av minst to ulike virksomheter
  - Et digitalisert multimedia arkiv

# Forutsetninger for det videre arbeidet

---

- › Tidlig og tett dialog med
  - › KS og eksterne interessenter
  - › Kjerneteamet i Innebygd arkivering, produktledere i Digitalarkivet og relevante fagmiljøene i AV
- › Tilgang til tilstrekkelig ressurser og rett kompetanse
- › Viktige føringer:
  - › Utgangspunkt i behov og muligheter, og ikke kun se på eksisterende standarder
  - › Fokus på "minimumsprodukter»
  - › Teste og verifisere tidlig for å vurdere og forbedre kontinuerlig
  - › Bruke hypoteser for å strukturere arbeidet med testing og verifisering

# Avgrensning for videre arbeid

## Innenfor

- Beskrivelse av metadatabehov som er nødvendige for å ivareta arkivfaglige hensyn for langtidsbevaring og tilgjengeliggjøring
- Forslag til hvordan metadatabehov skal uttrykkes
- Omfatter:
  - statlige, kommunale og private arkiver,
  - digitale og digitaliserte arkiver
  - Analoge arkiver som er registrert i katalog

## Utenfor

- Ikke tilbakevirkende kraft for materiale som allerede er mottatt
- I denne omgang ser vi ikke på regulering (riksarkivarens forskrift) og standarder

# Vedlegg

---



## Metadataelement (datasett)

Geografisk dekningsområde for datasettet

Språkene i selve datasettet

Unik kode for datasettet

Antall og oversikt over filer

Er det analoge dokumenter som er knyttet til uttrekket?

Kjente lakuner eller datatap, fullstendige data

Sjekksm for datasettet

Sjekksm algoritme for datasettet

Krypteringsnøkkel for datasettet

Krypteringsalgoritme for datasettet

Testrapporter

Endringslogg for datasett

Datoer for overføringer til tilgjengeliggjøring

Unik ID for publiseringsenhet

Tittel på publiseringsenhet

## Metadataelement (datasett)

Liste over alle arkivskapere

Navn på avleverende institusjon

Kontaktinformasjon til avleverende institusjon

Navn på system

Systemets formål og administrative brukssammenheng og historie

Hva er de primære informasjonsobjektene i datasettet/systemet?

Datamodell/informasjonsmodell for datasettet

Gradert informasjon

Skjerming som oppheves ved vedtak for hele datasettet

Skjerming som oppheves ved vedtak for elementer i datasettet

Skjerming som oppheves på dato for hele datasettet

Skjerming som oppheves på dato for elementer i datasettet

Startdato til innholdet i datasettet

Sluttdato til innholdet i datasettet

## Metadataelement (hovedobjektnivå)

Objekttype	Dette elementet er en referanse til en av objekttypene nevnt i rad 8. Helt nødvendig hvis rad 8 nevner flere enn én primær objekttype.
Entydig identifikator	En ID eller en annen identifikator som gjør det mulig å referere til (søke fram, hente ut) enkeltobjektene av denne typen i datasettet.
Aktør	Den personen eller institusjonen som gjør at objektene blir opprettet og videre behandlet eller registrert.
Funksjon/prosess	Den funksjonen, oppgaven eller prosessen hos arkivskaperen som fører til at objektene blir opprettet og viderebehandlet i systemet, og gjerne også hjemmelen for dette.
Utgangspunkt for objektet (tidsstempel)	Startpunktet for objektene, typisk en opprettelse, etablering, leveranse, fødsel e.a. med tilhørende data.
Endringer med tidsstempel	En kronologisk liste over de hendelsene eller endringene objektene blir utsatt for mellom et startpunkt og et sluttunkt i systemet, med en timestamp og en endringsbeskrivelse (logg).
Status (kan inngå i Endringer).	Et evt. sluttunkt for objektene i systemet, typisk nedleggelse, utmelding, salg, død e.a. med tilhørende data. Eller objektenes status når uttrekket produseres. Formålet her er ofte å avgjøre om objektet er avsluttet.
Klausulering	Opplysninger om at informasjonen om objekter er (eller kan være) unntatt