

SIARD i bruk i Arkivverket

Ane Hovdan

Seniorrådgiver

anehov@arkivverket.no

Samfunnet er i endring

Samfunnet i endring

- > Forvaltningen, næringsliv og innbyggere jobber, kommuniserer og bruker informasjon og data på nye måter
- > Tverrsektorielle prosesser og økt kompleksitet
- > Informasjonsmengden i samfunnet øker eksepsjonelt
- > Rettighetssamfunnet
- > Personvern (GDPR)

Hvordan treffer dette arkiv?

Datamengden som overføres fra offentlige virksomheter til Arkivverket og andre bevaringsinstitusjoner vil øke betydelig

Storstilt digitalisering av papirarkiver

Økt kompleksitet og hyppige endringer, mange ulike formater

Digitale arkiv vil bli overført tidligere

Endrede forventninger hos brukerne

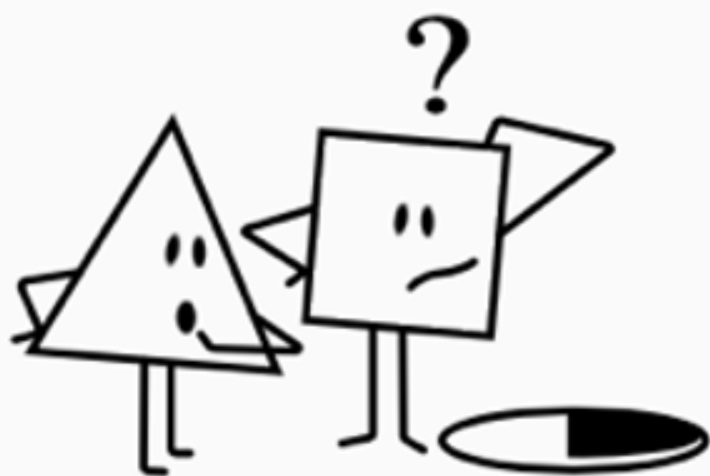
Nye bruksformål og brukergrupper

Gjeldende praksis for avleveringsformater

- Arkivverket fastsetter krav til formater for avlevering
- For statlig sektor beskrives krav knyttet til avlevering av digitalt materiale i Riksarkivarens forskrift kapittel 5
- For kommunal- og privat sektor står man friere til å sette egne krav



Stor variasjon i datasettene vi forventer å motta



- Ulike typer datasett
- Ulike datasett har ulikt behov for beskrivelser og bearbeiding for å kunne gjenbrukes
- I tillegg er det ulike måter et datasett kan gjenbrukes og det er ikke sikkert at alle datasett behøver like omfattende beskrivelser

Hvordan imøtekommer vi dette?

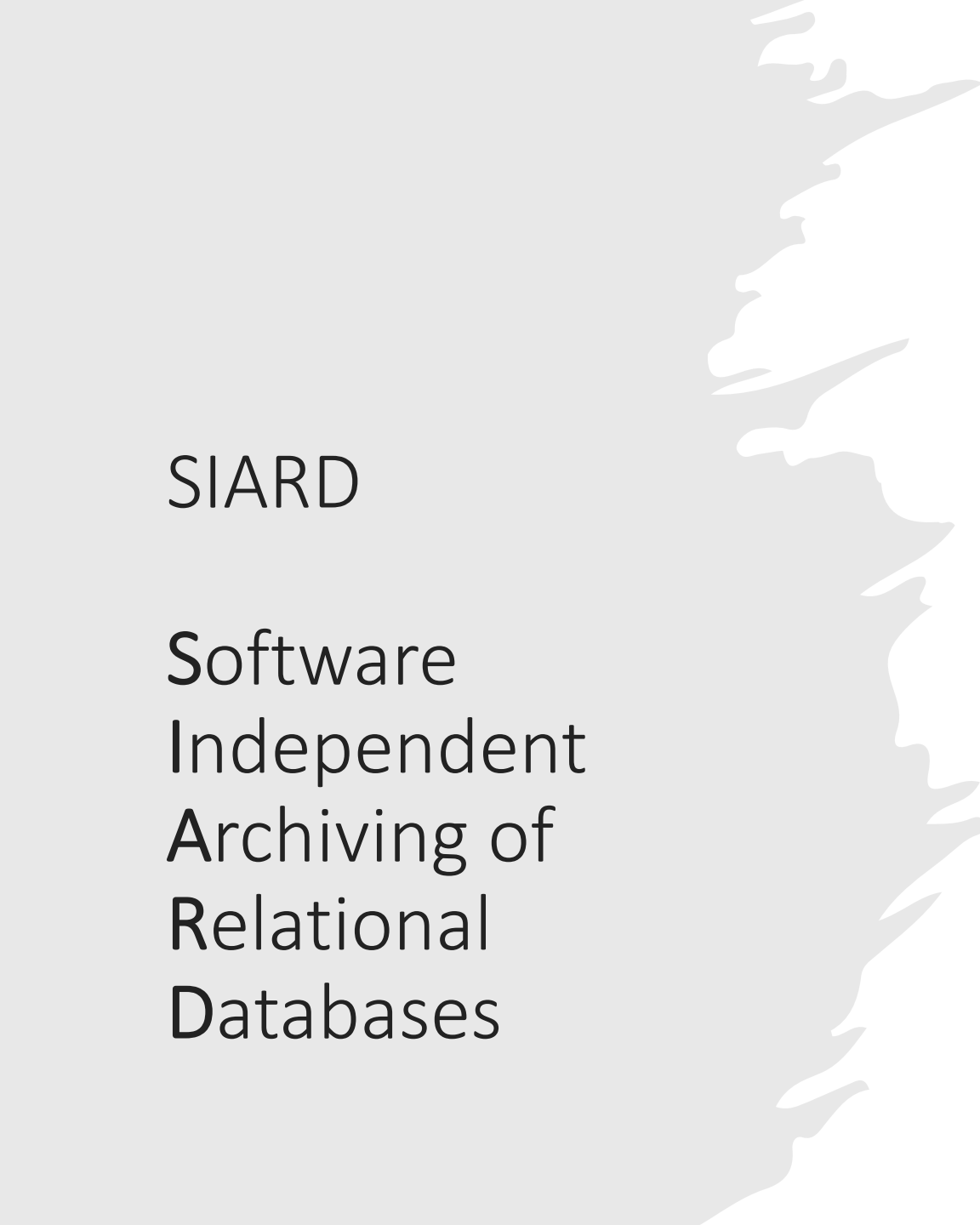
Hvilke behov har vi?

Minimumssett av metadata som ivaretar arkivfaglige hensyn og eventuelt andre vesentlige metadatabehov i prosessene for mottak, langtidsbevaring og tilgjengeliggjøring og samtidig legger til rette for automatisering, forenkling og fleksibilitet i prosessene

- > Hypotese 1: Vi tror at endrede behov, prinsippet om «kun-en-gang» og mulighetene teknologien gir oss gjør at vi kan forenkle kravene vi stiller
- > Hypotese 2: Vi tror at arkivene blir mer anvendbare og prosessene smidigere dersom vi ser behovene for metadata i sammenheng med dataenes helhetlige livsløp

Bevissthet om arkivfaglige hensyn

- Arkivteori og ISO-standarder: autentisitet, pålitelighet, integritet og anvendbarhet (APIA) og proveniens
 - NOU 2019:9: ekte, pålitelig, anvendbar, dekkende og satt i sammenheng
- Opphavssituasjonen er kjent
 - Informasjonsinnholdet er uendret
 - Informasjonen er forståelig og tilgjengelig for berettiget bruk
 - Informasjonen er satt i sammenheng
 - Informasjonen er dekkende



SIARD

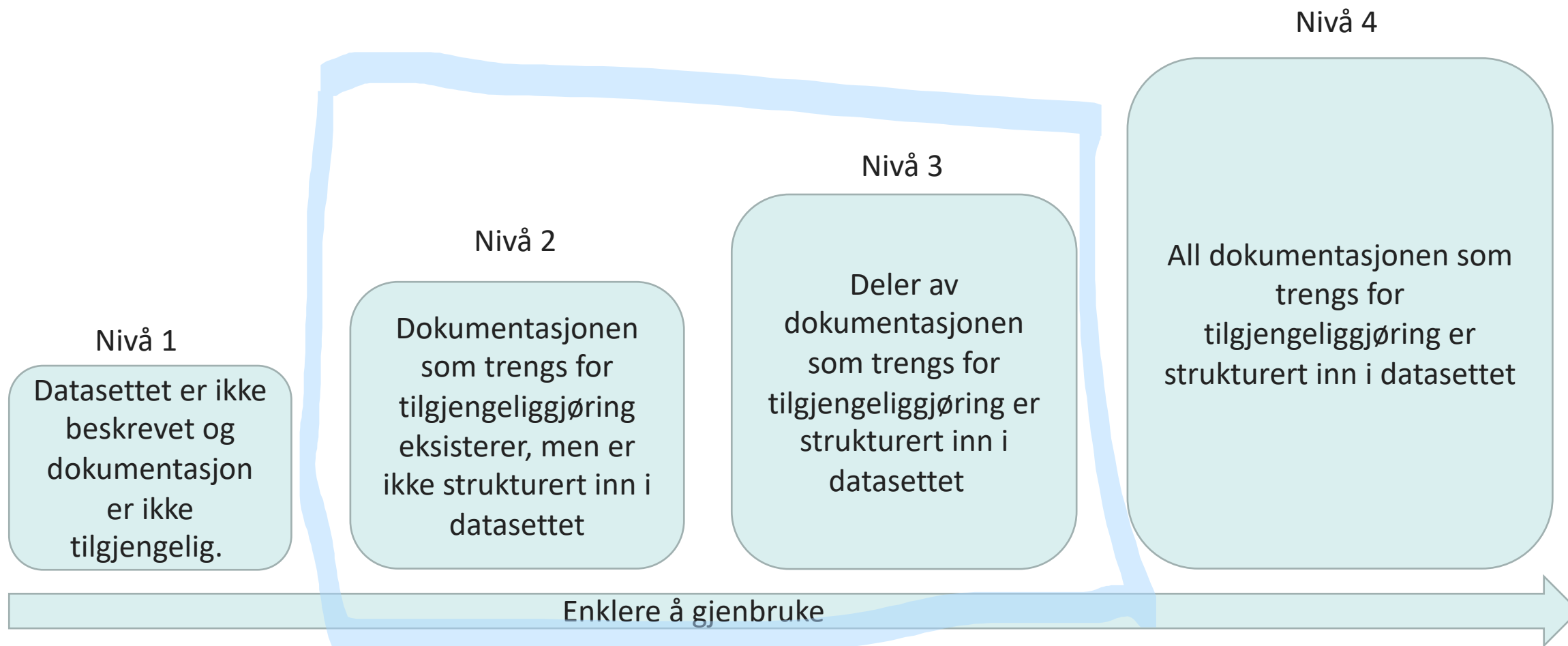
Software Independent Archiving of Relational Databases

- › Fremstille SIARD-versjon av databasen – enklere, raskere og billigere for forvaltningen
 - › Senker terskelen, fører til at arkivene kommer til depot
- Internasjonalt format med bred utbredelse
- Et SIARD-datasett er en speiling av databasen, men mangler applikasjonslogikken.
 - Måten systemet er bygget opp på har stor betydning av komplettheten/forståelsen av datasettet.
- **Hvor mye ekstra beskrivelser kreves for å kunne gjenbruke et SIARD-uttrekk??**

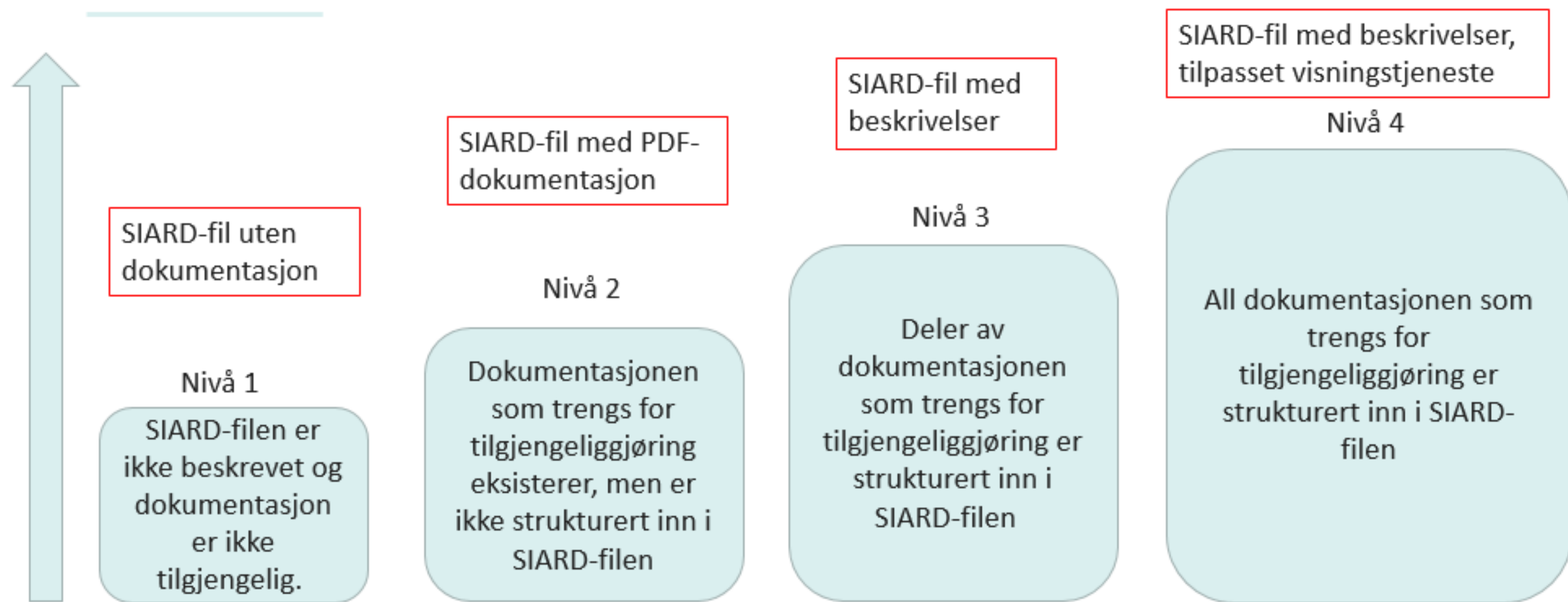
Vurdering opp mot arkivfaglige hensyn

- Tilliten til arkiv er avhengig av tilstrekkelig ivaretagelse av hensynene autenticitet, anvendbarhet, pålitelighet og integritet
- **Anvendbarhet**
 - Materialet finnes. Dette innebærer at materialet ikke er tapt.
 - Materialet er teknisk tilgjengelig. Dette innebærer at materialet kan fremstilles for videre bruk.
 - Materialet er forståelig. Dette innebærer at materialet kan tolkes.
 - Materialet kan gjenbrukes. Dette innebærer at materialet kan benyttes videre av andre virksomheter og i andre sammenhenger.

Nivå-inndeling av datasett basert på anvendbarhet



Forslag til tilnærming: Nivå-inndeling basert på anvendbarhet



Hva dekker retningslinjene?

- Legge føringer for hvor mye informasjon som trengs for å sikre anvendbarhet av datasett i SIARD-format og at det er godt nok for tilgjengeliggjøring
- Ta høyde for at databasene er ulike og dermed har ulike behov for beskrivelser
- Legge føringer for hvordan denne informasjonen skal struktureres, særlig med tanke på gjenbruk

Forslag: Hva kan inngå i de ulike nivåinndelingene?

- Nivå 1: Datasett uten beskrivelser eller dokumentasjon.
- Nivå 2: Datasett beskrevet på feltnivå. Kjenner til relasjoner, men de er ikke maskinlesbare. Datasett kan inneholde typefeil, datofeil, EBCDIC mv. Har tilgang til dokumentasjon som er nødvendig for videre anvendelse av arkivet. Hovedobjekt fra metadata for arkiv er identifisert på et overordnet nivå.
- Nivå 3: Datasett beskrevet på feltnivå. Relasjoner ligger i SIARD-filen. Datasettet inneholder kun mindre viktige tekniske feil. Har all dokumentasjon som er nødvendig for videre anvendelse av arkivet. Hovedobjekt fra metadata for arkiv er identifisert og man har identifisert korresponderende tabeller og felter i datasettet.
- Nivå 4: Datasettet er fullt ut tilpasset til tilgjengeliggjøringsløsning. Hovedobjekt fra metadata for arkiv er identifisert og mappet opp i uttrekket.



Søndre Buskerud politidistrikt d2541b7c-6f78-4bf4-9ebd-6fe25a49aec7

Project: 37.42% described 14 files

More ▾

Merge

New template

Save project

Find tables

Filter: All (164 package tables)

A-Z By priority high

- ✓ D dl_comment (30496)
- ✓ dl_company (1)
- ✓ dl_eldoklink (201506)
- ✓ dl_orqunit (55)
- ✓ dl_person (328)
- ✓ D edokfiles (0)
- ✓ edoktab (308381)
- ✓ oa_dok_adr (136664)
- ✓ **D oa_dokument (147447)**
- ✓ oa_sak (58618)
- ✓ Ca_eldok_history (40402)
- ✓ Ca_eldok_log (81623)
- ✓ ca_signoff_types (6)
- ✓ ca_user (291)
- ✓ dl_category (3)
- ✓ D dl_comment_files (0)
- ✓ dl_crossref (14)

Database name
15_sondre_buskerud_2003-2017_20190211_conv.siardMatch
37.42% ▾Matching template name
Doculive v7.0 MODARK 2019-04-16

Preview

Package table oa_dokument	Similarity 100.00%	Template table oa_dokument
begrensning {VARCHAR}		begrensning {VARCHAR}
avskjerming {INT}		avskjerming {INT}
gruppe {VARCHAR}		gruppe {VARCHAR}
avgradering {VARCHAR}		avgradering {VARCHAR}
gradvekt {INT}		gradvekt {INT}
nedgraddato {TIMESTAMP}		nedgraddato {TIMESTAMP}

oa_dokument (147447 rows) 33/74 (45%) described

More ▾

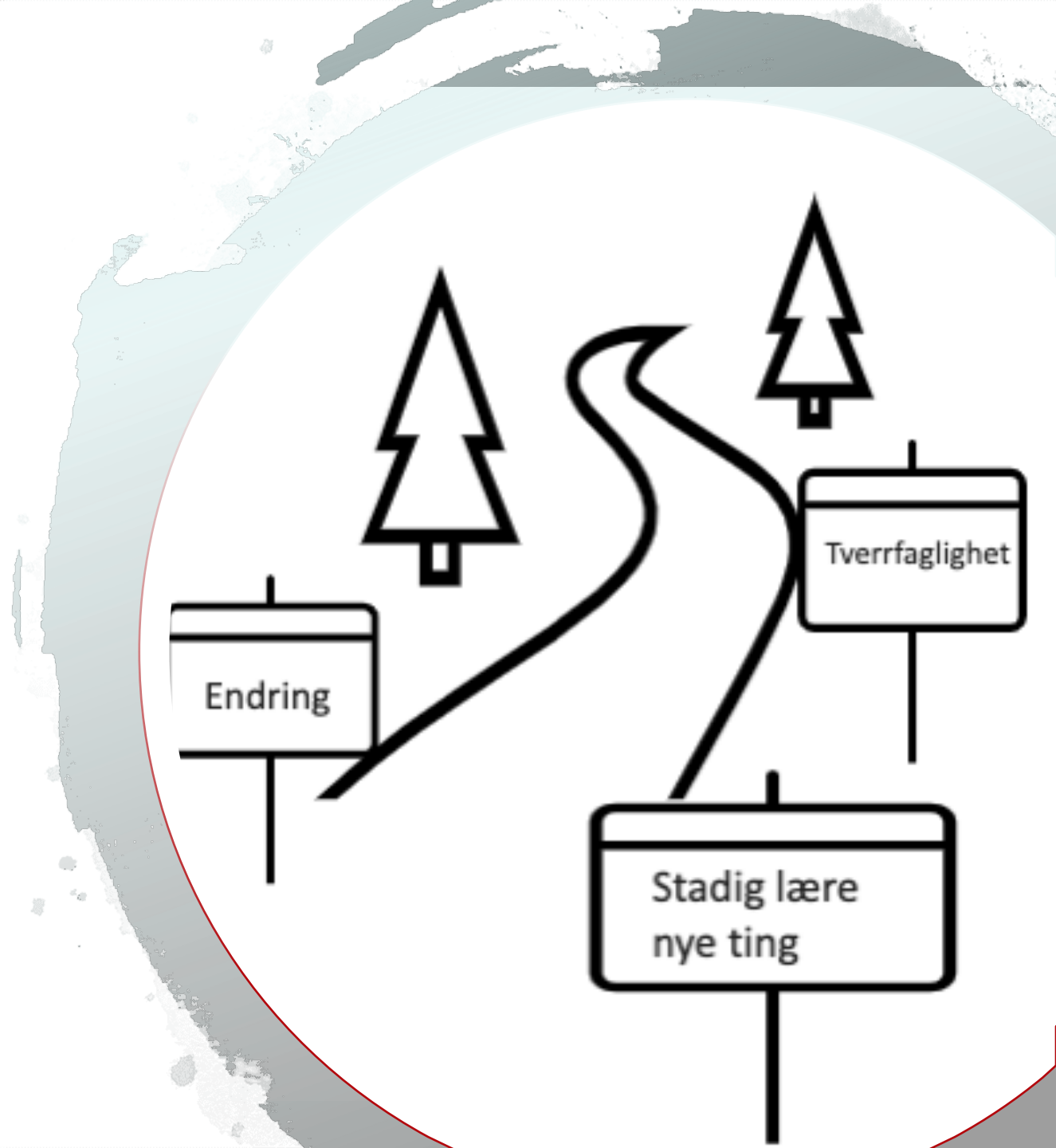
☒ Done

Columns diff Keys diff Data content

gradvekt	gradvekt
{n4:na} Graden av Begrensning	{n4:na} Graden av Begrensning
nedgraddato	nedgraddato
{n4:na} (?) Nedgraderingsdato	{n4:na} (?) Nedgraderingsdato
saksnr	saksnr
{n4:JP.SAID} {n5:sakaar+sakssekvensnummer#journalpost} Saksnr 4 siffer år + 5 eller 6 siffer løpenr + journalpostnummer =#RegistreringsID/{n4:JP.SAID} {n5:sakaar+sakssekvensnummer#journalpost} Saksnr 4 siffer år + 5 eller 6 siffer løpenr + journalpostnummer =#RegistreringsID	{n4:JP.SAID} {n5:sakaar+sakssekvensnummer#journalpost} Saksnr 4 siffer år + 5 eller 6 siffer løpenr + journalpostnummer =#RegistreringsID/{n4:JP.SAID} {n5:sakaar+sakssekvensnummer#journalpost} Saksnr 4 siffer år + 5 eller 6 siffer løpenr + journalpostnummer =#RegistreringsID
Expand	
journaldato	journaldato
{n4:JP.JDATO} {n5:journaldato#journalpost} Journaldato	{n4:JP.JDATO} {n5:journaldato#journalpost} Journaldato
type	type
{n4:JP.NDOKTYPE} {n5:journalposttype#journalpost} Dokumenttype Oppslag DOKTYPE.TYPE	{n4:JP.NDOKTYPE} {n5:journalposttype#journalpost} Dokumenttype Oppslag DOKTYPE.TYPE
uoff	uoff
{n4:JP.UOFF} {n5:skjermingshiemmel#journalpost} Hiemmel u off	{n4:JP.UOFF} {n5:skjermingshiemmel#journalpost} Hiemmel u off

Refleksjoner

- Det gjenstår en del arbeid og utprøving av det vi har presentert her
- For systemer med enkle datastrukturer ser dette ut til å være en veldig gunstig metodikk
- Vi har oppdaget at for systemer med mer komplisert datastruktur byr metodikken på flere utfordringer
- Vi tror at dette i fremtiden vil være en av flere måter å avlevere arkiv fra produksjonssystem som bygger på relasjonsdatabaser



Diskusjonsoppgaver

- Hva tenker dere er godt nok for å sikre anvendbarhet av et Noark-system i SIARD-format?
 - Hva er viktige søkeinnganger når man skal gjenbruke et Noark-system?
- Oppfatter dere at det er utfordrende å fremstille tradisjonelle Noark-uttrekk? Hva er de største utfordringene knyttet til dette?
- Ser dere noen svakheter med SIARD-metodikken som gjør at man ikke kan ha tillit til disse datasettene som arkiv?