



Felles datakatalog og arkiv

Arkiv, informasjonsforvaltning og det store bildet

Petr Vasilev

04.11.2020

Agenda

- › Bakgrunn
- › Tidligere arbeid
- › Hypoteser
- › Funn
- › Videre arbeid
- › Det store bildet

Bakgrunn

- › Informasjonsforvaltning og arkiv er to sider av samme sak
 - › Behov for mer innsikt om sammenhengen mellom arkiv og infoforvaltning
- › Bygge på kompetansen i informasjonsforvaltningsmiljøet
 - › Samspillet rundt Felles datakatalog

Tidligere og parallelt arbeid

- › Innebygd arkivering
 - › Baserer seg på samme arkivfaglige hensyn
 - › Piloter med NAV og Kulturtanken
- › Metadata for arkiv – del 1
 - › Forenkling av behov for metadata
 - › Baserer seg på arkivfaglige hensyn fra NOU
- › Diskusjoner rundt NOU 2019:9. Ny lov om samfunnsdokumentasjon og arkiver

Forutsetninger for arbeidet

- › Proof of concept
 - › Hovedfokus på å skaffe ny innsikt
 - › Deltakerne er ikke forpliktet for å følge opp videre arbeidet
- › Leveransen er utformet av arbeidsgruppen, og ikke Arkivverket
 - › Leveransen kan ikke tolkes som Arkivverkets offisielle føringer
 - › Det er fritt rom for deltakere i arbeidsgruppen for å utveksle meninger og tanker
- › Arbeidsgruppa konsentrert seg om behov, indirekte standarder

Spørsmålet vi stilte oss

- › Utfordring: Vi beskriver datasett i mange omganger i ulike sammenhenger
 - › Arkivplan, orden i eget hus, personvern og informasjonssikkerhet...
- › Er det mulig å se disse beskrivelsene i større sammenheng?
- › Gjøre arkiv til en tydeligere del av økosystem for deling av data, nasjonal samhandling og tjenesteutvikling?

Hypoteser og metadatabehov

- Det er mulig å bruke beskrivelsen av datasett i Felles datakatalog for å dekke metadatabehov for mottak, bevaring og tilgjengeliggjøring.
- Det er mulig å bruke standarder fra informasjonsforvaltning til dette.
- Metadatabehov på datasett
- Metadatabehov på enkeltobjekt
 - Nødvendig for å bygge skalerbare automatiske løsninger
 - Forklarer hensyn / mening med enkle egenskaper ved objekter

Våre 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

Funn : Datasett nivå

- DCAT-AP-NO 1.1 dekker mesteparten av metadatabehov på datasett
- DCAT-AP-NO 2.0 dekker de aller fleste metadatabehov på datasettnivå

Funn : Objektnivå

- › Ingen av metadatabehov på objektnivå kan tilfredsstilles verken med JSON Schema eller med ModelIDCAT-AP-NO
- › Formålet med dagens standarder for infomodeller
 - › Maskinell validering av datasett
 - › Menneskelig tolkning av modell og datasett
- › Formålet med metadatabehov på objektnivå
 - › Maskinell tolkning av modell og datasett

Funn : Begrepet hjemmel og oversikt over hjemler

- › Definisjon av hjemmel varierer med brukergruppe
 - › Er samtykke, kontrakt eller vedtak en hjemmel?
 - › Samtidig finnes det delvis enighet: Henvisninger til regeltekst (bl.a. lover og forskrifter) tolkes entydig som hjemmel
- › Referanser til hjemmel er ikke alltid varige
 - › Er det nok med Språkrådets skriveregler for lovhenvisninger?
 - › Trenger vi hjemmelskatalog?

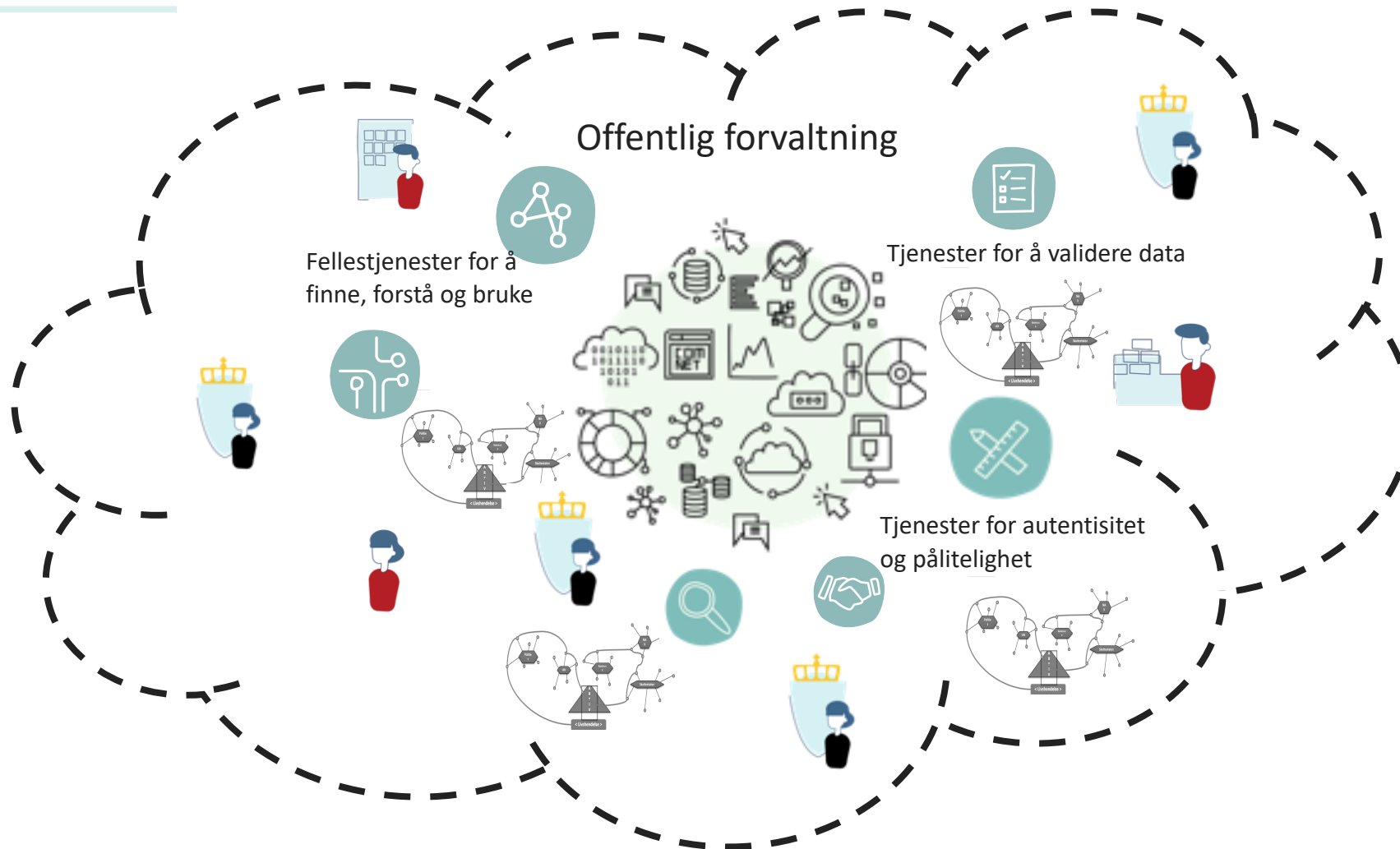
Funn : Prosesse og behandling

- Enkeltojekter må kunne kobles til prosesser eller behandlinger som skapte eller endret de
- Prosesskartlegging kan være ganske overordnet
 - foregår på overordnet nivå
 - er risikobasert
- Prosesskartlegging er del av arbeidet med orden i eget hus
- Veileder for orden i eget hus bør detaljere hvordan kobling mellom egenskaper til objekter kobles til de kartlagte prosessene og behandlingene

Arbeidsgruppens anbefalinger

- › Arkivverket bør ta innsikten med videre, for eksempel i prosjektet «Metadata for arkiv»
- › **Kun anbefalinger:**
 - › Digdir og Arkivverket: Videreutvikle veileder for orden i eget hus med kobling til prosess eller behandling
 - › Digdir / Informasjonsforvaltning miljø: Arbeide med ModellDCAT-AP-NO for maskinelt tolkning av modell og datasett
 - › Digdir / Nasjonalt ressurscenter for deling av data: Avklaring av hjemmelsbegrepet
 - › Digdir: Varig referanse til hjemmel

Felles økosystem for nasjonal digital samhandling og tjenesteutvikling



Takk!
Tid for spørsmål og svar

Petr Vasilev

04.11.2020

Eksempel av metadatabehov på datasett

Behovet	Beskrivelse
Hvilke virksomheter skapte datasettet?	Liste over alle arkivskaperene. Det offisielle navnet eller / og orgnr / personnr på den eller de virksomheter eller personer som har skapt arkivet. Unngå forkortelser.
Informasjonsmodell for datasettet	Maskinlesbar beskrivelse av hvordan data representeres og aksesseres. Eksempler på modellspråk som kan brukes er: ER, UML klassediagrammer, OWL, RDFS
Hvilke hjemler gjelder når datasettet skal behandles i sin helhet?	Er det ekstra hjemler som gjelder for datasettet i sin helhet. I dette tilfelle skrives disse hjemlene i dette feltet. Behovet skal hjelpe å ha utleveringshjemler for datasettet i sin helhet.
Hvilke geografiske områder er relevante?	Geografisk dekningsområde for datasettet. Landskode (ISO 3166-1 alpha-3), regionkode/fylkesnummer (ISO 3166-2:NO) og/eller kommunenummer (0301, 0411, 2030 osv)

Eksempel av metadatabehov på enkle objekter

Behovet	Beskrivelse
Entydig identifikator	En identifikator som gjør det mulig å referere til (søke fram, hente ut) enkeltobjektene av denne typen i datasettet. Mer info: difi.no/referanse katalogen/pekere-til-offentlige-ressurser-pa-nett
Aktør	Alle aktørene som er relaterte til funksjonen og prosessen
Prosess / Behandling	Den funksjonen, oppgaven eller prosessen hos arkivskaperen som fører til at objektene blir opprettet og viderebehandlet i systemet, og gjerne også hjemmelen for dette.
Hjemler / Behandlingsgrunnlag	Opplysninger om hvilke hjemler gjelder dette objektet.