



Beste praksis for å tilgjengeliggjøre åpne offentlige
data ved bruk av semantiske vevteknologier

av

Johan G. Eriksson

Masteroppgave i informasjonsvitenskap
Institutt for informasjons- og medievitenskap
Universitetet i Bergen

Våren 2011

Abstrakt

Organisasjoner i offentlig sektor innehar store mengder data som kan tilgjengeliggjøres som åpne data slik at andre aktører kan gjenbruke dataene. Tilgjengeliggjøring ved bruk av semantiske vevteknologier vil gi dataene utvidede muligheter for maskinlesbarhet og gjenbruk. Jeg har undersøkt ulike praksiser rundt tilgjengeliggjøring av åpne data i offentlig sektor og bruk av semantiske vevteknologier på åpne data. Jeg har utført elleve semi-strukturerte intervjuer med personer fra privat og offentlig sektor. Den kvalitative analysen av intervjuene gir innsikt i ulike fordeler, erfaringer og forutsetninger for å tilgjengeliggjøre åpne data. Oppgaven beskriver en beste praksis som vil veilede offentlige organisasjoner under arbeidet med å tilgjengeliggjøre åpne data, samt gi IT-leverandører en forståelse av hvilke fordeler semantisk vev og lenkede åpne data vil gi.

Forord

Jeg vil først og fremst si tusen takk til veilederen min Andreas Lothe Opdahl. Du har vært engasjert og gitt tydelige tilbakemeldinger. Jeg har lært masse av deg og setter utrolig stor pris på veiledningen du har gitt meg. Takk til Pia V. Jøsendal for entusiasme og tilbakemeldinger i oppgavens startfase. Takk til Roar Fjellheim for at du foreslo idé til oppgaven. Jeg vil selvfølgelig si tusen takk til dere som lot dere intervjuer og ville dele deres erfaringer. Det var kjempespennende å prate med dere!

Kjære Anette, tusen takk for at du er du! Tusen takk for at du har vært så tålmodig, støttende og forståelsesfull når jeg har jobbet med oppgaven! Og takk for at du passet på meg så jeg fikk i meg mat. Emilie, du er herlig og jeg er så stolt av deg! Pappa, jeg har hatt kjempestor glede av å diskutere oppgaven min med deg underveis! Du har motivert meg og alltid støttet meg. Takk til Cecilie og Mamma for at dere støtter meg og viser interesse for det jeg holder på med.

Takk til Rafael, Krister, Preben, Torgeir, Jonas og dere andre for alle høydepunktene de siste årene ved UiB.

Bergen, juni 2011
Johan G. Eriksson

Innhold

ABSTRAKT	2
FORORD	2
FIGURER	4
TABELLER	4
1. INNLEDNING	5
2. TEORI	7
2.1 DATADEFINISJONER	7
2.1.1 Informasjon og data.....	7
2.1.2 Metadata.....	7
2.1.3 Referansedata	8
2.1.4 Åpne offentlige data.....	9
2.2 SEMANTIKK	10
2.2.1 Viktigheten av at vi forstår dataene.....	10
2.2.2 Semantisk interoperabilitet.....	10
2.2.3 Begrepsapparater	11
2.2.4 Semantiske teknologier.....	11
2.2.5 Den semantiske veven.....	11
2.3 SEMANTISKE VEVTEKNOLOGIER.....	12
2.4 LENKEDE DATA	16
2.4.1 Lenkede åpne data	17
2.4.2 Fem stjerner ved åpne data	17
2.5 INNFØRING AV NY IT I OFFENTLIGE ORGANISASJONER.....	17
3. METODE	18
3.1 FORMULERING AV FORSKNINGSSPØRSMÅL.....	18
3.2 INNSAMLING AV BAKGRUNNSINFORMASJON	19
3.2.1 Utvalgelse av informanter.....	19
3.3 UTFØRING AV INTERVJUER.....	20
3.3.1 Intervjuguide.....	20
3.3.2 Gjennomføring av intervjuene	22
3.4 TRANSKRIBERING.....	22
3.5 KODING OG KATEGORISERING	23
4. RESULTATER	28
4.1 ÅPNE DATA	28
4.1.1 Fordeler.....	28
4.1.2 Hva har blitt gjort.....	29
4.1.3 Roller i tilgjengeliggjøringsarbeidet	30
4.1.4 Forutsetninger for at offentlige organisasjoner tilgjengeliggjør åpne data	31
4.1.5 Forutsetninger for at offentlige organisasjoner gjenbraker åpne data.....	33
4.1.6 Semantikkregisteret for elektronisk samhandling.....	35
4.2 SEMANTISKE VEVTEKNOLOGIER	36
4.2.1 Fordeler.....	36
4.2.2 Hva har blitt gjort.....	37
4.2.3 Forutsetninger for at offentlige organisasjoner bruker semantiske vevteknologier	40
4.3 LENKEDE DATA/LENKEDE ÅPNE DATA	43
4.3.1 Fordeler.....	43
4.3.2 Hva har blitt gjort.....	46
4.3.3 Forutsetninger for at offentlige organisasjoner tilgjengeliggjør Linked Open Data	47
4.3.4 Forutsetninger for at offentlige organisasjoner gjenbraker Linked Open Data	50

4.4	GENERELLE UTFORDRINGER TILKNYTTET LENKEDE ÅPNE OFFENTLIGE DATA	51
4.4.1	Tilknytning til lovverket	51
4.4.2	Tap av statlig inntekt	51
4.4.3	Politiske styringssignaler om å tilgjengeliggjøre åpne data	52
4.4.4	Begrensede midler til å utforske nye teknologier	52
4.4.5	Bedre informasjonsforvaltning må på plass før åpne data kan gjenbrukes i det offentlige	53
4.4.6	Spesialister til å velge datasett	54
4.4.7	Personvern	55
5.	DISKUSJON	55
5.1	BESTE PRAKSIS	55
5.1.1	Før tilgjengeliggjøring	56
5.1.2	Tilgjengeliggjøring	59
5.1.3	Etter tilgjengeliggjøring - Bruk av lenkede åpne data	62
5.1.4	Beste praksis fremstilt som en prosess	63
5.2	EVALUERING AV FORSKNINGSMETODEN	66
6.	KONKLUSJON OG VIDERE ARBEID.....	68
6.1	KONKLUSJON.....	68
6.2	VIDERE ARBEID	69
7.	KILDELISTE	70

Figurer

FIGUR 1:	EKSEMPEL PÅ REFERANSEDATA	8
FIGUR 2:	SEMANTIKK VED DELING AV INFORMASJON.....	10
FIGUR 3:	TEKNOLOGISTABELN FOR DEN SEMANTISKE VEVEN (SEMANTICFOCUS, 2010)	12
FIGUR 4:	ENKEL RDF-EKSEMPEL 1	13
FIGUR 5:	ENKELT RDF-EKSEMPEL 2	13
FIGUR 6:	ENKELT RDF-EKSEMPEL 3	14
FIGUR 7:	DIAGRAM AV LINKING OPEN DATA-SKYEN ANNO SEPTEMBER 2010	17
FIGUR 8:	5 STJERNER VED LINKED OPEN DATA	17
FIGUR 9:	6 STJERNER VED LINKED OPEN DATA	61
FIGUR 10:	BESTE PRAKSIS FREMSTILT SOM EN PROSESS.....	65

Tabeller

TABELL 1:	OVERSIKT OVER INFORMANTENE	20
TABELL 2:	INTERVJUMATRISSE	22
TABELL 3:	ITERASJON 1	24
TABELL 4:	ITERASJON 2.....	25
TABELL 5:	ITERASJON 3	26
TABELL 6:	OPPSUMMERING AV BESTE PRAKSIS	55

1. INNLEDNING

Offentlige data ønskes åpent tilgjengelig (Berners-Lee & Shadbolt, 2010; Myrseth, 2009; Øvrebø, Grene, & Reiersen, 2010). Tilgjengeliggjøring av åpne offentlige data vil øke bevisstheten rundt det offentlige i tillegg til at en slik åpning av data vil tillate myndighetene, landet og verden å fungere mer effektivt (Berners-Lee, 2009). Gjennomsiktbarhet har vært et av nøkkelordene bak den amerikanske regjeringens "Open Government Initiative". De forklarer gjennomsiktbarhet som at jobben de gjør skal være synlig for befolkningen (Orszag, 2009). Når offentlige data frigis, øker synligheten ved at alle kan få innsyn i hva myndighetene gjør. Det vil sørge for offentlig forsvarlighet og lik tilgang til offentlige tjenester (Kamal, 2006). I desember 2009 presenterte den amerikanske regjeringen et nytt direktiv som sa at hver offentlig virksomhet skal innen 45 dager åpent tilgjengeliggjøre minst tre datasett av høy verdi. Med høy verdi menes:

"Informasjon som kan bli brukt for å øke virksomhetens ansvarlighet og responstid; forbedre offentlige kunnskap om virksomheten og dens funksjoner og kjernevirksomhet, skape økonomiske muligheter, eller respondere på behov og etterspørsel som blir identifisert gjennom offentlig konsultasjon" (Orszag, 2009).

Direktivet forklarer at høy verdi på informasjonen vil legge grunnlaget for gjennomsiktbarhet i offentlig data. Etter at offentlighetsloven kom, har det vært større fokus på at offentlige organisasjoner tilgjengeliggjør data åpent. Offentlighetsloven har som formål å legge til rette for viderebruk av offentlig informasjon (Lovdata, 2011). Formålsparagrafen er gjengitt under:

"Formålet med lova er å leggje til rette for at offentleg verksemd er open og gjennomsiktig, for slik å styrkje informasjons- og ytringsfridommen, den demokratiske deltakinga, rettstryggleiken for den enkelte, tilliten til det offentlege og kontrollen frå ålmenta. Lova skal òg leggje til rette for vidarebruk av offentleg informasjon".

Innovasjon og nyskaping har ofte blitt nevnt blant de største motivasjonsfaktorene for å tilgjengeliggjøre åpne data. Kort tid etter at Avinor tilgjengeliggjorde sine flydata, kom det ulike systemer basert på disse dataene (Christensen, 2009). Eksemplet om Avinor underbygger Tim Berners-Lee sin påstand om at dersom det offentlige gjør sin jobb med å friggi data, vil det tillate bedrifter og andre uavhengige utviklere å bygge innovative tjenester som de kan levere til brukere. Frigivelse av data vil øke frivillige organisasjoner sin mulighet til å bidra med kreative løsninger til uløste samfunnsoppgaver som kan lede til en forbedring i det offentlige tjenestetilbudet.

Det er mange måter å lagre data på, som gjør at det finnes mange formater som dataene kan tilgjengeliggjøres i. I følge Aaby, Brenna, Jøsendal og Nilsen (2011) handler det om "å benytte gjenkjennelige godt støttede formater som ivaretar objektets opprinnelige verdi og mening". Hva slags formater som er passende for hvilke datasett avhenger av dataenes bruksformål. Det viktigste ved å tilgjengeliggjøre åpne data er å skille rådata fra presentasjonsinformasjon fordi det er rådata som gir verdiskaping i denne sammenhengen (Aaby, et al., 2011). Det er viktig at dataene tilgjengeliggjøres i formater som ikke er avhengige av et bestemt produkt eller leverandør for å kunne gjenbrukes. Formatet må maksimere potensielt gjenbruk, samt være gratis tilgjengelig til samfunnet uten lisenser, patenter eller andre begrensninger (Tauberer, 2011). Gjenbruksmulighetene blir begrenset når dataene mangler strukturerte beskrivelser som tillater brukeren å forstå dataene på samme måte som dataeier. Uavhengig av om dataene skal gjøres åpent tilgjengelig på nettet, eller om de skal deles mellom etater, er det derfor behov for at dataene er spesifisert utvetydig. Flere (Alani, et al., 2007; Berners-Lee & Shadbolt, 2010; Myrseth, 2009) argumenterer for at det offentlige bør tilgjengeliggjøre dataene sine ved bruk av semantiske vevteknologier. Den semantiske veven er et av flere initiativer som muliggjør strukturerte metadatabeskrivelser, og foretrekkes på grunn av muligheten til å spesifisere data i et åpent og gjenbrukbart format. Litteraturen (Alani, et al., 2007; Bizer, Heath, & Berners-Lee, 2009; Steve Harris, Gibbons, Davies, Tsui, & Crichton, 2008; Klischewski & Jeenicke, 2004) hevder at semantiske

vevteknologier og lenkede data kan bidra til mer effektiv håndtering og deling av informasjon. Dersom dataene beskrives som Linked Data vil de kunne kobles mot eksisterende vokabular som beskriver meningen i dataene. Semantiske beskrivelser vil heve gjenbruksmulighetene ved at dataene kan både forstås av andre menneskelige brukere og sluttapplikasjoner. Semantisk vevteknologi har mange bruksområder på åpne data når de først er gjort tilgjengelige. Hovedsakelig handler det om å bruke datamodellen RDF til å strukturere dataene og RDFS eller OWL til å beskrive sammenhengen mellom dataene. Når dataene gjøres åpent tilgjengelig i disse formatene, tilgjengeliggjøres de ved bruk av semantiske vevteknologier.

Det ligger utrolig mye spennende muligheter i at offentlige data blir åpent tilgjengelig. Om alle dataene som ulike aktører hadde, var åpent tilgjengelige i et økosystem basert på semantiske vevteknologier, ville det gitt et helt annet grunnlag for dataintegrasjon og korrekt gjenbruk. Hvis dataforvaltningen går i denne retningen, vil det kanskje være mulig etter hvert at offentlig samhandling kan dra nytte av disse åpne dataene. Første steg er å sette i gang en tilgjengeliggjøring av åpne data, og skape en forståelse av hvorfor bruk av semantisk vevteknologier og lenkede data vil heve gjenbruksmulighetene ved dataene.

Hvordan jeg har gått frem

Jeg har i denne oppgaven gjennomført et kvalitativt studium, basert på Eglene (2000) sin beskrivelse av nåværende- og beste praksis forskning. Metoden skal hjelpe organisasjoner til å lære fra andres erfaringer ved å kartlegge hva som fungerer og hva som er vanskelig ved problemområdet, for å kunne gjenta suksesser og unngå fallgruver. Jeg mener denne framgangsmåten er svært egnet til å besvare mine forskningsspørsmål fordi den gir mulighet til å studere dagens situasjon, samt innhente beskrivelser av hvordan og hvorfor ting gjøres. Grunnen til at jeg har valgt å studere problemområdet kvalitativt, er fordi jeg er interessert i beskrivelser og retningslinjer fremfor kvantitative resultater med statistiske resultater. Jeg har lest bakgrunnsinformasjon om åpne data og den semantiske veven som har ledet meg til kunnskapsrike mennesker jeg har fått muligheten til å intervju. Intervjuene har avdekket muligheter, fordeler, utfordringer og praksiser. Jeg vil beskrive en praksis som offentlige etater bør ta i betraktning når de vurderer å tilgjengeliggjøre åpne data ved bruk av semantiske vevteknologier. Oppgaven vil skape en forståelse av muligheter og utfordringer rundt tilgjengeliggjøring av åpne data og hvorfor den semantiske veven og lenkede data bør være tett knyttet til en slik tilgjengeliggjøring. Oppgaven åpner med å beskrive ulike datadefinisjoner før den presenterer viktigheten av felles forståelse. Deretter utredes det om semantiske vevteknologier. Teoridelen avsluttes med en introduksjon av lenkede data, og kort om innføring av ny IT. Videre følger en beskrivelse av metode og analyse. Oppgavens hoveddel er analyseresultatene som danner grunnlaget for diskusjonskapitlet hvor jeg beskriver beste praksis. Oppgaven avsluttes med konklusjon og videre arbeid.

2. TEORI

2.1 Datadefinisjoner

2.1.1 Informasjon og data

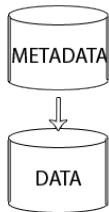
Data refererer til registrerte beskrivelser av ting, handlinger, aktiviteter og transaksjoner - deres størrelse, farge, pris, vekt, dato osv (Boddy, Boonstra, & Kennedy, 2008). Data er symboler og signaler som er bærere av informasjon (Braadland, 2002). Uten videre informasjon om dataene er de kun tall, bokstaver og tegn. Informasjon er data som er gjort meningsfulle og brukbare for mennesker (Braadland, 2002). Jeg kan ikke kinesisk, så en tekst bestående av kinesiske tegn vil for meg kun være data. Hadde jeg kunnskapen til å forstå kinesisk, ville jeg sett dataene som informasjon. Informasjon kan også sees som et subsett av data som betyr noe for personen som mottar dataene (Boddy, et al., 2008). Måten dataene har blitt prosessert på, avgjør dataenes verdi for mottakeren. I følge Engels og Myrseth (2011) er dataene maskinlesbare dersom de er:

”Representert på et åpent og veldefinert format slik at et mangfold av IT-systemer kan lese, manipulere og gjennomføre tapsfrie transformasjoner mellom ulike datarepresentasjoner. Dette uten at en benytter manuell datafangst / re-punching eller annen tapsbasert datafangst som skanning med OCR el”.

2.1.2 Metadata

Hva er metadata

Dataene trenger en kontekst for å bli forstått som meningsfulle data. En slik kontekst kan være informasjon om hvilket språk dataene er formalisert i, opprettsdato, hvem som har opprettet dataene, en tekstlig beskrivelse av dataene og så videre. Det er slik beskrivende informasjon som gir oss grunnlaget til å forstå dataene og muligheten til å bruke de. Metadata defineres som: *“Data attributes that describe, provide context, indicate the quality, or document other object (or data) characteristics”* (Greenberg, 2005). At metadata er dataattributter som beskriver, tilbyr kontekst, indikerer kvaliteten, eller dokumenter andre objekters- eller datas karakteristika, er en svært vid definisjon. En kortere definisjon blir presentert av McComb (2004) som forklarer at metadata bærer meningen til dataene de beskriver. Hva som menes med dataenes mening vil bli forklart under seksjonen 2.2 Semantikk. Wittenburg, Broeder, og Buitelaar (2004) forteller om to typer av metadata; *data om data* som inkluderer tekst om beskriver bilder, lyder, videoer og andre tekster, og *beskrivende nøkkelord* som former kataloger av store datasamlinger. Et eksempel på sistnevnte er Dublin Core. Lagoze, Lynch og Daniel Jr (1996) spesifiserer metadata til syv ulike typer metadata. *Termer og kondisjoner* beskriver ”reglene” for å bruke et objekt. Slike data inkluderer for eksempel en tilgangsliste med hvem som kan se objektet, kondisjoner for å bruke, gjenbrukspriser etc. Eksempler på *administrative data* er oppdateringsdato, opprettsdato, identiteten til administrator for å nevne noen. *Innholdsvurderinger* beskriver attributtene ved data, for eksempel hvor godt egnet dataene er til visse formål. *Opphavsdata* definerer kilden eller opphavet til et objekt, for eksempel når dataene ble digitalisert. *Lenking eller relasjonsdata* er relasjonen til andre data. Slike metadata til en journalartikkel kan ha referanser til hvor artikkelen kan finnes. Relasjoner til andre data bør gjøres ved bruk av en unik og varig identifikator som en URI. *Strukturelle data* er data som definerer de logiske komponentene ved komplekse eller sammensatte objekter, og hvordan man skal få tilgang til disse. Eksempel på slike er datatyper i et programmeringsmiljø (Lagoze, Lynch, & Daniel Jr, 1996). Det er viktig å vite at det finnes mange typer av metadata. Selv om det snakkes om metadata, kan det være ganske forskjellige data det snakkes om. Derfor har jeg valgt å beskrive en del ulike definisjoner av metadata. Det er viktig å forstå metadata for å kunne diskutere hvordan semantiske vevteknologier kan brukes til å beskrive data. Metadata er en sentral informasjonskilde for å kunne realisere den semantiske veven (Wittenburg, Broeder, & Buitelaar, 2004).



Figur 1: Metadata på toppen av dataene gir mening til dataene

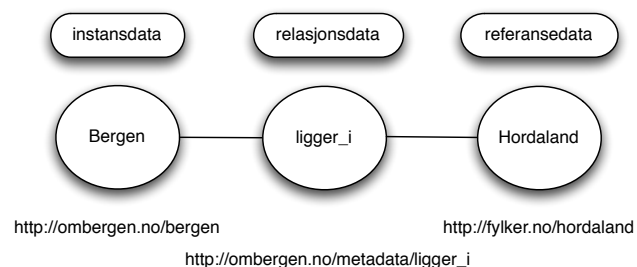
Bruk av metadata

Når internett har gjort informasjon så mye enklere tilgjengelig, vil de nødvendige verktøyene og ferdighetene skifte fra å øke tilgangen til informasjon, til kunnskap om hvordan å bruke den (Landsbergen Jr & Wolken Jr, 2001). Grimstad og Myrseth (2010) ser behovet for metodologier og verktøy for å formalisere metadata til et hjelpemiddel som kan bli håndtert og vedlikeholdt. De peker på flere måter en offentlig organisasjon kan nyttiggjøre seg av formaliserte metadata:

- Internt i organisasjonen vil metadata bidra til en oversikt over den nåværende situasjonen når det kommer til tilgjengelige data og meningen ved dem.
- Ved åpne offentlige data og Linked Open Data kan metadata brukes som en beskrivelse av meningen og de brukte identifikatorene.
- Veldefinerte metadata gjør det enklere å vedlikeholde en portefølje av internt tilknyttede systemer.
- En samling av metadata (metadatarепositorie) bestående av flere departementers metadata vil være til stor hjelp når det utføres prosesser innen harmonisering av metadata.
- Metadata kan brukes når det samarbeides på tvers av etater og tjenester, ved å tilby standardiserte modeller som de ulike aktørene tilpasser sine data mot.
- For å måle datakvalitet er det behov for at dataene måles i henhold til gitte regler. Metadata kan danne hoveddelen av slike regler.

2.1.3 Referansedata

Referansedata er data som blir brukt for å kategorisere andre data i en database, eller data for å relatere data i en database til informasjon utenfor virksomheten (Chrisholm, 2006). Dersom datasett *a* refererer til et annet datasett *c*, vil *c* være referansedata for *a*. En lignende definisjon av referansedata er data som representerer informasjon om klasser eller individer, som er felles for mange fasiliteter eller av interesse for mange brukere (Hoff, 2007). Denne definisjonen sier at referansedata er data som representerer informasjon om andre data. Er da referansedata en type metadata? Begrepet referansedata kan være litt vrient å plassere. I noen tilfeller kan referansedata være metadata som gir data om data, mens andre ganger er det data som kun refereres til og da blir det dataene som binder datasettene som blir metadata. Jeg vil beskrive et fiktivt eksempel for å forklare hva jeg mener med referansedata. I datasettet *OmBergen* finnes det ulik informasjon om Bergen. En av beskrivelsene sier at *Bergen* ligger i *Hordaland* fylke. I datasettet finnes ingen fylkesinformasjon, så for at *OmBergen* skal vise sammenhengen mellom disse dataene, må det refereres til datasettet *Fylker* som inneholder *Hordaland*. *Fylker* er i dette eksempelet referansedata for *OmBergen*. *Ligger_i* er metadata av typen relasjonsdata som bærer data for å lenke sammen de to andre dataene *Bergen* og *Hordaland*.



Figur 1: Eksempel på referansedata

2.1.4 Åpne offentlige data

Åpne data

Åpne data er ikke relatert til noen bestemt teknologi, men en paraplybetegnelse på data som er tilgjengelig slik at andre kan få tilgang til å gjenbruke de. Tilgjengeliggjøring, frigjøring og publisering vil brukes som synonymer for prosessen med å fristille dataene. Jeg presenterer noen ulike definisjoner av åpne data:

"Åpne data er data som er tilgjengelige for allmennheten" (Aaby, et al., 2011).

"Data som en borger, etat eller næringsdrivende kan få tilgang til ved innsyn og evt. en kopi. Formålet med bruken skal ikke være avgjørende for innsyn/tilgang" (Engels & Myrseth, 2011).

Offentlige data

Offentlige data blir både definert som data som det offentlige produserer, og data som det offentlige samler inn. Jeg har gjengitt noen definisjoner av offentlige data:

"Informasjon offentlige etater og kommuner produserer i forbindelse med arbeidsoppgaver utført på vegne av innbyggerne" (Aaby, et al., 2011).

"Informasjon som anvendes til intern saksbehandling i offentlig sektor, og som lagres elektronisk av en offentlig organisasjon eller myndighet" (IT-ogTelestyrelsen, 2010).

"Ikke-personlige data som myndighetene holder, og som blir innsamlet eller generert i forbindelse med offentlig tjeneste levering" (ChiefSecretarytotheTreasury, 2009).

"Data som offentlig sektor, alene eller sammen med andre, har finansiert eller opparbeidet, er ansvarlig for å forvalte, og har rettigheter knyttet til egen bruk og viderebruk" (Engels & Myrseth, 2011).

Felles for alle definisjonene er at offentlige data er knyttet til offentlig saksbehandling og at det er data som offentlig sektor lagrer og vedlikeholder.

Åpne offentlige data

Aaby et al. (2011) omtaler åpne offentlige data som data som er tilgjengelige i sin helhet for både mennesker og maskiner utenfor etaten. Tauberer (2011), forklarer åpne offentlige data som: *"informasjon offentlige myndigheter innehar som deles med samfunnet digitalt, over internett, på en måte som fremmer analyse og gjenbruk"*. Jeg vil bruke Open Government Working Group sine prinsipper for åpne offentlige data for å utdype hva det vil si at dataene er tilgjengelige i sin helhet og derfor fremmer analyse og gjenbruk. Offentlige data skal sees som åpne dersom dataene er gjort tilgjengelig i henhold til disse 8 prinsippene (OpenGovernmentWorkingGroup, 2007) :

1. Dataene må være fullstendige. Samtlige data gjøres tilgjengelig. Tilgjengelige data er ikke begrenset av sikkerhet og personvern.
2. Dataene må være grunndataene. De skal være så granulerte som mulig slik at dataeier ikke gir antagelser for hvordan dataene bør brukes. Dataene bør derfor ikke være sammenstilte.
3. Dataene må være tidsriktige. Dataene skal gjøres tilgjengelig så raskt som nødvendig for å opprettholde dataenes verdi.
4. Dataene må være mulige å aksessere for så mange som mulig til så mange formål som mulig.
5. Dataene må være maskinlesbare slik at datamaskiner automatisk kan prosessere dataene.
6. Dataene må være tilgjengelige for alle uten krav til registrering.
7. Dataene skal gjøres tilgjengelige i ikke-proprietære formater slik at bruksretten ikke begrenses

8. Dataene må være lisensfrie. Gjenbruket skal ikke begrenses av kopibeskyttelse og patenter.

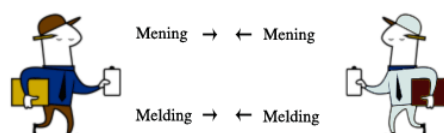
Engels og Myrseth (2011) definerer åpne offentlige data som summen av følgende egenskaper: offentlige data, beskrevne data, åpne data og maskinlesbare data. Jeg tolker definisjonene, og definerer åpne offentlige data som:

”Offentlige data som er åpent tilgjengelig i et maskinlesbart format med forklarende metadata”.

2.2 Semantikk

2.2.1 Viktigheten av at vi forstår dataene

Semantikk kan defineres som studiet av mening (McComb, 2004). Det sentrale er ikke kun hva dataene i systemene faktisk betyr i den virkelige verdenen, men hvorvidt menneskene, og andre programmer som bruker denne informasjonen forstår og er enige om hva slags mening systemet gir dataene. McComb (2004) beskriver fire typer av semantisk flyt: menneske til menneske (H2H), menneske til applikasjon (H2A), applikasjon til menneske (A2H) og applikasjon til applikasjon (A2A). Hovedforskjellen er i hvilken grad semantikken vedlikeholdes eller av hvem semantikken blir tolket. Semantisk opprinnelse referer til det tilfellet hvor enten mennesket eller en applikasjon er klar over semantikken i meldingen det/den produserer. Semantisk tolkning vil si at mottakeren av meldingen, enten menneske eller applikasjon, bestemmer seg for hva som menes i meldingen som sendes. I situasjoner der meldingen leveres uten kontekst eller en viss form av merking av meldingen, må mottaker tolke hva meldingen betyr. Tenk at du får meldingen ”10 til 15”. Betyr det mellom tidspunktet 10-15, 10-15 enheter av en bestilling, eller kanskje alderen 10-15? eller mener senderen en av de utallige andre forklaringene? Mottar vi en melding med et ord eller begrep vi har sett før, tror vi at vi vet hva begrepet betyr, til tross for at senderen mener noe helt annet. McComb (2004) peker på fire måter det har oppstått semantiske forskjeller i måten organisasjoner bruker begreper; *skapte ord*, *identifikatorer som ikke er ord*, *doble ord*, *alt for mange felles ord*. Homonymer eller like ord, er området som har mest potensial for misforståelse på grunn av måten forskjellige industrier, eller forskjellige deler av den samme industrien, bruker de samme begrepene ulikt. Når slike begreper ikke defineres eksplisitt fordi ”alle vet hva ordet betyr”, oppstår det kaos.



Figur 2: Semantikk ved deling av informasjon

2.2.2 Semantisk interoperabilitet

Problemstillingen omkring semantisk interoperabilitet oppstår i situasjoner der det er behov for å dele eller utveksle data, og hvor det ikke er krystallklart for de tiltenkte brukerne hvordan de delte eller utvekslede dataene skal tolkes (Grimstad, et al., 2008). Ved tilgjengeliggjøring av åpne data, og bruk av disse dataene er derfor semantisk interoperabilitet sentralt. Blant flere definisjoner av semantisk interoperabilitet, er noen gjengitt under:

”Organisasjonen, og dens informasjonssystemers evne til å oppdage nødvendig informasjon, eksplisitt beskrive meningen av de dataene den ønsker å dele med andre organisasjoner, samt prosessere mottatt informasjon på en måte som er konsistent med den originale hensikten av slik informasjon” (Ojo, Janowski, & Estevez, 2009).

“Interoperabilitet på kunnskapsnivå, som tilbyr samhandlende organisasjoner muligheten å brolegge semantiske konflikter, som oppstår fra forskjeller i implisitte betydninger, perspektiver, og påstander, for å skape et semantisk kompatibelt informasjonsmiljø basert på de gjensidig forståtte konseptene mellom ulike organisasjoner” (Park & Ram, 2004).

For å gjenbruke tilgjengeliggjorte data er det derfor avgjørende at dataene har tilstrekkelig med semantiske beskrivelser. Det er vanskelig å integrere ressurser som er utviklet basert på ulike begrepsapparater med forskjellige perspektiver på dataene (Heflin & Hendler, 2000). For å holde oversikten over hva som slags mening som legges til bestemte data, skapes begrepsapparater.

2.2.3 Begrepsapparater

Et begrepsapparat er en samling av begreper i en kontekst (Norheim, Udjus, & Smith-Meyer, 2010). Ølnes og Hove (2011) forklarer et begrepsapparat som en samling strukturerte metadata, og legger til at det også kan omtales som en strukturert ordliste eller vokabular. Begrepene i vokabularet har mening tilegnet seg, og det er denne meningen som skaper grunnlaget for felles forståelse eller misforståelse. Det finnes flere måter å strukturere et begrepsapparat på. En taksonomi er et vokabular som er organisert hierarkisk basert på attributtene ved tingene som blir klassifisert (McComb, 2004). Taksonomier brukes ofte til å skape begrepsapparater. De fungerer spesielt godt med biologiske data. Dette kan være fordi man vet at rasen Boarder Collie er en hundetype, så da kan man spesifisere at den arver egenskaper fra hund. Siden hund er et pattedyr, kan vi videre implisere at Boarder Collier også er et pattedyr. Slike ”er-type”-relasjon kalles hyponymer og fungerer som en gode basis for taksonomier. Til tross for slik semantikk spesifisert ved arv, kommer taksonomier til kort i forhold til ontologier når det gjelder mer detaljerte relasjoner. Gruber (1993) definerer en ontologi som en spesifisering av konsepter. Ontologier har et mye rikere sett av relasjoner, restriksjoner, og regler, slik at vi kan resonnerer på informasjonen i ontologien. Å resonnerer vil si at nye fakta kan utledes på bakgrunn av de gitte faktaene. Ontologier tilbyr en mekanisme for å introdusere et vokabular, for så å gi presis mening til termene i vokabularet (Horrocks, 2007). Selv om det ikke er noe formelt krav, er ontologier typisk organisert som et nettverk eller som en graf, fremfor et hierarki.

2.2.4 Semantiske teknologier

Semantiske teknologier kan beskrives som teknologier som tilbyr verktøy og metoder for å bygge mer tilpassningsdyktig og fleksibel programvare ved å utnytte meningen i den tilgjengelige informasjonen (Grimstad, et al., 2008). Kunnskapsrepresentasjon, ontologier, resonnering og regler, spørringer, og agenter og tjenester blir omtalt som kjernen av semantiske teknologier. Om dagen er arbeidet tilknyttet den semantiske veven en av de sterkeste driverne for semantiske teknologier.

2.2.5 Den semantiske veven

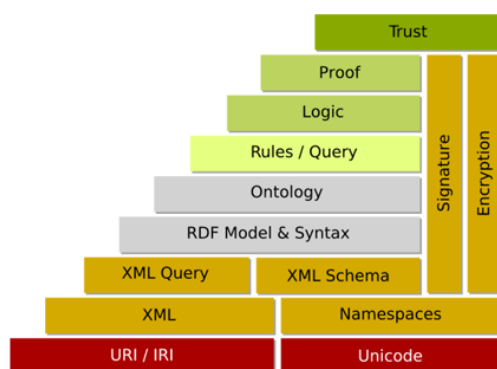
Tim Berners-Lee, James Hendler og Ora Lassila skrev den klassiske artikkelen ”The Semantic Web” i 2001. Her ble fremtidsutsiktene for den semantiske veven presentert, sammen med tanker og beskrivelser av hva som er de store forskjellene mellom den tradisjonelle verdensveven (WWW), og den semantiske veven. Verdensveven har gitt oss et globalt informasjonsområde for å dele og aksessere dokumenter. Den har utvilsomt vært en revolusjon når man tenker på hvor mye enklere det har blitt å gjøre dokumenter tilgjengelige for mennesker som ikke befinner seg på samme lokasjon. Måten brukerne har funnet informasjon på har vært å bruke nettlesere til å traversere gjennom hypertext-lenker, eller ved å la søkemotorer lete etter dokumenter som inneholder de nøkkelordene brukeren har oppgitt. Det er dokumentene som har blitt strukturert og lenket sammen, og brukerne finner nye dokumenter ved å følge lenker. Den semantiske veven skal utvide verdensveven med meningsfylt informasjon slik at maskiner og mennesker kan bedre samhandlingen (Berners-Lee, Hendler, & Lassila, 2001). Fordi den semantiske veven vil være en utvidelse av verdensveven og ikke en erstatning, vil også den semantiske veven være universal og desentralisert.

LinkedDataTools.com (2010) er et nettsted bestående av verktøy, informasjon og ressurser for den semantiske veven. Tiltaket har blitt satt i gang av flere profesjonelle utviklere som ønsker å hjelpe nettsamfunnet inn i den nye generasjonen av internett. Den semantiske veven defineres som *”en måte å lenke data mellom systemer eller entiteter som støtter en rik, selvbeskrivende interrelasjon mellom data tilgjengelig på tvers av den globale veven”* (LinkedDataTools.com, 2010). Matthews (2005) forklarer den semantiske veven ut i fra *visjon, programmet, og teknologiene*. Visjonen bak den semantiske veven dreier seg om tankene og forhåpningene Berners-Lee et al. (2001) har til denne omstruktureringen av internett. I visjonen har veven mening som oppnås ved å beskrive internettressursene på en meningsfylt måte. Vevven blir semantisk når vi fanger meningen bak koblingene. Med en gang veven har en mekanisme for å definere semantikken til ressursene og koblingene, vil det kunne være mulig å bruke automatisk maskinprosessering fremfor menneskelig mediering (Matthews, 2005). Den semantiske veven vil tillate maskiner å forstå dokumenter og data, men ikke menneskelig tekst og tale (Berners-Lee, et al., 2001). Et mye brukt eksempel til å forstå visjonen er søkemotorer. Tradisjonelle søkemotorer returnerer dokumenter som inneholder søkeordene. Dersom jeg søker etter artikler skrevet av en bestemt forfatter vil treffet som returneres inkludere alle dokumenter som inneholder forfatterens navn. Dette kan være svært tidskrevende hvis jeg søker etter en bestemt artikkel av for eksempel Tim Berners-Lee. Med semantiske beskrivelser av dokumentene, kan den semantiske veven resonnerer på koblinger, og returnere dokumenter basert på beskrivelser som forfatter, årstall, relevante arbeid, arbeidsgiver med mer.

Det semantiske vev-programmet startet som ”Web Metadata Working Group” i 1998, for så å bli til ”Semantic Web Activity” (Matthews, 2005). Det semantiske vev-initiativet er en del av den internasjonale organisasjonen World Wide Web Consortium (W3C) som setter standarder for teknologiene som underligger internett.

2.3 Semantiske vevteknologier

Teknologiene og standardene som danner byggeklossene for den semantiske veven er den tredje måten Matthews (2005) har valgt å forklare den semantiske veven på. Det semantiske vev-programmet har utviklet en lagvis arkitektur som ofte representeres ved følgende diagram. Teknologiene i stabelen er komponentene den semantiske veven består av.



Figur 3: Teknologistabelen for den semantiske veven (SemanticFocus, 2010)

URI/IRI

Internationalized Resource Identifier (IRI) er en utvidelse av Uniform Resource Identifier (URI), men de to blir ofte brukt om hverandre (Hebeler, Fisher, Blace, & Perez-Lopez, 2009). En Uniform Resource Locator (URL) lenker dokumenter, mens en URI lenker data. URI er en sentral del av infrastrukturen til verdensveven, og en standardisert måte å navngi ressurser på. Når hver eneste ressurs på nettet får sin URI, vil vi kunne lage en vev der alle data er unikt beskrevet, slik at de kan lenkes mot hverandre. Sammenlignet med relasjonelle databaser vil en bestemt rad i en tabell bli identifisert med en primærnøkkel som er unik for den tabellen innenfor én database. I kontrast, vil en URI ha et navn som er universelt unikt over hele

internett, noe som gjør URler valid uansett kontekst. En URI er den unike identifikatoren til hver enkelt ressurs på den semantiske veven.

XML: Namespace, XQuery og XSLT

XML (Extended Markup-Language) er etter hvert et modent og velbrukt språk for å utveksle informasjon på nettet. Heflin (2000) mener dette er fordi språket skiller innhold og presentasjon, som gjør det enklere å uthente og/eller gjenbruke dataene. I systemer bygget rundt semantiske vevteknologier, brukes det ofte termer fra flere vokabularer som RDF og OWL. Det er lett for at disse termene overlapper og definisjonen av et begrep i et RDF blandes med et begrep fra et OWL. XML Namespaces, heretter omtalt som navneområder, tildeler utvidede navn til elementene og egenskapene i et datasett slik at det ikke skal være tvil om hvilket vokabular elementet tilhører og da hvilken definisjon som skal brukes (W3C, 2009a). Et XML-navneområde blir identifisert ved bruk av en URI. <http://www.w3.org/2002/07/owl#> er navneområdet som indikerer at ressursen skal defineres i henhold til ontologispråket OWL. Når da ressursen blir beskrevet som *owl:ObjectProperty* vet maskinen at den skal bruke OWL-definisjonen av ObjectProperty.

XMLQuery eller XQuery er en måte å uthente data fra forskjellige typer XML-ressurser, både databaser og dokumenter. XQuery opererer på den abstrakte logiske strukturen ved XML, fremfor dokumentets overflatesyntaks (W3C, 2007). XSLT(Extensible Stylesheet Language Transformations) kan brukes på XML-dokumentene for å oversette XML til RDF.

RDF

Resource Description Framework (RDF) danner datasettet i den semantiske veven. RDF er ikke et språk, men en datamodell bestående av data (Heflin & Hendler, 2000). Der innholdet i et XML-dokument er strukturert i en hierarkisk trestruktur, er RDF-modellen en graf. Grafteori er en av de underliggende disiplinene bak den semantiske veven (Hebeler, et al., 2009). RDF-grafer inneholder setninger eller bestående av subjekt, predikat og objekt. På grunn av denne tredelingen, kalles setningene for tripler. Predikatet danner relasjonen mellom subjektet og objektet. Det finnes to typer noder: ressurser og literaler. Literaler representerer konkrete dataverdier som nummer eller strenger, og kan ikke være subjektet i tripler, kun objekter (Hebeler, et al., 2009). Ressurser representerer alt som kan få et navn, og kan enten være subjekt eller objekt. En ressurs er ikke noe annet enn et navn som representerer et objekt, handling eller konsept (Hebeler, et al., 2009). Ettersom strukturen på RDF-dokumentet er en graf, utgjør triplene to noder og predikatet binder nodene sammen. Når et RDF-dokument lenkes til et annet RDF-dokument, er det en node i hvert dokument som lenker datasettene. Et subjekt kan bli et objekt i en ny trippel og motsatt. Et predikat i en trippel, kan bli et predikat i en annen trippel. Figurene 3 og 4 demonstrerer konseptet bak en RDF-lenke med en forfatter, et publiseringsfirma og en bok.



Figur 4: Enkel RDF-eksempel 1



Figur 5: Enkelt RDF-eksempel 2

Disse eksemplene viser hvordan en subjekt-node i et RDF-dokument, blir en objekt-node i et annet RDF-dokument. Triplene i et RDF-dokument vil ha en URI for hver node, men dette er et forenklet eksempel. Eksempelet under viser en trippel der subjektet er en ressurs, mens objektet er et literal.



Figur 6: Enkelt RDF-eksempel 3

Grafstrukturen til RDF og det globale navneområdet, støttet gjennom bruk av URIs, gjør RDF til en svært egnet datamodell for deling og utveksling av data (Hebeler, et al., 2009). Ettersom grafer ikke har røtter, sammenlignet med en trestruktur, har ingen ressurser mer opprinnelig betydning enn andre. Å utveksle informasjon blir da lettere, fordi å kombinere grafene blir konseptuelt det samme som å plassere den ene grafen ved siden av den andre. Ettersom tripler kun er en samling av URIs og literaler, og hver URI og literal har et globalt omfang, fungerer tripler som et kraftig verktøy for informasjonsintegrasjon (Hebeler, et al., 2009). Bruken av globale navn er vitalt, fordi det medfører at triplene alltid kan flettes sammen uten å måtte oversette navnene. Videre siden hver setning i grafen kan flettes som den er, fører det til at hele grafer med informasjon kan integreres med en annen graf uten å måtte oversette noen navn. Siden RDF-setninger ikke trenger oversettelse når man flytter fra et system til et annet, er de valide i hvilken som helst kontekst. RDF-grafers abstrakthet gjør de verdifulle for menneskelig analyse. Det gjør det lettere å forstå hvordan de ulike ressursene er lenket og hva de forskjellige ressursene er for noe. Svakheten med en slik abstrakt modell, er at en datamaskin ikke vil forstå den. For at applikasjoner skal kunne bruke RDF til å utveksle informasjon, må RDF-grafen omgjøres til et konkret format. Denne prosessen kalles serialisering og gjør RDF praktisk for informasjonsutveksling ved at serialiseringen tilbyr måter for å konvertere mellom den abstrakte modellen og et konkret format (Hebeler, et al., 2009).

Flere ulike serialiseringsformater eller notasjoner finnes for RDF. Tre av de mest populære er RDF/XML, Turtle (the Terse RDF Triple Language) og RDFa. RDF tilbyr en måte å modellere informasjonen på, men hva selve informasjonen betyr, sier RDF-grafen ingenting om. For at RDF-grafen skal bli en semantisk informasjonsmodell, er det nødvendig med en metode for å spesifisere hva ressursene er, hvorfor predikatet binder akkurat disse to ressursene sammen og så videre. Måten den semantiske veven gjør dette på, er at den bruker en kombinasjon av et skjemaspråk og et ontologispråk som til sammen bruker mulighetene som finnes i vokabularer, taksonomier og ontologier.

RDFS

RDF-Schema tilbyr et spesifikt begrepsapparat for RDF som kan bli brukt for å definere taksonomier av klasser og egenskaper og enkle domener, samt predikatenes rekkevidde (Hebeler, et al., 2009). Det som forstås med at RDFS bruker navneområder, er at RDFS ikke prøver å definere disse vokabularene, men isteden tilbyr et språk så utviklere kan definere sitt eget delte vokabular. RDFS-vokabularer beskriver ressursene og predikatene som blir brukt i RDF-modellen. Spesifikasjonen definerer et antall egenskaper som har bestemt semantikk (Heflin & Hendler, 2000). Som at *rdf:type* blir brukt til å beskrive en instans sitt medlemskap av en gutt klasse, brukes *rdfs:subClassOf* for å poengtere at en klasse er en underklasse av en annen klasse. Hebeler et al. (2009) forklarer at skjemaspråket gjør at en kan arrangere klasser i arvehierarkier, definere domener og rekkevidde for predikater, påstå klassemedlemskap, samt spesifisere og tolke datatyper. RDFS er en av de fundamentale byggeklossene ved ontologier i den semantiske veven, og er det første steget i retning av å spesifisere semantikk i RDF (Hebeler, et al., 2009).

OWL

OWL (Web Ontology Language) tilbyr et uttrykksfullt språk for å definere ontologier som fanger semantikken i domenekunnskapen (Hebeler, et al., 2009). OWL utvider RDFS-vokabularet med tilleggsressurser som kan brukes for å bygge mer ekspressive ontologier for veven. OWL bruker RDF og RDFS, XML Schema-datatyper og OWL navneområder (Hebeler, et al., 2009). Nøkkelideen bak den semantiske veven er å uttrykke meningen av vevinnholdet, og det er annotasjonsverktøy som RDFS og OWL som gjør denne jobben (Horrocks, 2007). Dette skjer ved at OWL legger til flere restriksjoner tilknyttet strukturen og

innholdet i RDF-dokumenter, som gjør at maskiner kan gjøre mer prosessering og resonering (Hebeler, et al., 2009). Mens RDF er en enkel graf, og RDFS er et hierarkisk vokabular bestående av en liste med definerte termer, blir OWL ontologier brukt til å modellere domenekunnskap.

OWL er basert på beskrivelseslogikk eller description logics (DL). Beskrivelseslogikk (DL) er en familie innenfor språkene av kunnskapsrepresentering, som kan bli brukt for å representere kunnskapen i et applikasjonsdomene på en strukturert og formalisert måte (Baader, Horrocks, & Sattler, 2005). Et slikt domene blir beskrevet som individer, konsepter (klasser), og roller (relasjoner eller predikater i ontologispråk) (Horrocks, 2007). DL tillater at konseptbeskrivelser kommer fra atomiske konsepter og roller. En gitt DL kjennetegnes ved et sett av tilgjengelige operatører for å bygge konseptbeskrivelser. Ofte inkluderer disse minst snitt(\sqcap), union(\sqcup) og negasjon (\neg), samt begrensede former eksistensielle (\exists) og universelle (\forall) kvantifikatorer (Horrocks, 2007). Videre er det mulig å spesifisere at en klasse er en underklasse av en annen (Mann er underklasse av Menneske), og at en klasse er ekvivalent med en annen for eksempel at konseptet Menneske er det samme som Person. Slike roller finnes i OWL som *typeOf* og *sameAs*. Ettersom DL er logikk, har de en formalisert semantikk (Horrocks, 2007). Som tidligere nevnt må maskinene ha tilgang til strukturert informasjon, samt kunne resonere på denne informasjonen ut i fra et sett med inferensregler for at den semantiske veven skal fungere som tiltenkt med automatisk prosessering. I tillegg til å spesifisere utvetydig mening ved konseptene i domenet, kan resoneringsalgoritmer utvikles mot beskrivelseslogikk (Horrocks, 2007). Det er en formaliserte logikken som tillater resonering.

SWRL (Semantic web rule language) er utviklet som et tilleggslag over ontologien på teknologistabelen. I de tilfellene der OWL ikke har nok uttrykkskraft, kan SWRL brukes til å legge til mer detaljert semantikk. En *rule* eller regel i den semantiske veven er ofte kondisjoner som *if* \rightarrow *then*, og dersom et bestemt sett av påstander i datasettet viser seg å være sanne, blir ny kunnskap generert gjennom logisk resonering (Hebeler, et al., 2009).

SKOS - Simple Knowledge Organization System

SKOS er en datamodell for å dele og lenke ulike kunnskapsorganisasjonssystemer (W3C, 2009b). Tesauruser, taksonomier og ulike klassifikasjonsskjemaer deler en lignende struktur og blir brukt i lignende applikasjoner. SKOS fanger mange av disse likhetene, og gjør dem eksplisitt, for å tillate data og teknologiutveksling på tvers av ulike applikasjoner. SKOS tilbyr en standard, lavkostnadsmetode for å overføre eksisterende kunnskapsorganisasjonssystemer til den semantiske veven. I tillegg fungerer SKOS som et lettvektsspråk for å utvikle og dele nye kunnskapsorganisasjonssystemer. Den kan brukes alene, eller i kombinasjon med andre slike systemer for eksempel OWL (W3C, 2009b).

SPARQL og SQWRL

SPARQL Protocol and RDF Query Language er det standardiserte spørrespråket mot RDF anbefalt av W3C. Det er et spørrespråk fordi det definerer syntaksen en former spørringer mot RDF ut i fra, og en protokoll som beskriver hvordan en SPARQL klient kommuniserer med et SPARQL endpoint (Hebeler, et al., 2009). Et endpoint, heretter omtalt som tilgangspunkt, er en tjeneste som aksepterer og prosesser SPARQL-spørringer, og returnerer resultater i forskjellige formater avhengig av spørreformen. Ettersom RDF er en graf, og SPARQL er utviklet for å hente ut data fra RDF-modeller, fungerer SPARQL slik at man spør mot noder i grafen, og får returnert tripler. SPARQL brukes som spørrespråk mot OWL, fordi OWL kan serialiseres som RDF.

Tiltross for at SPARQL tilsynelatende fungerer som et spørrespråk mot OWL, har språket ingen forståelse av OWL og dets språkkonstruksjoner. Sånn som OWL er bygget på beskrivelseslogikk, er det behov for et spørrespråk som forstår slik logikk og semantikk som OWL gir datasettet. O'Connor og Das (2009) har foreslått SQWRL (Semantic Query-enhanced Web Rule Language), bygget på SWRL, som fungerer som et enkelt, men uttrykksfullt spørrespråk rettet mot OWL-ontologier. SQWRL tar en SWRL-setning og behandler den som en spørring mot OWL ved at regelen blir omgjort til en uthenting (O'Connor & Das, 2009). SQWRL bruker det innebygde miljøet i SWRL, og definerer så operatører som kan brukes til å uthente informasjon fra ontologien.

Proof and Trust

De øverste lagene i teknologistabelen dreier seg om bevis og tillit. Bevis er knyttet til den automatiske prosesseringen, og hvordan en skal kunne bevise at det svaret som blir funnet på den semantiske veven er korrekt (Henze, 2008). Slike svar blir utledet gjennom det logiske laget. Tillit adresseres i tilfeller der det skal sendes data mellom to aktører med forventinger til interoperabilitet og skalerbarhet. Følgende utfordringer er knyttet til interoperabilitet og skalerbarhet ved kommunikasjon mellom autonome aktører presenterer: sikkerhet, avhengighet, personvern, normer, metadata (i dette tilfellet informasjon om dataenes opprinnelse), autorisering, autentifisering, kryptering og kontrollmekanismer omkring hvilke data som sendes og hvem som får tilgang til dataene (Henze, 2008).

2.4 Lenkede data

Linked Data er W3C sitt navn på lenkede data og defineres som data i et maskinlesbart format med eksplisitt definert semantikk som lenker til andre datasett (Bizer, et al., 2009). Jeg bruker lenkede data synonymt med Linked Data. W3C sine ideer for Linked Data er å bruke veven til å lenke slike data fra ulike kilder. Resultatet kalles for dataveven, og forklares som en vev bestående av tingene i verden, beskrevet av dataene på veven. Hvis data tilgjengeliggjøres på denne måten, vil alle applikasjoner som er bygget rundt semantiske vevteknologier kunne gjenbruke disse datakildene uten å måtte transformere dataene. I den tradisjonelle utviklingen av informasjonssystemer har dataene vært lagret i en bestemt database, og blitt beskrevet spesielt for systemet som skal bruke dem. Linked Data er annerledes ved at brukerne av data, enten en person eller en applikasjon, lettere kan oppdage og integrere data fra et stort antall ulike kilder (Heath & Bizer, 2011). Forfatterne utdyper hvordan Linked Data skiller seg fra tradisjonelle tilgjengeligjøringsformer på flere områder:

En forent datamodell

Linked Data bruker RDF som én forent datamodell, i kontrast til andre metoder som er avhengig av mange ulike heterogene datamodeller.

En standardisert mekanisme for datatilgang

Linked Data bruker HTTP-protokollen som gjør at datakilder kan aksesseres ved å bruke generiske nettlesere. I tillegg kan summen av alle lenkede datasett indekseres og søkes gjennom av søkemotorer. For å få kunne bruke vevAPIer er man avhengig av tilgang til ulike proprietære grensesnitt.

Hyperlenkebasert dataoppdagelse

Ved å bruke URIer tillater Linked Data at det skapes hyperlenker mellom entiteter fra ulike datakilder. Disse datalenkene binder alle lenkede data til ét globalt dataområde, som tillater Linked Data-applikasjoner å oppdage nye datakilder i reell tid. I kontrast vil vevAPIer og datadumper i proprietære formater forbli isolerte datalagre med begrenset eller ingen tilgang til andre datasett.

Selvbeskrevne data

Lenkede data letter dataintegrasjonen fra ulike kilder ved å bruke delte vokabularer, gjøre definisjonene av disse vokabularene mulige å finne, og ved å tillate begreper fra ulike vokabularer å bli lenket til hverandre med vokabularlenker.

Disse egenskapene ved Linked Data-arkitekturen gjør det lettere for brukere av data å oppdage, aksessere, og integrere data (Heath & Bizer, 2011). Berners Lee (2009a) har presentert fire prinsipper som har blitt kjent som beste praksis for å tilgjengeliggjøre Linked Data:

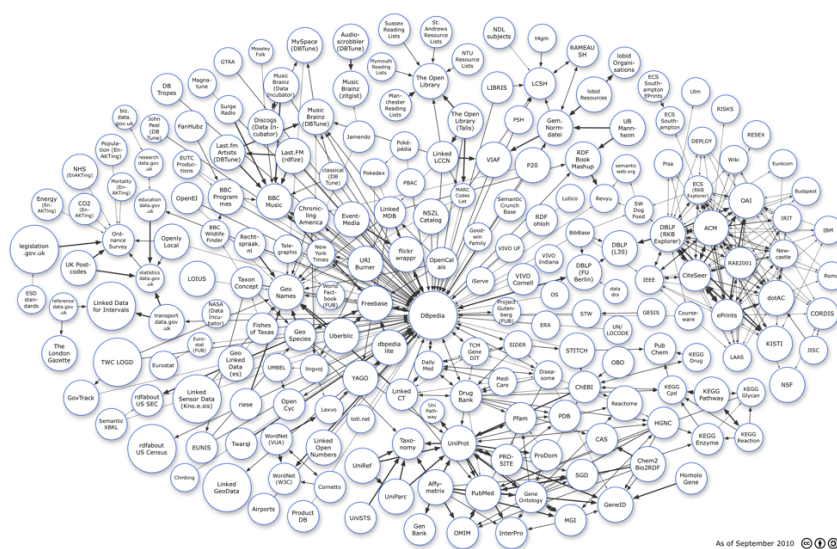
1. Bruk URIer som navn på ting
2. Bruk HTTP URIer slik at mennesker kan undersøke navnene
3. Når noen undersøker en URI, tilby nyttig informasjon ved å bruke standarder som RDF og SPARQL
4. Inkluder lenker til andre URIer, sånn at folk kan utforske flere ting

2.4.1 Lenkede åpne data

Linking Open Data-prosjektet startet opp i januar 2007 med støtte fra W3C sin ”Semantic Web Education and Outreach Group” (Bizer, et al., 2009). Hensikten og målet ved prosjektet er å skape dataveven ved å identifisere eksisterende datasett under åpen lisens, konvertere disse til RDF i tråd med Linked Data-prinsippene, for så å publisere dem på nettet (Bizer, et al., 2009). Linked Open Data defineres som:

”Data that is published under an open licence that allows unrestricted reuse, and that is marked up to identify the structure and meaning, making possible its automated collection for re-publishing and mashing up with other data” (Sargent & Taggart, 2010).

Linked Open Data er altså lenkede data som er gjort tilgjengelige med en åpen lisens for å tillate ubegrenset gjenbruk. Figuren under viser datasettene som har blitt tilgjengeliggjort og sammenlenket av Linking Open Data-prosjektet per september 2010 (Cyganiak & Jentzsch, 2011). Datasettene inneholder over 25 milliarder RDF-tripler, som er sammenlenket med ca 395 millioner RDF-lenker (W3C, 2011). Hver node i grafen er et datasett, og tykkelsen på pilen mellom to datasett indikerer antall lenker. De tykkeste pilene viser datasett med flest lenker.



Figur 7: Diagram av Linking Open Data-skyen anno september 2010

2.4.2 Fem stjerner ved åpne data

Berners Lee (2009a) kom med forslaget om å rangere tilgjengeliggjorte datasett fra 1 til 5 stjerner ut i fra deres åpenhet. Antall stjerner tilsvarer datasettets gjenbruksmuligheter. Linked Open Data belønnes med 5 stjerner.

- ★ Gjør dataene tilgjengelig i hvilket som helst format på nettet under en åpen lisens
- ★★ Tilgjengeliggjør strukturerte maskinlesbare data (Excel fremfor PDF)
- ★★★ Bruk ikke-proprietære formater (CSV fremfor Excel)
- ★★★★ Bruk URIs for å identifisere ting slik at andre kan peke til dataene
- ★★★★★ Lenk dataene til andre datasett (RDF)

Figur 8: 5 stjerner ved Linked Open Data

2.5 Innføring av ny IT i offentlige organisasjoner

IT blir sett på som en innovasjon når potensielle brukere ser på teknologien som relativt ny (Rogers, 1995 sitert i Kamal, 2006). I følge Pierce og Delbecq (1977, sitert i Kamal, 2006) og Cooper og Zmud (1990, sitert i Kamal, 2006) består den organisatoriske innføringsprosessen

av de tre stegene *initiering, adopsjon og implementering*. De samme forfatterne definerer initiering som presset for å endre, innsamle, og evaluere informasjonen knyttet til innføringen. Kamal (2006) beskriver denne fasen som en motivasjonstilstand der en organisasjon blir klar over en spesifikk teknologi og forsøker å innskaffe kunnskap om denne som resulterer i innføring av den nye teknologien. Adopsjon involverer beslutningen om å prioritere ressurser til å skulle innføre den ny teknologien, og implementeringen refererer til utvikling av og installeringsaktivitetene nødvendig for å sørge for at de forventede fordelene ved innovasjonen blir realisert (Kamal, 2006).

3. METODE

Vitenskapelige metoder utgjør et sett av retningslinjer som skal hjelpe forskeren med å sikre at den vitenskapelige virksomheten er faglig forsvarlig (Grønmo, 2004). Hvis ikke studier er basert på en metode, vil det være vanskelig å avgjøre kvaliteten på studiet og hvorvidt funnene er troverdige. Metoden forklarer hvordan forskeren har gått fram for å belyse problemstillingen og besvare forskningsspørsmålene. Grønmo (2004) forklarer videre at de vitenskapelige metodene i et bestemt fagområde sees som systematiske og planmessige fremgangsmåter for å etablere pålitelig kunnskap og holdbare teorier innenfor dette fagområdet.

Å forske på nåværende praksis skal være et organisert forsøk på å lære av andres erfaringer (Eglene, 2000). Evnen til å identifisere og evaluere løsningene utviklet av andre organisasjoner er et avgjørende steg i prosjektgjennomføring. Fremgangsmåten for å forske på nåværende praksis er å formulere spørsmål, identifisere mulige kilder med ekspertise og så undersøke hva de gjør (Eglene, 2000). Ved å kartlegge hvordan problemet har blitt løst tidligere, er intensjonen å bruke dataene til å danne en oversikt over hva som har fungert bra og hva som har fungert dårlig. Med fokus på faktorene som ledet til suksess, kan det dannes en beste praksis som blir en oppskrift for hvordan å best mulig løse problemet. Forskningsmetoden inneholder vanligvis tre steg: formulering av forskningsspørsmål, innsamling av bakgrunnsinformasjon, og utføring av intervjuer.

3.1 Formulering av forskningsspørsmål

Forskningsspørsmål har som hensikt å avgrense forskningen, og rette datainnsamlingen i en bestemt retning. Svært åpen forskning har vist seg å være risikofylt ved at for mye data samles inn, som leder til forvirring rundt fokuset ved forskningen når analysen skal skrives (Bryman, 2008).

Mine forskningsspørsmål

1. Hvordan bør offentlige organisasjoner gå frem for å tilgjengeliggjøre åpne data ved bruk av semantiske vevteknologier?
 - a. Hva er de ulike stegene i tilgjengeliggjøringsprosessen, og hvem er de involverte aktørene (ledelse, politikk, brukere, samarbeidspartnere)?
 - i. Hvilke organisatoriske faktorer blir prosessen påvirket av?
 - ii. Hva slags ansvar har de forskjellige rollene i prosessen?
 - iii. Hva slags ressurser kreves av transformering og opplæring?
 - b. Hvilke semantiske vevteknologier er sentrale å bruke ved tilgjengeliggjøring av åpne offentlige data?
 - i. Hva slags ontologier kan det offentlige bruke for å åpne dataene?
 - ii. Hvordan kan Linked Open Data brukes av det offentlige som et åpent begrepsapparat?

Kommentarer til forskningsspørsmål

Forskningsspørsmål 1.

Opprinnelig var forskningsspørsmål 1 ”Hvordan bør offentlige organisasjoner gå frem for å åpne dataene sine ved bruk av semantiske vevteknologier?”. Jeg omgjorde dette til å tilgjengeliggjøre fremfor åpne fordi jeg mener det blir mer presist.

Forskningsspørsmål 1.a

Tidligere i arbeidet så jeg for meg innføring av semantisk vevteknologi og lenkede data i offentlige organisasjoner. Derfor hadde jeg forskningsspørsmålet ”Hva er de ulike stegene i adopsjonsprosessen, og hvem er de involverte aktørene?”. Jeg har omgjort dette forskningsspørsmålet til tilgjengeliggjøringsprosessen, fordi det handler om å bruke teknologiene ved tilgjengeliggjøring, fremfor å innføre teknologiene direkte.

Forskningsspørsmål 1.b.i.

Jeg hadde som intensjon å undersøke Linked Open Data-datasettene for å anbefale hvilke ontologier som ville passet å bruke i offentlige organisasjoner for å skape felles begrepsforståelse ved elektronisk samhandling. Fordi jeg har valgt å ikke fokusere på åpne data for elektronisk samhandling mellom offentlige etater, har jeg valgt å ikke besvare dette forskningsspørsmålet.

Forskningsspørsmål 1.b.ii. Jeg har endret forskningsspørsmålet fra ”Hvordan kan LOD brukes av det offentlige som en åpen informasjonsmodell?” til ”Hvordan kan Linked Open Data brukes av det offentlige som et åpent begrepsapparat”. Dette var for å presisere informantenes argumenter.

3.2 Innsamling av bakgrunnsinformasjon

Det er avgjørende for en beste praksis at den som utfører studiet samler inn bakgrunnsinformasjon om hvordan ting er i dag. Eglene (2000) har kommet med noen forslag til hvordan man kan gå frem i denne fasen av forskningsprosjektet. Jeg starter med å *identifisere personer* jeg tror kan ha informasjon til å besvare forskningsspørsmålene. Når disse er funnet sender jeg e-poster eller ringer til de jeg tror har slik kunnskap, eller til de som kan sette meg i kontakt med personene som har det. Jeg kan også undersøke organisasjoner som utfører beste praksiser for å høre om de har noen erfaringer på det tema jeg undersøker. Neste trinn er å *utføre et bredt søk på internett* for å identifisere organisasjoner eller mennesker som prøver å løse lignende problemer. Det vil være lurt å tenke på teknologier eller metoder som kan være nyttige i mitt prosjekt og deretter søke etter organisasjoner som har oppnådd suksess ved å benytte seg av dem. Slik kan jeg bruke erfaring fra ulike steder til å gi innflytelse til mitt prosjekt. Søkeresultatene kan jeg bruke for å identifisere kunnskapsrike personer jeg bør snakke med. Å finne teoretisk litteratur å lese meg opp på vil være vesentlig for å utarbeide intervjuguiden jeg skal bruke når jeg skal snakke med personene jeg har kommet i kontakt med.

3.2.1 Utvelgelse av informanter

Jeg deltok på Semantiske Dager som ble avholdt i månedsskiftet mai-juni 2010 i Stavanger. I tillegg til å knytte kontakter fikk jeg innblikk i dagens situasjon når det gjaldt bruk av semantiske vevteknologier i offentlig sektor. Disse seminarene og presentasjonene ga meg innblikk i praksiser og ulike felter som jeg valgte å utforske nærmere. Flere av personene jeg bestemte meg for å intervjuer var sentrale mennesker på denne konferansen. Det at jeg var tilstede på deres foredrag og at jeg pratet med dem etterpå, ga meg grunnlag og mulighet til å ta kontakt med dem senere. I etterkant av konferansen leste jeg mer om personene, og opprettholdt kontakten med dem gjennom LinkedIn.

Informantene fra privat sektor er de jeg har intervjuet for å samle erfaringer rundt bruk av teknologiene. En av dem har jeg hatt kontakten med siden jeg startet arbeidet med oppgaven. Hun er svært aktiv i det norske miljøet tilknyttet bruk av semantiske vevteknologier og lenkede data. En av de andre har jeg kommet i kontakt med gjennom

henne. To av de andre ble jeg kjent med på konferansen. Den siste personen ble jeg anbefalt å snakke med gjennom et av de andre intervjuene. Samtlige av disse informantene har vært, eller er sentrale i de store semantisk vev-prosjektene i Norge. De har gjort mye for offentlig sektor, og kjenner derfor denne sektoren godt. De har i tillegg jobbet mye med bruk av semantiske vevteknologier på offentlige data. På grunn av deres erfaring og anerkjennelse ville jeg bruke dem for å forstå beste praksis. En av informantene fra offentlig sektor har jeg truffet under utarbeidelsen av prosjektskissen. Hun møtte jeg også igjen på konferansen. Tre av de andre holdt foredrag eller var tilstede på konferansen. De to siste ble jeg anbefalt å prate med gjennom et av de andre intervjuene. Felles for disse informantene er at de er tilknyttet det norske miljøet for bruk av semantiske vevteknologier. De har derfor grunnlag til å forstå den semantiske vev i tillegg til offentlig forvaltning. Da jeg i september var klar for å intervju, sendte jeg e-poster til disse personene. Da jeg hadde intervjuet elleve personer hadde jeg fått såpass mye resultater at jeg så datainnsamlingen som tilfredsstillende.

For å anonymisere informantene har jeg valgt å gjengi dem ved pseudonavn, men oppgi deres virkelige arbeidsområde og den sektoren der de er ansatt. For informantene fra offentlig sektor har jeg valgt å spesifisere hvilken organisasjon de tilhører. De valgte pseudonavnene fra privat sektor starter på P slik at leseren gjenkjenner at disse kommer fra privat sektor. Informantene fra offentlig sektor har navn som begynner på forbokstaven til den organisasjonen der de er ansatt.

Navn	Tilhørende sektor	Arbeidsområde
Unn	Offentlig (UDI)	E-samhandling og standardisering
Didrik	Offentlig (DIFI)	Standardisering, arkitektur og åpne systemer
Brage	Offentlig (Brønnøysundregistrene)	Semantisk og organisatorisk samhandling
Scott	Offentlig (Skatteetaten)	Modellering og standardisering
Siri	Offentlig (SSB)	Modellering, begreps og klassifikasjonssystemer
Sander	Offentlig(SSB)	Formidlingssystemer
Peder	Privat	Semantisk vev og lenkede data
Pontus	Privat	Semantisk vev og lenkede data
Petter	Privat	Semantisk vev og lenkede data
Pernille	Privat	Semantisk vev og lenkede data
Pål	Privat	Semantisk vev og lenkede data

Tabell 1: Oversikt over informantene

3.3 Utføring av intervjuer

Å forske på nåværende praksis blir ufullstendig dersom studiet mitt ikke inneholder intervjuer med de personene jeg identifiserte. Fordi skriftlig materiale har en tendens til å legge for lite fokus på det negative, er det veldig viktig å snakke med personer som har vært en del av slike prosjekter (Eglene, 2000). De som utfører slike studier utsetter ofte intervjuene fordi de føler at de må ha svært god kunnskap om emneområde før de kan snakke med andre om det. Dette forsinker læringsprosessen, og det er bedre å spørre kjente eksperter enn å selv prøve å bli ekspert (Eglene, 2000). Ved å studere litteratur og gjennomføre intervjuer med personer som har jobbet med lignende problemstillinger, vil jeg få et datagrunnlag å analysere som vil kunne resultere i en beste praksis.

3.3.1 Intervjuguide

Intervjuprotokoll

Da jeg kontaktet potensielle informanter sendte jeg ved denne teksten som forklarer hensikten med masteroppgaven, og hvorfor jeg ønsker å gjennomføre intervjuet.

”Jeg holder på med masteroppgave i informasjonsvitenskap ved UIB som fullføres i juni 2011. Tittelen for oppgaven er ”Beste praksis for åpning av offentlige data ved bruk av semantiske vevteknologier”. Jeg ønsker å intervju mennesker som er involvert i organisasjoner og eller prosjekter som har erfaring og eller interesse for elektronisk utveksling av informasjon og åpning av data. Målet mitt er å få vite hva slags ønsker/behov/erfaringer det finnes i organisasjonen for å skape felles begrepsapparater ved

elektronisk utveksling av informasjon, samt lære om prosessen, roller og motivasjon for slikt arbeid. I tillegg, organisasjonens ønsker/behov/erfaringer med åpning av dataene ved bruk av semantiske vevteknologier. Hensikten er å kartlegge hva som har fungert bra og hva som har fungert mindre bra i slike prosjekter. Resultater fra intervjuet vil bli anvendt som datagrunnlag i forskning på beste praksis. Jeg vil skrive ned intervjuet ut i fra lydopptaket, og deretter gjennomføre en kvalitativ analyse med formål om å kartlegge ønsker/behov og erfaringer. Informanter vil ikke bli navngitt. Resultatene fra analysen vil jeg diskutere mot teori, for så å foreslå prinsipper for beste praksis. Hensikten med masteroppgaven min er å komme fram til en beskrivelse av adopsjonsprosessen som må til for å åpne dataene som Linked Data i offentlige organisasjoner. Temaer jeg er interessert i å snakke om er organisasjonens forhold til og ønsker rundt åpne data ved utveksling av data, informasjonsmodellering og begrepsapparater, referansedata (Linked Data / Linking Open Data) og teknologiske standarder for semantiske vevteknologier”.

Jeg hadde opprinnelig tenkt å også ta for meg tilgjengeliggjøring av åpne data som et middel for å øke evnen til samhandling mellom offentlige etater. Derfor ønsket jeg å intervju informantene om elektronisk samhandling, og hvordan åpne data og semantiske vevteknologier ville gjøre utveksling av data mellom offentlige etater enklere og mer vellykket. Jeg snakket med informantene om e-samhandling, og fikk en del data om temaet. Dette har vist seg å være et svært komplisert og omfattende område. Derfor har jeg valgt å ikke analysere og diskutere åpne data for e-samhandling. Beste praksis for e-samhandling på bakgrunn av åpne data blir for omfattende i henhold til tidsrammen jeg har for denne oppgaven.

Intervjuskjema

Gjennom arbeidet med prosjektbeskrivelsen hadde jeg utarbeidet en del punkter jeg ville finne ut av. Disse var med å skape grunnlag for intervjuguiden. Da jeg arbeidet med å finne bakgrunnsinformasjon og litteratur, skrev jeg ned temaer og spørsmål jeg kunne tenke meg å stille til informantene. Jeg leverte spørsmålene mine som en punktliste til veileder. Vi kom fram til at jeg kunne presentere spørsmålene som en matrise, hvor strukturen ble tydeligere. Jeg jobbet med intervjumatriksen i iterasjoner, inntil veileder og jeg ble enige om at spørsmålene var klare for pilotintervjuer. Jeg gjennomførte et pilotintervju på en bekjent. Hensikten var å teste spørsmålene, og strukturen for å undersøke om informanten forstod spørsmålene, og om jeg stilte bra nok oppfølgingsspørsmål. Da jeg følte at intervjuguiden var tilfredsstillende, tok jeg den i bruk på en reell informant.

Under presenteres intervjuskjemaet formalisert som en matrise. Hver rute har et hovedspørsmål. For eksempel for å finne ut om utfordringer, muligheter, eller problemer tilknyttet en organisasjon/prosess – ”Ønsker din organisasjon å åpne noen av dataene sine?”. Spørsmålene i kursiv er tatt med som tillegsspørsmål.

	Organisasjon/prosess	Informasjonsmodellering/ begrepsapparater	Referansedata Linked Data/Linked Open Data	Teknologiske standarder og verktøy
Utfordringer/ muligheter/ problemer	Ønsker din organisasjon å åpne noen av dataene sine? <i>Hvilke muligheter ser du ved at din organisasjon åpner dataene sine?</i>	Hva slags problemer/utfordringer har det vært i forhold til felles begrepsforståelse i de prosjektene du har deltatt i? <i>Ulike aktørers behov for slik informasjonsmodellering</i>	Hvilke utfordringer/problemer har dere rundt referansedata og metadata?	Hva slags muligheter ser du ved å bruke semantiske vevteknologier for å åpne dataene?
Hva har blitt gjort?	Når det gjelder åpning av data, hva har blitt gjort? <i>Hva har fungert bra? Hva slags data deles med andre organisasjoner? Hvordan deles de?</i>	Hva har blitt gjort i din organisasjon/prosjekt for å skape/bruke felles begrepsapparater? <i>Hva slags modeller? Åpne/lukkede? Hvem har tilgang?</i>	Hvilke referansedata har blitt brukt i dine prosjekter? <i>Bruk av åpne data? Erfaringer med LOD?</i>	Hvordan har dere brukt semantiske vevteknologier?
Hvorfor ikke?	Hvorfor har det ikke vært arbeidet med åpning av data?	Hvorfor ønsker dere ikke en åpen informasjonsmodell?	Hva tror du er grunnen til lite bruk av referansedata?	Hva slags hindringer er i veien for innføring av

	<p><i>Problemer/utfordringer med å skulle åpne data? Personvern, politiske årsaker?</i></p> <p><i>Nødvendige forutsetninger for å åpne dataene/innføre åpne formater? Blir innføringsprosessen påvirket av eksterne krefter som organisasjonen ikke kan styre?</i></p>			<p>semantiske vevteknologier?</p> <p><i>Utfordringer ved bruk? Kostnad ved innføring i kontrast til andre teknologier?</i></p>
Hva synes du?	<p>Hva må til for at en slik innovativ teknologi får gjennomslagskraft og blir innført?</p> <p><i>Hvem tar beslutninger? Hva slags støtte i organisasjonen tror du må til for å muliggjøre semantisk åpning av data?</i></p>	<p>Howdan kan begrepsapparater bidra til mer effektiv håndtering og bruk av organisasjonens informasjon?</p> <p><i>Ønsker til innovasjon. Mer effektiv håndtering og bruk av informasjonen. Felles begrepsapparater.</i></p>	<p>Howdan kan offentlig datautveksling dra nytte av åpne referansedata?</p> <p><i>Bruk av LOD som åpent begrepsapparat?</i></p>	<p>Hvilke semantiske vevteknologier fungerte bra i dine prosjekter for å skape felles semantikk?</p> <p><i>RDF(RDFa, RDF/XML, Turtle, N3) ,OWL, RDFS osv</i></p>
Hvorfor bør dette gjøres?	<p>I en slik prosess for å åpne dataene/ finnes det antagelig ulike roller med forskjellig ansvarsområder. Hvem involverte du? Hvordan ble ansvar fordelt?</p>	<p>Howdan har dere gått frem for å skape felles forståelse av dataene på tvers av systemer/avdelinger?</p>	<p>Hva må til for å sammenknytte en offentlig organisasjon sine datasett mot slike Linked Open Data-ressurser?</p> <p><i>Howdan skape Linked Data?</i></p>	<p>Hvorfor bør akkurat semantiske vevteknologier brukes for å åpne dataene?</p>
Hvordan bør dette gjøres?	<p>Hvem ville du involvert? Hvordan ville du fordelt ansvar?</p>	<p>Hva har fungert bra/dårlig?</p>	<p>Hva slags hensyn må tas?</p> <p><i>Fordeler</i></p>	<p>Hva slags potensial har semantiske vevteknologier for integrasjon (åpenhet)</p>

Tabell 2: Intervjumatrise

3.3.2 Gjennomføring av intervjuene

Intervjuene ble gjennomført i perioden november 2010 til januar 2011 over telefon eller Skype. Lengden på intervjuene varierte fra 40 minutter til 1,5 time, men de fleste intervjuene varte ca 1 time. Jeg brukte CallRecorder for å ta opp Skype-samtaler. Dersom jeg hadde hatt telefonintervju, brukte jeg telefonens høyttalerfunksjon for å spille av samtalen i rommet, og tok samtalen opp ved bruk av Quicktime på datamaskinen. Informantene fra privat sektor er ekspertene på bruk av semantiske vevteknologier og lenkede data, så derfor startet jeg med spørsmålene mot høyresiden av matrisen. Dersom de hadde tid, stilte jeg også spørsmålene fra venstresiden av matrisen. Til informantene fra offentlig sektor stilte jeg spørsmål fra venstresiden av matrisen først. Dersom vi hadde god tid, og de ikke hadde så mye å si, stilte jeg spørsmål om lenkede data og semantisk vev. Jeg var ikke opptatt av å følge en bestemt rekkefølge på spørsmålene, men valgte spørsmål som passet bra ut i fra hvordan samtalen utviklet seg. Det var nyttig å ha utarbeidet mange spørsmål, slik at jeg hadde tilstrekkelig med oppfølgingsspørsmål.

3.4 Transkribering

Jeg transkriberte intervjuene etter hvert som jeg gjennomførte dem. Det var nyttig fordi da fikk jeg gått gjennom intervjuet i etterkant og sett hvilke spørsmål som fungerte mindre bra og hvordan jeg kunne bli bedre for hvert intervju. Dersom jeg hadde samlet opp intervjuene, og transkribert alle som en aktivitet, tror jeg det ville vært stor sjanse for at man blir ukonsentrert og ikke får med seg alt informantene sier. Jeg transkriberte så ordrett som mulig. Fordi jeg var opptatt av hva informantene sa fremfor hvordan de sa det, var jeg ikke veldig streng med henhold til transkribering av varighet på pauser, utropstegn osv. Der hvor jeg ikke forstod hva som ble sagt satte jeg et spørsmålstegn i transkripsjonen. Dette var for å vise at jeg ikke kunne trekke noen analyse ut av denne kommentaren. Jeg brukte iTunes til å spille av lydopptakene fra intervjuene. iTunes fungerer bra med den innebygde start/stopp knappen på

MacBook Pro. De fleste transkripsjonene skrev jeg i en helt enkel teksteditor. Jeg prøvde også å transkribere ved bruk av Nvivo, men jeg synes ikke Nvivo var like effektivt å jobbe med når det kom til transkribering. Dette var fordi start/stopp knappene ikke var like raske, og da gikk det seint å transkribere. Etter at jeg hadde transkribert en del intervjuer, så jeg at formateringen i Nvivo ble bedre med .doc-filer framfor .txt. De 3-4 siste intervjuene transkriberte jeg derfor i Word.

3.5 Koding og kategorisering

Seidel og Kelle (1995, som sitert i Basit, 2003) forklarer koding som en metode for å oppdage relevante fenomener, samle eksempler av disse fenomenene, og til slutt analysere dem for å finne likheter, forskjeller, mønstre og strukturer. Coffey og Atkinson (1996, referert i Graneheim & Lundman, 2004) forklarer koder som ”verktøy for å tenke” og ”heuristiske enheter”. Ved å feste slike merkelapper ved ulike ord, setninger, paragrafer og lignende vil de som skal analysere teksten få nye ideer og tenke annerledes om dataene. En kode kan festes til diskrete objekter, handlinger og andre fenomen, og skal forstås i sammenheng med konteksten (Graneheim & Lundman, 2004). Basit (2003) skriver at koding og kategorisering handler om å inndelegge dataene og tildele kategorier. Kategoriene blir sett på som et beskrivende nivå av innholdet. Fordi Nvivo kaller koder for noder bruker jeg disse to begrepene som synonymmer. Jeg koder intervjuene for å komme fram til kategorier jeg skal bruke i analysen av dataene. Når jeg sier at jeg koder og kategoriserer er det to deler av den samme prosessen.

Da jeg hadde gjennomført 5-6 intervjuer, begynte jeg å kode og lage kategorier fra transkripsjonene. Flere av kategoriene har sitt opphav i intervjumatrisen, og kategoriene som organiserer de ulike spørsmålene under lignende temaer. Spørsmålene jeg stilte under intervjuene, dannet også utgangspunkt for flere kategoriene. ”Organisasjon/prosess”-kategorien fra intervjumatrisen omgjorde jeg til ”Tilgjengeliggjøring av åpne data”. Svarene på disse spørsmålene omhandlet i stor grad åpne data og organisasjonens syn på å tilgjengeliggjøre åpne data. Derfor mener jeg at det var riktig å konkretisere min analyse rundt tilgjengeliggjøring av åpne data, og gå vekk fra ”Organisasjon/prosess” til ”Tilgjengeliggjøring av åpne data”. Kategorien ”Semantiske vevteknologier” representerer ”Teknologiske standarder og verktøy” fra intervjumatrisen. ”Lenkede åpne data” representerer ”Referansedata Linked Data/Linked Open Data”. ”Informasjonsmodellering/begrepsapparater” kodet jeg under kategorien ”Begrepsapparater”. Det var mange flere betraktninger rundt begrepsapparater enn informasjonsmodellering, så derfor valgte jeg å ha ”Begrepsapparater” som kategorinavn. Grunnen til at jeg diskuterer begrepsapparater i kontekst av tilgjengeliggjøring av åpne data er fordi begrepsapparater også kan tilgjengeliggjøres som åpne data.

Kodingen var en måte å organisere dataene jeg hadde samlet inn og for å få oversikt over hva de forskjellige informantene sa om de forskjellige temaene, og hva som ofte gikk igjen. Jeg leste gjennom transkripsjonene mine mange ganger, og streket under ord og uttrykk. Jeg brukte Nvivo til å sortere og strukturere dataene under koder som etter hvert skulle bli kategorier i analysen. Kodehierarkiet ble til ettersom jeg kodet transkripsjonene, og festet forskjellige deler av teksten under lignende koder. Prosessen har jeg utført i mange iterasjoner, der jeg har opprettet nye koder, og endret navn på koder og kategorier etter hvert som jeg har jobbet videre. Kategoriene på nivå 1 er de nevnte hovedkategoriene. Nivå 2 samler dataene under den bestemte hovedkategorien, og det er derfor viktig at omfanget av disse kategoriene blir presist, men samtidig så bredt at de ulike praksisene passer inn. Fordeler for eksempel er en vid kategori, men analysen blir mer presis enn om jeg hadde samlet fordeler under tre kategorier; Muligheter ved åpne data, Fordeler med åpne data, Gevinster ved åpne data. Kategoriene på nivå 3 er en presisering og detaljering av nivå 2. Navn på disse kategoriene bunner i koding som jeg gjorde i Nvivo, men under analysen har de gått gjennom flere iterasjoner. Det er disse kategoriene som er grunnlaget for beste praksis. Kategoriene på nivå 3 er ulike råd, anbefalinger, erfaringer som jeg har brukt til å diskutere beste praksis. Fordi disse kategoriene spisser analysen, er det nødvendig at navn på kategoriene gjenspeiler temaene informantene snakker om på best mulig måte.

Ved å vise den første og siste iterasjonen av nodehierarkiet i Nvivo, mener jeg det gir et godt bilde av kodingen jeg har gjort. Jeg viser ikke hver iterasjon jeg har gjort av

kategoriseringen, men inkluderer fire ulike versjoner. Hver iterasjon vises gjennom en tilhørende tabell med hvordan kodehierarkiet så ut ved enden av den iterasjonen. De to første tabellene presenterer første og siste versjon av kodehierarkiet jeg jobbet med i Nvivo. Den tredje viser hvordan kategoriene så ut mens jeg kondenserte og skrev analysen. Den siste tabellen viser den endelige kategoristrukturen. Under vises den første iterasjonen av kodehierarkiet jeg opprettet i Nvivo. Listen over noder er eksportert etter at jeg hadde opprettet en del noder.

Tilgjengeliggjøring av åpne data	Ønske om å tilgjengeliggjøre åpne data	Hvem ønsker å tilgjengeliggjøre åpne data
	Uplanlagt samhandling Roller i tilgjengeliggjøringsprosess Produktivitet Nødvendige forutsetninger Muligheter ved åpne data Kvalitet Hvorfor har det ikke vært arbeidet med tilgjengeliggjøring av åpne data Hva har blitt gjort Gevinster	
Semantiske vevteknologier	Deling mellom etater Utfordringer ved å bruke semantiske vevteknologier Samhandling Kostbart	Støtte i organisasjonen, bevissthet
	Hindringer for å innføre semantiske vevteknologier i en offentlig organisasjon Bruk av standarder Bruk av semantiske vevteknologier for å tilgjengeliggjøre åpne data	Utfordringer, politiske årsaker, hindringer, hvordan tilgjengeliggjøres åpne data, hva slags data
Lenkede åpne data	Åpne referansedata i offentlig samhandling Utfordringer rundt referansedata og metadata Nytte ved åpne referansedata i offentlig samhandling Muligheter som åpen informasjonsmodell Metadata Hindringer for overgang til Linked Open Data Forutsetninger for bruk i det offentlige Fordeler Erfaringer ved bruk	Utfordringer, politiske årsaker, hindringer, hvordan tilgjengeliggjøres åpne data, hva slags data
Begrepsapparater	Utfordringer med å skape felles forståelse Informasjonsmodell Hvem ønsker felles begrepsapparater Hva har blitt gjort	Hvilke fungerer bra, hva fungerer mindre bra Utfordringer, fordelen med semantiske vevteknologier ved tilgjengeliggjøring av åpne data
		Hvordan skape Linked Open Data
		Skape begrepsforståelse

Tabell 3: Iterasjon 1

Tabellen under viser neste iterasjon.

Tilgjengeliggjøring av åpne data	Ønske om å tilgjengeliggjøre åpne data	Hvem ønsker å tilgjengeliggjøre åpne data
	Uplanlagt samhandling Roller i tilgjengeliggjøringsprosess Nødvendige forutsetninger	
	Muligheter ved åpne data Kvalitet Hvorfor har det ikke vært arbeidet med tilgjengeliggjøring av åpne data	Tilrettelegging, Støtte i organisasjonen , Sammenstilling av åpne data , Samfunnsengasjement , Politiske styringssignal ,Minimale krav , Informasjonsforvaltning ,Bevissthet To forskjellige verdener i det offentlige
		Vanskelig for å se nytten , Utfordringer , Politiske årsaker,Organisasjonskultur,Maktperspektivet,Kontroll,Hindringer,Forretningsmodeller,Balansegang

	Hva har blitt gjort	Hvordan tilgjengeliggjøres åpne data, Hva slags data
	Gevinster	
	Deling mellom etater	Samhandlingsdimensjon, Hvordan ønskes informasjonen å utveksles, Hvordan utveksles informasjonen i dag
Semantiske vevteknologier	Utfordringer ved å bruke semantiske vevteknologier	
	Samhandling	
	Løfting av dataene	
	Kostbart	
	Hindringer for å innføre semantiske vevteknologier i en offentlig organisasjon	
	Bruk av standarder	Hvilke fungerer bra ,Hva fungerer mindre bra ,For å representere modeller
	Bruk av semantiske vevteknologier for å tilgjengeliggjøre åpne data	Utfordringer ,Fordelen med semantiske vevteknologier ved tilgjengeliggjøring av åpne data
Lenkede åpne data	Utfordringer rundt referansedata og metadata	
	Selvorganisering	
	Nytte ved åpne referansedata i offentlig samhandling	
	Muligheter som åpen informasjonsmodell og begrepsapparat	
	Metadata	
	Linked Enterprise Data	
	Hindringer for overgang til Linked Open Data	
	Forutsetninger for bruk i det offentlige	Veiledning ,Tilgangstyring, Ressurser , Lovverk , Hvordan skape Linked Open data , Hvordan bruke LOD , Autoritetsregistre
	Fordeler	
	Erfaringer ved bruk	
Begrepsapparater	Utfordringer med å skape felles forståelse	
	Semantic drift	
	Roller i begrepsdefinisjonsprosesser	
	Informasjonsmodell	
	Informasjonsforvaltning	
	Hvem ønsker felles begrepsapparater	
	Hva har blitt gjort	Skape begrepsforståelse,Skape begrepsapparater,SERES,Hva slags modeller,Bruk av eksisterende begrepsapparater
	Fordeler	

Tabell 4: Iterasjon 2

Etter hvert som jeg hadde kodet alle intervjuene, flyttet jeg kodene og den tilhørende teksten over til en tekstbehandler. Jeg brukte kodene som kategorier i analysen. Grunnen til at jeg forandret mange av kategoriene i tabellen over er fordi jeg synes sammenhengen og strukturen ble lite tydelig, og jeg ville derfor lage bedre kategorier. Som Dey (1993, som sitert i Basit 2003) sier:

”Når vi oppretter kategorier, tar vi avgjørelser om hvordan vi skal analysere dataene på en måte som er nyttig for analysen, og videre hvordan denne kategorien vil passe inn i analysens utvidede kontekst”.

Fordi oppgaven som helhet handler om tilgjengeliggjøring av åpne data ved bruk av semantiske vevteknologier, så jeg det som unødvendig at kategorien om åpne data het ”Tilgjengeliggjøring av åpne data”. Derfor endret jeg den til ”Åpne data”. Mens jeg analyserte, oppdaget jeg temaer i dataene som ikke passet naturlig under eksisterende kategorier. Da opprettet jeg nye kategorier. Etter nye iterasjoner av analysearbeidet, slo jeg sammen flere kategorier en kategori for å organisere og presisere analysen. Eksempelvis under ”Åpne data”-kategorien kombinerte jeg ”Fordeler” og ”Muligheter” til ”Fordeler” fordi mye av dataene som jeg kodet under ”Muligheter” kan sees på som fordeler ved åpne data. Tabellen under viser en versjon av kategoristrukturen jeg hadde mens jeg kondenserte og jobbet med analysen.

Åpne data	<p>Fordeler</p> <p>Hvem ønsker å tilgjengeliggjøre åpne data</p> <p>Roller i tilgjengeliggjøringsprosess</p> <p>Hva slags data</p> <p>Hvordan tilgjengeliggjøres dataene</p> <p>Nødvendige forutsetninger for å tilgjengeliggjøre</p> <p>Nødvendige forutsetninger for å bruke åpne data</p> <p>Hvorfor har det ikke vært arbeidet med tilgjengeliggjøring</p>	<p>Ulike gevinster, kvalitetsheving</p> <p>Bevisstgjøring, samfunnsengasjement, minimale krav, legge til rette for gjenbruk, støtte i organisasjonen</p> <p>Organisasjonskultur, kontroll og maktperspektiv, politiske årsaker, hvem har bruk for dataene, frykt for mangel på kontroll ved gjenbruk, tap av statlige inntekter, personvern</p> <p>Semantiske vevteknologier ved tilgjengeliggjøring av åpne data, ved tilgjengeliggjøring av begrepsapparater, ved tilgjengeliggjøring av begrepsapparater</p> <p>Konvertere til Linked Open Data</p> <p>Demonstrere muligheter</p> <p>Verktøystøtte, kostnader, proof and trust,</p> <p>Hva fungerer bra, hva fungerer dårlig</p> <p>Se sammenhenger, unngå redundans, muligheter som åpen informasjonsmodell og begrepsapparat</p>
Semantiske vevteknologier	<p>Fordeler med semantiske vevteknologier</p> <p>Hvordan gå frem for å få data på semantisk vevformat</p> <p>Forutsetninger for å skulle innføre semantiske vevteknologier i en offentlig organisasjon</p> <p>Hindringer for å bruke</p> <p>Bruk av standarder</p>	<p>Autoritetsregistre</p> <p>Veiledning</p> <p>Lovverk, ressurser, tilgangsstyring, metadata og informasjonsforvaltning</p> <p>Forskjellig informasjonsstyring, økonomi, juridiske aspektet, kvalitetsproblemer, kontekst forsvinner, bruk av andre teknologier, forståelse av Linked Open Data</p>
Lenkede åpne data	<p>Fordeler med lenkede data</p> <p>Nytte ved åpne referansedata i offentlig samhandling</p> <p>Erfaringer ved bruk</p> <p>Nødvendige forutsetninger for at det offentlige skal vurdere Linked Open Data</p> <p>Hvordan bruke Linked Open Data</p> <p>Hvordan skape Linked Open Data</p> <p>Hindringer for overgang til Linked Open Data</p> <p>Utfordringer rundt referansedata</p> <p>Linked Enterprise Data</p>	<p>Autoritetsregistre</p> <p>Veiledning</p> <p>Lovverk, ressurser, tilgangsstyring, metadata og informasjonsforvaltning</p> <p>Forskjellig informasjonsstyring, økonomi, juridiske aspektet, kvalitetsproblemer, kontekst forsvinner, bruk av andre teknologier, forståelse av Linked Open Data</p>
Begrepsapparater	<p>Viktigheten av begrepsapparater</p> <p>Ønske om å tilgjengeliggjøre begrepsapparatet åpent</p> <p>Roller i begrepsdefineringsprosess</p> <p>Hvordan skape begrepsforståelse</p> <p>Utfordringer med å skape felles forståelse</p> <p>SERES</p> <p>Skape begrepsapparater</p>	

Tabell 5: Iterasjon 3

Under kondenseringen valgte jeg å analysere kategorien "Begrepsapparater", og de påfølgende underkategoriene under de tre andre hovedkategoriene. Dette valget gjorde jeg for å presisere og trekke sammen analysen. Jeg endret også navn på noen underkategorier for å oppnå bedre konsistens i navngivningen på tvers av hovedkategoriene. Eksempler på dette er når jeg gikk bort i fra å ha kategorier som "Hindringer for overgang til Linked Open Data" og "Forutsetninger for å skulle innføre semantiske vevteknologier i en offentlig organisasjon". Det er en forutsetning at hindringer og utfordringer løses for at man kan få til en gjennomføring. Derfor valgte jeg "Forutsetninger for at offentlige organisasjoner gjenbraker Linked Open Data" og "Forutsetninger for at offentlige organisasjoner bruker semantiske vevteknologier". Når de forskjellige hovedkategoriene har lignende underkategorier er navngivningen mer konsistent og analysen blir er presis. Jeg endret navn på kategoriene gjennom flere iterasjoner helt til jeg hadde en kategoristruktur med god struktur og oversikt over analyseområdet. Teksten har blitt kraftig kondensert, og antall kategorier har blitt betydelig redusert. Etter iterasjoner med gjennomgang av kategoriene, kom jeg frem til kategoriene som vises under. Jeg har nummerert dette hierarkiet for at det skal bli enklere å lese. Det er ikke samme tallstruktur som brukes ellers i oppgaven.

1. Åpne Data
 - 1.1 Fordeler
 - 1.1.1 Ubegrensede gjenbruksmuligheter
 - 1.1.2 Grunnlag for nye tjenester
 - 1.1.3 Synliggjøring av offentlig sektor
 - 1.1.4 Kvalitetsheving
 - 1.2 Hva har blitt gjort
 - 1.2.1 Hvorfor etatene tilgjengeliggjør dataene sine
 - 1.2.2 Tilgjengeliggjøringspraksis
 - 1.3 Roller i tilgjengeliggjøringsarbeidet
 - 1.3.1 Borgere
 - 1.3.2 Dataeier
 - 1.3.3 Ledere
 - 1.3.4 Departementer
 - 1.4 Forutsetninger for at offentlige organisasjoner tilgjengeliggjør åpne data
 - 1.4.1 Endre holdninger knyttet til kontroll og gjenbruk
 - 1.4.2 Økt bevisstgjøring
 - 1.4.3 Utarbeide minimale krav for tilgjengeliggjøring
 - 1.5 Forutsetninger for at offentlige organisasjoner gjenbruker åpne data
 - 1.5.1 Tilgjengeliggjøring av begrepsapparater og metadata
 - 1.5.2 Semantikkregisteret for elektronisk samhandling
2. Semantiske vevteknologier
 - 2.1 Fordeler
 - 2.1.1 Smartere kombinerings av datakilder
 - 2.1.2 Gjenbruk av begrepsapparater
 - 2.1.3 Støtte for skalerbarhet
 - 2.2 Hva har blitt gjort
 - 2.2.1 Anbefalte standarder
 - 2.2.2 Utfordrende standarder
 - 2.2.3 Konvertering til RDF
 - 2.3 Forutsetninger for at offentlige organisasjoner bruker semantiske vevteknologier
 - 2.3.1 Skape forståelse av nytteverdi gjennom å demonstrere muligheter
 - 2.3.2 Øke kunnskapsnivået om semantisk vev for å håndtere kompleksiteten
 - 2.3.3 Bedre verktøystøtte
 - 2.3.4 Håndtere kostnadene
 - 2.3.5 Økt modenhetsgrad
3. Lenkede åpne data
 - 3.1 Fordeler
 - 3.1.1 Semantisk lag på åpne data
 - 3.1.2 Redusere overflødige kopier ved å referere til autoritative kilder
 - 3.1.3 Begrepsapparater som Linked Open Data
 - 3.1.4 Lenkede virksomhetsdata
 - 3.2 Hva har blitt gjort
 - 3.1.1 Erfaringer
 - 3.3 Forutsetninger for at offentlige organisasjoner tilgjengeliggjør Linked Open Data
 - 3.3.1 Øke forståelsen av Linked Open Data
 - 3.3.2 Forbedre evnen til å definere begreper
 - 3.3.3 Håndtere tilgangsstyring
 - 3.3.4 Ivareta dataenes opprinnelige kontekst
 - 3.3.5 Sørg for gjenbrukspotensial
 - 3.4 Forutsetninger for at offentlige organisasjoner gjenbruker Linked Open Data
 - 3.4.1 Villige til å gjenbruke andres data
 - 3.4.2 Informasjonsmodell
 - 3.4.3 Autoritetsregistre
 - 3.4.4 Flere ressurser

Fordi mange aspekter er felles for åpne data og Linked Open Data valgte jeg å analysere mange av disse dataene under en egen hovedkategori som jeg kalte ”Generelle utfordringer tilknyttet lenkede åpne offentlige data”. Tidligere i arbeidet hadde jeg kodet flere av disse kategoriene under ”Forutsetninger for tilgjengeliggjøring av åpne data”. Den store forskjellen på åpne data og Linked Open Data dreier seg om i hva slags format dataene gjøres

tilgjengelig og hvordan de gjøres tilgjengelig. Derfor vil åpne data og Linked Open Data ha mange grunnleggende fellestrekk. Jeg har valgt følgende kategorier:

4. Generelle utfordringer tilknyttet lenkede åpne offentlige data
 - 4.1 Tilknytning til lovverket
 - 4.2 Tap av statlig inntekt
 - 4.3 Politiske styringssignaler om å tilgjengeliggjøre åpne data
 - 4.4 Begrensede midler til å utforske nye teknologier
 - 4.5 Bedre informasjonsforvaltning må på plass før åpne data kan gjenbrukes i det offentlige
 - 4.6 Spesialister til å velge datasett
 - 4.7 Personvern

4. RESULTATER

4.1 Åpne data

4.1.1 Fordeler

Ubegrensede gjenbruksmuligheter

Didrik forklarer at dataene Direktoratet for forvaltning og IKT (DIFI) publiserer, kan gjenbrukes av andre. Når DIFI ikke har nok kapasitet til å gjennomføre analyser er det bedre at dataene gjøres tilgjengelig, slik at andre kan gjøre denne analysen. Dette synet deler Pernille hvor hun svarer at gjenbruk av åpne data gjør at man kan finne relasjoner mellom eksisterende data som man i utgangspunktet ikke trodde hadde noen tilknytning til hverandre. Det vil kunne gi overraskende momenter som en ikke hadde tenkt på i utgangspunktet. Peder snakker om et økosystem hvor den fremste karakteristikken ved å tilgjengeliggjøre åpne data er uplanlagt gjenbruk. Det vil si at noen kreative mennesker bruker data på måter man ikke hadde tenkt seg når man la ut dataene. Lar man andre gjenbruke sine egne data kan man få innspill til hvilke ting som er interessante å gjøre med disse dataene. Petter forteller at man legger opp til uplanlagt gjenbruk når man tilgjengeliggjør åpne data uten gjenbruksbegrensinger. Plukker noen opp dataene og gjenbraker de, har det ført til en uplanlagt samhandling som ikke var planlagt da dataene ble gjort tilgjengelige. Pontus mener tilgjengeliggjøring av åpne data bidrar til uplanlagt gjenbruk ved at en aktør gjenbraker en annens data uten at dataeier hadde planlagt denne gjenbruken. Han påpeker at det selv om det var fullstendig uventet for dataeier at noen ønsket å gjenbruke deres data til det de gjorde, kan utnyttelsen av dataene være veldig godt planlagt fra brukeren sin side. Åpne tilgjengelige data kan føre til nye samarbeid. Pernille tror det er viktig at organisasjoner forstår at dersom andre gjenbraker deres data så vil datatilbyderen bli primærkilden deres, og det gir dem som datatilbyder makten. Hun mener at det bør være et av de største insentivene for la andre gjenbruke dataene. Peder forteller at ansatte i offentlig sektor har opplevd økt motivasjon når de forstår at dataene de jobber med og tilgjengeliggjør blir gjenbrukt av andre: *"Plutselig følte de at jobben deres fikk en helt annen nytteverdi"*.

Grunnlag for nye tjenester

I dataene kan det ligge spennende muligheter til innovative tjenester som dataeier har oversett. Scott forteller at det er ved økt tilgjengeliggjøring av data at man kan lage nye gode tjenester basert på de dataene som tilgjengeliggjøres. Han sier at datasett fra flere kilder som kombineres, vil potensielt gi de distinkte datasettene høyere verdi enn de har hver for seg. Pernille mener at mengden åpne data, vil tilsvare antall innovative løsninger. Det finnes utrolig mye offentlige data som kunne vært tilgjengeliggjort og videre danne grunnlaget for nye tjenester. Pål forteller om forslaget til Fornyings-, administrasjons- og kirke departementet (FAD) der de ønsker å publisere åpne data om alle svømmehaller i Norge. Han mener offentlige organisasjoner må forstå at noen der ute kommer til å ta tak i disse dataene og lage nye tjenester; for eksempel iPhone- eller Android-applikasjoner. I tillegg må de ta inn over seg at et slikt gjenbruk av åpne data vil kunne være en fordel for de som tilgjengeliggjør dataene. Ved at dataene gjenbrukes i en innovativ tjeneste som eksempelvis kan være "Oversikt over svømmehaller i din kommune", vil det gi personer lettere tilgang til informasjon om denne svømmehallen, flere mennesker besøker svømmehallen, og

institusjonen øker billettinntektene sine. Han etterspør offentlige datasett tilknyttet kulturminner, tennisbaner, fotballbaner, åpningstider, museer og lignende. Han ser for seg at det kunne blitt skapt mange nyttige tjenester basert på slike datakilder. Brage ser ubegrensede muligheter ved åpne tilgjengelige data, og nevner Regnskapsregisteret som en datakilde han mener andre kunne hadde stor glede av å gjenbruke. Registeret kunne hjulpet næringsliv og andre til å fått en bedre forståelse av sin egen bedrift, og hvordan bedriften står i forhold til relaterte selskap.

Synliggjøring av offentlig sektor

Unn forteller at statistikk som offentlige organisasjoner tilgjengeliggjør, vil bidra til en mer transparent offentlig sektor fordi samfunnet får tilgang til produksjonstall fra de ulike etatene. Dersom Utlendingsdirektoratet (UDI) tilgjengeliggjør datasett som inneholder deres statistikk over antall asylsøkere og hvor mange saker UDI klarer å behandle i løpet av en bestemt periode, kunne disse datasettene bli sammenstilt og sammenlignet med statistikk fra andre land. Forutsatt at de andre landene gjør det samme, ville det indikert hvorvidt UDI burde effektivisere prosessene sine. Unn mener at tilgjengeliggjøring av slike data ville også vært en fordel for andre offentlige organisasjoner: *”Sett at Skattevesenet visste hvor mange som hadde søkt om opphold, da ville det gi noen indikasjoner på hva de må gjøre med sin produksjon”*. Pontus sier at i utlandet, og delvis i Norge så ser en på åpne data som et middel for økt synliggjøring av offentlig sektor. Siri tror det er nyttig å tilgjengeliggjøre åpne data slik at offentlige organisasjoner er mer bevisst på hva andre deler av myndighetene holder på med. Det vil kunne unngå at data samles inn flere ganger. Dersom en etat allerede har samlet inn bestemte data, kan dataene heller gjenbrukes av de andre organisasjonene som måtte ha behov for disse.

Kvalitetsheving

Pål sier at når organisasjoner skal vurdere å tilgjengeliggjøre åpne data fører det til bedre datakvalitet. Kvalitetsheving skjer ved at organisasjonene ser nærmere på sine databaser og sørger for at datakvaliteten på dataene som skal tilgjengeliggjøres er høy nok. De rydder opp i dataene deres fordi de er redde for at potensielle brukere vil oppdage feil og mangler. Han forteller at en annen form for kvalitetsheving kommer av at samfunnet kan bidra. Dette har skjedd i biblioteksektoren hvor kunnskapsrike mennesker har bidratt med ekstra kunnskap som ikke nødvendigvis bibliotekarere innad i organisasjonene innehar. Flere av informantene mener tilgjengeliggjøring av åpne data vil avdekke feil og heve datakvaliteten:

”Det ser vi på en del av de kilderegistrene som distribuerer data i dag. Der finnes det kvalitetsbrister, og de blir ofte identifisert av distributørleddet som videreformidler og sammenstiller dataene. De ser at det gir konsekvenser når det er feil informasjon som er tilgjengelig i kilderegistrene” (Brage).

”Når vi publiserer data på et aggregert nivå og flere får brukt dataene, er det flere øyne som ser på samme datasett. Da er det større muligheter for å oppdage feil eller inkonsistens med andre data samlet inn av andre mennesker. Så jo flere øyne som ser på dataene, jo bedre må kvaliteten bli” (Siri).

”Hvis man får tilbakemeldinger på at dataene ikke kan brukes fordi datakvaliteten er får dårlig, så kan jo det ha en innflytelse på datakvaliteten” (Peder).

4.1.2 Hva har blitt gjort

Hvorfor etatene tilgjengeliggjør dataene sine

Didrik forteller at DIFI ønsker å tilgjengeliggjøre åpne data og at de på en del områder har som politikk å åpne opp dataene. På andre områder er det de ansatte som tenker at dette kan være interessant for flere, så derfor blir dataene tilgjengeliggjort. Unn har ikke fått forståelse for at UDI har noen ønske om å tilgjengeliggjøre åpne data utenom statistikk. Scott forklarer at det står i Skatteetaten sin strategi at de skal tilby gode tjenester til det offentlige og hvis det offentlige ønsker informasjon fra Skatteetaten så vil de prøve å etterstrebe og dele ut det. De

ønsker å dele og bidra i offentlig sektor med den informasjonen de kan gi ut, og ønsker at andre etater tilgjengeliggjør sine data slik at Skatteetaten kan gjenbruke de dataene. Brønnøysundregistrene (BRREG) er i utgangspunktet interessert i å tilgjengeliggjøre data, men grunnet størrelsen på organisasjonen og spredningen mellom de ulike registrene, tror Brage at ønsket om å tilgjengeliggjøre varierer avhengig av hvilken driftsperson man snakker med. Andre forhold som påvirker villigheten er fokusområdet, kompetansesiden og profesjon. Det medfører at diskusjonen rundt tilgjengeliggjøring av åpne data utarter seg ulikt avhengig av om folk kommer inn fra teknisk, juridisk eller andre typer ståsted. Informantene fra Statistisk Sentralbyrå (SSB) mener at problemstillingen om hvorvidt de ønsker å tilgjengeliggjøre åpne data handler om i hvilken grad de er i stand til det fremfor om de ønsker det eller ikke. Brage forteller at arbeidet med åpne data som et samfunnsanliggende er relativt nytt, og til tross for at det har kommet noen nye tjenester basert på tilgjengelige data, er det først de to siste årene at det har blitt et tema. Han mener at det viktigste for å komme i gang med bruk av åpne data er å tilgjengeliggjøre mest mulig åpne data, men at en må ha et edruelig forhold til hva de kan brukes til og hva de kommer til å bli brukt til.

Tilgjengeliggjøringspraksis

Peder savner et åpent forum der folk kan komme med ønsker om hvilke data de har størst behov for. Da hadde dataeierne kunnet prioritere å tilgjengeliggjøre de datakildene som de får konkrete henvendelser om. Petter forteller at de som leverandør tilgjengeliggjør foretaksinformasjon og informasjon om steder som lenkede åpne data i form av URIs. Dataeier forvalter datasettene som URIen peker på, og URIen blir tilgangspunktet til de åpne dataene. Det meste av data hvor man kan få effekt av å aggregere basert på en URI, og referere ved en URI tilgjengeliggjøres. DIFI tilgjengeliggjør kartinformasjon, forskningsdata, statistikkdata og operasjonelle data som ikke er direkte tilknyttet saksbehandling i offentlig sektor. Resultatene fra prosjektene DIFI holder på med tilgjengeliggjøres på deres nettsider. Deres tilgjengelige data kan lastes ned eller aksesseres via direkteoppslag. Didrik sier at det er mange i DIFI sin arbeidsgruppe for formidling som ønsker vevtjenester som tilgangsmetoder fremfor Linked Open Data. Resultater fra Skatteetaten sine skatteberegninger tilgjengeliggjøres på deres nettsider og blir også trykket på papir. Scott forteller at Skatteetaten har hatt en diskusjon om deler av folkeregisteret kunne vært tilgjengeliggjort som åpne data. Unn forteller at statistiske data om hva slags søkere som har kommet, publiseres månedlig på UDI sine nettsider. SSB tilgjengeliggjør aggregert statistikk på sine nettsider i tillegg til anonymiserte data til forskningsformål. De publiserer også metadata og konseptvariabler. I Metadata-portalen kan en søke på alle variabler som er knyttet til deres data. Sander forteller at SSB mangler en tjeneste for å gjøre fornuftige søk på metadataene når en skal finne spesifikke tabeller. SSB jobber med et nytt formidlingssystem som vil gi datatilgang gjennom vevtjenester. Pontus mener data.norge.no som en portal med HTML-lenker til diverse tilgjengeliggjorte datasett, er en god begynnelse men tror at vi vil se en utvikling mot andre type tjenester på sikt. Dataene som BRREG har gjort tilgjengelig er ikke gjort frigjort slik at man kan hente ut hele datasett. Det er ingen formelle begrensninger på å gjøre det, men BRREG har ikke tilrettelagt for det i utgangspunktet. Hvis man ønsker hele datasett må man gjennom et distribusjonsledd som tar betaling. Dataene er gjort tilgjengelige slik at man kan gjøre enkeltoppslag på ulike register. Hvis en har et organisasjonsnummer, kan en gå inn å slå opp om det har skjedd noen endringer i for eksempel lokalisering og adressestrukturer. Brage forteller at det er veldig begrenset når det gjelder offentlige data som er totalt åpent tilgjengelig. Grunndataregistrene, adressebeskrivelser, matrikkelen til Kartverket, Enhetsregisteret og Folkeregisteret hos Skattdirektoratet er åpent tilgjengelig slik at en kan søke direkte i dataene.

4.1.3 Roller i tilgjengeliggjøringsarbeidet

Siri mener de forskjellige rollene i en tilgjengeliggjøringsprosess vil være jurister, dataeiere, vevspesialister og folk som jobber med semantikk. Brage forteller at jurister avgjør om det er mulig i henhold til lovverket å tilgjengeliggjøre dataene. I tillegg peker han på IT-avdelingen som de første man vil gå til for å undersøke muligheter og konsekvenser. Unn forklarer at i slike prosesser trenger hun en fasilitator, en som kan faget og prosessen, en jurist og en som kan informasjonen.

Borgere

Flere av informantene (Brage, Pernille, Peder og Pontus) mener dataene må åpnes og tilgjengeliggjøres fordi folk ønsker å gjenbruke dataene. Didrik ser at spesielt unge mennesker ønsker mer opplysninger. Derfor må DIFI kunne tilfredsstille kravet om å åpne dataene samt si hvor de kan finnes. Scott tror tilgjengeliggjøring av åpne data vil være litt behovsprøvd i forhold til hva samfunnet ønsker. I prosessen med å tilgjengeliggjøre data, tror Brage at folkets meninger og ytringer om åpne data spiller en viktigere rolle enn faglige tunge artikler om samfunnsgevinst i kroner og øre. Siri og Sander ser på SSB sine data som et offentlig gode. De bruker skattebetalernes penger til å samle inn dataene og da skylder de samfunnet å publisere dem.

Dataeier

Siri sier at det er dataeierne som bestemmer om dataene skal tilgjengeliggjøres og det avhenger da av hvilke data det er. Dataeier er de som avleverte mikrodata til SSB. Hun sier at de ansatte kan fremme at de er interessert i å tilgjengeliggjøre åpne data, men at de verken har beslutningsmyndighet eller makt til å gjøre det. Didrik sier også at det er dataeierne som har ansvaret for dataene og som tar en beslutning på om hvorvidt dataene skal gjøres tilgjengelig, og på hvilken måte de skal tilgjengeliggjøres. Brage nevner dataeier som en sentral person ved tilgjengeliggjøring.

Ledere

Resultatene viser at det er nødvendig med forankring i ledelsen for å tilgjengeliggjøre åpne data. Både Peder og Pål forteller at biblioteksektoren ønsker å tilgjengeliggjøre åpne data, og dele ressursene sine med hverandre. Pål sier at i ABM-verdenen finnes det mange ledere som har forstått nytten av åpne data, og ønsker å gå videre med det. Pernille mener en må få ledere til å ville tilgjengeliggjøre data ettersom teknikere ikke er beslutningstakere. Peder har erfart at det er enkeltpersoner i maktposisjoner som avgjør hvorvidt en organisasjon tilgjengeliggjør data. Brage sier det er avgjørende at lederne i organisasjonen forstår og ønsker å gjennomføre tilgjengeliggjøringen. Scott forteller at det er ledelsen som uttrykker hva som skal gjøres og i hvilken grad det skal gjøres. Pål presiserer at det må være vilje fra ledelsen for å få dette drevet videre.

Departementer

I offentlig sektor er det ofte myndighetene som avgjør om informasjonen skal tilgjengeliggjøres og deles. Flere av informantene snakker om at FAD er det departementet som er den sterkeste pådriveren for å tilgjengeliggjøre data. Gjennom opprettelsen av data.norge.no, og flere konkurranser der deltakerne skal lage programmer med åpne data, viser FAD at de ønsker å gå i denne retningen. Etatene får signaler fra departementsnivå om hvorvidt de får lov til å åpne dataene sine, men til nå finnes det ikke noen sentrale føringer på hvordan dette skal gjøres. Pål forteller at Rigmor Aasrud, Fornyingsministeren, er veldig positiv for tilgjengeliggjøringen av åpne data. Han sier at hun står på sitt og er virkelig villig til å legge føringer for de store organisasjonene. Didrik har mistanker om at det vil endre seg etter hvert og tror at det vil komme større trykk fra FAD og Nærings- og handelsdepartementet om å gjøre data tilgjengelig. Pontus poengterer at ettersom det er to tre stillinger i FAD og DIFI som jobber med åpne data, er det på langt nær et samfunnsengasjement på departementsnivå.

4.1.4 Forutsetninger for at offentlige organisasjoner tilgjengeliggjør åpne data

Endre holdninger knyttet til kontroll og gjenbruk

Det er snakk om en endring i hvordan ting gjøres, så Brage mener det må til en mental endring i organisasjonene. Han opplever at mange etater har en organisasjonskultur hvor det er frykt for å tilgjengeliggjøre og åpne opp dataene fordi det kan gi konsekvenser:

"Å se på at den informasjonen som man tidligere hadde et veldig sterkt eierskap til, og som man ønsker på mange måter å ha full kontroll over, skal begynne å

tilgjengeliggjøres og åpnes opp for strukturert og kanskje ustrukturert bruk. Det er skremmende. Det er mange som går rett i forsvarsstilling når de hører noe sånt”.

Når organisasjoner har sikret informasjonen sin i databaseløsninger kun for internt bruk, har de full kontroll over disse dataene. De må endre holdning når dataene skal gjøres tilgjengelig for andres bruk. Han tror organisasjoner må endre kontrollperspektivet. Når dataene skal bli åpent tilgjengelige og det ikke lenger skal ligge noen restriksjoner på hva dataene skal kunne brukes til blir mange tilbakeholdne. Pål tror det bunner i at etater har vært alene på sine databaser i mange år og vet hvem som bruker dataene. På store databaser må det offentlige ofte ha kontroll. Når dataene tilgjengeliggjøres vet en ikke lenger hvem som bruker dataene. Han forteller han spurte Forskningsrådet om å få gjenbruke dataene som ligger i deres prosjektarkiv. De svarte at de hadde problemer med å gi han tilgang til deres data. Begrunnelsen var at hvis de la den ut på nettet, så mistet de kontrollen over hvordan de ville brukes videre. Fra før av hadde de tilgjengeliggjort alle prosjekter som har fått finansiell støtte, samt navnet på tilhørende prosjektledere, på sin arkivside. Til slutt fikk han prosjektarkivet med prosjektbeskrivelser, prosjektidentifikatorer og ledende institutt, men ikke navnet på prosjektlederen. Det som var paradoksalt var at navnet på prosjektlederne allerede var mulig å finne på nettsidene, men de vil ikke gi bort dataene slik at folk kunne publisere dem. Pål forklarer dette som et gjennomgående holdningsproblem hos offentlige organisasjoner. Han snakker om en slik kontroll i sammenheng med maktperspektivet når toppledere bruker dataene som et konkurransefortrinn ovenfor sine kolleger i andre etater eller organisasjoner:

”Har du kontroll over dine data, så har du en maktfordel ovenfor andre som trenger disse dataene. Hvis du åpner opp, så gir du egentlig den fordelene bort. Så du må være veldig trygg på deg selv og hva du kan og din organisasjon når du gjør det. Noen etater er det. Men ikke alle”.

Økt bevisstgjøring

Pernille tror at å endre tilgangen til organisasjonens data fra organisasjons-til samfunnsnivå fører til skepsis blant offentlige organisasjoner. For at organisasjonene skal endre tankesettet tror hun det handler mye om å tørre å tilgjengeliggjøre noen data. Dersom organisasjonen ser nytten ved å tilgjengeliggjøre dataene sine vil det også gjøre endringen lettere. Hun opplever at flere av forskningsprosjektene deres hvor offentlige etater deltar bidrar til at etatene forstår nytteverdien ved åpne data. Hun tror at ytre påvirkning og oppmuntring fra andre som har tilgjengeliggjort data vil ha mye å si. Petter eksemplifiserer dette ved at programmeringsmiljøer ser mulighetene, og når de begynner å lage applikasjoner basert på åpne data vil det gjøre at andre ønsker å følge etter. Om ikke organisasjonene forstår og er bevisst på hva det skal endres til, er det naturlig at de ikke ønsker å gå bort fra den tilnærmingen de er vant til. Pernille mener det må jobbes med bevisstgjøring samt øke kunnskapen om hva det vil si å tilgjengeliggjøre dataene:

”Man har trodd at hvis man åpner dataene, så gir man bort forretningsideen sin. Det har først tatt noen år før folk har forstått at det ikke er tilfellet. At du heller kan skape merverdi for deg selv ved at du blir din egen og andres primærkilde”.

For Skatteetaten sin del, tror Scott at det går mye på bevisstheten og ønsket om å bruke ressurser på arbeidet. Både Brage og Scott sier det handler om bevisstgjøring. Dersom man oppfordrer etater til å tilgjengeliggjøre sine data, mener Scott det vil bidra til en mer kollektiv utdeling av data. En annen bevisstgjøring går på det at offentlige etater må forstå at andre ønsker tilgang til dataene deres. Unn forklarer at hun syns det er vanskelig å se for seg hvem som kunne gjenbrukt UDI sine produksjonstall. Didrik nevner at offentlig sektor ikke har vært klar over at andre i så stor grad har ønsket å benytte dataene i egne tjenester. Det er også flere i offentlig sektor som ikke tilgjengeliggjør data fordi de ikke ser nytten av å tilgjengeliggjøre dataene når de ikke er 100% i forhold til detaljerte beskrivelser. Brage forteller at det er viktig at etater tilgjengeliggjør data selv om ikke datakvaliteten er 100%. Han påpeker at mangelfulle beskrivelser ikke vil kunne gjenbrukes i alle tilfeller, men i mange

sammenhenger kan det være nettopp slike instansdata som danner grunnlaget for innovative tjenester. Når de offentlige organisasjonene vurderer å tilgjengeliggjøre data, skal de ta i betraktning at det er etterspørsel etter dataene.

Utarbeide minimale krav for tilgjengeliggjøring

Pål sier at den viktigste forutsetningen er at man har data digitalt tilgjengelig, og trekker frem et eksempel hvor han har digitalisert 560 000 papirark. Ellers mener både Pål og Brage at det bør stilles minimale krav til dataene som tilgjengeliggjøres. At de offentlige organisasjonene vet hvordan de skal gå frem og hva de skal tenke på når de skal tilgjengeliggjøre åpne data. Brage forklarer at departementene snakker om viktigheten av åpne data, men at de ikke er tydelige nok i nyanseringen av hva dataene kan og bør brukes til. Han mener det må stilles krav til metadatabeskrivelser når datasett tilgjengeliggjøres.

4.1.5 Forutsetninger for at offentlige organisasjoner gjenbraker åpne data

Tilgjengeliggjøring av begrepsapparater

Petter forteller at begrepsapparater handler om informasjonskvalitet og en bedre oversikt over dataene. Dersom man ikke vet hva man måler, er det vanskelig å måle og vurdere resultatene man får. Uten begrepsapparater vil det da være vanskelig å gjenbrake åpne data. Pontus mener at begrepsapparater er utslagsgivende for at dataene kan brukes:

”Hvis du ikke vet hva dataene dine betyr så vet du heller ikke hvordan du skal forvalte de. Du vet ikke hva du kan bruke de til og du har ikke noen mulighet for å måle kvaliteten på de så lenge du ikke vet hva de betyr. I det du begynner å beskrive hva dataene dine betyr mer og mer presist så har du mer og mer kontroll på hva du kan bruke det til, hvilken kvalitet de har, og hvordan dine data er i forhold til andres data. Så evnen til samhandling øker etter hvert som du mer og mer beskriver hva egentlig dine data betyr”.

For at SSB kan lage statistikk, er evnen til å gjenbrake begreper som brukes i skjemaer svært viktig. Petter forteller at klarhet i begreper er nødvendig dersom man ønsker å oppnå en stor grad av fleksibilitet i utveksling av informasjon mellom ulike parter. Begrepsapparater bidrar til å identifisere forenklinger som fører fram til effektivitet. Pernille deler en del syn med Petter når det gjelder fleksibilitet og oversikt:

”Det vil tydeliggjøre hva man faktisk snakker om og hva det er man tar en avgjørelse på bakgrunn av. Det vil gi likhet i saksbehandling. Pluss at du kan knytte sammen relaterte data. Du trenger ikke personer til å søke opp definisjonen av begreper fordi begrepet vil stå som relatert til et annet begrep og sånn plukker du med informasjon som er relevant for deg”.

Scott forteller at det er veldig viktig at man har et felles begrepsapparat med felles forståelse for å standardisere og forenkle kommunikasjonen både mellom mennesker og i forhold til mennesker og systemer. Manglende forståelse av dataene blir ofte trukket frem som et hinder for utstrakt gjenbruk. Slike beskrivelser sørger for at dataene faktisk har den meningen man tror de har. Denne semantiske dimensjonen må sikres slik at de potensielle brukerne av dataene tolker og forstår dataene på samme måte som dataeier. Gal tolking av dataene, vil føre til gale beslutninger. Han forteller om hensikten bak begrepsmodellering:

”Man prøver å lage en beskrivelse av dataene og begrepene på en måte som er forståelig både for menneske og maskin slik at maskiner er i stand til å gjøre automatisert saksbehandling og ha et korrekt grunnlag for å ta avgjørelser”.

Unn ser begrepsapparater som en nødvendighet for å kunne finne igjen informasjon. Man må kunne vite navnet på det begrepet man leter etter. Uten forståelse av informasjonen du skal bruke, er det vanskelig å gjenbrake den. Hun sier begrepsapparater kan bidra som en felles ordbok som vil legge til rette for bedre samhandling:

”Legger vi det åpent så kan Lånekassen gå å se på dette og mye tidligere være i stand til å se hva vi mener med begrepet. Hvis Lånekassen gjør det samme så kan vi bare gå til begrepsapparatet til Lånekassen og så kan vi sammenligne det. Det vil da si at vi allerede har et bilde av språket til den andre når vi møtes fysisk. Når forskjellige avdelinger i UDI har forskjellige saker, og noen av sakene glir litt over i hverandre, må man ha kontakt på tvers. Det er klart at har man det samme begrepsapparatet så snakker man jo det samme språket”.

Hun forteller at det var et vanvittig engasjement da de inviterte ansatte i UDI til å jobbe med en felles begrepsmodell. Brage forteller at DIFI er veldig opptatt av begrepsapparater. At offentlige aktører skal tilgjengeliggjøre sine begrepsstrukturer åpent er en målsetting som er innbakt i styringsdokumentene som hver etat får i 2011. Didrik bekrefter at DIFI ønsker åpne felles begrepsapparater. Sammen med Skattedirektoratet, Kartverket, NAV, KS, Riksarkivet, Utdanningssektoren, KITH og SSB jobber DIFI med hvordan man kan åpne opp begrepsapparater og hvem som burde ha ansvaret for disse begrepene. Målet er å komme fram til en strategi for hvordan en skal skape felles begreper i offentlig sektor, hvordan disse teknisk sett skal håndteres, og hvordan man eventuelt kan publisere åpne data.

Hvordan skape begrepsapparater

Pernille forklarer at det er viktig å få på plass en beskrivelse av hvordan begrepene skal defineres, hvem som skal eie definisjonene, og hvilke egenskaper som skal være tilknyttet disse begrepene. Scott mener man må begynne å lage modeller som tydeliggjør i hvilken kontekst begrepene skal brukes. Man må få oversikt over hvordan begrepene brukes forskjellig slik at man kan sammenligne og bli enige om en omforent bruk. Også han mener det er viktig med en registerløsning for å ta vare på begrepsdefinisjonene. Unn forteller at de definerer prosessene sine sammen med sine samhandlingsaktører. Det er ikke fokus på hele prosessen, men den delen av prosessen der en ser at en får grensesnittet mellom aktørene. Hun sier at når de skal lage en begrepsforståelse eller få den semantiske betydningen i noe informasjon som skal utveksles, så må de vite hvilken informasjon som benyttes av hvem i hvilke oppgaver. På den måten fanger de semantikken og begrepsforståelsen i det de ønsker å utveksle. Hun forteller hvordan de i UDI har jobbet med å skape en begrepsmodell:

”Vi har arbeidsmøter hvor man innkaller forskjellige fagavdelinger. Så begynner man med en del begreper og ord som man ofte bruker i saksbehandling. Hva er barn, flyktning, søker? Så sorterer man begrepene og prøver å putte det inn i en taksonomi. Det vil si at du sorterer begrepene og prøver å lage et rammeverk rundt dem i forhold til domene. Begreper gjøres på tre måter. Det er noe som kan skje på tvers gjennom det offentlige og da er det vel på metadata-nivå. Så gjøres det jobber internt hos forskjellige etater for å lage sine egne begrepsmodeller. Og når vi da skal samhandle med disse etatene så må vi se på hvilke informasjon vi bruker i forskjellige prosesser slik at vi kan pinne noen av disse begrepene lenger ned. Så det er de tre områdene man jobber med det på. Det er ikke gjort over natta. Det er en lengre prosess”.

Scott forteller at mye av utfordringen med å skulle definere begreper og skape felles begrepsforståelse går på at de ulike avdelingene bryr seg lite om hva andre gjør. Han sier det er vrient å formidle viktigheten av felles begreper til avdelinger som kun tenker på egne prosesser.

”Det er vanskelig å komme inn i prosjektene og si at vi ønsker å trekke ut det dere holder på med. Det som er spesielt for deres data, modeller og begreper sånn at vi kan representere det på en formell teknisk måte slik at vi kan sammenligne det med andre systemer”.

Pernille forteller at roller for hvem som bør være med i en slik prosess ikke er definert opp i offentlig sektor. Når et begrepsapparat skal utformes, må en involvere IT-arkitekter, folk som jobber med informasjonsformidling, de som er sterke til å kommunisere, lingvister og

filosofer. Når en skal sette sammen et nytt begrepsapparat er det viktig at en samler kompetanse utover de som er i organisasjonsledelsen. Når begreper skal defineres og knyttes til hverandre, er det viktig at man gjenbraker eksisterende begreper fremfor å finne på nye. En må ta utgangspunkt i virkelige begreper som brukerne, funksjonærer og resten av organisasjonen benytter. Hun anbefaler at en går forsiktig frem, og starter med noen begreper. Når det er flere enn to aktører som skal bruke et begrep, er det viktig at meningen i begrepene kommer tydelig frem slik at hvem som helst kan gjenbrake begrepet uten å måtte snakke med de som eier begrepet. Sander sier at innenfor semantikkområdet så forsøker de knytte sammen begreper med tesauruser og lignende måter å strukturere begreper på. Siri mener begrepsapparater og felles forståelse skapes ved at de som samarbeider snakker med hverandre.

4.1.6 Semantikkregisteret for elektronisk samhandling

Mange av informantene forteller om utfordringer knyttet til å skape, ivareta og gjenbrake begreper i offentlig sektor. SERES-prosjektet handler om å skape felles begrepsforståelse i offentlig sektor. Petter forteller litt om dagens situasjon, og hva som er hensikten bak SERES:

”Offentlig sektor forholder seg til en liste over de feltene som innrapporteres fra privat sektor til offentlig sektor. Det er ca 47000 felttyper og det kan være sånn som fornavn på person. Eller det kan være inntekt. Problemet med måten det gjøres på i dag er at man mister oversikten over om et inntektsbegrep som står på et skjema er det samme inntektsbegrepet som på det skjemaet. Målsettingen til SERES er å rydde opp for å forstå at ja, her skal inntekt skrives i nærmeste tusen og inntekt her skal skrives med komma og øre. Men, fortsatt er det samme inntektsbegrep. Det handler om å skille det strukturelle og det syntaktiske fra semantikken. Sånn at du får røyka ut hva det egentlig er du mener med inntekt. Er det bruttoinntekt, er det nettoinntekt, er det bruttoinntekt med eller uten sånn og sånn. Jeg tror Skatteetaten opererer med tjue-en forskjellige begreper bare på bruttoinntekt. Selv etter de har ryddet opp”.

Løsningen er et register, et felles sted, hvor begreper defineres og vedlikeholdes. Pontus forteller at de i SERES-prosjektet ser på metadata som data. Ulike etaters begrepsapparat legges inn i registeret uavhengig av hvor faktisk dataene finnes. Sander forklarer SERES som et forsøk på å sentralisere definisjonsmakten ved at du har et sted hvor definisjonene gjøres. Petter sier at SERES forvalter begrepsapparater, og skal gjøre det enklere for offentlige organisasjoner å gjenbrake og skape begreper:

”En del av det er jo nettopp begrepsapparatet. Det å tydeliggjøre hvilke definisjoner som ligger til grunn for hvert felt eller begrep i et skjema, og knytte disse definisjonene til hverandre sånn at man kan slå opp på begrepet. For eksempel å se i hvilke skjema akkurat denne definisjonen blir brukt” (Petter).

Brage forteller at de foreløpig ikke har begynt å jobbe i retning av å tilgjengeliggjøre SERES som Linked Open Data. Han sier det har like mye med ressurser som det har med ideologi å gjøre. De ønsker nemlig at modeller og beskrivelser skal være på en åpen løsning sånn at man har mulighet til å tilgjengeliggjøre det for allment bruk. Pontus forteller at de publiserer begrepsapparatet åpent ut i en serie formater, deriblant OWL og XMI. Formatene og hva dataene betyr kan tas over i andre typer språk som gjør de egnet til å bearbeide og egnet til å sammenstilles med andre begrepsapparater. Brage håper at SERES vil brukes som et begrepsapparat i offentlig sektor:

”Når vi først har gjort en så stor tung investering, så er det jo med tanke på at flest mulig innenfor offentlig sektor skal kunne få bruke SERES, og få en verdi av å bruke SERES. Det at flest mulig kan bruke det vil jo øke verdien av investeringen”.

Han håper SERES kan være med å supplere Linked Open Data med tyngre, og mer detaljerte beskrivelser av modellstrukturer rundt informasjonen. SERES må ivareta behovet som

offentlig sektor har for å dele informasjon seg i mellom. Han forteller at det er ønskelig å kunne koble sammen åpne data med de modellene og beskrivelsene som er tenkt i SERES.

4.2 Semantiske vevteknologier

4.2.1 Fordeler

Smartere kombinerings av datakilder

Flere av informantene trekker frem semantiske vevteknologier sin evne til å kombinere datakilder som teknologienes sterkeste fortrinn:

”Åpne data uten semantisk vev gir ikke noe nytte hvis du har tenkt å kombinere datakilder” (Peder).

”Hvis du har mange kilder og behov for å aggregere informasjon mellom disse kildene, så vil en semantisk vev-løsning være veldig veldig god. Det er dens styrke” (Pernille).

Peder forklarer at det er evnen til å spesifisere semantikk som gjør formatene unike. Fordi de gjør det mulig å eksplisitt definere hva som menes med de forskjellige feltene i datasettet, vil maskiner kunne forstå logikken, resonnerer og automatisk oppdage lenker mellom ulike datakilder. Det er slik semantiske vevteknologier kan sammenstille informasjon som man i utgangspunktet ikke visste var relatert. I Folkeregisteret finnes det mange ulike konsepter for adresse. Det kan være bostedsadresse, postadresse, jobbadresse osv. Med semantiske vevteknologier kan det spesifiseres hva de ulike konseptene betyr og hva som skiller de fra hverandre. Det vil gjøre at maskiner kan velge den riktige adressen når eksempelvis en automatisk tjeneste skal sammenkoble datasett fra NAV og Skattetaten. Slike egenskaper har ikke tradisjonelle informasjonsstrukturer som XML. Pål forteller om den logiske teorien bak den semantiske veven som muliggjør automatisk sammenstilling av ulike datakilder:

”Du kan bruke den informasjonen du får på den semantiske veven automatisk i applikasjoner og slå sammen ting basert på logiske og filosofiske antagelser som er implementert i rammeverket. Det er en stor fordel hvis du sammenligner det med Topic Maps og XML-standarder og andre lignende standarder”.

Peder forteller at programvareutvikling ved bruk av semantisk vevteknologi vil muliggjøre det som i dag er økonomisk umulig. Med tradisjonelle teknologier vil en trenge å betale en programmerer for å kombinere datakilder, som gjør at kostnaden ved å kombinere nye kilder øker for hver kilde man vil legge til. Med semantiske vevteknologier legges grunnlaget for gjenbruk når dataene struktureres, så brukere av systemet kan kombinere datakilder uten at en trenger å leie en programmerer for å gjøre den jobben. Måten det kan gjøres på er at en har grafiske grensesnitt hvor brukeren får se hvilke datasett som kan legges til, og velger å kombinere datasett etter behov. Det medfører at ansatte uten programmeringskompetanse kan legge til nye datakilder, og kostnaden ved å bruke en kombinasjon av forskjellige datakilder går dramatisk ned. Pål forteller om et arbeid de har gjort på kombinerings av databaser som Dun & Bradstreet har om firmavurdering. Han ser for seg at slike firmavurderinger og adresser fra BRREG og SSB-statistikk om bransje, kan integreres med semantisk vevteknologi og danne grunnlag for en ny tjeneste. Det vil skape merverdi for en kommersiell aktør som selger denne tjenesten videre.

Gjenbruk av begrepsapparater

Siri forteller at semantiske vevteknologier gir muligheter for å raskere identifisere om man snakker om de samme tingene. Pål sier det er slik fordi teknologiene tilbyr begrepsapparater som er universelt gyldig uten å spesifisere det for et bestemt domene. De semantiske beskrivelsene definerer domenemodellen slik at aktører på tvers av organisasjoner kan ha nytte av hverandres arbeid. Med semantiske vevteknologier blir dataenes kontekst definert på et globalt nivå som gir helt andre muligheter for gjenbruk. Pål forteller at de ser en trend til at

flere går vekk fra å definere begrepsapparatene sine på lokalt nivå. Peder mener begrepsapparater bør sammenstilles ved bruk av semantisk vevteknologi. Han tror ikke en skal tvinge et enkelt begrepsapparat på alle, men heller lenke begrepsapparater til andre begrepsapparater ved bruk av semantisk vevteknologi. Han mener det vil føre til en bedre semantisk interoperabilitet ved at man reflekterer over hva begrepsapparatene betyr, som gjør at problemer knyttet til forståelse kommer tydeligere frem. Når man får eksponert begrepsapparatene får de mer praktisk bruk, samtidig som man blir mer bevisst på hva som finnes i den enkelte organisasjons begrepsapparat. Pål støtter Peder sitt syn på å bruke semantiske vevteknologier for å lenke sammen ulike begrepsapparater. Ved at ulike organisasjoner publiserer sine begrepsdefinisjoner og knytter dem opp mot lignende definisjoner kan de sammenlignes. Han forteller at dette vil kunne lede til en kunnskapsdeling hvor aktører gjenbraker hverandres begrepsapparater. Dette vil gjøre at begrepsapparater kan gjenbrukes på tvers av kontinenter, som gir en helt annen måte å definere begrepsapparater på. Scott ser stor nytteverdi for offentlige organisasjoner ved at de gjenbraker hverandres begrepsapparater ut i fra felles utvekslingsformater:

”Vi lager noen beskrivelser her i Skatt og vi vet at NAV har begrepet også. De har faktisk bedre definisjoner enn oss på det, så da burde vi bare bruke deres. Hvis Statens Kartverk sitter på eiendomsinformasjon og denne er det faktisk de som er ekspertene på, det som går på matrikkel og eiendomsinformasjon, så bare referere til de. Ikke bruke tid på å finne opp ting på nytt. Hvis alle klarer å bruke bedre definerte ting som ligger ute og klarer å se litt hva som er nyttig og bruke felles utvekslingsformater så slipper man å gjøre så utrolig masse kostbare transformasjoner mellom data”.

Støtte for skalerbarhet

Pål forklarer at utgangspunktet i den semantiske veven er at en skal ha et rammeverk som tillater at alle sier alt om alt og at det skal være mulig å ha motsigelser. Den semantiske veven er bygget på en ”Open World Assumption”, som vil si at andre kan påstå andre ting enn meg, og det vil fortsatt være gyldig fordi hvert datasett kun er en del av en global database på veven. I relasjonsdatabaser kan jeg påstå at dersom jeg har kun ti hunderaser i min hundetabell, finnes det ingen flere hunderaser i verden. En slik lukket påstand om verden kan jeg ikke trekke i den semantiske veven. Han sier at vi aldri blir ferdig med å beskrive hva vi vet og hva vi kan, og et slikt åpen syn på verden gir semantiske vevteknologier uante muligheter for gjenbruk og utvidelser. Relasjonelle databaser er bygget i en trestruktur som krever modifikasjoner dersom forskjellige databaser skal lenkes sammen. Den semantiske veven bruker vevens grafstruktur som gjør det mulig for utviklere å gjenbruke og utvide andres datakilder. Pernille mener en slik grafstruktur er veldig fordelaktig fordi den eneste endringen en gjør er å hekte på en ekstra trippel i grafen. Det vil gi alle like store muligheter for å kunne bidra.

4.2.2 Hva har blitt gjort

Pål mener den akademiske satsingen har gitt en avkastning som har lagt et grunnlag og en annerledes forståelse for hvordan semantiske vevteknologier kan brukes på tvers av aktører, kulturer og implementasjoner. Det er en internasjonal bevegelse rundt den semantiske veven som gjør tilnærmingen veldig kraftig. Pål forteller at i kontrast til andre teknologier har W3C-standardene blitt utviklet i Europa, USA og Asia samtidig. I tillegg har hvert kontinent sin egen semantisk vev-konferanse, aktører som bruker dette, og utviklingsmiljøer som jobber videre med det.

Anbefalte standarder

Pernille forteller at de har brukt W3C sine semantiske vevstandarder. Også Petter forteller at det er disse standardene som brukes. Pernille anbefaler bruk av URIer som identifikatorer, som så lenkes til andre URIer i RDF. Didrik forteller at det er ønskelig at alle begrepene får egen URI slik at de kan identifiseres. Det vil legge grunnlaget for at man skal kunne lenke disse dataene som Linked Open Data. Han er sterk tilhenger av URIer som referanser fordi slike identifikatorer er teknologiavhengige. Selv om man velger en teknologi, så er ikke URI

en teknologi som hindrer ulike tekniske implementeringer ved senere anledninger. Petter mener målet bør være å bygge systemer hvor dataene er strukturert i RDF. Det vil gjøre det lettere for applikasjonene som skal bruke dataene.

”RDF-standarden gir deg mulighetene til å skape lenker mellom datasett. Den er den mest sentrale og den er helt stabil. Den har jo ikke endra seg på over ti år. Det betyr ikke at den ikke har svakheter, akkurat som HTML kommer i nye versjoner etter hvert, men den har vist seg å fungere”.

I SSB har de sett på bruken av RDF ved organisering av metadata, og da mulighetene ved å bruke RDF sammen med SDMX (Statistical Data and Metadata Exchange). Siri mener det er et trinn på riktig vei, og at de vil fortsette å vurdere bruken av RDF, men tviler på de tar det i bruk med det første. Pernille mener Turtle-serialiseringen av RDF er veldig bra fordi den har lett syntaks som gjør den enkel og oversiktlig å bruke. Også Peder mener Turtle er ryddig og enkel å bruke. Han oppfordrer offentlige etater til å tilgjengeliggjøre dataene sine i Turtle-format:

”Det kan hende at RDFa kommer til å bli den virkelige måten å eksponere RDF på, men jeg kan tenke meg at hvis man i offentlig sektor skal eksponere dataene sine, så er det like greit å eksponere det som Turtle og RDF/XML. Helst Turtle”.

Petter anbefaler SPARQL som tilgangspunkt til RDF. Når brukeren kjenner vokabularet til datasettet, gir SPARQL brukeren en kraftig metode for å gjenbruke dataene. Grunnen til det er spørrespråkets fleksibilitet som lar brukerne velge hvilke data de ønsker å gjenbruke og i hvilke formater de vil ha de i. I deres leveranser har de brukt SPARQL veldig aktivt, og de har tilgjengeliggjort sine åpne data via et SPARQL-tilgangspunkt. Han forteller at bruken av SPARQL er stadig økende. Sander ser for seg at SSB kunne bruke SPARQL som tilgangspunkt til deres data:

”Det optimale hadde egentlig vært å lage et SPARQL-tilgangspunkt som vi både kunne brukt internt som et spørreverktøy samtidig som vi og kunne modellert tradisjonelle tjenester på toppen av den. Da hadde du hatt mulighet til å tilby både de mer tradisjonelle REST-baserte vevtjenestene til sluttbrukere som ikke har lyst til å begynne med RDF, og så kunne du tjent de som har lyst til det”.

Han forteller at med måten SSB har indeksert metadataene sine på, ville det ikke vært komplisert å lagt et SPARQL-tilgangspunkt på toppen av disse dataene. Hva slags standard du bør bruke, avhenger i følge Peder av prosjektets kravspesifikasjon. Han sier at i mange tilfeller kommer man langt med å bruke RDFS. Dersom man bare skal drive med klassifisering av ting, er SKOS tilstrekkelig. Trenger man strengere klassifisering må man se på OWL.

Utfordrende standarder

Alle informantene som jobber med semantiske vevteknologier forklarer at full utnyttelse av OWL-funksjonalitet krever svært høy forståelse av OWL. Språket vil være utfordrende å bruke for folk som ikke er eksperter på området, noe som ikke er meningen i utgangspunktet. I tillegg krever full utnyttelse av OWL svært god kompetanse på logikk. Det har ført til at OWL ikke brukes på den mest uttrykksfulle måten. Petter sier det er mer RDFS ++ eller OWL --. Pernille tror det er lettere for folk å forholde seg til skjemaer enn til ontologier:

”OWL er et veldig kraftig språk. Det å klare for eksempel å generere mer data fordi du sier jeg har en far som heter sånn og sånn og min bror har en far som heter sånn og sånn så kan du da... automatisk resonnerer fram til ved hjelp av en regel at jeg og min bror er søsken. Uten at du eksplisitt har sagt det i systemet. Det er en ekstremt kraftfull mekanisme hvis du bruker det riktig. Bruker du det feil kan det bli rotete”.

Pernille forteller at en havner i mye filosofiske debatter når man skal definere opp hele verden i OWL. Det gjør det nødvendig med avgrensning. Peder mener at hvis man har behov for OWL-teknologi, så er det en del ting som vil bli vanskelig, men sannsynligvis mulig å løse. Petter sier det er mulig å bruke resonering, men at det vil være krevende:

”OWL Full har en stor uttrykkskraft, men lar deg ikke få svaret kjapt nok. Så er det sånn at bare et uttrykk i OWL-definisjonen din eller det som kalles aksiomer, det kan medføre at hele jobben går dobbelt så seint. Bruker dobbelt så lang tid”.

Peder forteller at de ikke brukte OWL til å resonere frem nye data fordi det tok for lang tid når datamengdene var såpass store. Petter forklarer at i praksis så bruker man OWL til å definere datamodellen, og svært lite resonering. Pernille forteller at hun har brukt OWL mest i modelleringsarbeidet fordi resonering på store datasett kan være tar svært tidskrevende. Petter sier at hele semantisk vev-miljøet har gått ganske langt vekk fra det å definere begrepene i OWL som aksiomer til mere det å bruke OWL som en datadefinisjonsmodell. Peder forteller om utfordringer med XSLT:

”Når du skal generere URIer, altså identifikatorene, så brukte vi en del data sånn som navn på en film. Så viser det seg at det er mange tegn inn i de titlene som ikke kan brukes i en URI, og så trenger man en eller annen strengmanipulasjon av de tegnene. Det blir veldig fine URIer av å gjøre det på den måten, men det var en ikke triviell greie. XSLT er ikke kraftig nok på strengmanipulering til at det var særlig vellykket. Vi brukte masse tid på noe forferdelige greier i XSLT til å gjøre det riktig. Sånn i etterpåklokskapens lys så er jeg ganske sikker på at SAXFilter i det tilfellet der ville vært det riktige å bruke”.

Pernille mener RDF/XML ikke er så bra fordi det gir en dårlig blanding mellom RDF og XML. Peder føler at RDFa er ganske komplisert, men ser at formatet er nyttig sammen med (X)HTML. Petter sier at Topic Maps er i realiteten død som standard fordi den er fullstendig overkjørt av W3C-standardene. Petter forklarer videre at det ikke er noe i veien for å bruke JSON og XML sammen med Linked Open Data, men at det ikke er det optimale.

Konvertering til RDF

Petter forteller at så fort de finner åpne strukturerte data, så konverterer de dataene til RDF med lenker til hverandre og eksterne kilder. Peder mener man som utvikler må se hva som finnes av åpne data, og så løfte dataene ut med en eller annen automatisk metode, for så å gå over etterpå å se om det blir riktig. Han mener det må finnes heuristikker for dette arbeidet. Dersom dataene eksisterer i en relasjonsdatabase kan man bruke R2M ved å lage en transformasjon mellom en relasjonsdatabase med eksisterende skjema, og en modell som kan brukes til å eksponere SPARQL. Pernille forteller at en kan legge et lag over den relasjonelle databasen som kopierer deler av den relasjonelle databasen til RDF. Når en lager en RDF-versjon så velger man hvilke data en ønsker synliggjort. Hun mener det er en av mer nyttige teknikkene. Da er hensikten å tilgjengeliggjøre den RDF-versjonen man får ut av konverteringsprosessen som Linked Open Data. Hun sier ulempen ved en slik seleksjonstilnærming kan være at man usynliggjør noen data som andre ønsker å gjenbruke. I tilfeller der det eksisterer SOAP vevtjenester, sier Peder at han ville brukt tilsvarende metode som med R2M. Han er ganske sikker på at dette vil i mange tilfeller være en god måte å gjøre dataene gjenbrukbare med semantiske vevteknologier. Han har brukt ulike typer RDF-templater for å løfte data ut av Excel-ark, og mener det fungerer tilfredsstillende så lenge det er statiske data. Peder mener det er riktig å konvertere statiske data til en trippel-struktur som støtter SPARQL-spørringer. Har man en godt strukturert XML-datakilde, vil XSLT kunne transformere XML til RDF. Dersom man ikke har krav til SPARQL-støtte, men bare ha data til et RDF-format mener han det PERL-baserte rammeverket RDF-Trine vil være den riktige metoden. Han påpeker at hvilken tilnærming man ønsker å bruke bør sees i samsvar med hva slags prosjekt og hvilke behov man har.

”Det er fryktelig mange måter å gjøre det på. Man må egentlig se på hvilken case man har og gjøre en vurdering ut i fra hva slags formater du skal løfte, hvilke type datasett og hvor ofte de oppdaterer seg. Kanskje de fremste kriteriene man bør se på”.

Peder sier det er vanskelig å gi råd for beste praksis på hvordan man helst bør løfte dataene ut av tradisjonelle formater og over til semantiske vev-formater. Pernille er enig i at det er mange måter å gjøre det på, og forteller at det selv om offentlige organisasjoner bør ha en RDF-versjon av dataene i den relasjonelle databasen, er det ikke nødvendig å slutte gå helt vekk fra slike databaser. De kan ta mye informasjon, er raske og nyttige til sitt formål.

4.2.3 Forutsetninger for at offentlige organisasjoner bruker semantiske vevteknologier

Skape forståelse av nytteverdi gjennom å demonstrere muligheter

Pernille mener at offentlige organisasjoner som skal vurdere å bruke semantiske vevteknologier er avhengig av å ha kjennskap til når det er nyttig og når det ikke er det. Det krever leverandører som vet hva de snakker om. Peder synes det er vanskelig å skulle overbevise folk i ledelsen. Han forteller at de tenker nytte ut fra hva de vet i dag, og ikke ut fra den kreativiteten som utløses ved gjenbruk. Scott tror at fordelene må beskrives. Det mangler forståelse for hva som ligger i semantiske vevteknologier, og hvordan bruk av disse kan hjelpe offentlige organisasjoner. Han etterspør demoer som viser gevinsten ved endringen og at det faktisk er nyttig og verdi i å gå den veien:

”Hva er det som ligger i en semantisk vev? Er det veldig annerledes enn det man egentlig har tenkt eller gjort tidligere? Hva er viktig i forhold til det man egentlig gjør? Det er store og tunge prosesser som pågår i forhold til data og systemer, og det at en skal tenke annerledes sitter langt inne”.

Petter tror en av de viktigste faktorene for å innføre semantiske vevteknologier, er å implementere noen løsninger som demonstrerer mulighetene. De må vise hva som er mulig med den semantisk veven, fremfor å problematisere hindringene. Også Peder mener at å demonstrere mulighetene vil bidra til å skape forståelse for nytteverdien: *”De eneste jeg føler jeg kan bidra med er jo egentlig å skrive ting som viser hvor nyttig det er”.* Pontus forklarer at demoene er viktige for å skape interesse og forståelse for at den semantiske veven er mulig å realisere. Han mener juristene og økonomene er nødt til å bli klar over nytteverdien. Pernille forteller at en må gjøre teknologene interessant, og få de til å forstå forskjellene på disse teknologiene og andre teknologier. Beslutningstakere må forstå nytteverdien, slik at de er villige til å investere og ta kostnaden. Scott mener det er vanskelig å forstå hva det dreier seg om fordi folk blir skremt av alle lovordene og mulighetene rundt den semantiske veven:

”Måten man snakker om en semantisk vev på gjør at det blir veldig høytflygende for mange. Man må se egentlig mer for seg at her ønsker man å legge ut informasjon hvor man klikker seg litt mer gjennom dataene og på en måte går de veiene dataene peker. Da er man også avhengig av at man klarer å knytte dette at man har identifikatorer uten at man skal tenke seg at veven skal være så intelligent”.

Scott mener det trengs en bevisstgjøring på det å se verdien av å bruke semantiske vevteknologier og Linked Open Data fremfor eksisterende teknologier. I dag fokuserer Skatteetatens på tjenesteorientering med datautveksling mellom brukersystemer og ulike integrasjonsplattformer. Pernille mener IT-skeptikere og andre som ikke har forstått nytteverdien av disse teknologiene må opplyses. Hun forteller at det er mange som har kjennskap til eksisterende teknologier og mener at de gir det samme utbytte. Pål mener det viktigste er at leverandører av programvare begynner å bruke disse standardene inne i systemene de produserer. Scott mener at offentlige organisasjoner må rydde opp i sin egen informasjonen før de kan ta i bruk semantiske vevteknologier for å tilgjengeliggjøre dataene.

Han ser utveksling av XML -dokumenter med bruk av identifikatorer som en måte å klargjøre informasjonen til senere bruk av semantiske vevteknologier.

Øke kunnskapsnivået om den semantiske veven for å håndtere kompleksiteten

Flere av informantene påpeker at å bruke semantisk vevteknologier kan være komplekst, og at det er nødvendig med et kunnskapsløft i programmeringsstanden. Pål syns det er lovende at mange flere universiteter har fått kurs innenfor semantisk vev slik at studenter begynner å skjønne, samt bruke teknologiene. Petter utfordrer norske universiteter til å bli bedre på semantisk vev, og mener Norge må utdanne flere programmerere som forstår den semantiske veven. Det innebærer at en må begynne å skrive deklarative datamodeller fremfor lukkede datamodeller. Pål tror at kompleksiteten ved den semantiske veven kan være en stopper:

”Å modellere kunnskap er ikke noen enkel sak. Å forstå hvordan man modellerer kunnskap er ikke enkelt i det hele tatt. Jeg har selv jobbet i kunnskapsmodellering siden 1988 og jeg kan ikke si at jeg skjønner alt. Det er så utrolig komplisert og hvordan skal du da få dette til slik at en vevutvikler i Indonesia forstår det like greit som en forsker i Sydney og en japaner som jobber i Samsung i Sør-Korea? Det kan være en stopper”.

Sander mener det er vesentlig høyere terskel for programmerere å bruke SPARQL fremfor en tradisjonell REST-basert vevtjeneste. Ettersom brukerbehovet er mye større for vevtjenester har ikke SSB lagt til rette for å støtte SPARQL som tilgangspunkt til deres data.

Bedre verktøystøtte

Peder mener at utviklerverktøyene for å utvikle programmer med semantiske vevteknologier må bli bedre slik at utviklere ønsker å bruke disse teknologiene fremfor tradisjonelle vevteknologier. Uten bedre verktøy får de ikke utviklet tjenestene som gjør at sluttbrukere uten programmeringskompetanse kan kombinere og gjenbruke datakilder.

”Utviklingsverktøyene som utviklere har tilgjengelige for å lage interessante ting med semantiske vevteknologier er for dårlige. Både i modenhet og funksjonalitet. Som utvikler føler man at man ikke får gjort de tingene som man med klassisk vevteknologi kan gjøre relativt enkelt, og dermed får man ikke noe nytte av det. Så prøver mange utviklere en gang, og så gir de opp. De verktøyene som kunne ha gjort det, er altfor tungvint i bruk og kan ikke konkurrere med de verktøyene som man bruker i dagens situasjon. Og dermed så føler ikke utviklere at de får noe igjen for å bruke RDF, kanskje tvert imot at det er bare en hindring”.

Han forteller at å utvikle disse verkøyene blir som å bygge en grunnmur man ikke kan selge. Kunden etterspør et ferdig hus, men problemet er at det ikke finnes noe grunnmur. Fordi man ikke kan selge og tjene penger på grunnmuren, er det ingen som lager denne grunnmuren. Peder forteller han har hatt mange ideer til hvordan grunnmuren skal se ut, men han har ikke fått den finansielle støtten til å utvikle den. Han mener det er kritisk at det utviklingen av slike verktøy blir tatt på alvor:

”Uten de får vi ikke demoene og da får vi ikke vist at det er økonomisk veldig lurt å gjøre det. Og så er det klart at når man faktisk har utviklerverktøyene på plass, så er det en vei å gå i det offentlige å forankre det i større institusjoner. Men, når det skjer så tror jeg at disse nettverkseffektene allerede begynner å slå inn såpass mye at det blir sånn at det må gjøres på lik linje som at man må ha en vevside. Det er liksom det som er bottom line da. Det er utviklerverktøyene som må vise at det helt tydelig er riktig å gjøre det”.

Petter mener også at verktøystøtten er for dårlig til at semantiske vevteknologier tas i bruk internt i etatene. Han tror det vil ta veldig lang tid før det skjer. Pål mener det ikke er noen grunn til å la verktøystøtte stoppe utviklingen av semantiske vevapplikasjoner. Han forteller at det finnes en del veldig gode kommersielle produkter som både er skalerbare og

velfungerende. De er ikke gratis tilgjengelig, men dette er gode, robuste, modne løsninger fra selskap som Oracle og HP. Han mener at en ikke bør trenge slike verktøy for å ta i bruk semantiske vevteknologier fordi det finnes mer en nok gode alternativ som er gratis tilgjengelig:

”I et prosjekt nå så bruker jeg Open Source-software fra Stanford. Protegé og Jena, og har kjørt i nesten atten måneder uten problemer. Det var en gratis programvare som kjører på en Linux-boks”.

Også Pernille forteller at Protegé fungerer helt fint.

Håndtere kostnadene

Transaksjonskostnaden ved å bytte ut eksisterende databaseteknologier med semantiske vevteknologier vil kunne være svært høy. Pernille forklarer at det er ingen som bytter de ut fordi det er såpass inntektsgenererende å administrere databaser. Dersom en skal betale kostnaden for å erstatte eksisterende databaserløsninger med semantiske vev-løsninger over natten vil det gjøre semantiske vevteknologier svært kostbare forklarer Pernille. Peder mener at dersom en tar kostnaden ved å utvikle en velfungerende verktøyplattform, vil semantiske vevteknologier på sikt senke utviklingskostnadene:

”Per i dag er det mye dyrere, men potensialet som denne teknologien har i kraft av sine unike egenskaper gjør at det kommer til å bli mye billigere å utvikle, og dermed så kan man realisere ting som i praksis er umulig å realisere i dag”.

Han mener at det økonomiske aspektet rundt den semantiske veven vil avgjøre om teknologiene får gjennomslagskraft. Han mener at semantiske vevteknologier vil senke utviklingskostnadene i fremtiden. Siri tror ikke det er mer kostbart, men at de samme ressursene brukes til andre ting. Pål sier at dersom en tar i bruk semantisk vev i dag vil det kunne være veldig kostbart. Kostnadsnivået avhenger av hvor godt data dine er strukturert og digitalisert, og hvor god forståelse du har for semantisk vev. Altså hvor mange feil du gjør underveis. Han mener at gode veiledninger, minimale krav å forholde seg til samt kurs vil begrense disse kostnadene. Han legger til at ”early adopters”, altså de første som tar i bruk en ny teknologi, vil alltid ha en høyere kostnad enn de som kommer etterpå. Han forteller av egne erfaringer med fri programvare og presiserer at bruk av semantiske vevteknologier ikke trenger å være dyrt, men at det kommer an på størrelse og viktighet av det du gjør, samt krav til stabilitet og robusthet. Han tror kostnadene ved å bruke semantiske vevteknologier reduseres jo fler utviklere som bidrar:

”Hvis de første dataene nå begynner å bli tilgjengelige og andre aktører ser den interessen. Da ligger det enda mere programvare åpent tilgjengelig på nett, da begynner utviklere å få øynene opp for det, og da kommer den ballen til å rulle, og da blir det billigere”.

Økt modenhetsgrad

Petter mener det gjenstår litt på modenhetsgraden til den semantiske veven. På modenhet forklarer Peder at det har blitt tilgjengeliggjort en del data som folk har begynt å gjenbruke, men at det gjenstår en del på modenhet som gjør at det er vanskelig å se veien videre. Unn har snakket med leverandører om å bruke semantiske vevteknologier, men har fått inntrykk av at det er et stykke igjen før de kan brukes i offentlig sektor. Pål ser teknologiene som tilfredsstillende modne, og mener det er et veldig godt grunnlag for å gå videre med det. Han snakker likevel om manglende modenhet ved de øverste lagene i teknologistabelen. Han tror at på de neste tre fire årene har folk jobbet med å utvikle de to tre øverste nivåene av teknologistabelen. Sånn som det er i dag, snakker Pål om utfordringene rundt å kunne bevise at dataene er reliable:

”Du kan jo tenke deg på folkehelse. De er jo livredd for at folk begynner å automatisk interpretere data om helse. Bare en liten feil skal til for at folk begynner å

dø. Hvem har ansvaret da? Er det folkehelsa som har publisert det, eller er det personen som har interpretert det feil? Da er det virkelig sånn at du må vite at dataene du bruker kommer fra en pålitelig kilde. At det virkelig er en person som har legitimitet til å si det. Og det vet man ikke på internett i dag. Det er nettvett egentlig ikke sant og kan man implementere nettvett automatisk? Det er et spørsmål”.

Videre nevner han utfordringer med å stole på at dataene er tolket riktig. Han mener det er nødvendig å få på plass enkle mekanismer som alle kan bruke:

”Hvis du har bestemt deg for at denne siden kan jeg stole på. Så må du også ha bevis på at det som står der virkelig er riktig. At det er riktig interpretert. Det gjelder lovdata for eksempel. Har jeg forstått definisjonen av samboer? Hvis jeg bruker en helt annen terminologi så kan det være at jeg rett og slett ikke har forstått den. Og da kommer en kompleksitet som er veldig høy. Der ligger det virkelig en utfordring”.

4.3 Lenkede data/Lenkede åpne data

4.3.1 Fordeler

Semantisk lag på åpne data

Petter diskuterer de 5 stjernene ved åpne data og fordelene med lenkede åpne data.

”Punkt en er i realiteten et kart som ikke kan gjenbrukes. Punkt to er at du legger det ut i en eller annen strukturert form. HTML er en god nummer to. Punkt tre er at man velger å legge ut i maskinlesbare formater. Men, fortsatt som en dump uten kontekst og uten identifikatorer. Punkt fire er at man fokuserer veldig mye på identifikatorene. For å kunne kalle det Linked Open Data, altså at man lenker til andres datasett så krever det nummer 5, og det er i realiteten kun mulig hvis du ser på standarder med RDF”.

Han forklarer at data.norge.no i dag befinner seg mellom to og tre stjerner. Han mener at det burde finnes en sjettestjerne hvor man ikke bare bruker RDF til å lenke til andre datasett, men sørger for at også vokabularet sitt er tilgjengelig for andre. Dette begrepsapparatet eller ontologien bør tilgjengeliggjøres som et navneområde. Det vil sørge for at man har en bevissthet rundt dataene man tilgjengeliggjør, som vil gi andre muligheten til å kunne forstå hva du mener med dataene dine. Uten Linked Data og semantiske vevteknologier vil dette ikke være mulig. Åpne data er uavhengig av teknologi, men for å maksimalisere gjenbruk og presisjon bør dataene tilgjengeliggjøres med 5 og 6 stjerner. Han sier videre at man får vanvittig mye bedre effekt av å bygge applikasjoner på toppen av Linked Open Data. Pernille sier at Linked Open Data gir andre dimensjoner for gjenbruk og entydighet fordi dataene er klart definert. Da vet man når man snakker om det samme, og når man ikke gjør det. Ved å fristille dataene ut av proprietære systemer og isteden strukturere de som Linked Open Data, vil det gi oversikt over de begrepene en forholder seg til. Hun forklarer nytten av slik entydighet med et eksempel om lik rettspraksis:

”Typiske eksempelet er Skatt og Nav. Hva faktisk bruttoinntekt betyr. Med en gang du åpent definerer og tydeliggjør hvilke type reglementer dataene er knyttet til, så vil det kanskje komme tydeligere frem hvilket av inntektsbegrepene man skal ta utgangspunkt i for at det blir likhet. Man alltid velger det samme begrepet i samme kontekst da. Noen ender opp med å få støtte, mens andre ikke får støtte fordi man bruker helt forskjellige bruttoinntektsbegreper”.

Pernille forklarer at måten denne entydigheten kan skapes på, er ved at en har en ontologi som heter NAV-ontologien hvor inntekt er definert i følge NAV sin kontekst, mens inntektbegrepet i en Skattekontekst er noe annet. Lesbarheten maskin-til-maskin i mellom, blir tydeligere ved at data tilgjengeliggjøres som Linked Data. Sander tror måten Linked Data

er logisk bygget opp vil gi fordeler. Og selv om man ønsker å bruke RDF og Linked Data på visse områder, er det ikke nødvendig at alt omstruktureres:

”Inferens gir muligheten til å finne sammenhenger på tvers av datasett som man tidligere ikke har sett. Du har og mulighet til å gjøre en mye bredere analyse av metadata på tvers av de datasettene du har. Jeg tror at jeg kan gjøre det her på forskjellige nivåer. En trenger ikke å legge om alt til å bruke for eksempel tripler eller Linked Data. En kan gjøre det på metadataene. En kan velge på hvilket nivå en her skal gjøre det. Konseptvariabler, kontekstvariabler, klassifiseringer osv”.

Pernille mener Linked Open Data er ideelt for organisasjoner som har behov for å kombinere og sammenstille informasjon fra flere kilder. Petter opplever at det er enklere å forklare hva slags effekter man får ut av å lenke data med semantiske vevteknologier til de som i utgangspunktet har et formidlingsbehov.

Redusere overflødige kopier ved å referere til autoritative kilder

Petter mener at Linked Open Data vil være nyttig for de som ønsker å redusere antall kopier, samt holde dataene sine oppdatert. Han vektlegger hovedfordelen med Linked Open Data som evnen til å peke på noe med en URL, og med et tilknyttet begrepsapparat forstå hva det er en får tilbake. Scott støtter dette synet når han sier at slike lenkede data vil gi en referanse til hva dataene betyr, og hvordan de er knyttet opp. Petter forklarer at Linked Open Data skiller seg fra dagens løsninger ved at man ofte referer til autoritative kilder. Det vil si at kilden har autoritet og er pålitelig. Hvis dataene er strukturert som Linked Open Data er det kun den autoritative kilden som må oppdateres. Det gjør at andre kan gjenbruke en bestemt URI, istedenfor at alle må oppdatere sine informasjonselementer. Unn forteller at dobbeltlagring av informasjon er en stor utfordring i offentlig sektor. Å ha færre kopier, som Linked Open Data tillater, vil kunne være en stor fordel for det offentlige. Petter forklarer fordelene med at kommunelister blir strukturert som Linked Open Data, og som lar aktører gjenbruke en felles kilde:

”Hvis den lå med lenkede URler ville man til enhver tid være sikker på om man pekte på Ås kommune i Akershus og ikke Buskerud. I dag brukes bare kommunenummer. Da kan en kommune ha endret navn eller ordfører, og den endringen vil bare ligge hos de som måtte være autoritative for kommuneinformasjon”.

Han tror at det vil bli opprettet en del sentrale punkter som mange trenger å referere til. Sannsynligvis vil et av formatene da være RDF med URler strukturert som Linked Open Data. Pontus forteller at Linked Open Data ofte er to ting: instansdata og metadata. Han snakker om Linked Data som en ny terskel når det gjelder informasjonsinfrastruktur:

”Lenkede data med semantiske vevteknologier er en samfunnsendring på linje med introduksjonen av internett og mobiltelefonen. Med en semantisk god nok beskrivelse til at maskiner kan gjøre seg nytte av semantikklaget. Det er det spranget som vi når er inne i. Det gjør at vi kan tenke systemutvikling og distribuerte informasjonssystemer på en måte som vi ikke har kunnet gjøre tidligere. Det vil påvirke måten vi skaper og forvalter informasjon på”.

Å lenke åpne data vil kunne gi felles referansepunkter slik at en kan henviser til de samme tingene på en digital måte. Fordi Linked Open Data er åpne data lenket til andre datasett ved bruk av semantiske vevteknologier, kan denne lenkingen gi kunnskap om ny informasjon. På eksemplifiserer dette ved at han jobber med sin database som han lenker til en annen database. Gjennom denne lenkingen finner han nye kilder som han da begynner å gjenbruke.

Begrepsapparater som Linked Open Data

Det finnes mange fordeler ved at veldefinerte begrepsapparater er åpent tilgjengelige. Akkurat som dataene kan gjenbrukes, vil det være fordelaktig om metadataene som beskriver de ulike dataene også tilgjengeliggjøres som Linked Open Data. Ettersom Linked Open Data-

strukturen har som hensikt å lenke sammen ulike datasett, bør det ikke være noe i veien for at Linked Open Data kan være en arena for deling av organisasjoners begrepsapparater. Siri mener bruk av Linked Open Data som et åpent begrepsapparat hadde vært en fordel. Hun ser for seg at ikke-sensitive metadata bør publiseres, og trekker frem fordelene med tilgjengeliggjøring av metadata på et semantisk og strukturert nivå. Peder har tro på at man eksponerer sitt begrepsapparat, men selv om en slik tilgjengeliggjøring ikke forutsetter Linked Open Data, ønsker han bruk av Linked Data-prinsippene:

”Man eksponerer sitt begrepsapparat i form av vokabularer som RDFS og OWL. Så ser man om det er en uoverensstemmelse mellom det du mener med adresse og det jeg mener med adresse. Det handler generelt om interoperabilitet i offentlig sektor. Når man eksponerer det for flere brukere så får man flere kommentarer og dermed så må man skjerpe seg. Jeg er veldig for at det skal være Linked Data-prinsipper som gjør at man får alle disse gunstige effektene som åpen dataintegrasjon gir”.

Pål er bestemt på at det er mulig å tilgjengeliggjøre offentlige organisasjoner sine begrepsapparater som Linked Open Data, men ikke på alle områder. De har ofte diskutert om norske lovdata kan være åpent tilgjengelige. At disse dataene kun finnes tilgjengelig i en norsktekstlig form, begrenser den maskinelle gjenbrukbarheten. Han mener man bør kunne bryte lovverket ned i deler, og publisere de forskjellige bitene med tilknyttede begrepsapparater. Petter synes det er svært interessant hvordan Linked Open Data vil kunne fungere som et åpent begrepsapparat med semantiske beskrivelser. Han mener Linked Open Data på sikt vil kunne erstatte den XML-baserte meldingsutvekslingen som i dag er praksis ved utveksling av informasjon mellom offentlige organisasjoner. Linked Open Data vil være gunstig i de tilfellene der man har registre av felles nytte, og ønsker å få tilgang til hverandres data. Dagens utvekslingspraksis er tungvint og ineffektiv. Man må gjøre avtaler med hver nye aktør som ønsker å få benytte seg av ulike registre. I XML-skjemaene er begrepssemantikken lite tydelig, som gjør at man må forklare hverandre hvordan man tenker rundt begrepene som benyttes. Petter tror Linked Open Data potensielt vil kunne føre til et ganske stort skifte i måten informasjon utveksles på.

Lenkede virksomhetsdata

Pernille forteller at Linked Enterprise Data handler om å bruke prinsippene for Linked Data internt i organisasjoner. Pål sier at det egentlig er det samme som Linked Open Data, men på virksomhetsnivå. En brannmur beskytter dataene gjennom intranettet, og gjør at dataene holdes internt i bedriften. Linked Enterprise Data fungerer som et lag over databasene slik at det blir lettere å gjenbruke data på tvers av systemer. Denne tilnærmingen skal sørge for at data i ulike databasesystemer kan kommunisere med hverandre. Pernille sier at Linked Enterprise Data vil kunne effektivisere informasjonsstrømmen internt, og ved å omorganisere dataene dine på denne måten vil det være muligheter for å optimalisere en del prosesser i bedriften. Peder sier at en får den samme problemstillingen som en får med åpne data, men at man åpner det for internt bruk. Han mener at Linked Enterprise Data vil være fordelaktig ettersom Linked Dataprinsippene vil ha like stor gyldighet internt som på Linked Open Data. Peder presenterer et scenario hvor bruk av Linked Enterprise Data vil være nyttig for organisasjonen som tar prinsippene i bruk:

”La oss avskaffe reiseregninger. Du har banken, du har interne systemer, HR-systemer som sier hvem som skal ha, hva blir refundert osv. Så har du eksterne flyselskapssystemer og lignende. Så skal du integrere det her når ingen av faktisk dataene er åpne. Men, det er samme type. Du skal bare samle inn data som viser at du har reist dit på firmaets regning. Du får refundert så og så mye etter hvilke regning som er, så henter du inn regningsinformasjon og så verifiserer hver av datakildene sin ting og så stoler du på at den integrasjonen har vært gjort riktig, og så bare utbetaler du de regningene uten at medarbeiderne faktisk trenger å gjøre noen ting. Hele greia blir automatisert. Jeg tror det ville vært en veldig deilig ting. Jeg tror det er et kjempeproblem hos de fleste firmaer at sånne ting går alt for lite strømlinjeforma”.

Selv om Linked Enterprise Data vil kunne være veldig nyttig, mener Pernille at man skal være litt forsiktig med å se det som den endelige løsningen. For selv om den strukturerer dataene internt, og tar bruk av Linked Data prinsippene, utelukker den åpenheten rundt definisjonene. Hun synes det er synd hvis dette er retningen man skal gå i, fordi en slik begrenset bruk av Linked Data ikke vil bidra til fellesskapet. Pål forteller at på sikt så er det meningen av Skatteetaten, Brønnøysund, NAV og alle etater skal ha tilgang til data som de må ha fra de steder de skal ha det fra. Hvis en nå begynner på virksomhetsnivå med organisasjonsdata på intranett, så løses ikke problemet med å åpne det opp mot andre organisasjoner. Han tror det er et logisk neste skritt, men støtter Pernille i at vi må tenke lenger enn at dette er den ønskelige løsningen. Pernille ser for seg at man kombinerer tilnærmingene slik at Linked Enterprise Data brukes inne i organisasjonen eller innenfor en etat til en viss grad, og så kombineres dataene med en Linked Open Data struktur som frigjør dataene eksternt også. Hun foreslår at man skal tenke på dataene som en underklasse slik at enkelte deler vil være mer nyttige som Linked Enterprise Data enn som Linked Open Data.

Det vil kunne sørge for at forretningskritiske og personsensitive data ikke tilgjengeliggjøres.

Brage er ikke direkte kjent med Linked Enterprise Data, men mener at en slik begrenset tilgjengeliggjøring av offentlige data vil kunne være et neste steg for å komme i gang med Linked Open Data. Å ta i bruk Linked Data--prinsippene internt og så evaluere intern gevinst vil mulig bidra til en trinnvis modning ved at man ser at bruk av disse prinsippene er verdifullt. Han tror en slik tilnærming kan fungere i de miljøene som er interessert i å ha bedre struktur på dataene sine, men svært skeptiske til å tilgjengeliggjøre dataene sine for hele omverdenen.

4.3.2 Hva har blitt gjort

Erfaringer

Petter forteller at bruken av Linked Open Data begrenses når det ikke finnes tilstrekkelig interessant informasjon å bruke. Han forklarer at det ikke er enkelt å se tydelig bruk av alle datakilder. Mens informasjon fra Wikipedia brukes hyppig, ser han ikke for seg hva slags bruk som skal komme av for eksempel Tjenestemannregistret de har tilgjengeliggjort. Det viktigste er at data er tilgjengelig for da vil de før eller siden bli brukt. Peder forteller at de brukte DBpedia som en kilde til tilleggsinformasjon på prosjekter der visse områder var mangelfulle. Pernille sier at de ønsket å gjenbruke eksisterende vokabular så mye som mulig. Måten de gjorde det på var ved å ta i bruk en "mix and match"-metode. Da tok de utgangspunkt i forskjellige vokabular, fremfor ett vokabular. Istedenfor å bruke hele vokabularet, gjenbrukte de kun definisjonene som passet for deres bruk. Det er viktig at folk som gjenbraker Linked Open Data ikke tilpasser definisjoner til andre formål enn de egentlig er tiltenkt. Pernille har erfart at vokabularene enten var ufullstendige, eller så var de så møysommelige at de ble for presise for deres bruk. I tilfeller der definisjonene ikke passer, anbefaler hun at man lager en egen ontologi som man lenker til de dataene man gjenbraker:

"Da definerte vi egentlig de begrepene som vi hadde behov for i kombinasjon med eksisterende. Så vi refererte til Dublin Core. Vi refererte til Music Ontology og så refererte vi til vår egen. Så sånn bygget vi egentlig opp".

Pernille forteller at de ikke hadde automatiserte prosesser når det gjaldt å finne hva slags Linked Open Data som passet til deres prosjekter. De hadde en del kjennskap til noen ontologier og så undersøkte de forskjellige typer ontologier på nett, og prøvde å se litt på hva man hadde brukt i andre løsninger.

Peder forteller at mangel på vedlikehold på identifikatorene kan føre til utfordringer med å bruke Linked Open Data. Istedenfor å følge lenken til et nytt konsept får man en 404-NOT FOUND hvor lenkingen stopper opp. Videre snakker han om at mange offentlige datasett ikke er tydelige beskrevet slik at man mangler oversikt over hvor de matcher i forhold til allerede eksisterende datasett.

Pernille forteller hvordan en offentlig etat bør gå frem for å tilgjengeliggjøre Linked Open Data. Det første etaten må gjøre er å identifisere lignende prosjekter, og finne en leverandør som er tilstrekkelig kunnskapsrik til å bruke Linked Open Data. Deretter må man få en oversikt over behovene, hvilke deler av organisasjonen som skal defineres, og så tilgjengeliggjøres de. Etter hvert bør det lages en RDF versjon av dataene med URIs og tilknytning til en ontologi med begrepsdefinisjoner.

4.3.3 Forutsetninger for at offentlige organisasjoner tilgjengeliggjør Linked Open Data

Øke forståelsen av Linked Open Data

Pernille forteller at omgjøring til Linked Open Data forutsetter at dataene er åpent tilgjengelige og det er derfor avgjørende at ulike organisasjoner tilgjengeliggjør dataene sine. Data kan tilgjengeliggjøres i RDF, og være klare for å lenkes til andre data, men dersom det ikke finnes noen tilgjengelige data, begrenser det selvfølgelig utviklingen av en lenket dataev. For at offentlige organisasjoner begynner å tilgjengeliggjøre data som Linked Open Data, må de modnes på forståelsen rundt Linked Open Data. Petter forteller at interessen er tilstede, selv om bevisstheten rundt bruk av formater mangler. Han jobber med data.norge.no og forteller at oppfatningen rundt dette er ganske lav. Det er lite fokus på formater og heller fokus på tilgjengeliggjøring av strukturerte data som det første fokuset. Pernille er enig i at det kreves mer kunnskap om hva Linked Open Data dreier seg om. Fordi Linked Open Data er en annen måte å tenke på, krever det modning og tid. Istedenfor å tenke tjenesteorientert arkitektur (SOA) med utbygging av flere programmeringsgrensesnitt (API), handler Linked Open Data om grunnlagsdataene. Leverandører må forstå hvorfor eksisterende teknologier hindrer tilgjengeliggjøring av data. Ved å fortsette å bruke vevtjenester og APIer som tilgangspunkter til grunndataene, skapes det regulerte bruksmønstre som begrenser tilgangen og gjenbruket av dataene. De som ønsker å bruke dataene må tilpasse seg etter dataeier sine begrensinger for dataene. Slike en -til-en forbindelser støtter ikke fleksibelt gjenbruk og deling av data. Didrik sier at mange ønsker å fortsette med tradisjonelle teknologier fordi de har en utviklingsstrategi på de eksisterende systemene som gjør at slike teknologier passer bedre inn i strategiene og i det utviklingsløpet de har laget for sine tjenester. Petter opplever offentlig sektor som ganske blandet når det gjelder hvor modne de er, men syns at de er på god vei:

”De siste tre årene har det vært kjørt et par forskningsprosjekter hvor bevisstheten rundt Linked Open Data har økt betraktelig og da snakker jeg om de organisasjonene i offentlig sektor som er modne. Sånne som SSB som forstår det her. Det kan jeg ikke si om for eksempel noen mindre kommuner eller for den saks skyld, etater som har mye mer fokus på det som går på planlagt samhandling mellom to aktører”.

I Semicolon II-prosjektet (Semicolon, 2011) er en stor del av etatene med som prosjektkunder eller eiere med hensikt om å teste ut Linked Open Data på egne systemer og data. Petter forteller at Linked Open Data er et av hvert fall tre temaer som kommer til å være sentralt over de neste tre år for de etatene. Han erfarer at etatene må få demonstrert effektene av Linked Open Data, før de ønsker å gå i den retningen. I tillegg må det en kritisk masse på plass, samt at det kortsiktige ambisjonsnivået må heves. Pål tror det vil være hensiktsmessig med slike veiledninger når det gjelder å skulle tilgjengeliggjøre Linked Open Data. Når det snakkes om offentlige data, så er det ikke sånn at alle institusjoner i det offentlige rom bare publiserer Linked Open Data sånn uten videre. Han oppfordrer til en dugnadsaksjon for å få i gang Linked Open Data. Alle må tilgjengeliggjøre litt hver. Scott mener det trengs en bevisstgjøring hos de forskjellige etatene ettersom teknologiene begynner å bli modne nok til å brukes. Han forteller at det ikke er noe i veien for at organisasjoner tilgjengeliggjør dataene sine som RDF og OWL. For å få gjennomslagskraft til å gå i denne retningen, tror han det er viktig at det holdes fokus på Linked Open Data og at det kommer påtrykk fra brukermassene ute slik at folk på høyere nivå i organisasjonene forstår viktigheten. Pontus mener at DIFI må se Linked Open Data som en sentral standard før andre etater tar disse i bruk. For at folk skal begynne å bruke standardene for Linked Open Data fremfor tradisjonelle vevformater, ser Pål

for seg at Linked Open Data må bli generert automatisk gjennom standardverktøy som alle bruker. Han trekker frem Drupal som et eksempel på et slikt rammeverk som lar brukeren eksportere RDFa.

Forbedre evnen til å definere begreper

Petter mener litt av problemet med den åpne uplanlagte samhandlingsverdenen, som Linked Open Data representerer, er at beskrivelsene og semantikken ved dataene ikke er tydelige nok. Når åpne data skal tilgjengeliggjøres for hele verden må datatilbyder være mer eksplisitt på hva forskjellige felter betyr. Prinsippene for hvordan det skal gjøres er ikke særlig tydelige. Han mener det er få organisasjoner i offentlig sektor som har god kontroll på sitt begrepsapparat. Også Brage ser at evnen til å ha beskrivelser av sin egen informasjon er variabel. Noen har godt definerte begreper, mens andre mangler beskrivelser fullstendig. Han lurer på hvordan en skal greie å lenke data til andres data når en ikke har kontroll over sine egne data. Han trekker frem SSB som et eksempel på godt definisjonsarbeid ved deres åpne metadatabase. Scott forteller at det er utfordrende å få lenket dataene opp med beskrivelser og riktig bruk av identifikatorer. Det er vanskelig å spesifisere hva dataene betyr og hvilken relasjon de har til hverandre. Brage er opptatt av at begrepsapparatene skal kunne bli maskinlesbare. Man kan ikke forvente at etater manuelt skal søke gjennom hverandres nettsider for å prøve å finne ut hvordan andre definerer begrepene sine. Han mener forskjellige aktører bør ha en lik måte å beskrive dataene sine på, og savner en enighet rundt metodikk for hvordan man skal få dette til. Han etterspør en standard å følge som gjør at organisasjoner beskriver dataene sine på samme vis. Mange ville hatt behov for en gitt måte å definere dataene sine på. Sander støtter Brage når han sier at måten dataene lenkes til hverandre må standardiseres. For at lenkingen skal gi mening er det nødvendig med en standardisert lenking på tvers av de dataene man har. Pernille opplever at det er mange som frykter de vil bruke mye tid på å definere begreper dersom de skal velge Linked Open Data, og fortsetter derfor isteden med eksisterende teknologier.

Håndtere tilgangsstyring

Petter tror Linked Open Data ville vært nyttig i offentlig sektor, men at man skal legge et lite parentes rundt o' en i Linked Open Data. Det er ikke alltid gitt at den informasjonen som snakkes om er åpen. Pål forteller om utfordringen med å skille mellom hva som skal tilgjengeliggjøres og hva som skal holdes tilbake:

”Deler av dataene Skatteetaten har, for eksempel Skattelistene, kan ligge åpent. Så er det en del som ikke er åpne, men åpne for eksempelvis NAV. Så er det en del som ikke er åpne for NAV heller, som bare er åpne internt mellom organisasjonsdeler. Og så er det en del som ikke er åpne, men bare tilgjengelig for de ansatte som trenger det. Og der ligger de sensitive personsopplysningene. Der ser du allerede fire-fem tilgangsnivåer på en sånn graf. Teknisk sett er det mulig å lage den grafen for Skattedirektoratet og gi dem alle disse gevinstene, men enn så lenge er det ikke mulig å skille mellom det som skal ut og ikke ut”.

Han spør seg om en bør ha tilgangskontroll på trippelnivå eller hva som vil være den beste løsningen. Han beskriver et eksempel som demonstrerer utfordringen:

”Jeg tok det prosjektarkivet som inneholder femten tusen prosjekter fra de siste ti årene i Norge. Når jeg begynte å generere RDF-tripler fikk jeg 800 000 tripler minst. Skal en legge en tilgangskontroll på hver eneste av de, eller på ontologier, så blir det en kompleksitet når man skal resonnerer etterpå. Når noen sender inn en SPARQL-spørring med et passord og et tilgangsnivå, så må jeg gå gjennom alle disse triplene og bestemme hvilke de kan få se på. Det blir forsinkelse gjennom hele systemet. Så her ligger det utfordringer”.

Den løsningen han mener det bør forskes på er å kunne konfigurere tilgangskontroll dynamisk. Didrik forteller at mange som sitter i DIFI sin arbeidsgruppe for formidling, ønsker å bruke vevtjenester som tilgangsmetoder for sine data. Det er fordi de føler de har bedre

kontroll med slike teknologier. De ser utfordringer rundt tilgangskontroll ved bruk av Linked Open Data. Fordi det er mye enkelttoppslag som gjøres på disse dataene, føler de at rettighetshåndtering og den type ting er enklere å håndtere med vevtjenester fremfor Linked Open Data. Selv om dataene kan være åpne, er tilgang til hele materialet begrenset. Pernille mener slike konkurrerende teknologier gjør det vanskelig å bruke Linked Open Data. Fordi mange føler mer trygghet i å jobbe med SOA, lukkede tjenester og XSDer, er det en forutsetning at Linked Open Data bedre ivaretar sikkerhet og tilgangsstyring.

Ivareta dataenes opprinnelige kontekst

Flere av informantene forteller at den konteksten dataene brukes i må ivaretas for at de ulike organisasjonene er i stand til å gjenbruke hverandres data. Når data tilgjengeliggjøres åpent, har datatilbydere utfordringer med å forvalte dataenes kontekst slik at dataene tolkes korrekt. Didrik mener litt av problemet med Linked Open Data er at en del av konteksten forsvinner. Han sier at man prøver å gjøre dette kontekstuevhengig, og at det blir problematisk fordi mye av den begrepsmessige informasjon i offentlig sektor er knyttet til en gitt kontekst. Han nevner utfordringen ved begrepet samboer. Skal det sees i konteksten samboer for stønad i NAV eller inntekt for Skattedirektoratet? Unn forteller at når UDI skal ha informasjon fra Skattedirektoratet, så må de sørge for at de får den samme konteksten som Skattedirektoratet bruker. Hun sier at det er først nå som dataene utveksles mellom etater at denne konteksten blir utfordrende. Når hver organisasjon tolker begrepet ut i fra sitt begrepsapparat, fører det til at de forskjellige begrepene har en forskjellig kontekst avhengig av hvor de blir benyttet. Det er derfor avgjørende at konteksten ivaretas, og at dataene tilgjengeliggjøres med et begrepsapparat. Brage mener det som har blitt gjort rundt Linked Open Data til nå, har tatt for lett på kulturelle behov. For at man skal overvinne skepsisen folk har tilknyttet bruk av andres data, må man etter hvert ha som ambisjon å tilgjengeliggjøre den konteksten dataene blir brukt i, beskrivelser av hvor de blir brukt og hva de blir brukt til.

Sander forteller at et tall i en celle vil oftest ikke gi mening med mindre konteksten er tydelig. Han er usikker på om han synes konteksten forsvinner ved bruk av Linked Open Data. Han tror heller folk må modellere dataene bedre slik at konteksten ivaretas. For at data skal kunne tilgjengeliggjøres gjennom RDF og SPARQL-tilgangspunkter, må man være i stand til å tilby konteksten som dataene tilhører.

Sørge for gjenbrukspotensial

Pål forteller at han har mest tro på selvorganiseringen som skjer i Linked Open Data-skyen. Han forklarer at Linked Open Data-grafen sin selvorganisering automatisk fjerner de nodene og kantene som blir lite referert til. Det vil da si at de datasettene som blir lite gjenbrukt automatisk forsvinner etter en viss tid. Han forklarer at slik som Linked Open Data skyen fjerner datasett som ikke blir brukt, må de som tilgjengeliggjør data sørge for at dataene er enkle å bruke slik at de ikke forsvinner:

"Du kan planlegge å publisere en ontologi som Linked Open Data og et datasett, men du kan ikke planlegge at den blir relevant og brukt mye. Det er en diskusjon jeg hadde med NRK. De ville gjerne lage en ontologi, og har jobbet med en metadatamodell på musikk i mange år. De har en veldig nøye modell, men den er såpass komplisert at det er usannsynlig at folk flest begynner å bruke den for å beskrive musikk. Musicbrainz på den andre siden, har modeller som er mye flatere og som er mye enklere å bruke. Men, som er mangelfulle på mange områder som klassisk musikk og sånt. Men, de begynner å bli brukt veldig mye av ungdommer og folk som utvikler iPhone applikasjoner og Android applikasjoner. Og da blir den brukt mest ikke sant og til slutt så må NRK forholde seg til Musicbrainz uansett. Da er spørsmålet, er det da ikke lurt å ta slike ting som utgangspunkt? Det prinsippet om å gjøre det enkelt i begynnelsen og se hvor du lander og så utvide etter hvert? Du lager forskjellige versjoner som kan knyttes sammen til en helhet som stemmer, men som også har en lightversjon som kan brukes av mange. Og da kommer du igjen i den diskusjonen om dette skal organisere seg selv basert på popularitet og bruk. Da blir det veldig spennende å se hvordan staten og disse organisasjonene tar det opp. På noen områder i det offentlige, som med SSB og Brønnøysund, må man være nøye og

da er kompliserte modeller unngåelig. Så de kommer sikkert til å ha sin verdi og autoritet, men nettopp på NRK og musikk er jeg veldig mye mindre sikker. Det er så mange som kan litt om musikk ikke sant. Det kan bli veldig spennende å se”.

4.3.4 Forutsetninger for at offentlige organisasjoner gjenbraker Linked Open Data

Villige til å gjenbrake andres data

Peder ser ikke noe i veien for å begynne å bruke Linked Open Data. Det forutsetter bare at man må være villig til å gjenbrake en del av de identifikatorene som allerede er tilgjengeliggjort. Selv om ikke alt er optimalt enda, mener han at ting kan gjenbrakes. Pål forteller at offentlig sektor ikke er vant til å bruke data som andre har laget.

Informasjonsmodell

Sander sier en informasjonsmodell er nødvendig for at SSB kan begynne å bruke Linked Open Data. Selv om man bruker tripler til å beskrive dataene, vil det ikke automatisk gi forståelse av dataene. Han sier at en må ha på plass en informasjonsmodell før semantikken gir noen mening:

”Du er på mange måter like langt da. Jeg har jobbet en del med semantisk interoperabilitet tidligere og hver gang stopper det opp fordi vi må ha en informasjonsmodell. Om den er felles eller ikke, det betyr ikke så mye, men du må ha den hvert fall”.

Pernille ser på bruken av en informasjonsmodell som et kart for hvordan du skal lage systemet ditt. Hun sier at et godt definert begrepsapparat gjør at informasjonsmodellen fungerer.

Autoritetsregistre

For at Linked Open Data kan fungere optimalt og rette lenker til de samme autoritative kildene, forteller Pål at det må på plass autoritetsregistre. Han ser for seg et register bestående av beskrevne URIer som lenker til ulike autoritative kilder. Det de jobber med nå er å se på muligheter for å lage felles autoritetsbeskrivelser av personer, organisasjoner, hendelser og steder, slik at utviklere har felles URIer å referere til.

”Vi må få et register på offentlige personer som er referert i pressen, bøker og sånne ting som er viktige i historien til Norge, sånn at vi kan referere til dem på et sentralt sted. Hvis jeg skal referere til Dag Frøland, hvor ligger den? Hvor skal jeg referere? Enten den norske Wikipedia eller Nasjonalbiblioteket eller hvor som helst. Da hadde vi hatt noe som er veldig bra. Og det samme gjelder steder. Utsteder og konsertsteder skal vi også ha en beskrivelse på slik at vi kan referere til dem” .

Han poengterer også at det må arbeides med hvem som skal publisere slike lenker etter hvert som registrene kommer på plass. Scott etterspør metadata om hvem som eier dataene og hvem som er autoritativ kilde. Pål mener det bør utføres en behovsanalyse der det kommer frem hvilke data en trenger åpent for å komme i gang med Linked Open Data.

Flere ressurser

Sander forteller at spørsmålet om hvorvidt de skal gå over til Linked Open Data som tilgjengeliggjøringsform handler i bunn og grunn om tid og ressurser: *”Det må prioriteres. Det er egentlig dét det koker ned til. Vi har ingen motvilje mot det i det hele tatt. Snarere tvert imot”.* Siri mener det finnes mange gode ideer her i verden, men at de mangler ressurser til å gjennomføre dem: *”Alle våre utviklere er ganske opptatt med for eksempel å lage ny vevside for hele SSB”.* I tillegg kreves det prioriteringer fra forretningsiden.

4.4 Generelle utfordringer tilknyttet lenkede åpne offentlige data

4.4.1 Tilknytning til lovverket

Scott sier at begrepene som brukes i skattedomenet må brukes riktig fordi de er knyttet til lovverket. Lovverket kan være et hinder når det gjelder å skape felles forståelse. Når det finnes så mange forskjellige begreper, for eksempel samboer, er det uten presisering i lovverket svært vanskelig å vite hvilket begrep en skal benytte. Didrik forteller at tradisjonell forvaltning har vært å behandle informasjonen litt forsiktig. Data tilgjengeliggjøres ikke til noen som ikke har hjemmel for å ha dataene. Pernille nevner at dataenes tilknyttede lovhemler gjør tilgjengeliggjøringen utfordrende. Også Unn legger vekt på lovverkets innflytelse på offentlige data. Fordi dataene er offentlige og hjemlet i lovverket må de knyttes til presise juridiske koder. Den juridiske tilknytningen gjør det enda viktigere å vite hvor dataene kommer fra:

”Når du tenker på offentlige data, så er det mye data som er resultat av en saksbehandling. Det vil si at du har knyttet data til noe lovverk, til noen hjemler, til noen retningslinjer, til noen forskrifter. Og da må vi ha tunga rett i munn og vite at vi snakker om det samme begrepet. Ja, offentlige data er ofte.. hvert fall sånn som har med person å gjøre er, et resultat av en tolkning”.

Mange av begrepene UDI bruker, er i en juridisk kontekst som gjør det nødvendig med begrepsapparater for å unngå juridiske feiltolkninger. Sander forklarer at SSB tilbyr mikrodata for forskere, men at man havner midt inne i det juridiske området med en gang en skal begynne å tilby data på det nivået. Brage forklarer at problemstillinger rundt åpne data er veldig ofte knyttet til et lovverk. Det gjør at juristene blir involvert enten ved at de er en del av driftsavdelingen, eller støtter opp under ulike fagmiljø. Deres jobb blir ofte å utrede om konsekvenssiden. Brage sier at han opplever juristene som veldig skeptiske til alt som har med tilgjengeliggjøring av data å gjøre. De ser mange trusler ved slik tilgjengeliggjøring. Lovverket er i enkelte tilfeller med på å videreføre en konservativ holdning mot å tilgjengeliggjøre åpne data. Fordi lovverket ble skrevet før elektronisk distribusjon av data var et tema, må man av og til søke om lovendringer før man for lov til å distribuere informasjon til andre. Når organisasjoner ikke ønsker at andre skal ha tilgang til deres data, finner de paragrafer i lovverket som fjerner muligheten for å tilgjengeliggjøre. Pål mener tilgjengelighet av data står og faller på det juridiske og organisatoriske aspektet i offentlig sektor. Dersom data er tilgjengelig tror han at de fleste vil gå i denne retningen rimelig kjapt. Men, dersom data er lite tilgjengelig med rare lisensmodeller vil det være en stopper for utviklingen av Linked Open Data.

4.4.2 Tap av statlig inntekt

Statens forretningsmodeller er noe av det som sees som en åpenbar hindring for tilgjengeliggjøring av offentlige data. I samme regjering er det ulike syn på om staten skal ta betalt for å tilgjengeliggjøre åpne data. Selv om mange data er åpent tilgjengelig, forteller Brage at åpenheten rundt data blir begrenset fordi staten tar seg betalt gjennom private distributører. Brønnøysundregistrene genererer hundrevis av millioner i forskjellige typer gebyrer ved å tilgjengeliggjøre offentlig informasjon. Han spør seg hvorfor BRREG som råvarekilde skal være nødt til å kanalisere deres informasjon gjennom et fordyrende mellomledd, og ikke bare legge data rett ut. Brage forteller at diskusjonen med å gi ut dataene gratis er allment for alle departementer, men at de i forskjellig grad har ulike prismodeller som sikrer inntektssiden. I de fleste tilfeller ligger det en forretningsmodell knyttet til distribusjon av informasjon fra store offentlige register. Den forretningsmodellen er det noen som har ansvar for å ivareta og håndheve. Didrik forteller at hvis den inntekten forsvinner, så må den kompenseres på en eller annen måte. Enten gjennom økte skatter eller ved at man reduserer tjenestenivået i offentlig sektor. Fordi det er vanskelig å se når staten vil tjene penger som en konsekvens av at offentlige organisasjoner tilgjengeliggjør data, er det vanskelig å frigi data åpent. Han tror Finansdepartementet vil være tilbakeholdne nettopp fordi en tilgjengeliggjøring av åpne data vil redusere det offentliges inntekter direkte. Scott uttrykker at det ikke bare er enkelt å frigi data uten noen form for kostnad til de som ønsker å

bruke dataene. Han forklarer hvorfor Kartverket ikke ønsker å tilgjengeliggjøre sine kartdata gratis:

”Det pågår ganske store avtaleverk mellom interessentene hvor det faktisk er ganske masse penger involvert. Det gjør at disse dataene har denne kvaliteten og er så gode som de er. Så det vil si at hvis de åpner opp alle disse dataene, så faller jo dette avtaleverket i fisk og da må man finne andre modeller for å sørge for at dataene har den kvaliteten og er så gode som de er. Det er vel det som er en del av problemet for Kartverket”.

Pål forklarer at Kartverket som en offentlig organisasjon ikke tilgjengeliggjør sine data på grunn av kommersielle interesser. Han ser Kartverkets fremtid som et av to alternativer. Enten vil de fortsette å ta betalt for kartdataene, eller så vil det komme et initiativ som ”Open Street Map” som langsomt vil overta og bortmarginalisere deres kartvirksomhet. Når organisasjoner over tid har generert inntekt ved å selge data, må de finne andre inntektsmodeller dersom de skal tilgjengeliggjøre dataene gratis. Dette vil være en nødvendig forutsetning for i det hele tatt å kunne tilgjengeliggjøre data. Pål trekker frem et eksempel fra NRK der de har fått inn krav fra staten om å ha en egen inntekt. NRK spør om høyere lisenser, men hvis de ikke får det sier departementet at da må NRK finne andre måter å tjene penger på. De får ikke være kommersielle.

”Der sitter NRK i en skvis. Da ser de på sine databaser og så sier de ja, men nå har vi masse musikk og det er vår ressurs. Vi har mer musikk enn de fleste, så vi lager en egen base på det og så skal vi selge det”.

Didrik tydeliggjør at hovedinsentivet ved å tilgjengeliggjøre åpne data ikke bør være økte inntekter. I hvert fall ikke fra den enkelte virksomhet. Brage forstår ikke hvordan departementene kan si at alle skal åpne opp sin informasjon, men samtidig så har de klare forretningsmessige begrensninger i forhold til å gjøre det. Tilgjengeliggjøres data gebyrfritt, går det utover inntektssida til staten. Han mener det paradokset må delvis overvinnes før offentlige organisasjoner begynner å tilgjengeliggjøre data.

4.4.3 Politiske styringssignaler om å tilgjengeliggjøre åpne data

Brage forteller at beslutningen om å få lov til å tilgjengeliggjøre data i offentlig sektor ligger på politisk nivå. Viljen og interessen fra organisasjoner og personer rundt omkring i offentlig sektor til å tilgjengeliggjøre åpne data er ikke er nok. Brage forteller at det finnes ildsjeler som brenner for saken, men at det må være enda tyngre og tydelige signaler fra regjeringsnivå for at en slik tilgjengeliggjøring skal prioriteres. På den ene siden, hos FAD, er åpne data en sentral sak og noe som FAD flagger veldig høyt. Til tross for dette mener Brage at verken FAD eller DIFI er tunge nok som fagdepartement til å fatte tydelige krav. Parallelt sitter de andre departementene og har retningsmodeller som tilsier at driften av de ulike etatene er avhengig av inntektene som kommer gjennom mellomdistributører. Brønnøysundregistrene må gå til Næringsdepartementet og ha en dialog om nødvendige endringer som må gjøres før de kan legge ut informasjonen åpent. Han tror forståelsen omkring forretningsdimensjonen må selges inn på mot regjeringsnivå slik at de kan fatte et tydelig politisk krav om at det her skal skje. Både Pernille og Petter mener politiske årsaker har hindret tilgjengeliggjøring av åpne data.

4.4.4 Begrensede midler til å utforske nye teknologier

Pernille mener det fortsatt mangler sterke påstander for den økonomiske besparelsen man gjør ved å bruke semantiske vevteknologier og Linked Open Data fremfor tradisjonelle teknologier. Mye av grunnen til at offentlige organisasjoner sliter med å utforske nye teknologier, er den årlige utdelingen av midler gjennom statsbudsjettet. Fordi finansieringen kommer årlig er vanskelig for offentlige organisasjoner å planlegge langt frem i tid. Offentlige organisasjon har et visst behov der og da som de får midler til å dekke. For å få midler til å utforske semantiske vevteknologier og Linked Open Data, må det fremmes et behov og innvilges midler til det. En annen ting hun ser som nødvendig er at etatene får

avsatte midler for å tilgjengeliggjøre data sine. Petter mener det finnes økonomiske årsaker til manglende tilgjengeliggjøring. Brage mener det må diskuteres fram til hva det må suppleres med for at Linked Open Data i det offentlige får mest mulig samfunnsgevinst:

”Hvis man sier at Linked Open Data potensielt har så mange milliarder i verdi, så består den x antall milliarder av en del som er knyttet til det med å lage ulike produkter for både oss borgere og næringsliv. En del av de milliardene blir knyttet til det å rasjonalisere og effektivisere informasjonsflyt innad i offentlig sektor. Og en del av de milliardene går til å kunne rasjonalisere prosessene ute i næringsliv og andre plasser. Det finnes mange ledd i denne bevilgningen. For å få gevinst på alle de områdene der da er spørsmålet; hva kreves for å få til det?”

4.4.5 Bedre informasjonsforvaltning må på plass før åpne data kan gjenbrukes i det offentlige

Brage mener det må på plass bedre strukturer og beskrivelser av informasjonen før de store offentlige etatene vil begynne å bruke andres tilgjengeliggjorte data. Også Scott mener at åpne data uten metadatabeskrivelser blir lite gjenbrukbare. Hvis en ikke er fullstendig trygg på hvor dataene kommer fra, og til hvilket formål de kan brukes, senker det gjenbrukmulighetene. Brage forteller at generelle beskrivelser på mer eller mindre ustrukturert måte ikke gir den kvaliteten som kreves hvis de offentlige aktørene skal gjenbruke informasjonen seg i mellom. Han mener det må finnes veldig tydelige eierskapsbeskrivelser samt forvaltningsbeskrivelser rundt informasjonen. Disse dataene må inneholde informasjon om dato, oppdateringer, etableringsdato, samt dataenes gyldighet. Han forteller at det dreier seg om en tilknyttet informasjonsforvaltning. For at noen skal tørre å gjenbruke dataene, så må de være sikre på strukturene og de må kunne skjønne hvordan dataene har oppstått. Man må være sikker på at forvaltningssiden blir ivaretatt og at dataene har jevnlig kvalitetsoppdateringer. Han mener at det er slik informasjonsforvaltning som vil gjøre at dataene kan brukes i flere sammenhenger enn til utvikling av iPhone-applikasjoner. Han ønsker at disse beskrivelsene gjøres på en metodisk sammenfallende måte som gjør det enklere å fange dem og samle den type beskrivelser av dataene i systemet. Det må komme fram hva slags modeller og prosesser dataene er tilknyttet. Scott etterspør metadata knyttet opp til når informasjonen ble laget og hvilken versjon det er. Han etterspør standardiserte felter som sier noe om hvem som er ansvarlig for dataene, når de ble laget, hva de inneholder, hva hovedtemaet er og hvor de finnes. I den nasjonale strategien for metadata jobber de blant annet med hvordan man skal beskrive dataene, hvordan dataene skal utveksles, og hvilke felter en trenger for å søke opp informasjon på data.norge.no. Scott sier det er nødvendig med en standard som beskriver feltene man bør ha:

”Vi kommer ikke så mye lengre enn det som går på definisjoner, informasjonsmodeller og den semantiske forståelsen eller betydningen av ting. På det nivået, men vi trenger også det. Også trenger man for data.norge.no at man har en enhetlig måte å beskrive referansedata på et datasett og tjenester ikke minst. Det er mulig en burde hatt en slags nasjonal meta- og datamodell, men det er ikke så lett å bli enige om den”.

For å sikre gjenbrukbarhet av de åpne dataene som tilgjengeliggjøres, ser han for seg en utvidelse av data.norge.no som legger til rette for at alle datasett tilgjengeliggjøres med tilhørende metadata. Ved at åpne metadata publiseres, kan man gå inn å se på andres informasjon og deres prosesser. Han nevner data i forhold til rettighetsinformasjon, administrativ informasjon, forvaltningsinformasjon og andre ting som man ønsker å knytte opp mot andre informasjonsdeler av de dataene. Siri peker på strukturering av informasjonsmengden som en viktig forutsetning for å kunne bruke åpne data. Hun sier det må finnes tilstrekkelig metadata så du kan vurdere kvaliteten ved de svarene du får:

”Det finnes uendelig mengder med data og legger vi alt ut i en portal så kommer du til å drukne i det. Det må legges tilrette for at du klarer å finne de dataene du er ute etter. At du får den kvaliteten du er ute etter. Ikke alle trenger samme kvalitet.

Det er mye kvalitetssikring ved tilgjengeliggjøring av metadata for å ikke gå i fallgruvene. Det er sikkert mye spennende man kunne finne hvis man slo sammen feil data på feil grunnlag. Men, der vil det bare bli helt teit og den dataportalen vil få dårlig rykte på seg”.

Siri forteller videre at de i SSB tilgjengeliggjør metadata i tillegg til instansdata slik at andre kan gjenbruke de publiserte dataene. Pål ønsker en mer synlig informasjonsforvaltning, slik at de som ønsker å gjenbruke dataene kan få innsyn i hvordan dataeierne behandler dataene. Når folk har gjenbrukt ulike datasett med semantiske vevteknologier, kan man gjøre det usynlig hvor dataene kommer fra. Hvordan skal en kontrollere at alle lisenser er i orden? Pontus mener den tekniske diskusjonen rundt det å bruke Linked Open Data er interessant å starte med, men at den bare utgjør en liten del av det totale bildet. Han er opptatt av at informasjonsforvaltningen og poengterer at det er mange flere spørsmål å ta hensyn til enn om datasettene er beskrevet i RDF eller OWL.

”For at du skal være trygg på at du faktisk kan bruke de dataene jeg har til det du ønsker, så bør du være opptatt av hvordan oppstod mine data, hva er kvaliteten på de, hva er de juridiske betingelsene for at du kan viderebruke de? Hvilke data har jeg om de ulike kildene som gjør at jeg kan stole på de i forbindelse med den oppgaven jeg skal ha løst? Hvilken informasjonsforvaltning er kilde X underlagt? Hva er kvalitetskriteriene de måler sine data på? Hvor sikker er jeg på at kilde X vil være der de neste årene hvis jeg skal basere min virksomhet på den?”

4.4.6 Spesialister til å velge datasett

Pontus forklarer at det er vanskelig å trekke noe helt sikkert ut av distribuerte datakilder fra forskjellige informasjonsregimer. For hver ny kilde en trekker inn i en sånn informasjonskube, så øker usikkerheten om det en skal analysere. Han synes Linked Open Data-skyen er kjempespennende, men tror at det vil komme opp et hav av misforståtte konklusjoner etter hvert. Han mener det må på plass en infrastruktur for hvordan man skal bruke åpne data. Det må bli bedre mekanismer som sikrer at dataene har tilstrekkelig gjenbrukskvalitet. Dårlig kvalitet på de åpne dataene vil føre til problemer for de som gjenbraker dem. Han spår at offentlig sektor vil tilgjengeliggjøre flere titusener av datasett, og mener det vil bli vanskelig å vite hvilke datasett man skal velge når man skal løse en oppgave. Når folk har manglende forutsetninger for å sette sammen datakilder, vil det fort kunne føre til at datakilder som hver for seg er meget bra, vil bli helt feil når de sammenstilles. Den tekniske dimensjonen er løsbart ettersom mange sitter med kunnskap om hvordan ulike datasett kan kombineres rent teknisk. Men, det må fokuseres mer på de semantiske, organisatoriske og juridiske dimensjonene ved bruk av åpne data. Uten spesialister til å gjøre denne jobben, frykter han at tolkningen og bruken av de åpne dataene blir helt feil:

”Hvilken av disse dataene er de beste for mitt formål? Jeg trenger hjelp til å finne ut hvilke data jeg skal bruke for å gjøre min jobb. Når jeg vet hvilke datasett jeg skal bruke, så trenger jeg hjelp til å sette de sammen på en sånn måte at jeg er sikker på at summen av disse datasettene gir meg verdifull informasjon og at jeg ikke ødelegger grunnlaget for gode resultater. Hvis jeg sitter på en rolle i en offentlig etat eller i næringslivet og skal ha opp en tjeneste, eller trenger svar på noe for å ta en forretningsmessig beslutning så har ikke jeg tid til å sette meg inn i titusener av datasett. Da vil jeg gå til de spesialistene som er best på å sette dette sammen, og så få rådgivning”.

Han mener teknisk tilrettelegger for å få tak i mangfold av datakilder, roller for å hjelpe med semantisk tilrettelegging og rolle med å velge ut de riktige datakildene må på plass før Linked Open Data kan brukes i offentlig sektor. Også Peder mener det bør forskes mer på hvordan etatene kan vite hva slags datasett som passer til deres bruk. Pontus forteller at han ikke kan bruke eksempelvis Folkeregisteret som basis for sin virksomhet dersom han ikke er sikker på at dataene er korrekte. Han har vanskeligheter for å se for seg hvordan politikere og Norges Bank skal potensielt kunne bruke åpne data til å styre samfunnet uten en slik kvalitetssikring.

Scott er avhengig av en viss kvalitetssikring for å kunne være trygg på at han tar riktige beslutninger.

4.4.7 Personvern

Pål mener at de store mengdene med personsensitiv informasjon vil kunne være en hindring for å tilgjengeliggjøre åpne data. I DIFI legger de ikke ut personidentifiserbar informasjon. Petter forteller at alle aktiviteter rundt Linked Open Data, hvertfall i Norge, prøver å unngå personinformasjon. Det er veldig lett for å komme i konflikt med personvernloven dersom man tilgjengeliggjør slike data. Siri forklarer at SSB er forsiktige med å legge ut opplysninger som går på enkeltperson:

”Jeg tror ikke du hadde ønsket at SSB hadde lagt ut alt de vet om deg på deres hjemmeside. Da tror jeg du hadde mistet tillit til oss, og nektet å svare på neste undersøkelse. Vi har samlet inn din inntekt og sagt til deg at vi lover å ikke si den til noen andre, men kun bruke opplysningene til statistisk formål. Da har vi inngått en avtale med deg som vi er forpliktet til å holde”.

I UDI holdes mye data tilbake fordi de er låst av personvern. Ulike menneskeskjebner og deres personlige tragedier er ikke informasjon man ønsker å synliggjøre. Unn gir uttrykk for at UDI er en veldig lukket organisasjon. Hun sier det også er store juridiske aspekter rundt det med deling av personinformasjon innen det offentlige. Pernille tror det er en skepsis i offentlig sektor for å tilgjengeliggjøre data på grunn av personvern.

5. DISKUSJON

5.1 Beste praksis

Beskrivelsen av beste praksis er tredelt, og vil gi anbefalinger for tre faser ved åpne data. Den begynner med råd å følge før dataene skal tilgjengeliggjøres, deretter anbefalinger for valg av formater og avveininger ved tilgjengeliggjøring, og avslutter med ting å ta hensyn til når etatene skal ta i bruk åpne data som er gjort tilgjengelige. Prinsippene oppsummeres i tabell 6.

Før tilgjengeliggjøring	Tilgjengeliggjøring	Etter tilgjengeliggjøring
<ul style="list-style-type: none"> • Forstå organisasjonskulturen • Demonstrer nytteverdi • Skap forankring i ledelsen • Politiske styringssignaler • Andre inntektsmodeller • Hvilke data ønskes • Avklar juridiske betingelser • Sett sammen riktig kompetanse • Se etter lignende prosjekter • Bedre mekanismer for tilgangsstyring • RDF bør genereres automatisk • Lenk de interne dataene 	<ul style="list-style-type: none"> • Engasjer teknisk kompetanse • Skap bedre begrepsapparater • Gi dataene URIer • Bruk RDF som datamodell • Bruk Turtle som RDF-notasjon • Spesifiser mening med RDFS • Gi tilgang via SPARQL • Gjør begrepsapparatet tilgjengelig • Gjør dataene enkle å gjenbruke • Gjør mest mulig tilgjengelig 	<ul style="list-style-type: none"> • Økte midler til etatene • Bedre informasjonsforvaltning • Spesialister finner datasett • Bedre verktøy • Etablere kritisk masse • Bruk begreper korrekt

Tabell 6: Oppsummering av beste praksis

5.1.1 Før tilgjengeliggjøring

Forstå organisasjonskulturen

Kartlegg organisasjonskultur og hindringer som finnes internt. Mange organisasjoner innehar en kultur som tilsier at gjenbruk uten restriksjoner gir negative konsekvenser for datatilbyder. De blir tilbakeholdne når de ikke lenger har kontroll over hvordan dataene benyttes. Mange ledere bruker informasjon som makt og er redd for å miste denne konkurransefordelen ved å tilgjengeliggjøre data. Del kunnskap om hva det vil si å tilgjengeliggjøre dataene. Det er ikke alle offentlige organisasjoner som forstår at å frigjøre data kan være verdiskapende ved at organisasjonen som datatilbyder blir andres primærkilde. Det har også vist seg at tilgjengeliggjøring av offentlige data kan føre til økt jobbglede og trivsel hos de ansatte. De opplever jobben sin som mer givende når dataene de produserer blir nyttige for andre.

Huijboom og Van den Broek (2011) omtaler den lukkede kulturen i offentlig sektor som den sterkeste hindringen for å tilgjengeliggjøre åpne data. Myndighetene ønsker ikke å dele data fordi de er redde for hva den åpne databruken skal medføre. Det offentlige er tvilende til å gi bort rådata med frykt for at kommersielle aktører eller private personer kan bruke dataene til å tjene penger (Chaffey, 2010). Aaby et al. (2011) skriver at det er viktig å identifisere frykt og motforestillinger omkring dataåpenhet, og at slike motforestillinger omkring eksempelvis dataeierskap og personvern må bli tatt på alvor og diskutert i organisasjonen.

Demonstrer nytteverdien

Den offentlige organisasjonen må ha kjennskap til når det er nyttig med semantiske vevteknologier, og når det ikke er det. De må modnes på forståelsen rundt Linked Open Data. Juristene og økonomene er nødt til å forstå nytteverdien. Mange som har kjennskap til eksisterende teknologier, mener at disse gjør den sammen nytten. Gjør teknologene innad i etatene interessant! Om de ikke er klar over og bevisste på hva det skal endres til, er det naturlig at de ikke ønsker å gå bort fra den tilnærmingen de er vant til. Opplys om forskjellene fra eksisterende teknologier, og vis at semantiske vevteknologier vil bidra til bedre gjenbruk av de åpne dataene. Skap interesse ved å demonstrere bruk av teknologiene på åpne data. Demoer vil vise at den semantiske veven er mulig å realisere. Presenter mulighetene og teknologiene på en forståelig måte. Involver etatene i forskningsprosjekter slik at de ser nytten. Etatene må få demonstrert effektene av Linked Open Data før de ønsker å gå i den retningen. Start en dugnadsaksjon for å få i gang tilgjengeliggjøringen av Linked Open Data hvor alle tilgjengeliggjør litt hver. Digitaliser dataene dersom de kun finnes i papirformat.

Basisforståelse av lenkede data og stegene for å tilgjengeliggjøre offentlige data på denne måten er svært begrenset (Sargent & Taggart, 2010). Dette skyldes at fokuset på gjenbrukbare åpne data i maskinlesbart format er relativt nytt, og har bare i de senere år beveget seg fra academia til næringslivet. Teknologer som har jobbet med andre teknologier, har derfor lite erfaring med Linked Open Data og må opplyses. Alani et al. (2007) erfarte at det var vitalt å vise organisasjonene eksempler fra hvor og hva slags verdi som skapes ved bruk av semantiske vevteknologier.

Skap forankring i ledelsen

Få ledelsen overbevist om å tørre å la andre gjenbruke etatens data. For å få gjennomslagskraft til å gå i denne retningen, må det holdes fokus på Linked Open Data. Da er det viktig med ildsjeler i organisasjonen som viser engasjement og sprer entusiasme for Linked Open Data. At det kommer påtrykk fra brukermassene der ute er av betydning, slik at folk på høyere nivå i organisasjonene forstår viktigheten. Å tenke annerledes sitter langt inne. Mange offentlige ledere tenker nytte ut i fra hva de vet i dag, men de må tenke nytte ut i fra den kreativiteten som utløses ved gjenbruk. Det kortsiktige ambisjonsnivået må heves. Hvorvidt personen er mottakelig for Linked Open Data og ønsker om å tilgjengeliggjøre avhenger av personens fokusområde, kompetanse og profesjon.

Ildsjeler for IT (IT champion) er ledere som aktivt promoterer deres personlige visjon for å bruke informasjonsteknologi, og som gjør hva de kan for å fremme prosjektet tross for manglende tillatelser og andre implementasjonshindringer; de risikerer ofte sitt eget rykte for

å sikre at innovasjonen blir gjennomført (Beath, 1991). Slike ildsjeler kan være nødvendig for at offentlige organisasjonen tilgjengeliggjør data som Linked Open Data. Vellykkede ildsjeler kan overkomme byråkratiske barrierer og drive frem endringer i organisasjonen, mens organisasjonens konkurrenter står fast ved å tvile på om endringen er mulig å gjennomføre (Beath, 1991). Sargent og Taggart (2010) erfarte at det ofte måtte initiativ fra ildsjeler i organisasjonen før åpne data skulle bli et faktum. Frigjøringen skjedde fordi personer som jobbet og hadde ansvaret for vevteknologiene innad i organisasjonen tok initiativ om å tilgjengeliggjøre lenkede data, fremfor interesse fra ledelsen.

Politiske styringssignaler

For at etater skal ha lov til å tilgjengeliggjøre data må de få klare styringssignaler fra politisk nivå. Til tross for at personer og organisasjoner rundt om kring i offentlig sektor ønsker å tilgjengeliggjøre data, kan de ikke gjøre dette uten tillatelse fra departementene. Det må opprettes flere stillinger i offentlig sektor knyttet til åpne data. Dagens situasjon med 2-3 stillinger gir ikke sterke indikasjoner på at åpne data er av stor betydning.

Politisk lederskap sees på som en viktig faktor for å kunne tilgjengeliggjøre åpne data (Huijboom & Van den Broek, 2011). I følge Sargent og Taggart (2010) er det nødvendig med begeistring og støtte fra en anerkjent myndighet. President Barack Obama og den tidligere britiske statsministeren Gordon Brown har begge kommet med klare styringssignaler for at USA og England skal åpne opp offentlige data. Å innføre ny bruk av IT i offentlige organisasjoner avhenger av støtte fra administrative myndigheter (Kamal, 2006) og skjer raskere når staten krever at endringen skal gjennomføres (Tolbert & Zucker, 1983). Det er derfor sannsynlig at sterke styringssignaler fra regjeringen vil være nødvendig for å tilgjengeliggjøre åpne data ved bruk av semantiske vevteknologier.

Andre inntekstmodeller

Offentlige organisasjoner må skape andre inntekstmodeller. Når staten i dag budsjetterer med salg av data, er det vanskelig å skulle gi bort offentlige data gratis. For at offentlige etater skal kunne tilgjengeliggjøre åpne data må disse frafall av disse inntektene dekkes på andre måter.

Huijboom og Van den Broek (2011) skriver at spesielt i europeiske land får flere offentlige organisasjoner sin inntekt ved å selge data. Derfor er det vanskelig for dem å tilgjengeliggjøre dataene åpent uten noen form for gebyrer.

Hvilke data ønskes

Offentlige organisasjoner bør lytte til innspill fra borgere om hvilke datasett de ønsker å gjenbruke. Andre der ute ønsker tilgang til deres data, og etterspørselen er stor! Fordi det finnes mange mennesker som ønsker et bestemt datasett til å utvikle en ny tjeneste, kan det være nyttig å utføre en behovsanalyse der det kommer frem hvilke data samfunnet trenger. Opprett et åpent forum der folk kan komme med ønsker om hvilke datasett de har størst behov for. Da kan dataeierne prioritere å tilgjengeliggjøre de datakildene som de får konkrete henvendelser om. Tilgjengeliggjøring av åpne data vil være litt behovsprøvd i forhold til hva samfunnet ønsker. Dersom det kommer sterke ytringer fra befolkningen om at det offentlige må åpne data, tar politikerne slike samfunnsengasjement i sterkere betraktning enn faglige tunge artikler om samfunnsgevinst i kroner og øre.

Borgerne er i den beste posisjonen til å avgjøre hva slags teknologier som vil være best egnet til å skape merverdi av de offentlige dataene som tilgjengeliggjøres (Tauberer, 2011). Derfor er innspill fra samfunnet vesentlig ved tilgjengeliggjøring av åpne offentlige data. Data som organisasjonen anser som lite interessant for andre enn seg selv, kan fort være dataene som andre ønsker å gjenbruke (Aaby, et al., 2011). Offentlige organisasjoner bør derfor ikke vurdere kun etter hva de selv synes er spennende data å dele, men må lytte til samfunnet om hva slags formater og datasett som etterspørres.

Avklar juridiske betingelser

Sørg for at dataens juridiske kontekst er bevart, og at etaten har lov til å tilgjengeliggjøre dataene. Offentlige data er tilknyttet til lovverket som gjør at de må brukes riktig. Det er i tillegg en del data som ikke etatene har lov til å tilgjengeliggjøre. Når dataene skal tilgjengeliggjøres er det viktig at den juridiske tilknytningen ikke forsvinner slik at de som gjenbraker dataene er kjent med hva slags juridiske betingelser disse dataene har.

Engels og Myrseth (2011) anbefaler å bruke åpne og velprøvde lisenser og i tillegg vurdere juridiske sider, inkludert rettigheter og personvern før dataene blir tilgjengeliggjort.

Sett sammen riktig kompetanse

Engasjer en jurist for å undersøke om lovverket tillater å tilgjengeliggjøre dataene. Det må avklares med dataeier om hvorvidt dataene kan gjøres tilgjengelig, og på hvilken måte. Vevspesialister spiller en viktig rolle med å faktisk gjøre dataene tilgjengelige på nettet. IT-avdelingen vil være sentral for å undersøke muligheter og konsekvenser. For å lenke sammen data, og skape riktige definisjoner må det være roller med personer som kjenner dataene, prosessen og semantikken.

Sargent og Taggart (2010) erfarte at det krevdes ulike ressurser med ulike ferdigheter for at offentlige organisasjoner skal tilgjengeliggjøre åpne data i lenkede formater. Blant rollene som ble nevnt var mennesker som kjenner de offentlige dataene fra innsiden og teknisk ekspertise omkring vevkoding og tilgjengeliggjøring av dataene.

Se etter lignende prosjekter

Offentlige organisasjoner bør se etter lignende prosjekter og hva som har blitt gjort tidligere. Lignende tilfeller av ditt prosjekt kan gi verdifull innsikt i fallgruver og suksessfaktorer. For hver innovative idé organisasjonen kommer opp med, kan du nesten være sikker på at det finnes andre med slike ideer og relevant erfaring (Eglene, 2000). Fenomenet dataåpenhet er nokså nytt, så det kan være nyttig og inspirerende å besøke andre organisasjoner sine nettsteder:

”Eksempler gir både et innsyn i hvilke datasett man bør publisere, hvordan en ny åpenhetspraksis kan markedsføres, og hvordan dataåpenhet kan skape nye og innovative løsninger i samspill med enkeltpersoner, medier og teknologiskaper”
(Aaby, et al., 2011).

Kjennskap til, forståelse for, og evnen til å ta lærdom fra andres forsøk kan gi organisasjonen en god start på prosessen (Eglene, 2000).

Bedre mekanismer for tilgangsstyring

Det må løses utfordringer rundt tilgangskontroll om Linked Open Data skal bli måten offentlige organisasjoner tilgjengeliggjør sine data på. Det må jobbes mer med å håndtere tilgangsstyring. Det er i dag vanskelig å skulle skille mellom hvilke data som skal gjøres åpent tilgjengelig, og hvilke som må holdes lukket. Personvern er en stor utfordring når det gjelder åpne data. Det er viktig å tilgjengeliggjøre dataene slik at personsensitive opplysninger ikke kommer på avveie. Det er en forutsetning at Linked Open Data bedre ivaretar sikkerhet og tilgangsstyring. Flere offentlige organisasjoner fortsetter å tilgjengeliggjøre data ved bruk av vevtjenester og APIer fordi de føler at rettighetshåndtering er enklere å håndtere med vevtjenester fremfor Linked Open Data. Det må forskes på å kunne konfigurere tilgangskontroll dynamisk.

For at etater skal i bruk semantiske vevteknologier ved tilgjengeliggjøring av åpne data, må de være sikre på at teknologiene gir dem muligheten til å velge hvilke data som skal deles, og hvilke som skal holdes lukket (Alani, et al., 2007).

RDF bør genereres automatisk

Det er svært fordelaktig om standardverktøy som offentlige etater bruker til å generere data gir muligheten til å strukturere data som RDF. Dersom sluttbrukere innad i etatene kan eksportere data i maskinlesbare formater, vil det kreve mindre ressurser til datatransformasjon. Drupal tillater brukeren å eksportere RDFa. Slike verktøy gjør det enklere for brukeren å tilgjengeliggjøre lenkede data.

Det er få verktøy for å publisere lenkede data, som medfører et manuelt og tidskrevende arbeid (Sargent & Taggart, 2010). Offentlige organisasjoner som bruker systemer for å håndtere innholdet på nettstedene (CMS) med manglende støtte for RDFa og XHTML, vil ha vanskeligheter for å tilgjengeliggjøre data i åpne lenkede formater. Tiltross for fordelene med en slik automatisk generering av RDF, vil det medføre utfordringer knyttet til oppgradering av slik nødvendig programvare (Sargent & Taggart, 2010).

Lenk de interne dataene

Lenk virksomhetsdataene internt med Linked Enterprise Data. Linked Enterprise Data kan være et logisk første skritt mot at offentlige data er lenket sammen. En slik tilnærming kan fungere i de miljøene som er interessert i å ha bedre struktur og skalerbarhet på datasystemene sine, men skeptiske til å tilgjengeliggjøre dataene sine for hele verdenen. Kombiner Linked Enterprise Data med Linked Open Data slik at Linked Enterprise Data brukes inne i organisasjonen og så kombineres dataene med en Linked Open Data-struktur som frigjør dataene eksternt. Det vil kunne sørge for at forretningskritiske og personsensitive data ikke tilgjengeliggjøres.

Flere bedrifter har begynt å ta i bruk prinsippene til lenkede data internt i sine informasjonssystemer (Engels & Myrseth, 2011). RDF gir systemer muligheten til å strukturere data fra forskjellige steder i et fleksibelt og skalerbart trippelformat, som blir presentert som hovedgrunnen for hvorfor systemer oppnår suksess ved å ta i bruk semantiske vevteknologier (S. Harris, Ilube, & Tuffield, 2010). Alani et al. (2007) poengterer at etater ikke trenger å gå vekk fra eksisterende databaser ved at de lenker dataene i RDF. Selv om dataene er konfidensielle og ikke skal tilgjengeliggjøres åpent på nett, er det sett på som en fordel å tilgjengeliggjøre begrepsapparatene. Engels og Myrseth (2011) forklarer dette ved at en organisasjon bruker andres ontologier og henter inn åpne data som beriker den ”interne” informasjonen.

5.1.2 Tilgjengeliggjøring

Engasjer teknisk kompetanse

Ta kontakt med leverandører dersom nødvendig kompetanse ikke finnes internt i organisasjonen. Engasjer leverandører som kjenner teknologiene, og vet hva de snakker om. Konverteringsmetode er avhengig av prosjekttype, formater en skal løfte, hva slags type datasett og hvor ofte dataene må oppdateres. Bruk R2M for å lage en transformasjon fra relasjonelle databaser til RDF. Selektér hvilke deler av den relasjonelle databasen man vil ha til RDF, og kopier til RDF. Tilgjengeliggjør så RDF. XML til RDF kan gjøres ved å bruke XSLT. SAXFilter kan være et godt alternativ til XSLT når det gjelder strengmanipulering. Dersom man ikke har krav til SPARQL-støtte, men bare transformere data til et RDF-format kan dette gjøres med PERL-rammeverket RDF-TRINE.

I Norge er ESIS (ESIS, 2011) og Computas (Computas, 2011) to leverandører av tjenester og løsninger basert på semantisk vevteknologi. De har svært god kunnskap om bruk av teknologiene på offentlige data. Linked Open Data Consultancy og Linked Open Data Around-The-Clock (LATC, 2011) er to europeiske forskningsprosjekter som tilbyr ekspertise for å hjelpe andre med å tilgjengeliggjøre åpne lenkede data.

Skap bedre begrepsapparater

Offentlige etater må bli bedre til å definere dataene sine og hva de ulike begrepene betyr. Det er viktig at organisasjonene får bedre kontroll på sine begrepsapparat, og rydder opp i informasjonen de har internt. Når dataene skal gjøres åpne for automatisk gjenbruk som Linked Open Data, må beskrivelsene være tydeligere. Det er da nødvendig at begrepsapparatene er maskinlesbare i formater som RDFS, SKOS eller OWL. Gå frem ved å samle kompetanse utover de som sitter i organisasjonsledelsen. Involver IT-arkitekter, personer som jobber med informasjonsforvaltning, jurister, lingvister, filosofer samt og folk som er sterke til å kommunisere. Kall sammen forskjellige fagavdelinger. Begrepsapparatet skal ta utgangspunkt i reelle begreper som brukes, fremfor å skape nye. Start i det små med et par begreper som ofte brukes i saksbehandling. Sorter begrepene og strukturer de i forhold til hverandre. I dag er taksonomier vanlig å bruke for å lage et rammeverk rundt begrepene i relasjon til konteksten. Ontologier anbefales fordi de kan gi beskrive rikere relasjoner mellom dataene. Det er vesentlig at meningen kommer tydelig frem slik at hvem som helst kan gjenbruke begrepet. Oversikt over hvordan begrepene brukes i forskjellige kontekster gjør at man kan sammenligne og bli enige om et omforent bruk. For å ta vare på disse begrepsapparatene er det viktig med en registerløsning for å ta vare på begrepsdefinisjonene. SERES er en slik løsning.

Ved å dokumentere dataene med gode beskrivelser, øker brukbarheten. Det gir organisasjonen et bedre overblikk over dataene (IT-ogTelestyrelsen, 2010). Åpne tilgjengelige begrepsapparater med tilknyttede informasjonsressurser gir det beste utgangspunktet for utvikling av nye tjenester og applikasjoner (Ølnes & Hove, 2011). Alani et al. (2007) forteller at utviklingen av små, datasentrerte ontologier var et relativt oppnåelig mål for utviklere i offentlige organisasjoner, selv med lave budsjetter.

Gi dataene URIer

Gi dataene en egen URI. En slik identifikator gjør dataene unike, og legger opp til at dataene kan lenkes til andre data. Med en slik URI vil man raskere kunne identifisere dataelementer. Sørg for at det er mulig å finne URIene. Si fra at de har blitt tilgjengeliggjort, og hvor de er tilgjengelige.

Hvert datasett som skal tilgjengeliggjøres som åpne lenkede data, må tilgjengeliggjøres med en passende URI (Sargent & Taggart, 2010).

Bruk RDF som datamodell

Bruk RDF som format og datamodell. Ved at dataene tilgjengeliggjøres i en RDF blir dataenes kontekst gyldig på et globalt nivå. Dataene blir maskinlesbare, og det er mulig å skape lenker til andre datasett slik at dataene kan prosesseres automatisk med semantiske vevteknologier. Dataene som tilgjengeliggjøres kan da lettere kombineres og sammenstilles med andre data. Dette vil heve gjenbruksmulighetene. Etater bør ta utgangspunkt i 5 stjerners data-modellen og tilgjengeliggjøre dataene i RDF med URIer slik at dataene oppnår så mange stjerner som mulig. Analysen avdekket at det var kun åpne data med RDF som ville gi stjerne nummer 5.

Aaby et al. (2011) argumenter for at RDF er det formatet som best representerer datasett med semantisk definerte relasjoner og unik representasjon av alle objekter ved hjelp av URIer. Det gjør at RDF gir særdeles gode muligheter for automatisk viderebehandling av data (IT-ogTelestyrelsen, 2010).

Bruk Turtle som RDF-notasjon

Bruk Turtle-serialiseringen som notasjon i RDF. Den er oversiktlig og har lett forståelig syntaks som gjør den enkel å lese. RDFa er et godt alternativ som kan brukes sammen med XHTML direkte på nettsteder.

Harris et al. (2010) anbefaler å bruke Turtle fordi den er lettere å lese. Det gjør det mye enklere å rette opp feil i programvarekoden. Heath og Bizer (2011) argumenterer for at RDF/XML bør velges dersom dataene kun skal tilgjengeliggjøres ved bruk av et serialiseringsformat. Grunnen er at dette formatet støttes av de fleste verktøyene som konsumerer lenkede data.

Spesifiser mening med RDFS

Spesifiser semantikken i RDF-modellen ved bruk av RDFS. SKOS og OWL er alternativer etter behov. SKOS er tilstrekkelig dersom man skal drive med taksonomier. OWL er svært uttrykksfull, men utfordrende å bruke.

Linked Open Data-initiativet anbefaler å beskrive meningen av data ved bruk av RDFS eller OWL; modellene kan anvendes på mange bruksområder for informasjon som lagring, utveksling, filtrering, søking og sammenstilling av data (Engels & Myrseth, 2011). I en Linked Data-sammenheng, er det ofte tilstrekkelig å uttrykke vokabularene i RDFS (Heath & Bizer, 2011). Når RDFS brukes sammen med noen egenskaper fra OWL, for eksempel sameAs, blir kombinasjonen ofte kalt for RDFS++ (Heath & Bizer, 2011) En annen fordel med å representere relasjoner i eksterne RDFS-baserte metadataontologier, er at brukere kan tilføye egne eller reformulere eksisterende relasjoner (Wittenburg, et al., 2004).

Gi tilgang via SPARQL

Gjør dataene tilgjengelig gjennom et SPARQL-tilgangspunkt. Ved at dataene ligger åpent i RDF med et SPARQL-tilgangspunkt gir det brukeren svært fleksible gjenbruksmuligheter.

Hvis databaser gjøres tilgjengelig på internett i RDF-formater, anbefales det å bruke en vevtjeneste med SPARQL som generisk grensesnitt (Engels, 2010)

Gjør begrepsapparatet tilgjengelig

Tilgjengeliggjøres dataene som Linked Open Data må den som publiserer sørge for at dataenes kontekst ikke forsvinner. Mye offentlige data er resultat av en saksbehandling og svært kontekstavhengig. Frigjør derfor begrepsapparatet på samme måte som dataene og sørg for at begrepsapparatet er mulig å finne for brukere. En stjerne nummer 6, som dreier seg om å tilgjengeliggjøre datasettets begrepsapparat, bør være med i denne modellen. Dette vokabularet, som beskriver det som finnes i datasettet, bør tilgjengeliggjøres som et navneområde, slik at andre igjen kan forstå hva du mener med dataene du har tilgjengeliggjort. Det er for å maksimalisere gjenbruk og presisjon at dataene bør tilgjengeliggjøres med 5 og 6 stjerner.

★★★★★★ Tilgjengeliggjør dataenes begrepsapparat som et navneområde

Figur 9: 6 stjerner ved Linked Open Data

Dersom begrepsapparatet er felles og blir brukt av flere aktører, muliggjør det utveksling av ressurser basert på felles emneord (Ølnes & Hove, 2011). En felles semantisk datamodell kan hjelpe med å danne en felles forståelse av termer og definisjoner. Dette gjelder ikke bare på datanivå, men en slik modell kan i tillegg fostre interoperabilitet også på ledelsesprosessen og det organisatoriske nivået. Mangel på en felles informasjonsmodell kan lede til dataduplisering, ulike representasjonsskjemaer, samt vanskeligheter med å håndtere data (Sarantis, Charalabidis, & Askounis, 2009).

Gjør datasettet enkelt å gjenbruke

Linked Open Data er bygget slik at datasettene som gjenbrukes minst, automatisk forsvinner. Derfor må datatilbyder sørge for at dataene som tilgjengeliggjøres er enkle å bruke som sørger for at mange gjenbruke dem. Dersom datasettet er komplisert med mangelfulle beskrivelser er det vanskelig å gjenbruke dette, og den potensielle gjenbruksverdien synker.

Gjør mest mulig tilgjengelig

Flere offentlige organisasjoner vegrer seg for å tilgjengeliggjøre sine data fordi de føler at kvaliteten ikke er bra nok. For at offentlige data skal komme til nytte, er det ikke nødvendig at kvaliteten på dataene er 100%. Mangelfulle beskrivelser vil være et hinder for at andre offentlige etater kan gjenbruke dataene, men dataene vil fortsatte være nyttige for samfunnet. Ved at dataene er åpne, kan andre gi tilbakemeldinger som hever datakvaliteten.

Aaby et al. (2011) skriver at dataene vil bli gjenbrukt så lenge de er gode, men det må ikke overvurderes hvor mye tid som skal legges i for å gjøre et datasett komplett. ”*Det er ikke slik at et datasett må være komplett for å kunne deles, så lenge dokumentasjonen beskriver hvorvidt det mangler data eller finnes dårlige data*” (Aaby, et al., 2011). Flere land har uttalt at deres offentlige data er av for lav kvalitet, og dermed ikke tilgjengeliggjøres (Huijboom & Van den Broek, 2011). Begrenset datakvalitet har vært sett på som en hinder for publisering av åpne data. Tilgjengeliggjør dataene slik de er. Paul Chaffey (2010) mener det er en stor skepsis hos statlige etater og kommuner mot å offentliggjøre data som ikke er bearbeidet. Dataene kan gjøres bedre ved at andre får gjenbruke dem (IT-ogTelestyrelsen, 2010). Et eksempel på en offentlig norsk virksomhet som opplevde en positiv effekt ved å frigi data er Avinor. Ved sommeren 2009 gjorde Avinor deler av sine flydata tilgjengelig (Christensen, 2009). Etter at dataene var fri, mottok selskapet tilbakemeldinger fra systemutviklere vedrørende inkonsistens og mangler i datasettene. Feil ble håndtert og resulterte i en økt kvalitet på datasettene til Avinor.

5.1.3 Etter tilgjengeliggjøring - Bruk av lenkede åpne data

Statlige midler til å utforske nye teknologier må økes

Det vil være kostbart å endre til semantiske vevteknologier i starten. De første som tar i bruk en ny teknologi, vil ofte ha en høyere kostnad enn de som kommer etterpå. Likevel vil det bli billigere å utvikle på sikt. Kostnadsnivået avhenger av hvor godt organisasjonens data er strukturert og digitalisert. Grad av forståelse for den semantiske veven vil ha mye å si. Måten etaten får årlige midler gjennom statsbudsjettet fører til vanskeligheter for å utforske ny teknologi og andre måter å gjøre ting på. Penger er satt av til visse prosjekter der og da som gjør det vanskelig for etatene å bruke slike midler på tilgjengeliggjøring. Det må diskuteres fram til hva det må suppleres med for at Linked Open Data i det offentlige får mest mulig samfunnsgevinst. Det kortsiktige ambisjonsnivået må heves slik at det bevilges penger til å utforske nye teknologier. I tillegg må det tildeles tid til å bruke nye standarder.

Den store forskjellen på offentlige og private organisasjoner når det gjelder IT-innovasjon kan forklares ved det tilgjengelige budsjettbeløpet for å adoptere ny informasjonsteknologi (Kamal, 2006) Videre forklarer Kamal (2006) at et høyt nivå av finansiell støtte kan ha en positiv innflytelse på adopsjon av IT-innovasjon i offentlige organisasjoner. Uten mer finansiell støtte er det vanskelig for offentlige organisasjoner å ta i bruk Linked Open Data.

Bedre informasjonsforvaltning

På grunn av kulturen som sitter i organisasjonene knyttet til frykt for å gjenbruke andres data, er det vanskelig å skulle bruke Linked Open Data. Skal Linked Open Data fungere sammen med offentlige data må det tas bedre hensyn til slike kulturelle behov. For at man skal overvinne organisasjoners skepsis tilknyttet bruk av andres data, må man etter hvert ha som ambisjon å tilgjengeliggjøre mer strukturer rundt dataene som beskrivelser av hvor de blir brukt og hva de blir brukt til. Uten tilstrekkelige metadata, er det umulig for offentlig sektor å gjenbruke åpne data. Data.norge.no bør være tilknyttet en metadata.norge.no med ulike typer metadata. For at åpne data kan gjenbrukes av andre offentlige organisasjoner må brukeren ha tilgang til dataenes begrepsapparat. Gjenbruken blir mer effektiv når brukeren raskere forstår hvordan hun skal gjenbruke begrepene korrekt. Det vil være vanskelig for offentlige etater å gjenbruke åpne data dersom det ikke finnes en data som beskriver hvordan dataene er strukturert og hvordan de forventes å brukes.

IT-og Telestyrelsen (2010) skriver at den ultimate dokumentasjonen på offentlige data ville være at de tilgjengelige dataene kobles mot en informasjonsmodell oppbygget av semantikkdefinisjoner. Det anbefales å etablere en oversikt over hvor metadata for offentlig sektor finnes, ikke bare data (Engels & Myrseth, 2011). Metadata letter bruken av informasjonen ved å tilby en kontekst til dataene slik at de kan enklere kan deles på tvers av ulike informasjonssamfunn, i tillegg til at metadata inneholder nødvendige pekere for å gjenfinne informasjonen (Landsbergen Jr & Wolken Jr, 2001). Det presiseres at offentlige administrasjoner ønsker å bli bedre på informasjonshåndtering, og det viktigste insentivet for å nå dette målet er å forbedre effektiviteten og kvaliteten på de elektroniske tjenestene de tilbyr (Klischewski & Ukena, 2010). For å bedre informasjonsforvaltningen er det nødvendig med strukturerte data som gir en detaljert beskrivelse av egenskapene til et datasett.

Spesialister finner datasett

Det må på plass eksperter som kan velge ut hvilke datasett som passer til en gitt oppgave. Det vil være svært mange tilgjengelige datasett, så det må på plass spesialister som kan sette sammen datasett slik at både de semantiske, organisatoriske og juridiske aspektene ivaretas.

Dette synet støttes av Engels og Myrseth (2011) som forklarer at enkeltetater som SSB i dag allerede har åpent lagt ut flere tusen datasett, som fører til at terskelen for å velge optimalt datasett for en oppgave fort blir en spesialistoppgave.

Bedre verktøy

Utviklingsverktøyene må bli bedre for at flere leverandører vil tilby semantisk vevteknologi. Flere verktøy må legges i hendene på utviklere slik at mange flere enn bare utviklere er i stand til å kombinere datakilder. Da vil det utvikles bedre utviklingsverktøy slik at etatene kan bruke disse teknologiene internt. Det er utviklerverktoyene som må vise at det helt

tydelig er riktig å bruke semantiske vevteknologier. Semantiske vevteknologier må prioriteres mer på Universitetene i Norge slik at antall spesialister på disse teknologiene øker. Den semantiske veven er komplisert, og ekspertene må bli enda bedre. Det er fortsatt en del igjen på modenhetsgraden til de øverste nivåene ved den semantiske vevstabelen. Kan nettvev implementeres? Områdene som innebærer tillit og bevis ved bruk av semantiske vevteknologier har et stykke igjen. Vanskelig i dag å stole på at beslutninger som har blitt inferert et eller annet sted i informasjonskuben er korrekt. Utviklere må begynne å skrive deklarativer datamodeller fremfor lukkede. Det handler om hva som modelleres, fremfor hvordan det modelleres.

I følge IT-og Telestyrelsen (2010) er RDF enda ikke så utbredt i Danmark, i tillegg til at det er begrenset verktøystøtte for å bruke teknologiene. Engels og Myrseth (2011) skriver at det er viktig med bedre verktøy slik at næringslivet, journalister, studenter og forskere og andre får et operativt miljø hvor de kan sette sammen, filtrere og visualisere data uten det kreves mye utstyr og kompetanse. Målet er å få på plass bedre verktøy slik at ikke bare store mediehus og næringsdrivende med egen IT-stab skal kunne bruke åpne data (Engels & Myrseth, 2011). Det er likevel verdt å ta i betraktning at data i RDF krever en større investering i tid og ressurser hos brukerne av dataene (Sargent & Taggart, 2010). Derfor er det viktig med bedre verktøy som gjør denne jobben enklere.

Etablere kritisk masse

Før offentlige organisasjoner begynner å bruke Linked Open Data må det en kritisk masse på plass. Etater må se at andre har brukt slike data og at det er nyttig for dem å bruke det. DIFI som fremmer IKT-standardene de offentlige etatene skal bruke, må se Linked Open Data som en sentral standard før andre etater tar disse i bruk.

En ny teknologi har oppnådd kritisk masse på det tidspunktet så mange individer har tatt den i bruk at videre adopsjon av teknologien blir selvdreven (Rogers, 2003). Når mange nok offentlige organisasjoner bruker Linked Open Data, vil de resterende organisasjonene også ta i bruk Linked Open Data og semantiske vevteknologier.

Bruk begreper korrekt

Når åpne data skal gjenbrukes, er det viktig å bruke begreper slik de er ment å brukes. En anbefaling er å få på plass autoritetsregistre slik at man vet hvilke data man skal bruke. Linked Open Data vil gi en felles måte å beskrive dataene på, men det må opprettes felles steder å referere til. Dersom en definisjon ikke passer til bruksformålet, må man bruke en annen fremfor å tilpasse denne til den aktuelle oppgaven. Finnes det ingen passende definisjon, anbefales det at man lager en egen ontologi som man lenker til de dataene man gjenbraker.

5.1.4 Beste praksis fremstilt som en prosess

Jeg har visualisert mange av prinsippene for beste praksis i en prosessmodell. Modellen viser en foreslått prosess for å tilgjengeliggjøre åpne offentlige data ved bruk av semantiske vevteknologier. Fordi at ønsket er å tilgjengeliggjøre åpne data flere ganger ønsker jeg å skape en prosess, fremfor et prosjekt. Denne avveiningen støttes av Aaby et al. (2011) som i tillegg skriver at tilgjengeliggjøring av åpne data handler om å etablere en varig praksis, der målet er å bli stadig flinkere.

1. Avklar behov og krav til endring

En offentlig organisasjon starter med å avklare ulike aktørers behov, og hva slags krav som må tilfredstilles før offentlige data kan gjøres åpne. Det er viktig å forstå kulturen som sitter i organisasjon for å vite hva som kan gjøres. En slik kartlegging gir grunnlag til å arbeide med bevisstgjøring og demonstrasjon av teknologienes nytteverdi. Dette vil kunne skape forankring i ledelsen. I tillegg til demoer, er det viktig med ildsjeler og samfunnsengasjement. Lytt derfor til samfunnet og hvilke data de ønsker. Når organisasjonen har valgt hvilke datasett som skal tilgjengeliggjøres, må de juridiske betingelsene til dataene avklares. For at organisasjonen kan gå videre med å tilgjengeliggjøre roller for prosessen, må regjeringen gi styringssignal om at de får lov til å tilgjengeliggjøre dataene åpent og at dette ikke går utover inntekstmodellen.

2. Fordel roller og ansvar

Når etaten har fått lov til å tilgjengeliggjøre data åpent, må det settes sammen riktig kompetanse. Slik kompetanse er beskrevet i beste praksis. Se etter lignende prosjekter, og ta lærdom av deres erfaringer. Det må forskes mer på hvordan Linked Open Data håndterer hva slags data som skal gjøres tilgjengelig og hvilke som skal holdes igjen. Dersom RDF genereres automatisk, vil det være lettere for ansatte uten programmeringskompetanse å skape lenkede data. En slik generering vil kunne støtte en intern lenking av virksomhetsdataene.

3. Bruk semantiske vevteknologier

Engasjer teknisk kompetanse om nødvendig. Slike eksperter kan hjelpe med å skape bedre begrepsapparater. Gi dataene URIer slik at de kan lenkes i en RDF-modell. Turtle vil være et gunstig format å bruke. Spesifiser semantikken i datamodellen ved bruk av RDFS. Gi tilgang til datasettet via et SPARQL-tilgangspunkt.

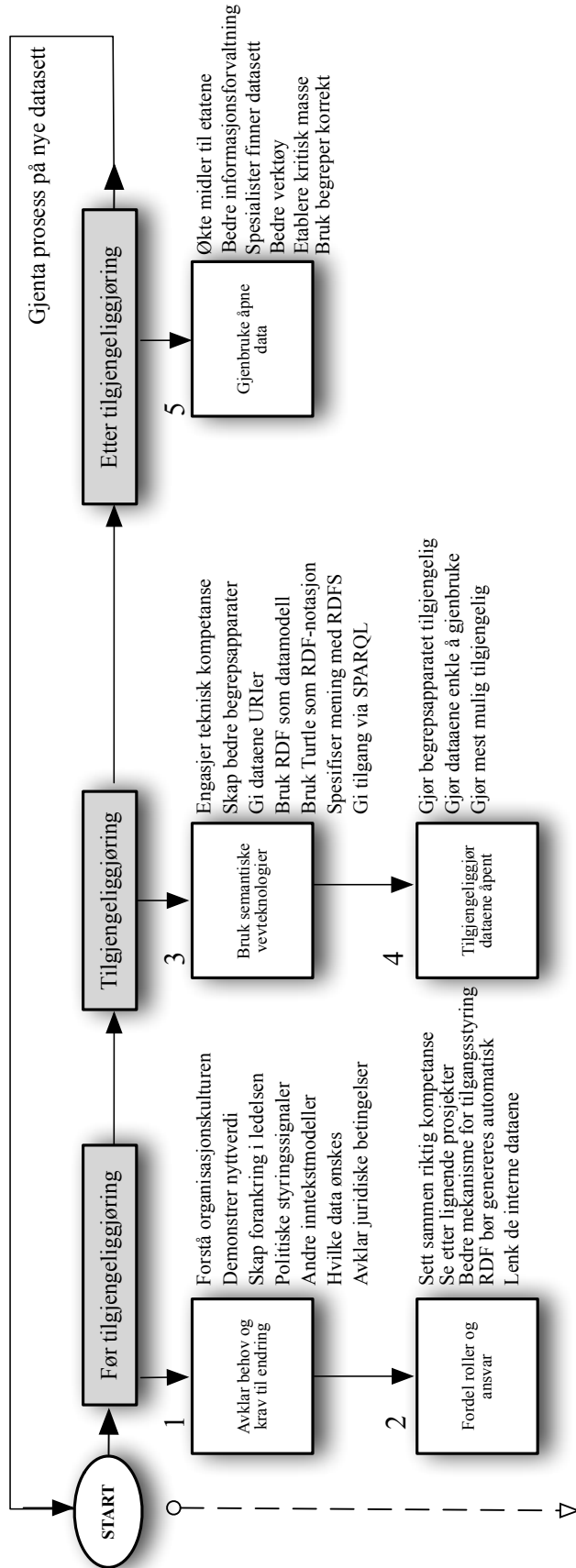
4. Tilgjengeliggjør data åpent

Gjør begrepsapparatene tilgjengelig som Linked Open Data. Er dataene enkle å bruke med et forståelig vokabular, vil det øke bruken av datasettet. Selv om data med mangelfulle beskrivelser ikke vil kunne gjenbrukes i alle settinger, er de likevel verdifulle som åpne tilgjengelige data. Gjør derfor mest mulig data tilgjengelig, tross for manglende kvalitet. Kvaliteten kan heves etter hvert. Når dataene er gjort tilgjengelige, gjenta prosessen på nye datasett.

5. Gjenbruk åpne data

Når det finnes åpne tilgjengelige data, må etatene få mer midler til å utforske bruken av disse. Dataene må ha en tydeligere informasjonsforvaltning slik at potensielle brukere forstår dataene, og hvordan de kan brukes. Fordi det etter hvert vil være mange datasett tilgjengelig, må det på plass spesialister som kan hjelpe de offentlige organisasjonene med å finne passende datasett. Det vil være vanskelig å bruke semantiske vevteknologier internt i offentlige organisasjoner før det kommer bedre verktøy som gjør det enklere å ta teknologiene i bruk. Kommer slike verktøy vil det vil sette i gang bruken, og bidra til at en kritisk masse gjenbrukes åpne data. Da vil mange flere gjenbruke åpne data. Når datasett skal gjenbrukes er det viktig at begreper brukes korrekt i tråd med den opprinnelige betydningen av begrepet.

BESTE PRAKSIS FOR Å TILJENGELIGGJØRE ÅPNE OFFENTLIGE DATA VED BRUK AV SEMANTISKE VEVTEKNOLOGIER



Figur 10: Beste praksis fremstilt som en prosess

5.2 Evaluering av forskningsmetoden

Jeg vil diskutere forskningsmetoden mot Guba (1981) sine kriterier for å vurdere troverdigheten ved naturalistiske studier. Tiltross for at jeg har utført kvalitative intervjuer, og ikke kvalitative observasjoner som naturalistiske studier er, mener jeg at prinsippene passer til å evaluere mitt studium. Hensikten er å avgjøre om funnene er sannsynlige, overførbare til andre kontekster, stabile og om de er i tråd med andres tolkninger (Guba, 1981). Måten jeg vil gjøre det på, er å diskutere aspektene over mot kredibilitet, generaliserbarhet, pålitelighet og bekreftbarhet.

Kredibilitet

Høy grad av kredibilitet vil gjøre funnene sannsynlige. Før og under studiet er det flere metoder for å stadfeste kredibilitet. *Forlenget engasjement på området som skal studeres* handler om å tilbringe lenger tid enn planlagt på området som skal studeres. *Utholdene observasjon* er tett knyttet til metoden over og skal gi forskeren mer tid til å finne karakteristika ved personene som observeres. Metodene passer for observasjonsstudier, og ikke så godt for intervjuer. Derfor vil jeg ikke diskutere disse mot mitt studium. *Diskusjoner med kollegaer* om hva som har blitt gjort og studiets fremgang vil bidra til kredibilitet. Jeg hadde jevnlig kontakt med veileder. Omtrent hver tredje uke sendte jeg arbeidet mitt til han og vi diskuterte framgangen i studiet. *Triangulering* er når studiet brukere data fra flere kilder, ulike forskere, ulike teorier og ulike metoder for å kryss-sjekk data tolkninger mot hverandre. Mine elleve intervjuer er fordelt på privat og offentlig sektor. Det gir litt ulike bakgrunner som snakker om de samme tingene, som gjør dataene mer troverdige. Jeg har ikke brukt andre personer til å gjøre noe av studiet mitt. Ulike teorier har jeg ikke testet. At jeg har lest litteratur i tillegg til intervjuene gjør at jeg kan sammenligne resultatene fra intervjuene mot det som står i faglitteraturen. Jeg har *samlet tilstrekkelig referansemateriale* slik at de som måtte ønske det kan teste funnene mot de opprinnelige datakildene. Jeg har valgt å inkludere mange sitater når jeg presenterer resultatene. Transkripsjonene fra lydopptakene har jeg tatt vare på, og kan fås tilgang til etter forespørsel. Lydopptakene er ikke gjort tilgjengelige noe sted, men jeg har de fortsatt. *Medlemstester* går ut på at de som har blitt observert får gi tilbakemeldinger om funnene stemmer. Jeg har ikke tatt kontakt med informantene etter at jeg har analysert intervjuene. Det skulle jeg ha gjort dersom jeg hadde hatt tid. Etter studiet er gjennomført er det viktig å *etablere sammenheng*. Jeg har gjort dette ved å teste mine tolkninger mot deler av analysen for å sørge for at det ikke er noen interne konflikter eller motsigelser. Oppgaven min har en rød tråd ved at funnene og diskusjonen er konsekvent med andre deler av oppgaven. *Å teste konklusjoner mot innsamlede data* handler om å gå tilbake til rådataene slik de var før analysen, og sørge for at det ikke har blitt trukket konklusjoner ut i fra data som ikke finnes. Jeg har i ettertid lest gjennom transkripsjonene, men ikke hørt lydopptakene om igjen. Jeg har ikke utført *medlemstester* med informantene for å få deres kommentarer på funnene.

Generaliserbarhet

Generaliserbarhet vil gjøre funnene nyttige i ulike kontekster. Metoden *meningsfull utvelging av informanter* vil si at kildene er valgt ut i fra viktighet og relevans for området som studeres. Jeg ville undersøke både tekniske og organisatoriske faktorer, så jeg valgte informanter både fra leverandørsiden og fra de offentlige organisasjonene som skal tilgjengeliggjøre data. Jeg har beskrevet deres arbeidsområde og hvilken sektor de tilhører. Grunnen til at akkurat disse er valgt er fordi ekspertene er flere av de mest anerkjente fagpersonene i Norge. Informantene fra offentlig sektor har alle relevant erfaring med semantikk og åpne data. Det som vil være et diskusjonselement er mitt utvalg på elleve informanter. De svarene jeg får vil være farget av de jeg snakker med. Hadde jeg snakket med ansatte i offentlig sektor som ikke kjente til utfordringer og fordeler med åpne data, ville jeg fått andre resultater. Derfor er det litt vanskelig å generalisere de offentlige informantene sine svar som svarene til de offentlige organisasjonene. Dersom jeg hadde snakket med enda flere personer, ville studiet hatt høyere generaliserbarhet. Ved å *samle tykke beskrivende data* vil det være mulig å sammenligne om funnene kan overføres til andre kontekster. Resultatene fra analysen er fyldige med mye informasjon. Selv om funnene beskriver hvordan noen offentlige organisasjoner jobber med åpning av data, vil bruk av semantiske vevteknologier

være like relevant i andre offentlige organisasjoner. Organisasjonene som har blitt sett på, deler mange fellestrekk. Å skape tykke beskrivelser av konteksten kan også gjøres i etterkant av studiet for å gjøre funnene mer generaliserbare.

Pålitelighet

Har studiet høy grad av pålitelighet vil funnene sees som stabile. Holdbare funn kan skapes ved å bruke *overlappende metoder*. Det vil si at ulike metoder brukes parallelt slik at svakheter ved en metode kompenseres med styrkene til en annen. Dersom flere metoder gir samme resultater, vil funnene være stabile. Jeg har lest teori i tillegg til å intervju sentrale personer. Det vil si at jeg har lest om hvilke praksiser som finnes, og blitt fortalt om hvordan ting gjøres. Fordi skriftlig materiale har en tendens til å ikke vektlegge det negative, har det vært svært nyttig å gjennomføre intervjuer. Litteraturen støtter mange av funnene. *Etablering av endringslogger* vil gjøre det mulig for andre å undersøke prosessene for datainnsamling og analyse, og hvordan forskeren kom frem til funnene. Jeg har beskrevet hvordan jeg jobbet med datainnsamling og analyse i metodekapitlet. Intervjumatrisen med spørsmål jeg brukte under intervjuene er vedlagt i oppgaven, som gjør det mulig å se hvilke spørsmål jeg stilte. Jeg presenterer flere iterasjoner av kodene mine for å vise hvordan analysen har utviklet seg. Med min fyldige beskrivelse og forklaring av hvordan jeg har kommet fram til resultatene, blir prinsippene for beste praksis mer pålitelige. Etter studiet er gjennomført kan funnernes pålitelighet heves ved at noen *kompetente personer vurderer endringsloggen* og gir kommentarer på i hvilken grad fremgangsmetoden passer med generelt aksepterte metoder. Min veileder har diskutert analysemetoden sammen med meg.

Bekreftbarhet

Høy grad av bekreftbarhet vil fjerne skepsis rundt funnene slik at andre ikke trenger å repetere analysen. For at resultatene skal kunne bekreftes å være i tråd med andre kilder, kan *triangulering* være en metode å bruke. Da samler forskeren inn data fra ulike perspektiv, bruker varierte metoder og varierte kilder for å vise at ulike kilder kommer til de samme konklusjonene. Etter at jeg kom fram til prinsipper for beste praksis, diskuterte jeg funnene mot andre kilder som viste seg å bekrefte mange av mine påstander. Å *praktisere refleksivitet* vil si at det kommer tydelig frem hvordan studiet har blitt utført, hvordan endringer har blitt utført og hvorfor beslutninger har blitt tatt. Det er viktig å dokumentere eventuelle skift i forskningsfokus. Jeg har forklart i metodekapitlet hvordan jeg har jobbet med forskningsspørsmål, datainnsamling, og analyse. Fordi jeg skriver om avveininger og endringer jeg har gjort underveis, viser jeg hvordan jeg har tenkt. Etter studiet er utført kan forskeren *arrangere en revisjon av bekreftbarheten* ved å la en ekstern person gå i gjennom de innsamlede rådataene for å gjøre en tilsvarende analyse. Dersom den andre personen kommer fram til de samme resultatene, er det lettere å bekrefte funnene. Med tidsrammen jeg har for min oppgave har jeg ikke hatt tid eller ressurser til å forberede en slik revisjon.

6. KONKLUSJON OG VIDERE ARBEID

6.1 Konklusjon

Hensikten med oppgaven var å undersøke *hvordan offentlige organisasjoner bør gå frem for å tilgjengeliggjøre åpne data ved bruk av semantiske vevteknologier*. For å finne ut av det, intervjuet jeg seks ansatte i fem forskjellige offentlige etater og fem eksperter på semantiske vevteknologier fra fire forskjellige leverandører i privat sektor. Parallelt med de kvalitative intervjuene, leste jeg faglitteratur for å få bedre forståelse av emneområdet. Dataene jeg samlet inn gav meg grunnlaget for å forstå nåværende praksis. Analysen av nåværende praksis har resultert i en beskrivelse av beste praksis med 28 anbefalinger fordelt på tre faser ved prosessen: før tilgjengeliggjøring, tilgjengeliggjøring, og etter tilgjengeliggjøring (som tilsvarer bruk av de åpne dataene). For å gjøre prinsippene mer anvendelige, har jeg laget en modell som fremstiller prosessen grafisk.

Modellen fester anbefalingene til *de ulike stegene i tilgjengeliggjøringsprosessen*. Stegene er: avklar behov og krav til endring, fordel roller og ansvar, bruk semantiske vevteknologier, tilgjengeliggjør data åpent og gjenbruk åpne data. *Hvem som blir de involverte aktørene og hva slags ansvar de forskjellige har i prosessen* vil avhenge av hvilke steg som skal utføres. Dataeierne er en svært sentral aktør. Det er de som bestemmer hva som skal tilgjengeliggjøres og hvordan. En forankring i organisasjonsledelsen vurderes som svært viktig for å sette i gang arbeidet. Videre vil slike ledere ta beslutninger ut i fra styringssignaler og føringer fra departementene. Brukerne vil kunne være med å skape engasjement og påvirke de ulike lederne. En ildsjel som brenner for åpne data og bruk av semantiske vevteknologier er viktig for å etablere et trykk mot organisasjonsledelsen og departementene. Brukerne vil være ansatte i den offentlige organisasjonen, private organisasjoner, andre etater, forskere, studenter og andre borgere. Juristene avgjør dataenes juridiske betingelser, og hvorvidt etaten har lov til å tilgjengeliggjøre disse dataene. IT-avdelingen, eller andre som kan informasjonen, må underrettes for å avklare muligheter og hvordan dataene er lagret. Vev- og semantikkspesialister vil gjøre jobben med å bruke semantiske vevteknologier.

Det er kartlagt flere *organisatoriske faktorer som vil påvirke prosessen*. Etater sin organisasjonskultur vil kunne være en hindring. Mange ledere vil ikke tilgjengeliggjøre åpne data fordi dataene er deres konkurransefordel, og de tror de mister denne dersom de gjør dataene tilgjengelig for andre. De er i tillegg redd for å tilgjengeliggjøre dataene uten restriksjoner for videre bruk fordi de frykter negative konsekvenser i form av kvalitetsproblemer. Fordi mange organisasjoner ikke er bevisste i forhold til hva det vil si å tilgjengeliggjøre åpne data og bruke semantiske vevteknologier, vil de ikke bruke ressurser på slikt arbeid. *Hva slags ressurser som kreves av transformering og opplæring* varierer ut i fra hva slags hva slags arbeid som skal utføres. Må det engasjeres teknisk kompetanse eller finnes det personer internt i organisasjonen som kan utføre denne jobben? Det kreves mindre arbeid til å tilgjengeliggjøre data ved bruk av semantiske vevteknologier når dataene allerede er digitalisert som lenkede data. Om systemer genererer RDF automatisk, vil det kreve minimalt med transformering. Slike systemer gjør at organisasjonene ikke trenger like mye opplæring. Det er kostbart å skulle forandre IT infrastrukturen i offentlige organisasjoner og gå vekk fra eksisterende databaseløsninger. Dette er ikke nødvendig ettersom semantiske vevteknologier og lenkede data kan brukes sammen med disse teknologiene. Hvor dyr transformasjonen blir, avhenger av faktorene som nevnes over.

Hvilke semantiske vevteknologier som er sentrale å bruke ved tilgjengeliggjøring av åpne offentlige data må samsvare med datasettet som skal tilgjengeliggjøres. Dataene bør likevel gis URIer som lenkes sammen i RDF-modeller skrevet i Turtle eller RDFa. Begrepsapparatet bør modelleres i RFDS, eventuelt i SKOS eller OWL ved behov. Dataene i RDF-modellen bør kunne gis tilgang til via et SPARQL-tilgangspunkt. *Linked Open Data kan brukes av det offentlige som en åpent begrepsapparat* ved å bidra til felles forståelse av ulike begreper. Ved at hver etat beskriver sine data i RDFS eller OWL og tilgjengeliggjør disse begrepsapparatene som lenkede åpne data vil vokabularene kunne sammenlignes. Å bli bevisst på hva hver enkelt organisasjon legger i et begrep vil kunne bidra til økt begrepsforståelse og mer effektiv gjenbruk. I tillegg til begrepsforståelse vil måten dataene er lenket på, opplyse om andre datakilder etaten kan gjenbruke. Ved å følge lenker kan

organisasjonen finne verdifull informasjon som ellers hadde vært vanskelig å få tak i.

Opgaven har gitt innsikt i offentlige organisasjoners dataåpenhet, semantikk, lenkede data, og hvorfor semantiske vevteknologier bør brukes ved tilgjengeliggjøring av åpne offentlige data. Funnene har bidratt med en beste praksis for hvordan offentlige organisasjoner kan gå frem i denne prosessen. Til tross for visse hindringer for å få fullt utbytte av å bruke disse teknologiene på offentlige data, er det i dag tilstrekkelig modenhet for å ta de i bruk. Mange land har brukt slike teknologier på offentlige data, og antall datasett i RDF vil trolig fortsette å øke med tiden.

6.2 Videre arbeid

Kasusstudium

Det kunne vært spennende å utføre prosessen som et kasusstudie på en offentlig organisasjon sine data. Å gjennomføre prinsippene ville vært veldig nyttig for å evaluere hvordan prosessen fungerer. Et slikt kasusstudie havnet utenfor tidsrammene av denne masteroppgaven.

Semantiske vevteknologier og åpne data i offentlig e-samhandling

Det foregår i dag elektronisk samhandling innenfor offentlig sektor, men omfanget av slik samhandling er begrenset grunnet manglende kunnskap om hvor dataene er lokalisert, eksistensen til dataene, manglende tilgjengeliggjøring, samt manglende systemstøtte for lesing av dataen (SERES, 2010). Rapporter (Grimstad & Myrseth, 2010; Grimstad, et al., 2008; Norheim, et al., 2010) viser at det er et økende behov for felles formater ved utveksling og deling av informasjon på tvers av organisasjoner i offentlig sektor. For å fremme interoperabilitet, som best forstås som deling av informasjon (Landsbergen Jr & Wolken Jr, 2001), er det avgjørende at informasjonen som deles forstås likt av samtlige aktører som skal bruke informasjonen. Sann som datautvekslinger i det offentlige skjer i dag er det mest mellom parvise aktører (SERES, 2010). En aktør sender data til en mottaker i såkalte "batcher" med grunndata, som mottaker lagrer i det formatet som passer med det eksisterende systemet. Dataene blir gjerne oppdatert ettersom det oppdages kvalitetsfeil, men disse endringene kommer ikke tilbake til de som sitter på den originale datakilden. En slik måte å gjennomføre elektronisk samhandling, fører til redusert kvalitet på dataene, og videre til redusert saksbehandlingskvalitet. Ved å ha åpne data som deles og administreres felles på tvers av etater kan slike situasjoner begrenses. Gjenbrukskostnadene vil reduseres dersom dataene er åpent tilgjengelige. Brønnøysundregistrene for eksempel, betaler årlig ca. kr.1 400 000 for å motta ukentlige data fra Folkeregisteret (SERES, 2010).

Semantiske vevteknologier kan være nøkkelen for å løse problemer knyttet til interoperabilitet i offentlig sektor fordi slike tilnærminger kan se ut til å støtte systematisk håndtering av alle tenkelige elektroniske ressurser, gi støtte for administrative prosesser på tvers av organisasjoner, systemer og infrastrukturer, samt gi forbedret tjenestekvalitet (Klischewski & Jeenicke, 2004). Ved interorganisasjonssamarbeid er felles standarder og formater en forutsetning for effektiv samhandling. Bruken av felles IT definerer en klar mening ved den pågående kommunikasjon, gir en forstått standard for å lagre og aksessere relatert arbeid, samt det skapes et repositorie med alle mål, regler og prosedyrer som deltakere kan bruke som en referanse (Dewett & Jones, 2001). I den grad de delte dokumentene er semantisk forstått av begge parter vil det strømførme kommunikasjonen og standardiserer koordineringen. De involverte partene vil trenge mindre tid til artikuleringarbeid som fører til mer tid til å jobbe med den faktiske oppgaven. Det ville vært veldig spennende som et videre arbeid å se på bruk av semantiske vevteknologier og åpne data for elektronisk samhandling i offentlig sektor.

7. KILDELISTE

- Aaby, A., Brenna, A., Jøsendal, P. V., & Nilsen, A. W. (2011). Viderebruksveileder. Retrieved 21.05, 2011, from <http://no.wikibooks.org/wiki/Viderebruksveileder>
- Alani, H., Dupplaw, D., Sheridan, J., O'Hara, K., Darlington, J., Shadbolt, N., et al. (2007). *Unlocking the potential of public sector information with semantic web technology*.
- Baader, F., Horrocks, I., & Sattler, U. (2005). Description logics as ontology languages for the semantic web. *Mechanizing Mathematical Reasoning*, 228-248.
- Beath, C. M. (1991). Supporting the information technology champion. *Mis Quarterly*, 355-372.
- Berners-Lee, T. (2009). Putting Government Data online *Design Issues*, from <http://www.w3.org/DesignIssues/GovData.html>
- Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. *Scientific american*, 284(5), 28-37.
- Berners-Lee, T., & Shadbolt, N. (2010, 25.02). Our manifesto for government data. <http://www.guardian.co.uk/news/datablog/2010/jan/21/timbernerslee-government-data>
- Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data-the story so far. *International Journal on Semantic Web and Information Systems*, 5(3), 1-22.
- Boddy, D., Boonstra, A., & Kennedy, G. (2008). *Managing Information Systems: Strategy and Organisation* (3.edition ed.): Financial Times/ Prentice Hall.
- Braadland, T. R. (2002). *Innføring i informasjonsbehandling*: Fagbokforlaget.
- Bryman, A. (2008). *Social Research Methods* (3 edition ed.). New York: Oxford University Press Inc.
- Chaffey, P. (2010). Slipp offentlig data fri. *Computerworld* Retrieved 15.02, 2010, from <http://www.idg.no/computerworld/article158680.ece>
- ChiefSecretarytotheTreasury. (2009). *Putting the Frontline First: smarter government*. London.
- Chrisholm, M. (2006, April). Master Data versus Reference Data. Retrieved 28.05, 2011, from <http://www.information-management.com/issues/20060401/1051002-1.html?pg=1>
- Christensen, A. (2009). Hva skjer når man gir bort flydata gratis? [IT-rådgivning, Kvalitet og testing, Virksomhet 2.0]. <http://open.bekk.no/2009/12/13/hva-skjer-nar-man-gir-bort-flydata-gratis/>
- Computas. (2011). Computas. *Share Your Knowledge* Retrieved 27.05, 2011, from www.computas.com
- Cygniak, R., & Jentzsch, A. (2011). Linking Open Data. Retrieved 06.05, 2011, from <http://lod-cloud.net>
- Dewett, T., & Jones, G. (2001). The role of information technology in the organization: a review, model, and assessment. *Journal of Management*, 27(3), 313.
- Eglene, O. (2000). *Conducting Best and Current Practices Research: A Starter Kit*. Albany: Center for Technology in Government.
- Engels, R. (2010). *Anbefalinger for en vellykket tilstedeværelse i den digitale kulturelle verden* (No. ABM-skrift #66): ABM.
- Engels, R., & Myrseth, P. (2011). *Åpne og lenkede data* (No. 2011-276). Høvik: Det Norske Veritas AS.
- ESIS. (2011). ESIS Norge. Retrieved 27.05, 2011, from www.esis.no
- Graneheim, U., & Lundman, B. (2004). Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. *Nurse education today*, 24(2), 105-112.
- Greenberg, J. (2005). Understanding metadata and metadata schemes. *Cataloging & classification quarterly*, 40(3), 17-36.
- Grimstad, T., & Myrseth, P. (2010). *Information Governance and Metadata Strategies as a Basis for Cross-sector e-Services*. Paper presented at the eChallenges e-2010.

- Grimstad, T., Myrseth, P., Meyer, H. S.-., Lasse Udjus, Hans Solli-Sæther, Dag Belsnes, et al. (2008). *Semicolon: State Of The Art* (No. 2008-0996). Høvik: Det Norske Veritas AS.
- Grønmo, S. (2004). *Samfunnsvitenskapelige metoder*. Bergen: Fagbokforlaget Vigmostad & Bjørke AS.
- Guba, E. (1981). Criteria for assessing the trustworthiness of naturalistic inquiries. *Educational Technology Research and Development*, 29(2), 75-91.
- Harris, S., Gibbons, J., Davies, J., Tsui, A., & Crichton, C. (2008). *Semantic technologies in electronic government*. Paper presented at the Proceedings of the 2nd International Conference on Theory and Practice of Electronic Governance.
- Harris, S., Ilube, T., & Tuffield, M. (2010). Enterprise Linked Data as Core Business Infrastructure. *Linking Enterprise Data*, 209-219.
- Heath, T., & Bizer, C. (2011). *Linked Data: Evolving the Web into a Global Data Space* (1st ed.): Morgan & Claypool.
- Hebeler, J., Fisher, M., Blace, R., & Perez-Lopez, A. (2009). *Semantic Web Programming*: Wiley.
- Heflin, J., & Hendler, J. (2000). *Semantic Interoperability on the Web*. Maryland: University of Maryland.
- Henze, N. (2008). The Proof and Trust Layers of the Semantic Web. Unpublished Presentation. IVS Semantic Web Group.
- Hoff, Ø. (2007). *Informasjonsmodell for Skip-Land Kommunikasjon (ISLK)* (No. 2007-0076): DNV.
- Horrocks, I. (2007). *Semantic Web: The Story so far*. Paper presented at the 16th International World Wide Web Conference.
- Huijboom, N., & Van den Broek, T. (2011). Open data: An International Comparison Of Strategies. *European Journal of ePractice*, 12, 13.
- IT-ogTelestyrelsen. (2010). *Teknisk vejledning til at sætte Offentlige Data I Spil*. København: Ministeriet for Videnskab, Teknologi og Udvikling.
- Kamal, M. (2006). IT innovation adoption in the government sector: identifying the critical success factors. *Journal of Enterprise Information Management*, 19(2), 192-222.
- Klischewski, R., & Jeenicke, M. (2004). Semantic web technologies for information management within e-government services.
- Klischewski, R., & Ukena, S. (2010). E-Government Goes Semantic Web: How Administrations Can Transform Their Information Processes. In T. Vitvar, V. Peristeras & K. Tarabanis (Eds.), *Semantic Technologies for E-Government* (pp. 99-125): Springer Berlin Heidelberg.
- Lagoze, C., Lynch, C. A., & Daniel Jr, R. (1996). *The Warwick Framework: A Container Architecture for Aggregating Sets of Metadata*: Cornell University.
- Landsbergen Jr, D., & Wolken Jr, G. (2001). Realizing the promise: Government information systems and the fourth generation of information technology. *Public Administration Review*, 61(2), 206-220.
- LATC. (2011). Linked Open Data Around-The-Clock. *Support you to publish and consume Linked Data on the Web* Retrieved 27.05, 2011, from <http://latc-project.eu/>
- LinkedDataTools.com. (2010). LinkedDataTools. Retrieved 10.09, 2010, from <http://www.linkeddatatools.com>
- Lovdata. (2011). Offentleglova. Retrieved 06.05., 2011, from <http://www.lovdata.no/all/tl-20060519-016-001.html>
- Matthews, B. (2005). Semantic web technologies. *E-learning*, 6(6), 8.
- McComb, D. (2004). *Semantics in Business Systems*. San Francisco: Morgan Kaufmann Publishers, Elsevier.
- Myrseth, P. (2009). *Offentlige data - hvem, hva hvor?* Paper presented at the EDOK 2009.
- Norheim, D., Udjus, L., & Smith-Meyer, H. (2010). *Utredning om egnede standarder for definisjonsarbeid i offentlig sektor*.
- O'Connor, M., & Das, A. (2009). *SQWRL: a query language for OWL*.
- Ojo, A., Janowski, T., & Estevez, E. (2009). Semantic interoperability architecture for electronic government, *Proceedings of the 10th Annual International Conference on Digital Government Research: Social Networks: Making Connections between Citizens, Data and Government* (pp. 63-72 %@ 978-971-60558-60535-60552 %U

- <http://portal.acm.org/citation.cfm?id=1556176.1556192&coll=ACM&dl=ACM&CFID=76476863&CFTOKEN=49825665>): Digital Government Society of North America.
- OpenGovernmentWorkingGroup. (2007). 8 Principles of Open Government Data. Retrieved 04.05, 2011, from <http://www.opengovdata.org/home/8principles>
- Orszag, P. R. (2009). *Open Government Directive*.
- Park, J., & Ram, S. (2004). Information systems interoperability: What lies beneath? *ACM Transactions on Information Systems*, 22(4), 595-632.
- Rogers, E. M. (2003). *Diffusion of innovations* (5th ed.). New York: Free Press.
- Sarantis, D., Charalabidis, Y., & Askounis, D. (2009). An ontology for stakeholder collaboration and knowledge exploitation in e-government project management, *Proceedings of the 3rd International Conference on Theory and Practice of Electronic Governance* (pp. 61-67 %@ 978-971-60558-60663-60552). Bogota, Colombia: ACM.
- Sargent, V., & Taggart, C. (2010, 22.05). Publishing Local Open Data - Important Lessons from the Open Election Data project. <http://data.gov.uk/blog/publishing-local-open-data-important-lessons-open-election-data-project>
- SemanticFocus. (2010). Semantic Web Layer Cake. Retrieved 07.09, 2010, from <http://www.semanticfocus.com/media/insets/semantic-web-layer-cake-2.png>
- Semicolon. (2011). Semicolon II. *Samhandling i offentlig sektor*, 2011, from <http://www.semicolon.no>
- SERES. (2010). Semantikkregisteret for elektronisk samhandling. Retrieved Mai 2, 2010, from www.brreg.no/samordning/semantikk/index.html
- Tauberer, J. (2011). Open Data is Civic Capital: Best Practices for “Open Government Data”. Retrieved 18.04, 2011, from <http://razor.occams.info/pubdocs/opendataciviccapital.html#bestprac>
- Tolbert, P. S., & Zucker, L. G. (1983). Institutional Sources of Change in the Formal Structure of Organizations: The Diffusion of Civil Service Reform, 1880-1935. *Administrative Science Quarterly*, 28(1), 22-39.
- W3C. (2007). XQuery 1.0: An XML Query Language. *W3C Recommendations* Retrieved 27.09.10, from <http://www.w3.org/TR/xquery/>
- W3C. (2009a). Namespaces in XML 1.0 (Third Edition). *W3C Recommendations* Retrieved 27.09, 2010, from <http://www.w3.org/TR/REC-xml-names/>
- W3C. (2009b). SKOS CURRENT STATUS. Retrieved 04.05, 2011, from http://www.w3.org/standards/techs/skos#w3c_all
- W3C. (2011). Linking Open Data. Retrieved 09.05., 2011, from <http://esw.w3.org/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>
- Wittenburg, P., Broeder, D., & Buitelaar, P. (2004). Towards metadata interoperability, *Proceedings of the Workshop on NLP and XML (NLPXML-2004): RDF/RDFS and OWL in Language Technology* (pp. 9-16 %U <http://portal.acm.org/citation.cfm?id=1621066.1621068&jmp=cit&coll=ACM&dl=ACM&CFID=76476863&CFTOKEN=49825665#>): Association for Computational Linguistics.
- Ølnes, S., & Hove, N. A. (2011). *Begrepsapparat for reiselivet – Forslag til utvikling av felles ”ordliste” for reiselivet* (No. 9/2011). Sogndal: Vestlandsforskning.
- Øvrebø, O. A., Grene, G. T., & Reiersen, L. T. (2010). *Viderebruk av datakilder i offentlig sektor: potensial og hindringer*. Bergen: Institutt for informasjons- og medievitenskap.