

Byantikvarens LODne liste

Erfaringer med modellering av byantikvarens gule liste i RDF

Audun Stolpe

Seniorforsker ved Forsvarets Forskningsinstitutt (FFI)

7. mars 2014

Outline

- 1 Verdien av åpne data
- 2 Hvordan gjøre offentlige data mer tilgjengelige?
- 3 LOD som publiseringsparadigme
- 4 Byantikvarens gule liste
- 5 Fremblikk

Flere bobiler på norske veier!

- Tenk deg at du ønsker å starte bobilutleie



Flere bobiler på norske veier!

- Tenk deg at du ønsker å starte bobilutleie
 - Du bør være der turistene er, og



Flere bobiler på norske veier!

- Tenk deg at du ønsker å starte bobilutleie
 - Du bør være der turistene er, og
 - så vidt mulig, langt fra konkurrenter



Flere bobiler på norske veier!

- Tenk deg at du ønsker å starte bobilutleie
 - Du bør være der turistene er, og
 - så vidt mulig, langt fra konkurrenter
- Du trenger å vite hvor turistene drar,



Flere bobiler på norske veier!

- Tenk deg at du ønsker å starte bobilutleie
 - Du bør være der turistene er, og
 - så vidt mulig, langt fra konkurrenter
- Du trenger å vite hvor turistene drar,
 - f.eks. <http://statbank.ssb.no/statistikbanken/>



Flere bobiler på norske veier!

- Tenk deg at du ønsker å starte bobilutleie
 - Du bør være der turistene er, og
 - så vidt mulig, langt fra konkurrenter
- Du trenger å vite hvor turistene drar,
 - f.eks. <http://statbank.ssb.no/statistikbanken/>
- .. samt å vite hvorvidt disse stedene allerede har slike utleiefirmaer



Flere bobiler på norske veier!

- Tenk deg at du ønsker å starte bobilutleie
 - Du bør være der turistene er, og
 - så vidt mulig, langt fra konkurrenter
- Du trenger å vite hvor turistene drar,
 - f.eks. <http://statbank.ssb.no/statistikbanken/>
- .. samt å vite hvorvidt disse stedene allerede har slike utleiefirmaer
 - f.eks. <http://www.brreg.no/>



Flere bobiler på norske veier!

- Tenk deg at du ønsker å starte bobilutleie
 - Du bør være der turistene er, og
 - så vidt mulig, langt fra konkurrenter
- Du trenger å vite hvor turistene drar,
 - f.eks. <http://statbank.ssb.no/statistikbanken/>
- .. samt å vite hvorvidt disse stedene allerede har slike utleiefirmaer
 - f.eks. <http://www.brreg.no/>
- Du trenger altså å kombinere data fra flere kilder



Flere bobiler på norske veier!

- Tenk deg at du ønsker å starte bobilutleie
 - Du bør være der turistene er, og
 - så vidt mulig, langt fra konkurrenter
- Du trenger å vite hvor turistene drar,
 - f.eks. <http://statbank.ssb.no/statistikbanken/>
- .. samt å vite hvorvidt disse stedene allerede har slike utleiefirmaer
 - f.eks. <http://www.brreg.no/>
- Du trenger altså å kombinere data fra flere kilder
 - i dette tilfelle BRReg og SSB



Hva gjør du?

- Venter på at noen (BRReg? SSB?) publiserer et distribusjonskart?

Hva gjør du?

- Venter på at noen (BRReg? SSB?) publiserer et distribusjonskart?
- Sammenstiller dataene manuelt?

Hva gjør du?

- Venter på at noen (BRReg? SSB?) publiserer et distribusjonskart?
- Sammenstiller dataene manuelt?
- hva om du ønsker å gjøre tilsvarende analyser flere ganger, dvs.

Hva gjør du?

- Venter på at noen (BRReg? SSB?) publiserer et distribusjonskart?
- Sammenstiller dataene manuelt?
- hva om du ønsker å gjøre tilsvarende analyser flere ganger, dvs.
 - over tid, og

Hva gjør du?

- Venter på at noen (BRReg? SSB?) publiserer et distribusjonskart?
- Sammenstiller dataene manuelt?
- hva om du ønsker å gjøre tilsvarende analyser flere ganger, dvs.
 - over tid, og
 - mhp. andre næringer?

Hva gjør du?

- Venter på at noen (BRReg? SSB?) publiserer et distribusjonskart?
- Sammenstiller dataene manuelt?
- hva om du ønsker å gjøre tilsvarende analyser flere ganger, dvs.
 - over tid, og
 - mhp. andre næringer?
- Hva med andre typer data i andre kombinasjoner, eksempelvis

Hva gjør du?

- Venter på at noen (BRReg? SSB?) publiserer et distribusjonskart?
- Sammenstiller dataene manuelt?
- hva om du ønsker å gjøre tilsvarende analyser flere ganger, dvs.
 - over tid, og
 - mhp. andre næringer?
- Hva med andre typer data i andre kombinasjoner, eksempelvis
 - boliger til salgs + byggesaker

Hva gjør du?

- Venter på at noen (BRReg? SSB?) publiserer et distribusjonskart?
- Sammenstiller dataene manuelt?
- hva om du ønsker å gjøre tilsvarende analyser flere ganger, dvs.
 - over tid, og
 - mhp. andre næringer?
- Hva med andre typer data i andre kombinasjoner, eksempelvis
 - boliger til salgs + byggesaker
 - idrettsanlegg + luftforurensning

Hva gjør du?

- Venter på at noen (BRReg? SSB?) publiserer et distribusjonskart?
- Sammenstiller dataene manuelt?
- hva om du ønsker å gjøre tilsvarende analyser flere ganger, dvs.
 - over tid, og
 - mhp. andre næringer?
- Hva med andre typer data i andre kombinasjoner, eksempelvis
 - boliger til salgs + byggesaker
 - idrettsanlegg + luftforurensning
 - åpne turisthytter + sjenkebevillinger (privat + offentlig)

Hva gjør du?

- Venter på at noen (BRReg? SSB?) publiserer et distribusjonskart?
- Sammenstiller dataene manuelt?
- hva om du ønsker å gjøre tilsvarende analyser flere ganger, dvs.
 - over tid, og
 - mhp. andre næringer?
- Hva med andre typer data i andre kombinasjoner, eksempelvis
 - boliger til salgs + byggesaker
 - idrettsanlegg + luftforurensning
 - åpne turisthytter + sjenkebevillinger (privat + offentlig)
 - etc.

Informasjon og innovasjon

Informasjon er en avgjørende faktor i innovasjon:

Informasjon og innovasjon

Informasjon er en avgjørende faktor i innovasjon:

- bruksområdene for et datasett er en 'fremskridende åpenbaring'

Informasjon og innovasjon

Informasjon er en avgjørende faktor i innovasjon:

- bruksområdene for et datasett er en 'fremskridende åpenbaring'
- potensialet i et datasett forsterkes gjennom

Informasjon og innovasjon

Informasjon er en avgjørende faktor i innovasjon:

- bruksområdene for et datasett er en 'fremskridende åpenbaring'
- potensialet i et datasett forsterkes gjennom
 - oppfinnsom gjenbruk

Informasjon og innovasjon

Informasjon er en avgjørende faktor i innovasjon:

- bruksområdene for et datasett er en 'fremskridende åpenbaring'
- potensialet i et datasett forsterkes gjennom
 - oppfinnsom gjenbruk
 - kombinasjon med andre datasett

Informasjon og innovasjon

Informasjon er en avgjørende faktor i innovasjon:

- bruksområdene for et datasett er en 'fremskridende åpenbaring'
- potensialet i et datasett forsterkes gjennom
 - oppfinnsom gjenbruk
 - kombinasjon med andre datasett

Informasjon og innovasjon

Informasjon er en avgjørende faktor i innovasjon:

- bruksområdene for et datasett er en 'fremskridende åpenbaring'
- potensialet i et datasett forsterkes gjennom
 - oppfinnsom gjenbruk
 - kombinasjon med andre datasett

Fra EUs gjenbruksdirektiv:

Content production has given rise to rapid job creation and continues to do so (...) Public sector information is an important primary material for digital content products and services and will become an even more important content resource with the development of wireless content services.

Åpne data og samfunnsøkonomi

Forholdsvis velstuderte effekter:

Åpne data og samfunnsøkonomi

Forholdsvis velstuderte effekter:

- bidrar til å sikre ansvarlighet i forvaltningen, f.eks i forhold til.

Åpne data og samfunnsøkonomi

Forholdsvis velstuderte effekter:

- bidrar til å sikre ansvarlighet i forvaltningen, f.eks i forhold til.
 - offentlig forbruk

Åpne data og samfunnsøkonomi

Forholdsvis velstuderte effekter:

- bidrar til å sikre ansvarlighet i forvaltningen, f.eks i forhold til.
 - offentlig forbruk
 - finansiering av partier og organisasjoner

Åpne data og samfunnsøkonomi

Forholdsvis velstuderte effekter:

- bidrar til å sikre ansvarlighet i forvaltningen, f.eks i forhold til.
 - offentlig forbruk
 - finansiering av partier og organisasjoner
 - budsjettering

Åpne data og samfunnsøkonomi

Forholdsvis velstuderte effekter:

- bidrar til å sikre ansvarlighet i forvaltningen, f.eks i forhold til.
 - offentlig forbruk
 - finansiering av partier og organisasjoner
 - budsjettering

- bidrar til å opprettholde en velinformert offentlighet

Åpne data og samfunnsøkonomi

Forholdsvis velstuderte effekter:

- bidrar til å sikre ansvarlighet i forvaltningen, f.eks i forhold til.
 - offentlig forbruk
 - finansiering av partier og organisasjoner
 - budsjettering
- bidrar til å opprettholde en velinformert offentlighet
- er vesentlig for overvåkning av samspillet mellom miljø og helse

Åpne data og samfunnsøkonomi

Forholdsvis velstuderte effekter:

- bidrar til å sikre ansvarlighet i forvaltningen, f.eks i forhold til.
 - offentlig forbruk
 - finansiering av partier og organisasjoner
 - budsjettering
- bidrar til å opprettholde en velinformert offentlighet
- er vesentlig for overvåkning av samspillet mellom miljø og helse
 - luftforurensing,

Åpne data og samfunnsøkonomi

Forholdsvis velstuderte effekter:

- bidrar til å sikre ansvarlighet i forvaltningen, f.eks i forhold til.
 - offentlig forbruk
 - finansiering av partier og organisasjoner
 - budsjettering
- bidrar til å opprettholde en velinformert offentlighet
- er vesentlig for overvåking av samspillet mellom miljø og helse
 - luftforurensing,
 - strålingsfare, spredningen av radiomaster

forts..

... ennvidere:



forts..

... ennvidere:

- understøtter analytisk/kritisk journalistikk;



forts..

... ennvidere:

- understøtter analytisk/kritisk journalistikk;
 - brenner verneverdige bygninger oftere enn andre?



forts..

... ennvidere:

- understøtter analytisk/kritisk journalistikk;
 - brenner verneverdige bygninger oftere enn andre?
 - hvilke bygninger eies av registrerte foretak?



forts..

... ennvidere:

- understøtter analytisk/kritisk journalistikk;
 - brenner verneverdige bygninger oftere enn andre?
 - hvilke bygninger eies av registrerte foretak?
 - koblinger mellom politikere og næringsliv



forts..

... ennvidere:

- understøtter analytisk/kritisk journalistikk;
 - brenner verneverdige bygninger oftere enn andre?
 - hvilke bygninger eies av registrerte foretak?
 - koblinger mellom politikere og næringsliv

- Stimulerer vekst og gründervirksomhet



forts..

... ennvidere:

- understøtter analytisk/kritisk journalistikk;
 - brenner verneverdige bygninger oftere enn andre?
 - hvilke bygninger eies av registrerte foretak?
 - koblinger mellom politikere og næringsliv
- Stimulerer vekst og gründervirksomhet
 - skattbar virksomhet



Outline

- 1 Verdien av åpne data
- 2 Hvordan gjøre offentlige data mer tilgjengelige?**
- 3 LOD som publiseringsparadigme
- 4 Byantikvarens gule liste
- 5 Fremblikk

Generelle prinsipper for gjenbrukbarhet

Noen grunnregler:

Generelle prinsipper for gjenbrukbarhet

Noen grunnregler:

- Gjør ingen antagelser om brukeren av datasettet

Generelle prinsipper for gjenbrukbarhet

Noen grunnregler:

- Gjør ingen antagelser om brukeren av datasettet
 - maskin eller menneske? Lesbart uansett!

Generelle prinsipper for gjenbrukbarhet

Noen grunnregler:

- Gjør ingen antagelser om brukeren av datasettet
 - maskin eller menneske? Lesbart uansett!
 - søk, sortering, transformering—prosa er poesi for maskiner!

Generelle prinsipper for gjenbrukbarhet

Noen grunnregler:

- Gjør ingen antagelser om brukeren av datasettet
 - maskin eller menneske? Lesbart uansett!
 - søk, sortering, transformering—prosa er poesi for maskiner!
- benytt et åpent ikke-proprietært format

Generelle prinsipper for gjenbrukbarhet

Noen grunnregler:

- Gjør ingen antagelser om brukeren av datasettet
 - maskin eller menneske? Lesbart uansett!
 - søk, sortering, transformering—prosa er poesi for maskiner!
- benytt et åpent ikke-proprietært format
 - minst risiko for avleggshet

Generelle prinsipper for gjenbrukbarhet

Noen grunnregler:

- Gjør ingen antagelser om brukeren av datasettet
 - maskin eller menneske? Lesbart uansett!
 - søk, sortering, transformering—prosa er poesi for maskiner!
- benytt et åpent ikke-proprietært format
 - minst risiko for avleggshet
- benytt terminologi som

Generelle prinsipper for gjenbrukbarhet

Noen grunnregler:

- Gjør ingen antagelser om brukeren av datasettet
 - maskin eller menneske? Lesbart uansett!
 - søk, sortering, transformering—prosa er poesi for maskiner!
- benytt et åpent ikke-proprietært format
 - minst risiko for avleggshet
- benytt terminologi som
 - er utvetydig, også utenfor sin opprinnelige kontekst

Generelle prinsipper for gjenbrukbarhet

Noen grunnregler:

- Gjør ingen antagelser om brukeren av datasettet
 - maskin eller menneske? Lesbart uansett!
 - søk, sortering, transformering—prosa er poesi for maskiner!
- benytt et åpent ikke-proprietært format
 - minst risiko for avleggshet
- benytt terminologi som
 - er utvetydig, også utenfor sin opprinnelige kontekst
 - .. slik at man kan peke på samme ting fra flere steder

Generelle prinsipper for gjenbrukbarhet

Noen grunnregler:

- Gjør ingen antagelser om brukeren av datasettet
 - maskin eller menneske? Lesbart uansett!
 - søk, sortering, transformering—prosa er poesi for maskiner!
- benytt et åpent ikke-proprietært format
 - minst risiko for avleggshet
- benytt terminologi som
 - er utvetydig, også utenfor sin opprinnelige kontekst
 - .. slik at man kan peke på samme ting fra flere steder



Outline

- 1 Verdien av åpne data
- 2 Hvordan gjøre offentlige data mer tilgjengelige?
- 3 LOD som publiseringsparadigme**
- 4 Byantikvarens gule liste
- 5 Fremblikk

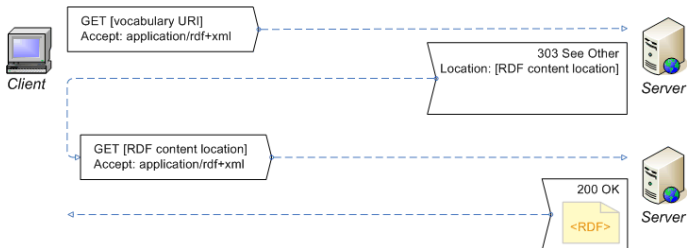
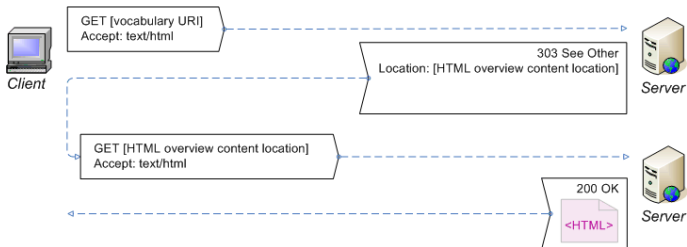
LOD som publiseringsparadigme

Prinsipper for lenkede data:

- 1 bruk URLer som navn på ting
- 2 oppslag på en URL skal gi en beskrivelse av tingen
- 3 gi en RDF beskrivelse av den samme tingen på den samme URLen
- 4 inkluder lenker til relaterte data



Innholdsforhandling



Fordeler ved LOD

- Data overflyter applikasjonsbarrierer (eksempel senere)
- Data er maskinleselig (til forskjell fra feks PDF)
- Meta-data flyttes oppstrøms (eksempel senere)
- Data er selvdokumenterende (eksempel senere)
- Data kan traverseres med HTTP mekanismer (eksempel senere)
- RDF kan kombineres med beregnbare begrepsmodeller (ontologier)

Outline

- 1 Verdien av åpne data
- 2 Hvordan gjøre offentlige data mer tilgjengelige?
- 3 LOD som publiseringsparadigme
- 4 Byantikvarens gule liste**
- 5 Fremblikk

Et bittelite utdrag av den gule listen

Eiendom	Gate	Art	Nord	Øst
209/367	7. juni plassen	Kontor	6643181	596689
30/150	Abbedikollen	Bolig	6644446	593153
233/435	Svingen	Arbeiderbolig	6641752	599503
218/13	Akersbakken	Gravplass	6644221	597721
208/4	Akersgata	Teater	6643156	597504
207/1	Akershus festning	Vaktstue	6642303	597164

Figur: Horisontalt og vertikalt beskåret utdrag av den gule listen

Modellering: Transformasjon av listen til RDF



En 'naiv' fremgangsmåte som fungerer for alle tabulære data:

Modellering: Transformasjon av listen til RDF



En 'naiv' fremgangsmåte som fungerer for alle tabulære data:

- 1 la hver rad *s* modelleres som en distinkt entitet/unik URL

Modellering: Transformasjon av listen til RDF



En 'naiv' fremgangsmåte som fungerer for alle tabulære data:

- 1 la hver rad s modelleres som en distinkt entitet/unik URL
- 2 la hver kolonne p modelleres som et distinkt predikat/unik URL

Modellering: Transformasjon av listen til RDF



En 'naiv' fremgangsmåte som fungerer for alle tabulære data:

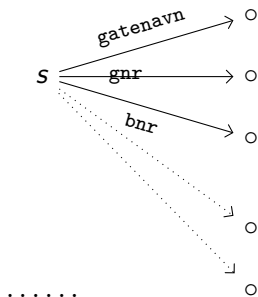
- 1 la hver rad s modelleres som en distinkt entitet/unik URL
- 2 la hver kolonne p modelleres som et distinkt predikat/unik URL
- 3 la enhver celle o i matrisen som ikke er en konkret datatype (streng, desimaltall etc.) representeres som en distinkt entitet/unik URL

Modellering: Transformasjon av listen til RDF



En 'naiv' fremgangsmåte som fungerer for alle tabulære data:

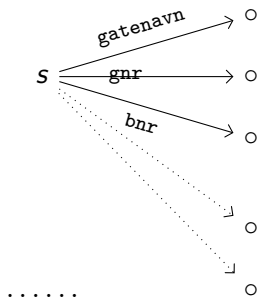
- 1 la hver rad s modelleres som en distinkt entitet/unik URL
- 2 la hver kolonne p modelleres som et distinkt predikat/unik URL
- 3 la enhver celle o i matrisen som ikke er en konkret datatype (streng, desimaltall etc.) representeres som en distinkt entitet/unik URL
- 4 relatér enhver celle o i matrisens posisjon (s, p) ved hjelp av en RDF trippel/påstand (s, p, o)



Figur: Resultat av den naive fremgangsmåten

Transformasjon

- Rad \mapsto subject S, eks. rad 23 \mapsto `http://.../{matrikelkode}/{lokasjon}`
- Kolonne \mapsto RDF predikat, eks. gate \mapsto `http://{et stedsvokabular}/gatenavn`
- Celle \mapsto eksempelvis streng; Borger Withs gate \mapsto "Borger Withs gate"



Figur: Resultat av den naive fremgangsmåten

Transformasjon

- Rad \mapsto subject S, eks. rad 23 \mapsto `http://.../{matrikelkode}/{lokasjon}`
- Kolonne \mapsto RDF predikat, eks. gate \mapsto `http://{et stedsvokabular}/gatenavn`
- Celle \mapsto eksempelvis streng; Borger Withs gate \mapsto "Borger Withs gate"

En mer informert modellering

Den gule listen ...

- har en implisitt *mereologisk* struktur:

En mer informert modellering

Den gule listen ...

- har en implisitt *mereologisk* struktur:
 - forskjellige fysiske strukturer utgjør deler av den samme eiendommen

En mer informert modellering

Den gule listen ...

- har en implisitt *mereologisk* struktur:
 - forskjellige fysiske strukturer utgjør deler av den samme eiendommen
 - ... trenger et vokabular for å uttrykke del-helhet relasjoner

En mer informert modellering

Den gule listen ...

- har en implisitt *mereologisk* struktur:
 - forskjellige fysiske strukturer utgjør deler av den samme eiendommen
 - ... trenger et vokabular for å uttrykke del-helhet relasjoner
- inneholder implisitt typeinformasjon

En mer informert modellering

Den gule listen ...

- har en implisitt *mereologisk* struktur:
 - forskjellige fysiske strukturer utgjør deler av den samme eiendommen
 - ... trenger et vokabular for å uttrykke del-helhet relasjoner
- inneholder implisitt typeinformasjon
 - hver rad angir en fysisk struktur

En mer informert modellering

Den gule listen ...

- har en implisitt *mereologisk* struktur:
 - forskjellige fysiske strukturer utgjør deler av den samme eiendommen
 - ... trenger et vokabular for å uttrykke del-helhet relasjoner
- inneholder implisitt typeinformasjon
 - hver rad angir en fysisk struktur
 - hver eiendom angir en juridisk entitet

En mer informert modellering

Den gule listen ...

- har en implisitt *mereologisk* struktur:
 - forskjellige fysiske strukturer utgjør deler av den samme eiendommen
 - ... trenger et vokabular for å uttrykke del-helhet relasjoner
- inneholder implisitt typeinformasjon
 - hver rad angir en fysisk struktur
 - hver eiendom angir en juridisk entitet
 - dette kan gjøres eksplisitt vha. eksisterende vokabularer

En mer informert modellering

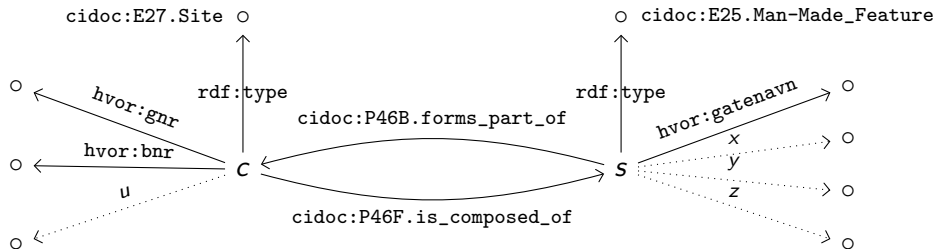
Den gule listen ...

- har en implisitt *mereologisk* struktur:
 - forskjellige fysiske strukturer utgjør deler av den samme eiendommen
 - ... trenger et vokabular for å uttrykke del-helhet relasjoner
- inneholder implisitt typeinformasjon
 - hver rad angir en fysisk struktur
 - hver eiendom angir en juridisk entitet
 - dette kan gjøres eksplisitt vha. eksisterende vokabularer
 - fysiske strukturer \mapsto `cidoc:E25.Man-Made_Feature`

En mer informert modellering

Den gule listen ...

- har en implisitt *mereologisk* struktur:
 - forskjellige fysiske strukturer utgjør deler av den samme eiendommen
 - ... trenger et vokabular for å uttrykke del-helhet relasjoner
- inneholder implisitt typeinformasjon
 - hver rad angir en fysisk struktur
 - hver eiendom angir en juridisk entitet
 - dette kan gjøres eksplisitt vha. eksisterende vokabularer
 - fysiske strukturer \mapsto `cidoc:E25.Man-Made_Feature`
 - eiendommer \mapsto `cidoc:E27.Sites`



Figur: En mer eksplisitt modellering

Outline

- 1 Verdien av åpne data
- 2 Hvordan gjøre offentlige data mer tilgjengelige?
- 3 LOD som publiseringsparadigme
- 4 Byantikvarens gule liste
- 5 Fremblikk**

LOD har utviklet seg mye siden dette:

- Praktisk:
 - Definisjon av The Linked Data API
 - Implementasjoner av The Linked Data API, eksempelvis ELDA
 - Visualiseringsverktøy, eksempelvis sqvizler
- Teoretisk:
 - Identifisering av et OWL-ontologifragment for LOD
 - Forskning på datafederering og -integrasjon over LOD
 - Teori som understøtter konservativ modellering
 - eks. Skjæveland og Stolpe "Expressing reservation about incoming data" (COLD 2013)