

Semantisk web, emnekart og ontologier

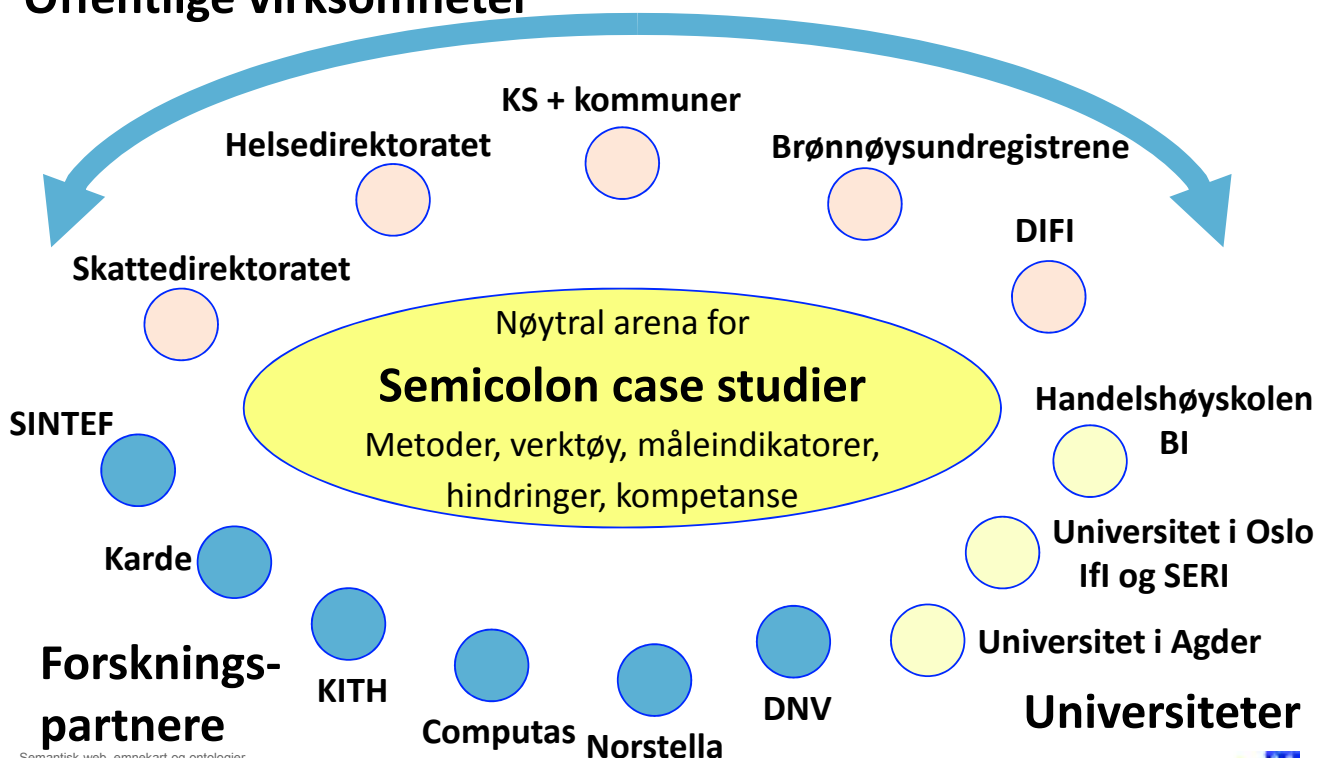
Forelesning hos AFIN i kurs DRI3010 - Høst 2011 (Systemutvikling, offentlig styring og demokrati)

Per Myrseth
28. september 2011



Semicolon II-deltakere

Offentlige virksomheter



Rationale for Semicolon

- Full interoperability is highly prioritised in Norway
- Norwegian public sector is increasingly user-oriented, aiming at the provision of electronic services on a 24/7 basis
- Citizens and businesses should report information only once and thus make life easier for citizens and businesses
- EU's Lisbon strategy claims that a good public sector represents a competitive advantage for businesses
- Collaboration between public organisations, citizens and businesses is necessary to provide a more effective and efficient public sector

Main topics in Semicolon II

- **Metadata and semantics**
 - Internal information governance
 - Linked Open Data and Public Data
 - Improving the metadata repository SERES
- **eDialoger**
- **Measuring eCollaboration**
- **Juridical issues on eCollaboration and metadata**
- **Plattform for collaboration**

Fra forelesningsplanen

28.9 Semantisk web, emnekart og ontologier. (Per Myrseth)

Pensum:

- Tim Berners-Lee, James Hendler & Ora Lassila: The Semantic Web, 2001. Artikkel fra Scientific American Magazine.
- Engels R og Myrseth P: Åpne og lenkede data. En informasjonsinfrastruktur for elektronisk samhandling. Rapport nr. 2011-276, revisjon nr. 1. Det Norske Veritas, februar 2011, kap. 3 og 4 (10 s).
- Om XML: <http://www.w3.org/TR/REC-xml/>
- <http://www.brreg.no/samordning/semantikk/>
- Om semantisk web: http://no.wikipedia.org/wiki/Semantisk_web
- Miazine, K: Emnekart og juridisk kunnskapsforvaltning

Kilde: http://www.uio.no/studier/emner/jus/afin/DRI3010/h11/undervisningsmateriale/Forelesningsplan_3010_H2011_dws_aj.pdf

Intro og trender

Collaboration anyone?



Semantic Web og Linked Open data

- *The first step is putting data on the Web in a form that machines can naturally understand, or converting it to that form. This creates what I call a Semantic Web – a web of data that can be processed directly or indirectly by machines.*

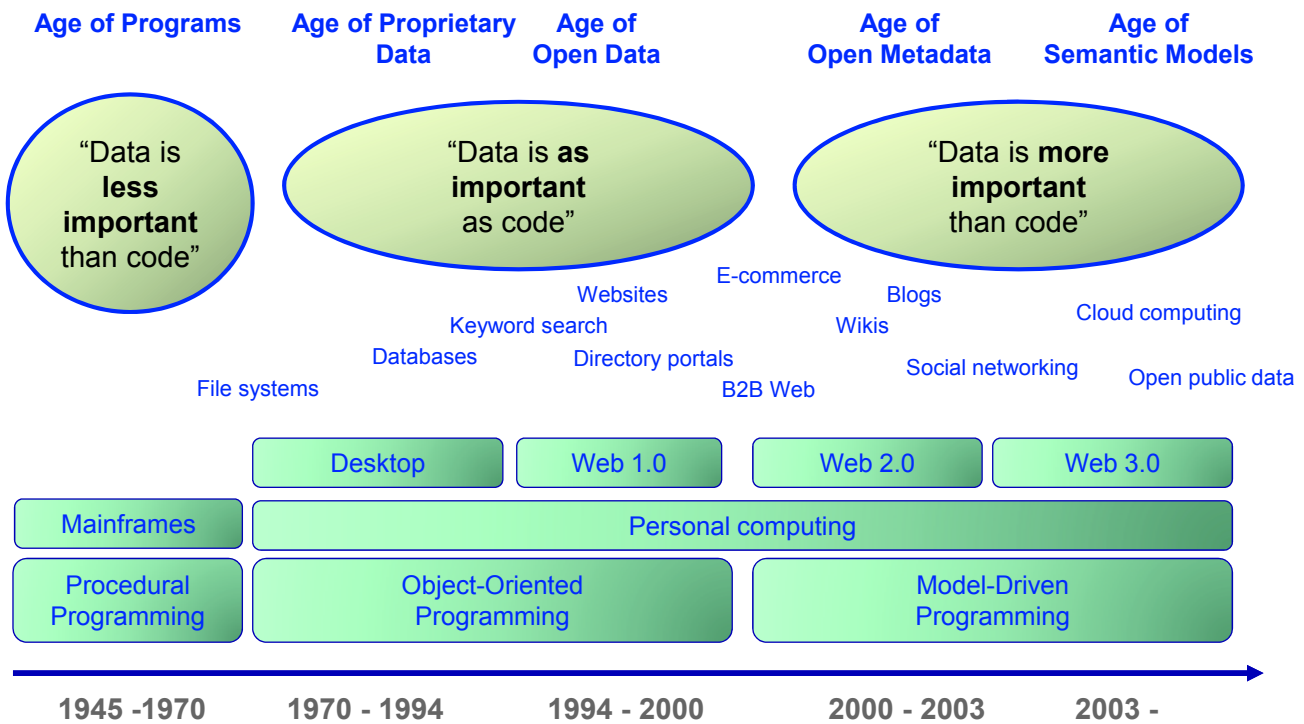
Tim Berners Lee – Weaving the Semantic web



- Building a web for open, linked data that could do for numbers what the Web did for words, pictures, video etc.
- Unlock our data and reframe the way we use it together. *1

*1 http://www.ted.com/index.php/talks/tim_berniers_lee_on_the_next_web.html

Semantic Technologies: From code to meaning



Semantisk web, emnekart og ontologier
28. september 2011
© Det Norske Veritas AS. All rights reserved.

Inspired by: Nova Spivack (2007): How the WebOS evolves?
See http://novaspivack.tylenad.com/nova_spivacks_weblog/2007/02/steps_towards_a.html



Volume of data increases

Most people in IT already have read that researchers believe digital business data to be stored is growing at a 40 to 60 percent clip each year.

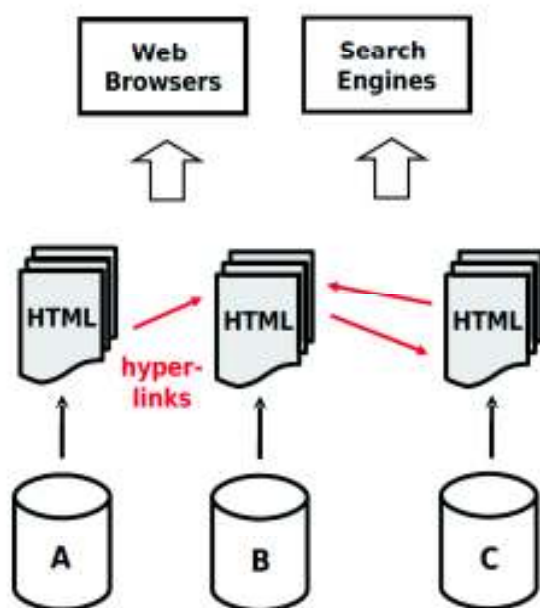
So it makes sense that this explosion in stored files, logs and other data has become one of biggest challenges to solve for large enterprises.

Kilde: eksempelvis Gartner:
<http://www.eweek.com/c/a/Data-Storage/Data-Growth-Now-a-FirstTier-Challenge-for-Enterprises-Gartner-Reports-273073/>

Semantisk web, emnekart og ontologier
28. september 2011
© Det Norske Veritas AS. All rights reserved.



The Classic Web



Single global information space

2. URLs as
 - globally unique IDs
 - retrieval mechanism
3. HTML as shared content format
4. Hyperlinks

Shortcomings

- Content is not well structured
- You can not ask expressive queries
- You can not process content within applications

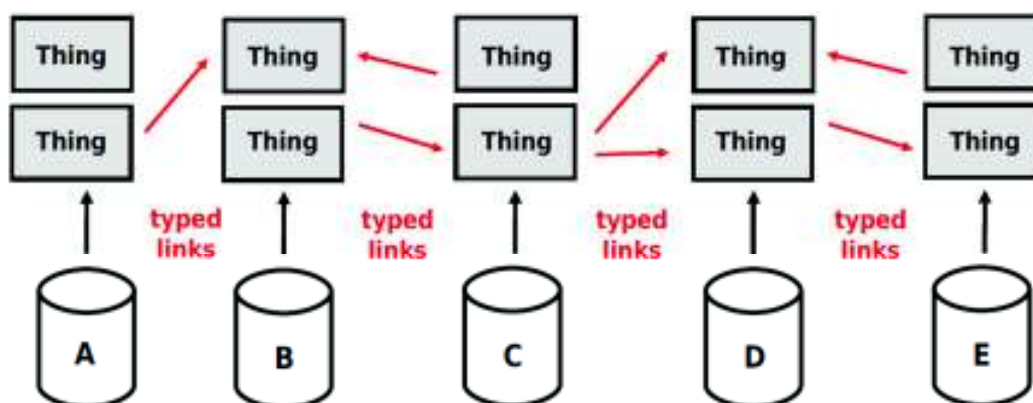
Christian Bizer: How to Publish Linked Data on the Web - Introduction (10/27/2008)

© Del Mundo Ventos A.S. All rights reserved.

MANAGING RISK

Linked Data

- Use Semantic Web technologies to
2. publish structured data on the Web,
 3. set links between data from one data source to data within other data sources.

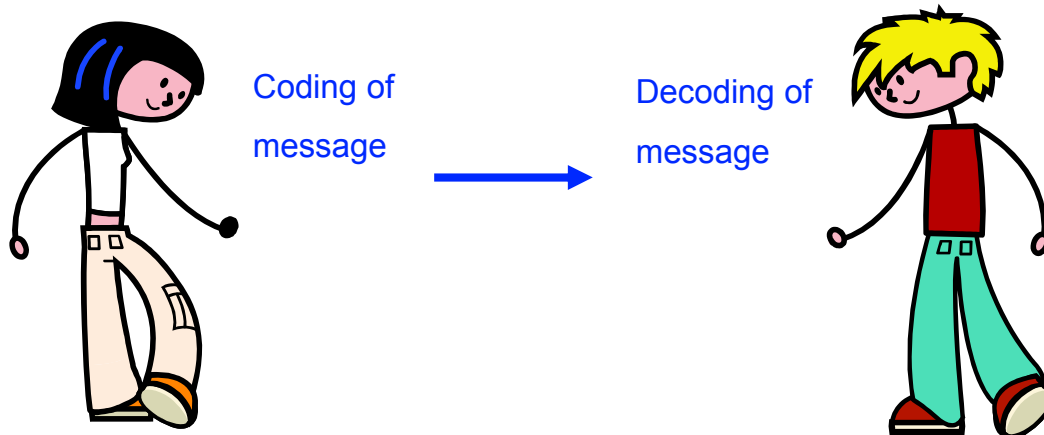


Christian Bizer: How to Publish Linked Data on the Web - Introduction (10/27/2008)

© Del Mundo Ventos A.S. All rights reserved.

MANAGING RISK

Communication between persons



Exchange: Sound, in a sequence

What is exchanged: Terms in a common language

Rules for using terms: Grammar

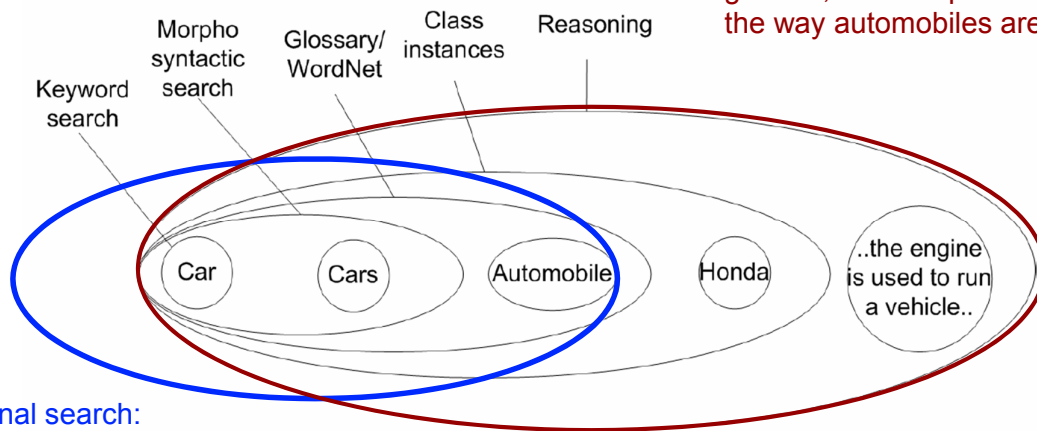
Context: Know to both sender and receiver

Types of interoperability

- Technical
 - Communication protocols and security, exchange formats
- Semantic
 - The meaning of exchanged data is well understood. Trust and quality in data.
- Organizational
 - Goals, budgets, knowledge and processes are aligned
- Juridical
 - Legal compliance. Related to processes, actions, agreements, data exchange, data governance etc.
- Political
 - Goals and frame conditions are aligned

Semantic vs Traditional Search: Concepts instead of Keywords

Semantic search:
A search for "car" returns documents discussing cars in general, Honda specifically, or the way automobiles are built



Traditional search:
"car" returns documents containing the keywords *car*, *cars*, *automobile*

Begrepsdannelse, kvalitet og forvaltning

Back Search Favorites Go Links
 Address http://www.dn.no/forsiden/borsMarked/article1867613.ece



Makroøkonom Bjørn-Roger Wilhelmsen i First Securities ble svært overrasket da han oppdaget NAV-feilen. Foto: Per Thrana

- Vi ble tatt på sengen

I halvannet år har Nav rapportert feil arbeidsledighetstall. Det kan ha påvirket Norges Bank, tror makroøkonom.

Cecilie Langum Becker
 Publisert: 26.03.2010 - 11:20 Oppdatert: 29.03.2010 - 07:21

MANAGING RISK

Semantisk web, em
 28. september 2011
 © Det Norske Veritas

First ontology brought to court

Trying Leviathan: The Nineteenth-Century New York Court Case That Put the Whale on Trial and Challenged the Order of Nature

by D. Graham Burnett



- Synopses & Reviews
- Table of Contents
- Comment on this title and you could win free books!
- See more like this:

ISBN13: 9780691129501
 ISBN10: 0691129500
 All Product Details

Only 1 left in stock at \$29.95!



Synopses & Reviews

Publisher Comments:

In *Moby-Dick*, Ishmael declares, Be it known that, waving at upon holy Jonah to back me. Few readers today know just I sides in one of the great controversies of the early nineteenth

In *Trying Leviathan*, D. Graham Burnett recovers the strang against the then-popular (and biblically sanctioned) view th was fish oil and therefore subject to state inspection. But th nature, and how we know it, was at stake. Burnett vividly re- portous philosophers, Jacobin lawyers, look the witness si middle of the century between Linnaeus and Darwin, the tri comfortable biblical categories, and in came near sorting me

The trial in 1818 lasted for three days.

The jury finding that before the law, at least, Mr. Judd was in the wrong.

"A whale is a fish," wrote both the *New-York Gazette* and the *Evening Post*.

Semantisk web, em
 28. september 2011
 © Det Norske Veritas AS. All rights reserved.

<http://www.powells.com/biblio/1-9780691129501-0>



Skattebetalings- håndboken

1. utgave 2010

- Forord:
-Boken er for disse ment som et bidrag til a forenkle arbeidet og til å oppnå en mest mulig ensartet praksis i og på tvers av etatene. En annen målgruppe er eksterne aktører og deres rådgivere. Vi håper at stoffet kan bidra til en enkel forståelse av regelverket.....
- Boken lister 64 relevante lover.

Systematisering av betydningen av termer, relatert til kjøretøy

Hjelpemiddelsentralen

Toll og avgift

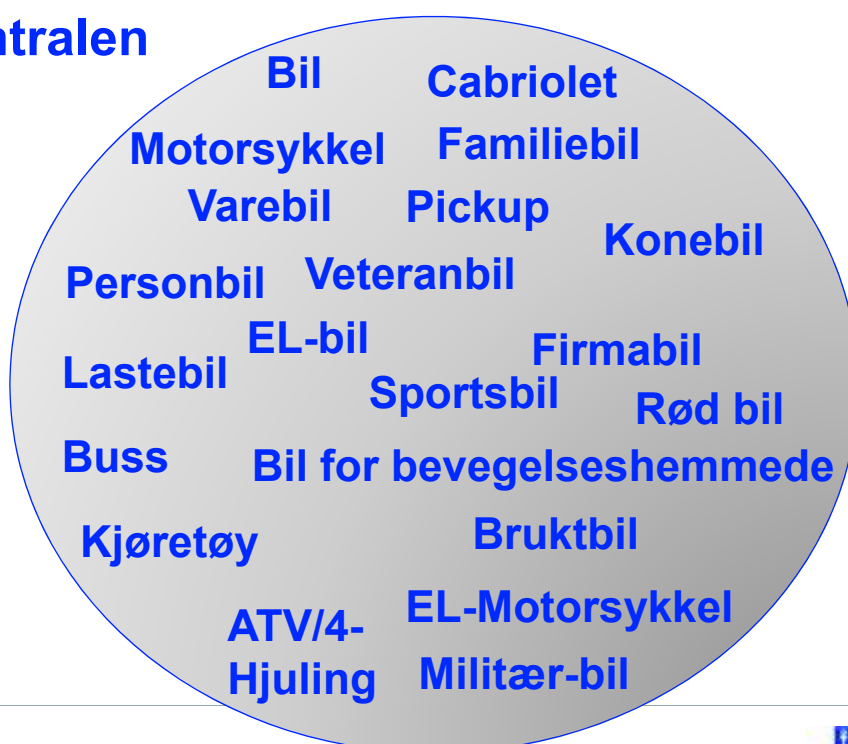
Forsikring

Autofil, TV

Bompenger

Skattelegging

Bil-produsenter

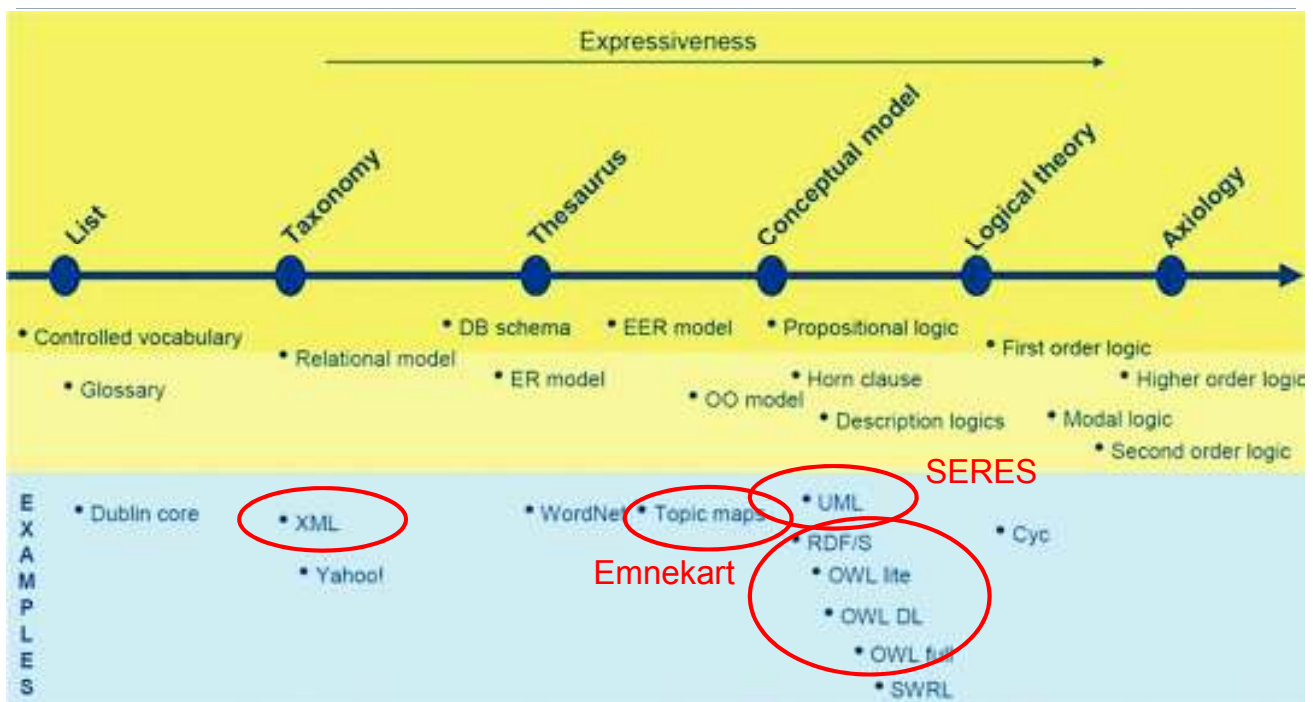


Uttrykkskraft og formalisme i begrepsmodeller

Ordliste	Taksonomi	Tesaurus	Ontologi
<u>a</u> <u>b</u> <u>c</u> <u>x</u> <u>y</u> <u>z</u>			<p>Logikk for knyttet til assosiasjoner</p>
Kontrollert liste av begreper	Hierarki av begreper, vist med dobbel arv over	Som taksonomi, men tillater flere typer relasjoner mellom begreper	Som tesaurus, men med formelle skranker

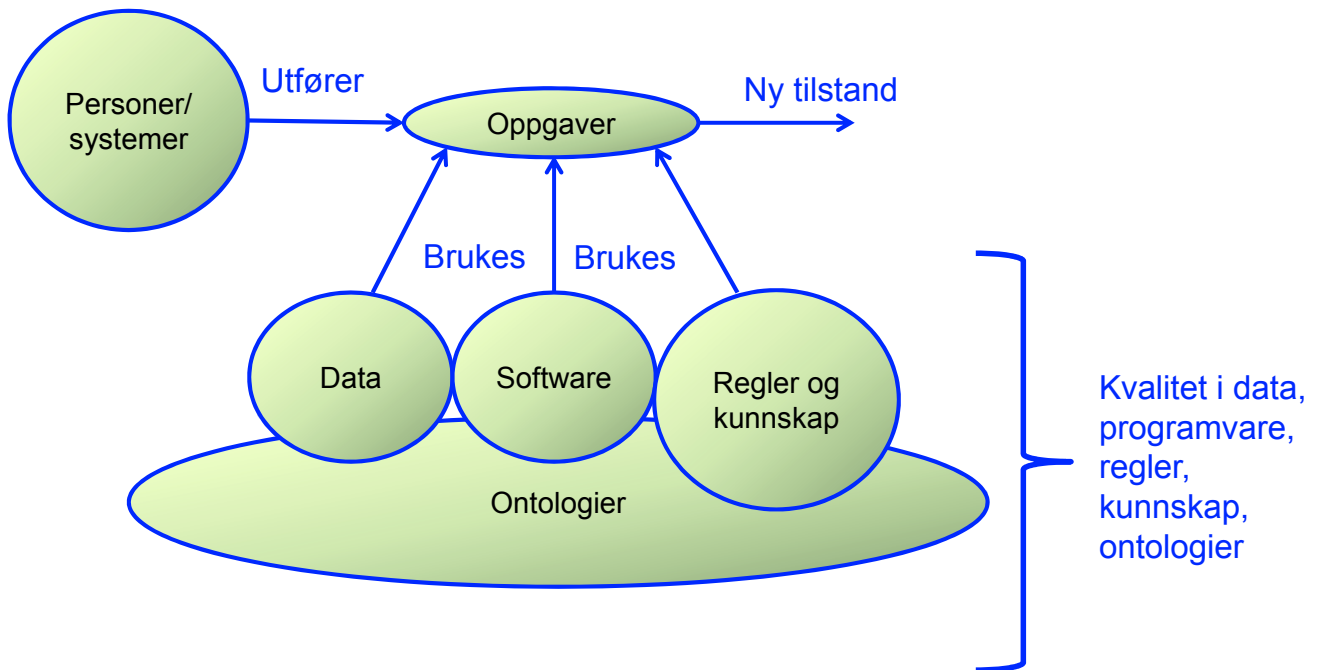
Økende kompleksitet og kostnad

Uttrykkskraft og formalisme i begrepsmodeller II

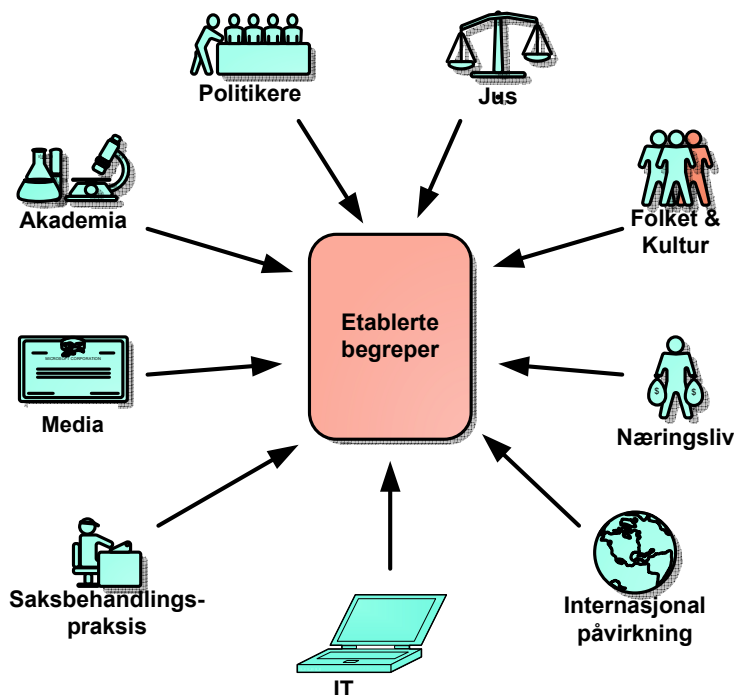


Source: Davis, M. 2006, Semantic Wave 2006 [Davis M]

Hvor passer ontologier inn?



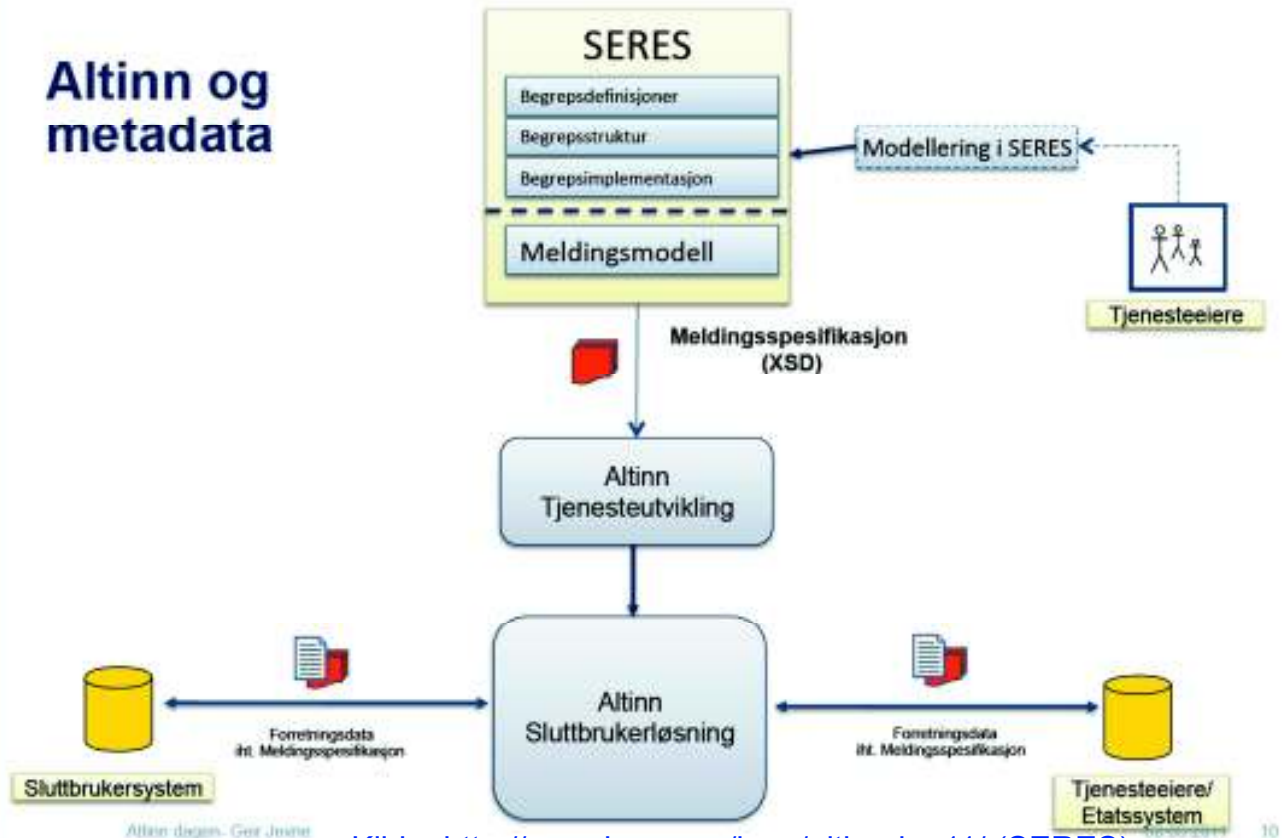
Hvem har definisjonsmakt for hva?



Påvirkes av :

- Samhandling
- Samfunnstrender som globalisering,
- Teknologit utvikling
- Anvendelse

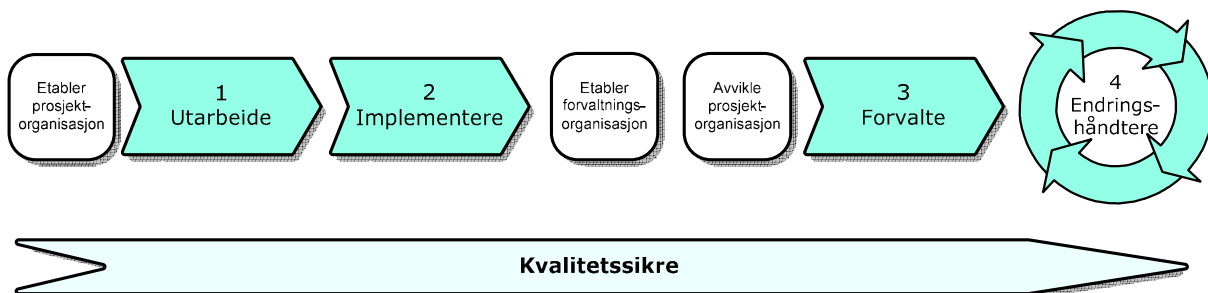
Altinn og metadata



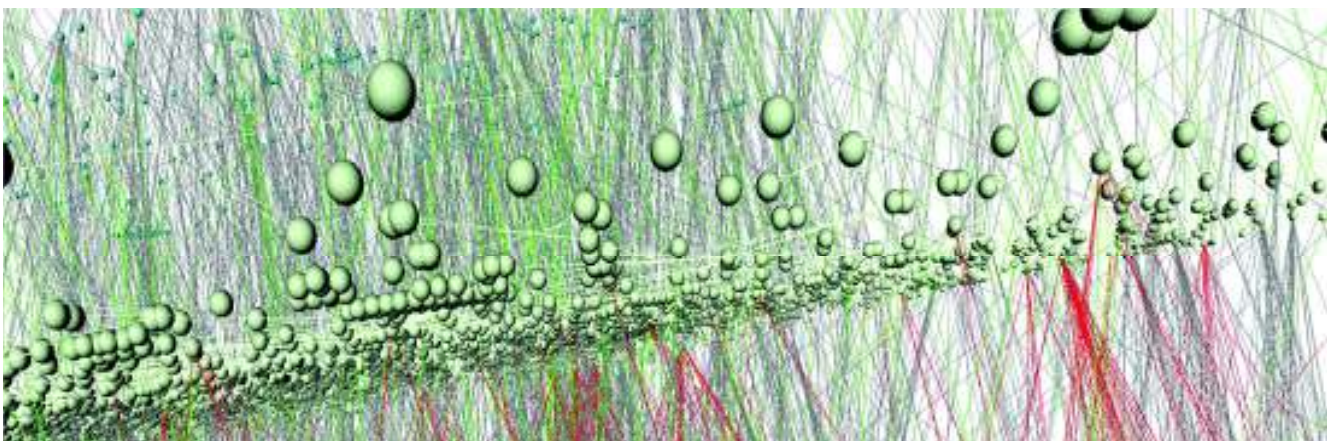
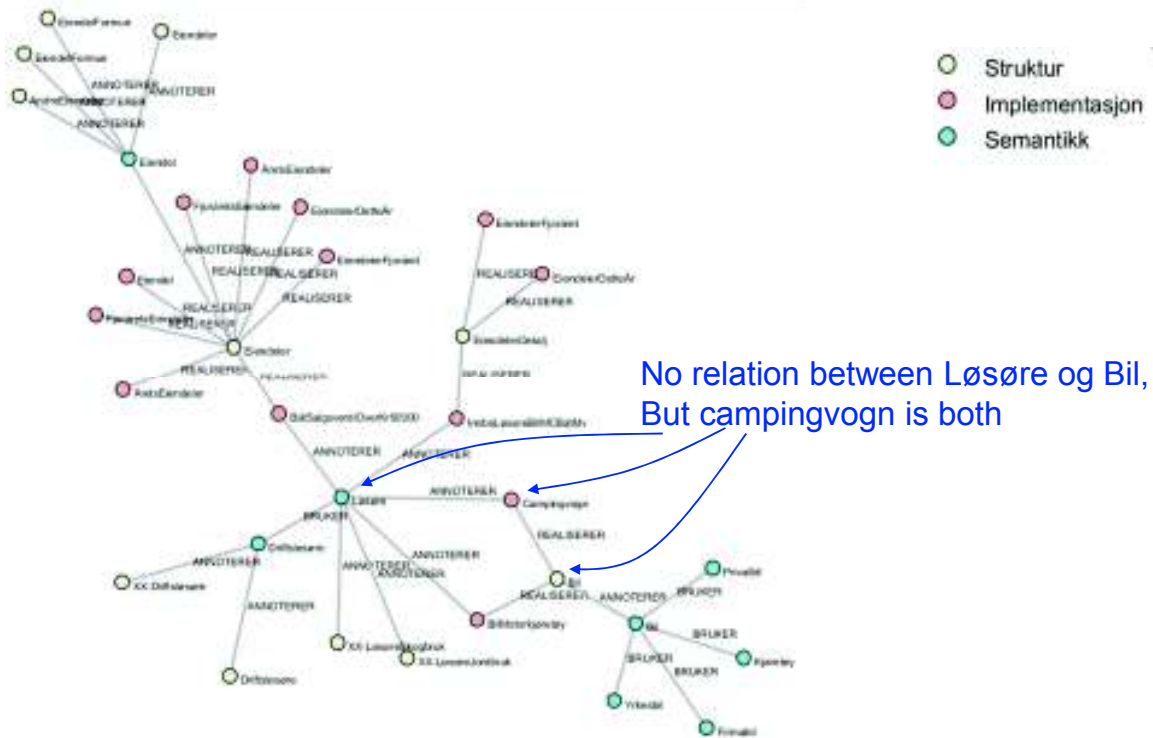
Kilde: <http://www.brreg.no/kurs/altinndag11/> (SERES)

Lifecycle management of terms and concepts used in eGov services

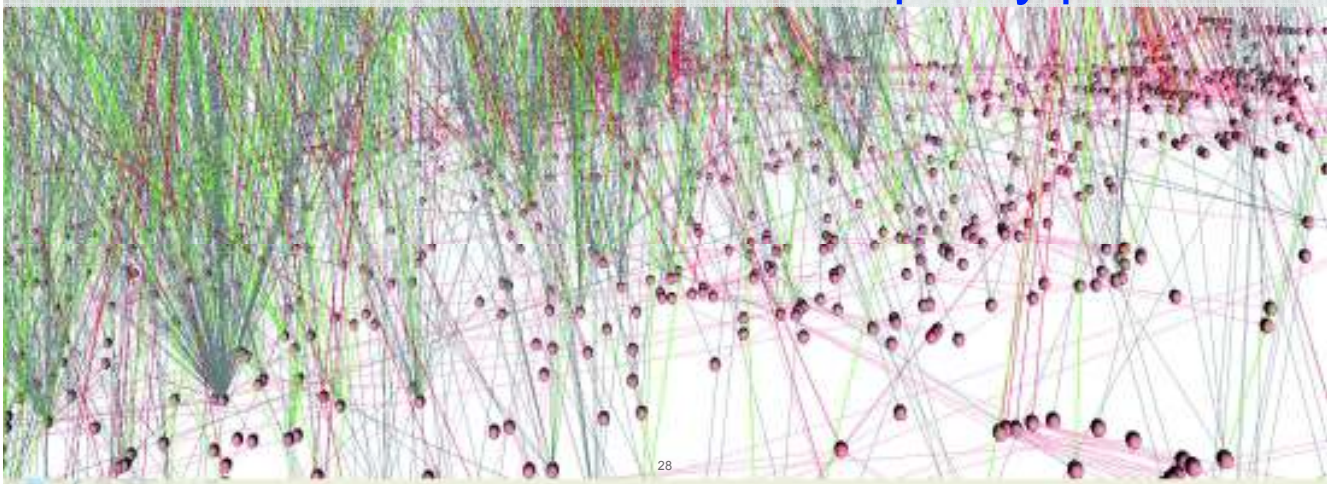
- Establish project
- Define acceptance criteria held by the result
- Perform the life cycle process
- Perform quality assurance of the result, during engineering and by the end



Eksempelgraf fra SERES modell

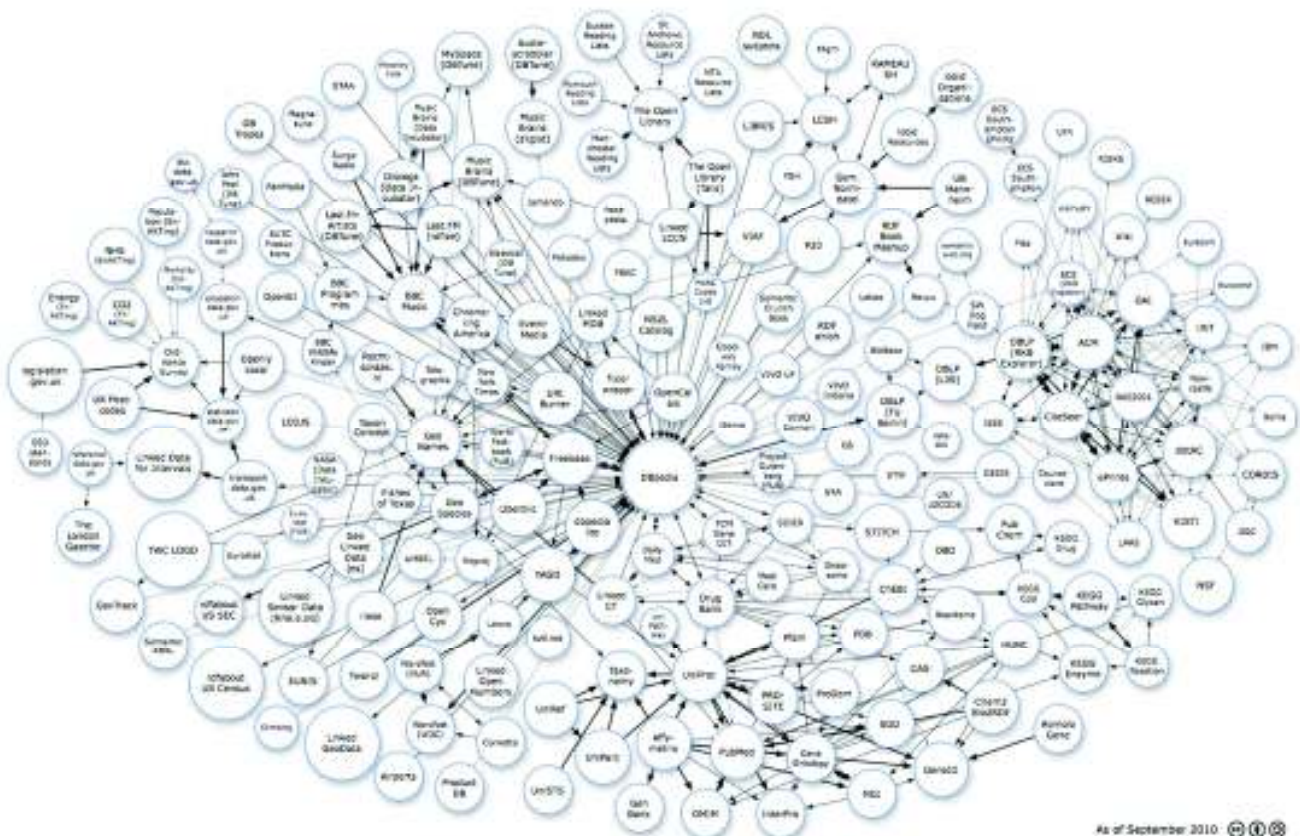


Is manual validation of metadata quality possible?

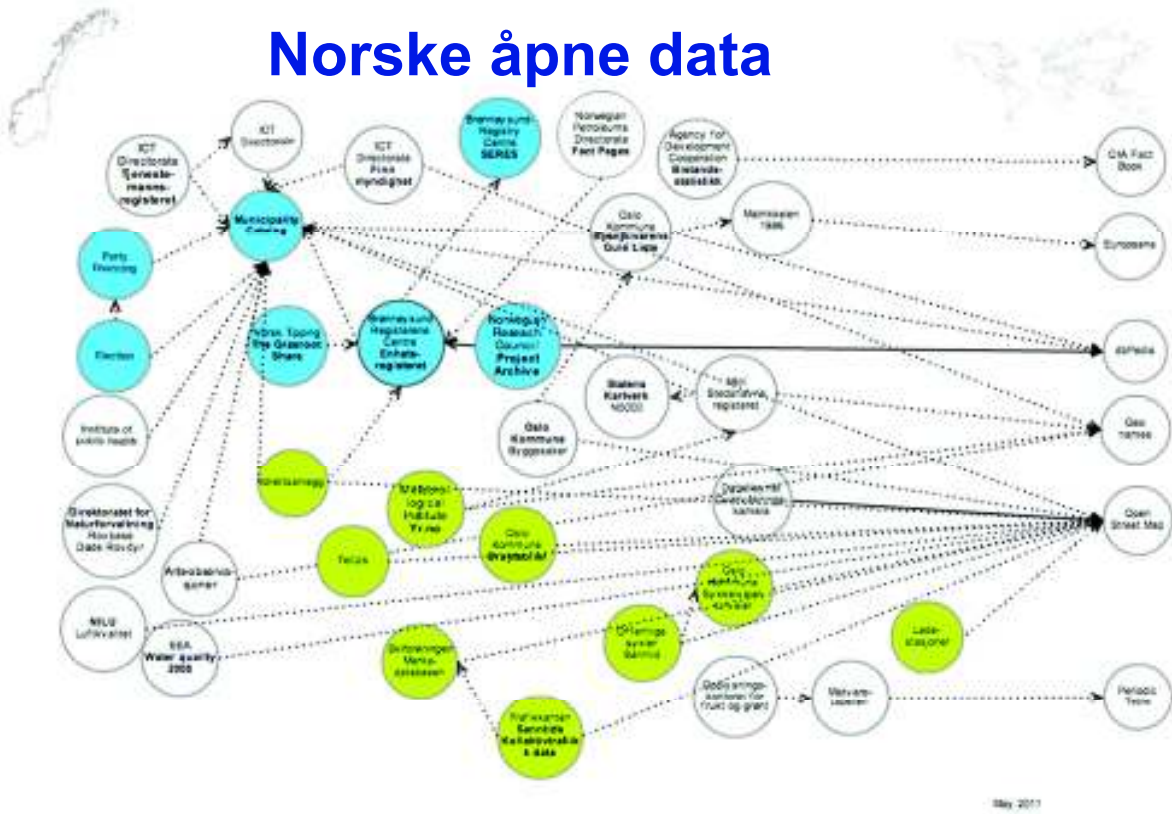


Åpne data

Linked open data / del av web 3.0

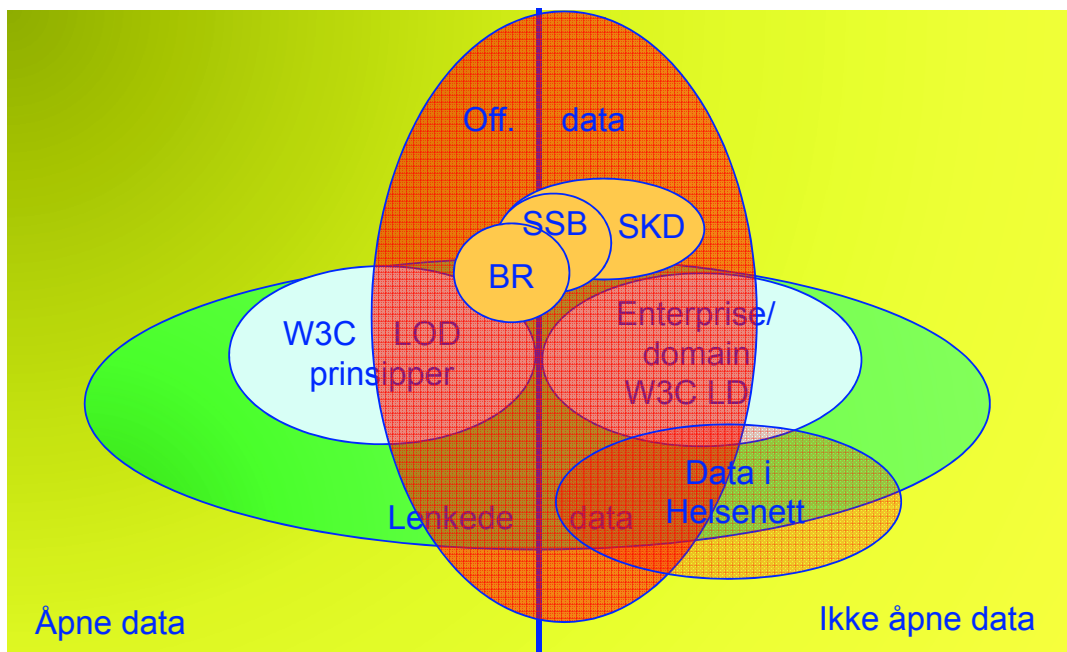


Norske åpne data

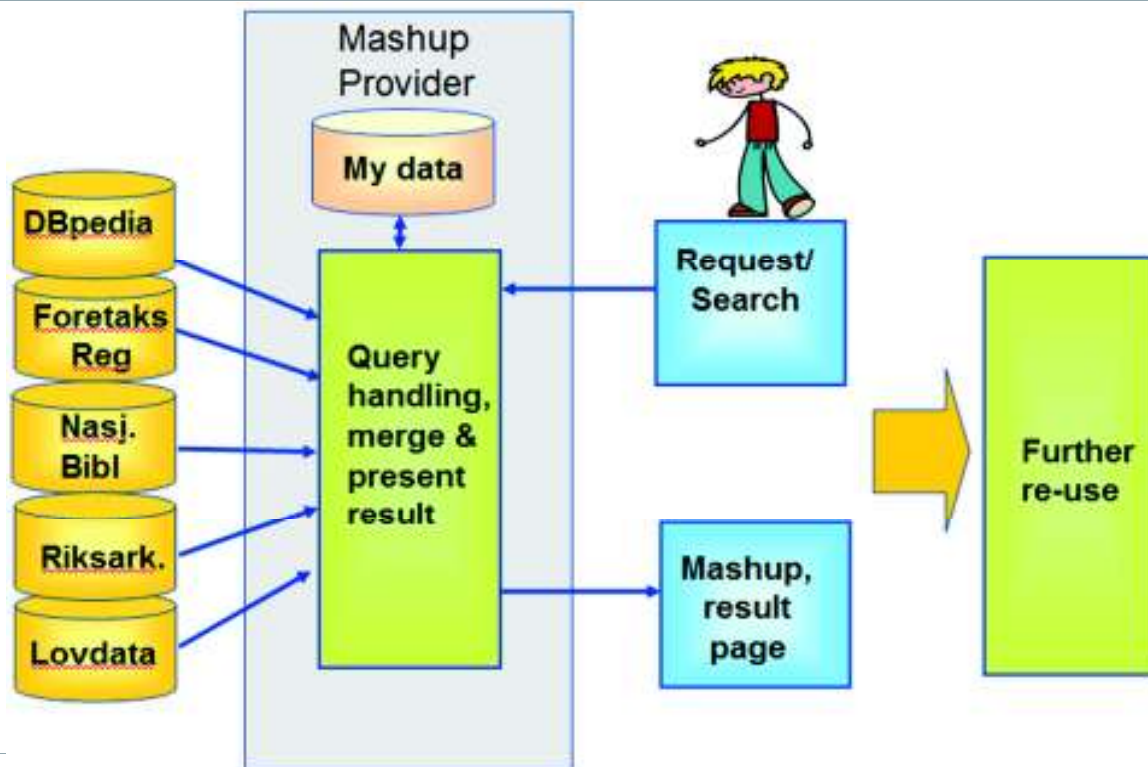


May 2011

Visualisering av typer åpne og lenkede data



Viderebruk, en mulig skisse



Åpne data initiativer fra FAD/DIFI

- data.norge.no
- DIFI datahotell

Metadata og åpne data, tiltak i offentlig sektor

Tildelingsbrev og statsbudsjett (i)

Fra Statsbudsjettet post 904, BRREG:

Prioriteringer for 2011

Etaten skal prioritere å være à jour med innregistrering av dokumenter og sørge for kontinuerlig kontroll og oppfølging av angitte resultatindikatorer for hvert register. Det fastsettes måltall i tildelingsbrevet for 2011 for saksbehandlingstid, avgivelse av registerinformasjon, andel elektronisk innrapportering, svarrespons på publikums telefonanrop og kvalitetsmål for saksbehandlingen. **Det er en generell målsetting å øke den elektroniske innrapporteringen i 2011.**

Tildelingsbrev og statsbudsjett (ii)

Fra tildelingsbrevet til SSB 2011:

Samarbeid for bedre statistikk

- Internasjonalt statistiksamarbeid og oppfylging av internasjonale krav og pålegg skal tillegges stor vekt.
- Statistisk sentralbyrå arbeider for økt **elektronisk samhandling**, både ved samarbeid mellom etater om felles innrapportering, og for **enkler gjenbruk av data**.

Kvalitet i alle ledd

- Satsingen på å **heve kvaliteten** på virksomhetens produkter skal videreføres. Statistisk sentralbyrå skal utføre sine oppgaver kostnadseffektivt, faglig kompetent og med sterk vekt på integritet og faglig uavhengighet.

Tildelingsbrev og statsbudsjett (iii)

▪ Fellesføringer i tildelingsbrev for 2011:

Tilgjengeliggjøring av offentlige data

”Etatene skal gjøre egnede og eksisterende rådata tilgjengelige i maskinlesbare formater. Dette gjelder informasjon som har samfunnsmessig verdi, som kan viderebrukes, som ikke er taushetsbelagte og der kostnadene ved tilgjengeliggjøring antas å være beskjedne (bortfall av inntekter ved salg av data anses som en kostnad). Formater og bruksvilkår må være i overensstemmelse med [Referanse katalogen](#) og FADs føringer på nettstedet data.norge.no. Informasjon om hvilke rådata som er tilgjengelige, skal publiseres på virksomhetens nettside. Dette bør ses i sammenheng med registrering av datasett på nettstedet data.norge.no.

Etater som vurderer å etablere nye eller å oppgradere eksisterende publikumstjenester med utgangspunkt i rådata, skal normalt gjøre disse rådataene offentlig tilgjengelige i maskinlesbare formater, dersom ingenting er til hinder for slik tilgjengeliggjøring av rådata. Før virksomheten eventuelt selv etablerer nye publikumløsninger basert på rådata, må det vurderes om det er mer kostnadseffektivt å tilgjengeliggjøre rådata i maskinlesbare formater som grunnlag for at andre kan utvikle tjenester. Det skal fremkomme av årsrapporten hvilke data som er gjort tilgjengelige. Dersom publikumstjenester blir etablert uten tilgjengeliggjøring av rådata, skal dette begrunnes i årsrapporten”.

- <http://www.regjeringen.no/nb/dep/fad/dok/rundskriv/2010/fellesforinger-tildelingsbrevne-for-2011.html?id=624792>

Tildelingsbrev og statsbudsjett (iv)

Fra tildelingsbrevet til SKD 2011:

Kvalitetsheving av folkeregisteret

Skatteetaten skal i 2011 fortsette arbeidet med **kvalitetsheving av folkeregisteret**, blant annet gjennom forebyggende aktivitet i form av informasjonskampanjer og ved å utvikle hjelpemidler som gjør det enklere å få riktig informasjon fra innbyggere og andre meldingsleverandører.

Gevinstrealisering til 9,1 milliarder, relasjon til etablering av samhandlingstjenester

- **SAM: NAV og Arbeidstilsynet, Melding om yrkesskade**
- **SAM: Skattedirektoratet (SKD), eDialog i forbindelse med dødsfall**
- **Uklart: Statistisk Sentralbyrå (SSB)**
 - Formidlingstjenester (kopi av off. registre mm)
 - IDUN (160 spørreskjema)
 - KOSTRA (60 Skjem)
- **Trad: Klima- og forurensningsdirektoratet (KLIF), Rapportering fra konsesjonsbelagte virksomheter**
- **SAM: Fiskeri- og kystdepartementet (FKD), Effektiv informasjonsutveksling i norsk havbruksnæring**
- **SAM: Brønnøysundregistrene (BR), Samordnet registermelding på ny skjemamotor**
- **Uklart: Brønnøysundregistrene, Produkter og tjenester**
- **Uklart: Senter for statlig økonomistyring (SSØ), Elektronisk datafangst**
- **SAM: Felleskontoret for LO- og NHO-ordningene, Elektronisk søknad om avtalefestet pensjon**
- **SAM: SKD, NAV, SSB og BR, EDAG – elektronisk dialog med arbeidsgiver**
- **Utgår? Direktoratet for forvaltning og IKT (DIFI), Elektronisk faktura**
- **SAM: Dataflytprosjektet, Elektronisk informasjonsflyt og beslutningsgrunnlag i landbruket**
- **SAM: Statens Vegvesen (SVV), Autosys**
- **SAM: Statens Innkrevingsentral (SI), Elektronisk samhandling**
- **SAM: NAV, Sykemelding**

Realisering forutsetter interoperabilitet på alle nivåer

De 24 tjenestene som er vurdert i den reviderte nyttekostanalysen gir en netto nåverdi på ca 9,1 milliarder kroner for analyseperioden 2008 – 2026 og en årlig gevinst på 1,4 milliarder kroner ved full effekt i 2020

Types of interoperability (repeated)

- Technical
 - Communication protocols and security, exchange formats
- Semantic
 - The meaning of exchanged data is well understood. Trust and quality in data.
- Organizational
 - Goals, budgets, knowledge and processes are aligned
- Juridical
 - Legal compliance. Related to processes, actions, agreements, data exchange, data governance etc.
- Political
 - Goals and frame conditions are aligned

Safeguarding life, property and the environment

www.dnv.com

