

# Evaluering Av Klienter For Semantisk Vokabular

**SEMICOLON**

SAMHANDLING I OFFENTLIG SEKTOR

## Innholdsfortegnelse

English summary .....	5
Leserveiledning .....	5
Evaluering Semantisk Vokabular klienter .....	6
Evaluering av Use Case .....	7
TopBraid <b>Enterprise Vocabulary management (EVN)</b> .....	9
Dokumentasjon .....	9
Innledende testing .....	9
Erfaringer med leverandør og tekniske utfordringer .....	9
Informasjon om test-instans .....	10
Roller og Brukerstøtte .....	10
Støtte for DIFI-Ontologi .....	10
Versjonering og mulighet for å spore endringer .....	11
Import og Eksport av Data .....	12
Eksportmuligheter .....	16
Navnerom .....	17
<b>Søke- og sammenlikningsmuligheter</b> .....	17
Evaluering av UseCase .....	19
<b>UC1: Vedlikehold av terminologi / vokabular for et statlig organ</b> .....	19
Opprette et nytt konsept .....	20
Lese konsepter .....	20
Oppdatere konsepter .....	20
Søke, navigere og sammenlikne konsepter .....	21
Lese konsept og relaterte attributter .....	21
Eksportere til andre verktøy .....	21
Endre status på konsepter .....	21
Opprettholde konsistens for en gruppe begreper .....	22
<b>UC2: Publisering av terminologi / vokabular</b> .....	22
<b>UC3: Registrere et konsept med tilhørende relasjoner</b> .....	22
Legge til kommentarer på relaterte konsepter .....	22
<b>UC4: Traversere og se på konsepter fra et annet statlig organ</b> .....	23
Avgrense en domenemodell for en etat .....	23
<b>UC5: Sammenlikne konsepter</b> .....	23
Andre observasjoner .....	23
Konklusjon og inntrykk .....	23
Poolparty (PP) .....	25
Dokumentasjon .....	25
Innledende testing .....	25

Erfaringer med leverandør og tekniske utfordringer .....	25
Informasjon om test-instans .....	27
Rolle, workflow og brukerstøtte .....	27
Støtte for DIFI-Ontologi.....	27
Versjonering og mulighet for å spore endringer .....	28
Import og eksport av Data .....	29
Eksportmuligheter .....	31
Navnerom.....	32
Søk- og sammenlikningsmuligheter .....	32
Evaluering av UseCase .....	34
UC1: Vedlikehold av terminologi / vokabular for et statlig organ .....	34
Opprette et nytt konsept.....	34
Lese konsepter .....	34
Oppdatere konsepter.....	35
Søke, navigere og sammenlikne konsepter .....	35
Lese konsepter og relaterte attributter.....	36
Eksportere til andre verktøy.....	36
Endre status på konsepter .....	36
Opprettholde konsistens for en gruppe begreper .....	36
UC2: Publisering av terminologi / vokabular .....	36
Publisere konsepter .....	36
Søke, navigere og sammenlikne konsepter .....	37
Lese konsept og relaterte attributter .....	38
Endre en eller et set med attributter.....	38
Export concept to other tools .....	38
UC3: Registrering av et konsept med tilhørende relasjoner .....	38
Validere konsistens av graf.....	39
Enter comments on related concepts.....	39
UC4: Traversere og se på konsepter fra et annet statlig organ.....	39
UC5: Sammenlikne konsepter .....	39
Andre observasjoner .....	39
Konklusjon og inntrykk.....	39
Collibra Bussiness Glossary (CBG) .....	41
Dokumentasjon.....	41
Innledende testing.....	41
Erfaringer med leverandør og tekniske utfordringer .....	41
Informasjon om test-instans .....	42
Roller og brukerstøtte.....	42

Støtte for DIFI-Ontologi.....	44
Versjonering og mulighet for å spore endringer.....	45
Import og eksport av data.....	47
Eksportmuligheter .....	51
Navnerom.....	51
Søke- og sammenlikningsmuligheter.....	51
Andre features.....	53
Evaluering av UseCase .....	54
UC1: Vedlikehold av terminologi / vokabular for et statlig organ .....	54
Opprette et nytt konsept.....	55
Lese konsepter .....	55
Oppdatere konsepter.....	55
Søke, navigere og sammenlikne konsepter .....	56
Lese konsepter og relaterte attributter.....	56
Eksportere til andre verktøy.....	56
Endre status på konsepter .....	57
Opprettholde konsistens for en gruppe begreper .....	57
UC2: Publisering av terminologi / vokabular.....	57
UC3: Registrere et konsept med tilhørende relasjoner.....	58
Legge til kommentarer på relaterte konsepter.....	58
UC4: Traversere og se på konsepter fra et annet statlig organ.....	59
Avgrense en domenemodell for en etat .....	59
Konklusjon og inntrykk.....	59
Annet.....	59

## English summary

This report is a review of several ontology tools.

The evaluation is based on several use-cases identified for this evaluation by several public agencies in Norway.

As a result of the evaluation of these tools, input for requirements for a tool for managing metadata and terminology for Norwegian public sector will be developed.

We expect this work to be concluded by the end of 2014.

The tools mentioned in this report have been presented at several conferences in Norway to collect input and to encourage Norwegian public sector to identify needs and requirements for how they would work with their business critical vocabularies.

We would also like to thank the vendors for their support and willingness to answer questions raised during the evaluation, and for providing a environment we could use for testing our use cases.

## Leserveiledning

Vi kan advare om at det fort kan være endel forvirringer i dette dokumentet der ulike ord som betyr det samme brukes litt om hverandre. Eksempler på dette er Concept, konsept, Asset og Begrep. Eller Ontologi, Terminologi og metamodel.

Dette dokumentet er skrevet av teknologer og ikke pedagoger. Det vil derfor kunne være (endel) tilfeller der ikke alt blir forklart i tilstrekkelig detalj, eller ikke blir presentert på en intuitiv og informativ måte.

Noen av påstandene i dette dokumentet er basert på prøving og feiling, og ikke etter dypdykk i dokumentasjon eller spesifikk testing.

Ikke alt blir presentert sekvensielt. Vi kan derfor referere til funksjonalitet et sted som ikke blir forklart før senere i dokumentet.

Noen av figurreferansene kan være feil. Kryssrefereringen i OpenOffice forsvant på et tidspunkt. Hvis referansen er feil gjelder regelen referanse+1.

# Evaluering Semantisk Vokabular klienter

Dette dokumentet vil teste og sammenlikne tre mulige kandidater for en sentralisert løsning for et felles semantisk vokabular applikasjon for norske etater. De tre aktuelle kandidatene er:

- Collibra
- Topbraid Enterprise Vocabulary Net (EVN)
- PoolParty (PP)

Vi vil teste følgende egenskaper for de forskjellige klientene:

**Erfaringer med leverandør, og tekniske utfordringer:** Har leverandøren gitt oss den hjelpen vi trenger? Hvordan har de besvart henvendelser? Hvor god har oppfølgingen vært? Har det vært noen spesielle tekniske utfordringer eller bugs med verktøyet?

**Roller – Brukerstøtte:** Hvilke roller støtter administreringen av vokabularet? Og har vi mulighet til å lage nye roller eller navn på eksisterende roller? Dette omfatter også om vokabularet kan publiseres for ikke-autoriserte brukere.

**Import-støtte:** I denne testen vil vi ta utgangspunkt i eksporter fra SERES og hvordan vi kan importere alt interessant innhold fra dem til systemet vi tester. Dette omfatter attributter såvel som relasjoner og relasjonstyper.

Det er også interessant å se om det er mulig å importere ontologier i tillegg til vokabularet.

**Støtte for DIFI- og SERES-:** Kan vi definere DIFI sin ontologi i verktøyet? Er det noen datatyper som ikke støttes? Eller andre konvensjoner som gjør importen vanskelig?

Grunnleggende attributter som kreves av DIFI er:

Attributt	Type	Multiplikator	Kommentar
Identifikator	<String> eller <URL>	1	Globalt unik
Anbefalt term	<String>	1	Gjerne flere språk
Definisjon	<String>	1	Gjerne flere språk
Kilde	<String> eller <Term>	0 - flere	Lover, forskrifter, registre etc. Som er relevant for opphav til definisjon og alternativ formulering.
Fagområde	<String> eller <Term>	1 - flere	
Gyldig Fra	<Dato>	1	
Ansvarlig Virksomhet	<String>	1	
Språk	<String>	1	Enten som attributt, eller som type for hvert attributt som er språkavhengig.

Tabell 1: DIFI-ontologi

I tillegg har SERES noen attributter og relasjoner vi tar med. SERES deler inn Begrepsrelasjoner som en egen type, og ikke bare som et attributt. For eksempel:

*Navn* omfatter *Fornavn*

I relasjonen over er *Navn* og *Fornavn* Begreper, og *omfatter* er en relasjon. I seres vil *omfatter* være en BegrepsrelasjonType. I tillegg til begrepene *Navn* og *Fornavn* vil vi få en Begrepsrelasjon som

har type *omfatter* og start *Fornavn* og slutt *Etternavn*. Det er mulig det er gjort slik på grunn av at SERES er mof-basert. Når vi lager ontologi for SERES/DIFI i begrepsklienter vil vi prøve en versjon med disse tre typene, og en versjon hvor relasjonen er en del av ontologien, uten å ha egen type for Begrepsrelasjon.

Vi har to eksporter fra SERES tilgjengelig for Test. Disse er Konkursregisteret og Brønnøysundregistrene. I disse eksportene eksisterer BegrepsrelasjonTypene *er*, *harRolle*, *omfatter* og *synonym*. Vi tar derfor disse med i ontologiene i første omgang. Hvis andre relasjoner påkreves kan de opprettes i klientene.

Attributtene for SERES blir da:

Attributt	Type	Multiplikator	Kommentar
Navn	<String>	1	SERES implementasjon av anbefalt Term
GUID	<String>	1	Global unik identifikator for SERES. Tas med i tilfelle man skal eksportere data tilbake til SERES.
Definisjon	<String>	1	Gjerne flere språk
Er	<Begrep>	0 - flere	Relasjon mellom to Begreper.
Har Rolle	<Begrep>	0 - flere	Relasjon mellom to Begreper.
Synonym	<Begrep>	0 - flere	Relasjon mellom to Begreper.
Omfatter	<Begrep>	0 - flere	Relasjon mellom to Begreper.

Tabell 2: SERES-attributter

I tillegg er det kanskje ønskelig med et attributt *Versjon* som kan brukes for å vise hvilken versjon av Begrepet som vises, og som kan brukes til å spore tilbake til tidligere versjoner.

**Eksportmuligheter:** Hvilke formater støttes? Og hvordan kan instansene konfigureres opp med hensyn på muligheter for web-services eller REST?

**Navnerom:** Dette kom opp som et viktig felt under demonstrasjon 24.09.13. Dette punktet tar for seg hvordan modellen kan deles inn i navnerom, og hvordan dette kan brukes til å snevre modellen

**Versjonering og muligheter til å spore endringer:** Hvor god støtte har verktøyet for versjonering? Er det mulig å se hvilke endringer som er gjort på et nivå/Begrep? Samt hvem som utførte endringen?

**Søke-, sammenliknings- og publiseringsmuligheter:** Hvor god støtte har verktøyet for søk? Hvilke filtreringsmuligheter finnes? Er det mulig å sammenlikne Begreper på tvers av domene?

**Støtte for språk og egen tilpasning:** Er det støtte for norskt grensesnitt? Kan vi endre layout til noe mer norskt?

## Evaluering av Use Case

Evaluering av usecase fra [Survey on vocabulary and ontology tools](#).

Testene baserer seg på krav nevnt under *Functionality needed*.

Vi har fem Use Case i undersøkelsesdokumentet. Disse er:

- UC1: Vedlikehold av terminologi / vokabular for et statlig organ
- UC2: Publisering av terminologi / vokabular

- UC3: Registrere et konsept med tilhørende relasjoner
- UC4: Traversere og se på konsepter fra et annet statlig organ
- UC5: Sammenlikne konsepter



# TopBraid Enterprise Vocabulary management (EVN)

Topbraid Enterprise Vocabulary Net er en web-basert løsning for forenklet utvikling og vedlikehold av web-controllerte vokabular. Det hjelper foretningsinteressenter med å samarbeide med å definere og linke bedriftsvokabularer og ontologier brukt for informasjonsintegrering, tilpasning og søk.

Verktøyet er tenkt brukt sammen med en applikasjon *Composser* som er et eclipse-basert verktøy.

## Dokumentasjon

**Brukerdokumentasjon:** <http://topquadrant.com/topbraid/evn/43doc/userguide.html>

**Annen dokumentasjon:** <http://topquadrant.com/topbraid/evn/43doc/>

## Innledende testing

Web-sidene til Topbraid tilbyr to måter å teste EVN. Den ene er å laste ned en versjon av Composer som kjører opp en egen instans av EVN lokalt på maskinen. Den andre er å registrere seg for en testkonto, og få en instans hos Topbraid som de kontrollerer.

Vi har testet begge to. Det tar litt tid før man får en instans satt opp, så vi har brukt tiden til å teste på lokal installasjon. Vi har hatt endel problemer blant annet med å konfigurere database for instansen. Det meste av uttestingen har derfor blitt gjort på ekstern instans av EVN.

## Erfaringer med leverandør og tekniske utfordringer

EVN har i stor grad bydd på problemer når det gjelder tekniske utfordringer. I perioden før vi fikk egen instans for testing hos Topbraid forsøkte vi å kjøre testversjon lokalt. Oppsettet av MySQL-database for denne instansen viste seg å være vanskelig. Vi fulgte brukerdokumentasjonen til punkt og prikke. Etter å ha gjort dette sendte vi forespørsel til support hos Topbraid. Denne ble etterhvert besvart. På grunn av at Topbraid har base i USA gikk denne dialogen noe sent, en forespørsel/svar per dag. Etter 3 forespørsler kom fikk vi ikke landet en løsning. Da stoppet også support å svare. Det skal nevnes at det står på Topbraid sine sider at de prioriterer betalende kunder.

Når det gjelder støtte for MySQL så må man legge et bibliotek (jar-fil) i en folder ved workspacet man bruker i composer. Dette står det ingenting om i dokumentasjonen, og det er på ingen måte intuitivt å gjette seg til dette eller hvor composer i det hele tatt har sin classpath. Vi fikk denne informasjonen av support. Denne fiksen sørget for at vi fikk en annen feilmelding enn den vi hadde fått tidligere.

Det tok ca en uke å sette opp en testserver for oss. Dette er ok. Serveren er lukket for kun oss, og vi fikk 4 brukere å teste med.

Bob DuCharme fra TopQuadrant tok kontakt med oss 16 oktober, ca en og en halv måned etter at vi fikk satt opp test-instansen. Han lurte på hvordan det gikk, om vi hadde noen spørsmål og om vi var interessert i et web-møte. Vi var i utgangspunktet interessert i et web-møte, men måtte takke nei på grunn av mangel på tid.

Vi fikk også svar på endel spørsmål i denne henvendelsen. Blant annet fikk vi tilbud om nyere versjon av test-instans. Vi fikk også svar på hvordan man kan sammenlikne konsepter. Skulle gjerne hatt svar på dette litt tidligere da dette dokumentet i skrivende stund tar høyde for at dette IKKE er mulig.

Dokumentasjonen (ovenfor) er oversiktlig. Det er mulig noe kunne vært mer utdypet. Vi får også

gode kom i gang- tutorials.

Det kan virke som det er litt konsistensproblemer av og til. Når jeg søkte på dato (med to vokabularer importert) og trykket litt frem og tilbake på alternativene som ble funnet fikk plutselig Brønnøysundregistrene samme innhold som Konkursregisteret i viewet *Concept Hierarchy*. Når jeg gikk ut og inn av edit funksjonen ble alt normalt igjen. Men dette er helt klart en bug.

Man kan importere et vokabular inn til et annet ved å på prosjektet trykke *Add Import* og velge det vokabularet man vil importere. Dette er en prosess som er enkel og som går raskt å utføre. Hvis man derimot ønsker å kansellere en import av et vokabular, for eksempel hvis man har importert feil vokabular, er det ingen funksjonalitet for dette. Det burde det vel strengt tatt vært? Spesielt med hensyn på at importen skjer på produksjonsnivå, og ikke i en *Working Copy*. Spørsmål sendt til TopQuadrant.

## Informasjon om test-instans

Siterer mail fra Topbraid:

Accesss your EVN demo at the following URL:

<http://tellu.topquadrant.net>

===== USER INFO =====

username:

password:

username:

password:

username:

password:

username:

password:

---

## Roller og Brukerstøtte

EVN har et sett på tre roller som man kan tilegne brukere. Vi kan ikke rename eller legge til roller ut over dette. De tre rollene som støttes er: *Viewer*, *Editor* og *Manager*.

Rollene gis per ressurs, altså vokabular eller Working Copy. For et vokabular kan en bruker ha rolle som editor, mens et workspace av samme vokabular har brukeren view-only eller ingen rettigheter.

Ut fra hva vi forstår av dokumentasjonen opprettes brukere ved (for versjonen uten Composer) å legge brukere til Tomcat-instansen som hoster serveren. Det er ingen online-støtte for å legge til eller endre brukere.

En *Viewer* kan lese alt av data og rapporter for et vokabular, eksportere data fra vokabularet i forskjellige formater samt lage en *Working Copy* av produksjonsvokabularet og gjøre endringer på den. *Working Copy*-instansen må godkjennes av en *Editor* for det vokabularet.

En *Editor* kan gjøre alt en *Viewer* kan pluss gjøre endringer på et produksjonsvokabular.

Men *Manager* kan Opprette vokabular, kontrollere tilgang for andre brukere til et vokabular og importere data til vokabularer. En *Manager* kan også fryse en *Working Copy* for review (eller avfryse for videre arbeid hvis den ikke blir godkjent).

Når det gjelder publisering av vokabular vil disse kun være synlige for innloggede brukere. Man kan lage ubegrenset mengde *Viewers*.

## Støtte for DIFI-Ontologi

EVN har en egen enkel editor for ontologi. Det er tre elementer som brukes. Disse er :

- **Concept:** Et Skos:concept. Dette struktureres hierarkisk, med Concept i bunnen. I vår Ontologi har vi et Concept *Begrep*.

- **Attribute:** Et attributt som har en eller flere verdier basert på en enkel datatype. Det er ikke mulig å sette multiplisitet for attributter. Alle instanser av attributter har et språk. En liste med forslag til språk er tilgjengelig. Norsk er ikke i denne listen. Det er mulig dette kan konfigureres opp. Datatypene som støttes er vist i Figur.
- **Relationship:** Relasjoner mellom Concepts. Det er ikke mulig å sette multiplisitet på relasjoner. Relasjoner har en range som indikerer hvilken Concept som er target.

The screenshot shows a web interface for editing an object property named 'omfatter'. It includes a 'label' field with the value 'omfatter', a 'comment' field with a rich text editor, and a 'Relationship Characteristics' section with 'domain' and 'range' both set to 'Begrep'.

Figur 1: Egenskaper for relasjon

## Versjonering og mulighet for å spore endringer

EVN viser gode muligheter for å se endringer, når de ble gjort og hvem som gjorde det.

I eksempelet over har vi valgt et *Begrep* fra en *Working Copy*. Vi kan enkelt spore endringer, og se når og av hvem endringene ble gjort. Vi kan også enkelt revert endringene.

### Labels and Definition

preferred label:	Corea del Sur (es)
	Added by Administrator on Aug 29, 2013 8:16:27 AM <a href="#">Revert</a>
	Republic of Korea
alternative label:	South Korea
	Deleted by Administrator on Aug 29, 2013 8:16:27 AM <a href="#">Revert</a>
	South Korea
hidden label:	South Korea
	Added by Administrator on Aug 29, 2013 8:17:23 AM <a href="#">Revert</a>
	Republic of Korea
	Deleted by Administrator on Aug 29, 2013 8:17:23 AM <a href="#">Revert</a>

Figur 2:

### Endringer for et Begrep

Vi kan også spore endringer på høyere nivå. I en *Working Copy* har vi mulighet til å generere en rapport for å vise forskjeller mellom kopien og produksjon. Se Figur nedenfor.

Comparison Report				
Vocabulary Test2, working copy tttt				
Adresstype ( <a href="http://semicolon2.no/vocab/Begrep/Br???nn???ysundregistre#Adresstype">http://semicolon2.no/vocab/Begrep/Br???nn???ysundregistre#Adresstype</a> )	Property	Production		Working Copy
definisjon			Dette er en definisjon for Adresstype.	
Ansatt ( <a href="http://semicolon2.no/vocab/Begrep/Br???nn???ysundregistre#Ansatt">http://semicolon2.no/vocab/Begrep/Br???nn???ysundregistre#Ansatt</a> )	Property	Production		Working Copy
definisjon		Person som er i l??nnet arbeid hos en arbeidsgiver.	Person som er i l??nnet arbeid hos en arbeidsgiver. haba haba.	
EktepaktEndring ( <a href="http://semicolon2.no/vocab/Begrep/Br???nn???ysundregistre#Ektepakt">http://semicolon2.no/vocab/Begrep/Br???nn???ysundregistre#Ektepakt</a> )	Property	Production		Working Copy
definisjon				
preferred label		Ektepakt	EktepaktEndring	

Figur 3: Sammenlikning mellom Working Copy og produksjon

Arbeidsflyten i EVN er organisert på følgende måte: Vi har som nevnt tre roller. En *Manager* lager en ny *Working Copy* og tildeler *editor* eller *viewer* roller på den instansen for relevante brukere. En *editor* gjør alle endringer på instansen og sender denne til review når vedkommende er ferdig med endringer. Kopien er nå fryst. *Manageren* vil enten sende den tilbake til *editoren* eller godkjenne instansen. Instansen blir i så fall merget med produksjon.

Manager kan også velge å publisere vokabularet. Da blir det tilgjengelig for alle readere av systemet.

## Import og Eksport av Data

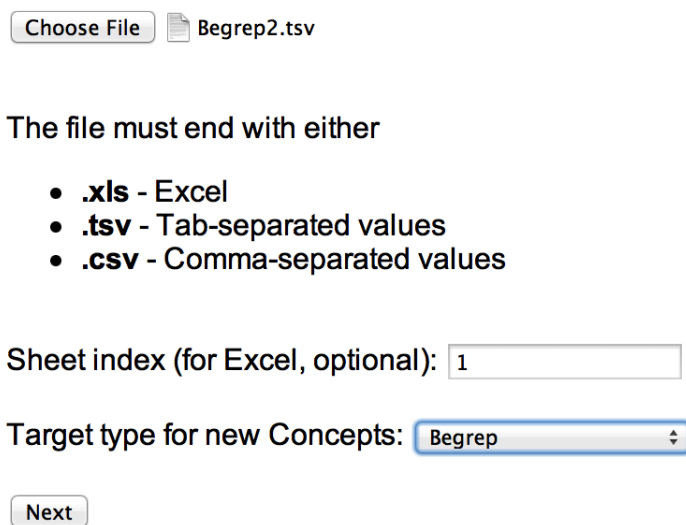
Det er litt forvirrende hvordan man importerer data til EVN. Hvis man velger et vokabular får man valgene Import External Data:

Hvis man derimot velger en instans av *Working Copies* får man ikke opp alle disse mulighetene. Da er det kun *Import Concepts from Spreadsheet* som kommer opp. Denne lar deg velge mellom Excel, Tab-separated og CSV. Vi tester først med å importere en CSV-fil.

### Import External Data

[Import RDF File](#) [Import Concepts from Spreadsheet](#) [Import MultiTes File](#)

Figur 4: Importalternativer for EVN



The screenshot shows a web-based import dialog. At the top, there is a 'Choose File' button next to a file icon and the name 'Begrep2.tsv'. Below this, a text label states 'The file must end with either' followed by a bulleted list: '.xls - Excel', '.tsv - Tab-separated values', and '.csv - Comma-separated values'. Underneath the list is a text input field labeled 'Sheet index (for Excel, optional):' with the value '1' entered. Below that is a dropdown menu labeled 'Target type for new Concepts:' with 'Begrep' selected. At the bottom left is a 'Next' button.

Figur 5: Importdialog for stylesheets

Figur viser dialogen for å importere stylesheet-filer. Legg merke til at vi velger Begrep som Target. Begrep er et Concept definert i ontologien vår. Når vi importerer via stylesheets kan vi ikke importere ontologien, kun mappe til en allerede eksisterende ontologi.

Figur viser et valg man får opp. EVN viser oss utdrag av stylesheet vi skal importere, og lar oss velge hvordan stylesheetet er strukturert. Dette er mest hensiktsmessig dersom man skal dele modellen inn i flere nivåer. I første omgang bruker vi en flat struktur, som SERES, med alle begrepene i en mappe.

## Select Spreadsheet Type

If your spreadsheet contains a hierarchy that shall be mapped into (SKOS) broader terms, then select one of the first five design patterns below, otherwise select the "No Hierarchy" pattern.

### 1. Self-Join

Term	Parent
World	
Europe	World
France	Europe
Germany	Europe
Pacific	World
Australia	Pacific

### 2. Path with Separator

Term	Hierarchy
World	World
Europe	World > Europe
France	World > Europe > France
Germany	World > Europe > Germany
Pacific	World > Pacific
Australia	World > Pacific > Australia

### 3. Path with fixed-length Segments

Id	Term
01	World
0101	Europe
010101	France
010102	Germany
0102	Pacific
010201	Australia

### 4. Column-based Tree

Level1	Level2	Level3
World		
World	Europe	
World	Europe	France
World	Europe	Germany
World	Pacific	
World	Pacific	Australia

### 5. Column-Pair-based Tree

Id1	Level1	Id2	Level2	Id3	Level3
01	World				
01	World	02	Europe		
01	World	02	Europe	03	France
01	World	02	Europe	04	Germany
01	World	05	Pacific		
01	World	05	Pacific	06	Australia

### 6. No Hierarchy

Term	Size
World	Huge
Europe	Medium
France	Small
Germany	Small
Pacific	Large
Australia	Medium

Cells with light color are optional.

### Your Source Spreadsheet (First 6 rows)

guid	navn	Begrepsnivå_Id
http://seres.no/guid/Begrep/Abandonering/427861	Abandonering	0
http://seres.no/guid/Begrep/Beløp/427860	Beløp	0
http://seres.no/guid/Begrep/Bomasse/427859	Bomasse	0
http://seres.no/guid/Begrep/Rapportering/427858	Rapportering	0
http://seres.no/guid/Begrep/Politidistrikt/427857	Politidistrikt	0
http://seres.no/guid/Begrep/SensitiveOpplysninger/427856	SensitiveOpplysninger	0

Figur 6: Egenskaper for stylesheet

## Import Spreadsheet

### Column Mapping

Column	Target Property
Begrep	<input type="text"/>
navn	<input type="text"/>
definisjon	<input type="text"/>
guid	<input type="text"/>

### Hierarchy (Self-Join)

Column containing the parent ids	<input type="text"/>
Column containing the child ids	<input type="text"/>

### Unique Identifiers

Id column #1

Id column #2

Start of URIs

### Your Source Spreadsheet (First 6 rows)

Begrep	navn	definisjon	guid
Fornavn	Fornavn	Fornavn er den første delen av navnet på en person	
Etternavn	Etternavn	Siste del av navn på person	http://semicolon.no/etternavn
Personnavn	Personnavn	Identifikator brukt for person. Vanligvis bestående av fornavn og etternavn. Men kan også inneholde et eller flere mellomnavn	http://semicolon.no/personnavn
Mellomnavn	Mellomnavn	Alternativ del av navn på person som står mellom fornavn og etternavn	http://semicolon.no/mellomnavn

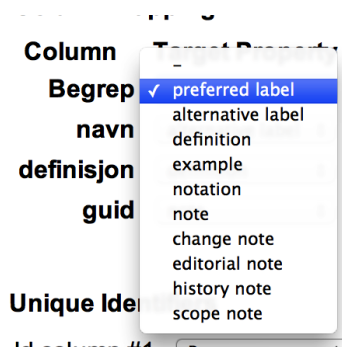
Figur 7: Mapping av attributter

Etter at vi har valgt egenskaper for stylesheets må vi Mappe feltene i stylesheet'et til et attributt tilhørende konseptet vi importerer (Figur). Dette fungerer ikke når vi importerer CSV, men ser ut til å fungere for Excel.

Når vi importerer Stylesheets har vi ikke mulighet til å importere relasjoner. Dette betyr at vi manuelt må sette disse inn selv etter at importen er gjort.

Vi har heller ikke mulighet til å definere ID. Dette er ikke kritisk da vi kan ha et eget attributt vi kontrollerer for unik ID.

Vi kan også importere fra andre formater som tidligere nevnt. Disse formatene er litt kraftigere. Hvis vi velger å importere en turtle- eller rdf-fil får vi mye mer kontroll over importen.



Figur 8: Mapping av attributter

Vi kan velge strukturen til modellen med hensyn på hierarki, importere selve ontologien og selv definere konseptenes unike ID.

I første omgang velger vi å ikke importere ontologien, men definere denne i EVN. Dette for at vi kan eksportere den og studere formatet for å lage en Transformer som binder til riktige verdier. Vi kommer til å teste på samme måte for de andre klientene i testen.

Vi gjør to tester på dette. En som er basert på SERES-struktur, og en som kutter ut typene Begrepsrelasjon og BegrepsrelasjonType fra SERES.

For å utføre testene lager vi nye vokabularer i EVN med namespace <http://semicolon.no/vocab/>.

Begrepene vil få id etter konvensjon [<Etat>#<Begrepsnavn>](http://semicolon.no/vocab/).

Vi har nå laget en parser i java som parser seres-eksporter og oversetter dem til turtle-format som kan importeres direkte i EVN. Vi deler dette inn i tre overordnede nivåer som vist i Figur.



Figur 9: Hierarki, SERES-basert

Under *Begreper* ligger alle begrepene. For eksempel blir det importerte begrepet *Utlodning* ha følgende data som vist i Figur.

Definisjonen er tom fordi den ikke er lagt til i SERES.

Utlodning (Begrep)	
<b>Labels and Definition</b>	
preferred label:	Utlodning (en)
type:	<a href="#">Begrep</a>
<b>Standard Relationships</b>	
has broader:	<a href="#">Begreper</a>
<b>Custom Properties</b>	
definisjon:	
guid:	<a href="http://seres.no/guid/Begrep/Utlodning/427828">http://seres.no/guid/Begrep/Utlodning/427828</a>
navn:	Utlodning

Figur 10: Begrepsattributter

Figur 14 viser data tilhørende begrepsrelasjonen *rt5*. Til og fra har linker til begrepene de representerer. Feltet type refererer til en definert begrepsrelasjonstype.

Figur viser data for Begrepsrelasjonstypen *er*.

## rt5 (Begrepsrelasjon)

### Labels and Definition

**preferred label:** rt5  
**type:** [Begrepsrelasjon](#)

### Standard Relationships

**has broader:** [Begrepsrelasjoner](#)

### Custom Properties

**fra:** [Sluttutlodning](#)  
**til:** [Utlodning](#)  
**type:** [er](#)

*Figur 11: Begrepsrelasjonsattributter*

Vi tester også med å legge relasjonene direkte på Begrepet.

I Figur er relasjonene *omfatter* og *er* lagt direkte på konseptet.

## Personnavn (Begrep)

### Labels and Definition

**preferred label:** Personnavn  
**type:** [Begrep](#)

### Custom Properties

**definisjon:** Personnavn best??r av fornavn og ett enkelt eller dobbelt etternavn og kan i tillegg ha mellomnavn.  
**er:** [Navn](#)  
**guid:** <http://seres.no/guid/Br???ysundregistrene/Begrep/Personnavn/4997>  
**gyldigFra:** 2013-09-20  
**navn:** Personnavn  
**omfatter:** [Etternavn](#)  
[Fornavn](#)  
[Mellomnavn](#)

*Figur 12: Uten egen type for relasjoner*

Vi lager et nytt vokabular kalt Test1 med namespace <http://semicolon.no/vocab/>.

Når vi skal importere fra andre kilder enn stylesheets må vi importere direkte til prod-versjonen, ikke via *Working Copies*. Man kan importere både vokabularet og ontologien. Hvis man gjør endringer på eller legger til nye begreper kan man importere på nytt for å få med endringene. Dette fungerer derimot ikke for å slette elementer. Samt at slik det er lagt opp nå er navnet på begrepet med i identifikatoren, for eksempel <http://semicolon.no/vocab/Begrep/Skatteetaten#Selskap>. Hvis Selskap skal endres til for eksempel Bedrift må det fortsatt ha den gamle identifikatoren for at oppdateringen skal bli riktig. Hvis ikke blir det opprettet et nytt Begrep med den nye identifikatoren (dette er kanskje ikke så realistisk). Hvis vi gjør endringer kommer oppdateringene i tillegg til data



som var lagt til tidligere. I Figur nedenfor har vi oppdatert definisjonen til et Begrep og importert på nytt. EVN lager da et nytt entry for definisjon, uten å slette den gamle verdien (som i eksempelet er tom). Ut fra rendringen ser det ut som om det er den nye endringen som vises default i viewene, men vi må manuelt gå inn og slette den gamle definisjonen.

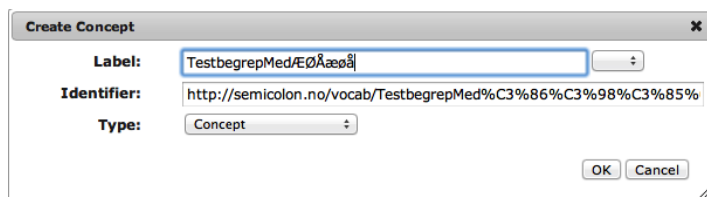
Foreløpig ser det ut til at serveren vi går mot hos Topbraid har litt problemer med kodingen. Dette er ikke tilfelle når vi kjører lokalt.



Figur 13: Re-importert Begrep

Tegnene i Figur til venstre som har to spørsmålstegn er importerte begreper. Vi kan muligens komme rundt dette med å ta med encoding i import-filen.

Begrepet TestbegrepMedÆØÅæøå ble laget via dialog. Kodingen ser ok ut i dialogen, men blir noe FUBAR når begrepet opprettes. Dette kan helt sikkert endres vha. Innstillinger på serveren. Vi hadde ikke dette problemet da vi kjørte lokal instans. Bob DuCharme i Topquadrant mener dette er fikset i nyere versjoner.



Figur 14: Norske tegn

Det ser ut som at `&#10;` (linebreak) gir litt problemer når det er brukt i definisjonene. Vi trimmer bort alle linebreaks fra input, slik at vi unngår problemer med importen i fremtiden. Det samme gjelder definisjoner der " (quotes) er brukt i teksten. Disse erstattes med '. SERES støtter både linebreaks og quotes i tekst for attributter, det gjør ikke EVN (eller turtle).

## Eksportmuligheter

Eksportformatene som støttes av EVN er

- Turtle – Aksesserbart over nett.
- N-Triple – Aksesserbart over nett
- RDF/XML – Aksesserbart over nett
- Spreadsheet (Hierarchy og Concept Overview)
- Sharepoint-fil - Sharepoint er et Microsoft-verktøy.

Eksporterte spreadsheets inneholder alle detaljer for Konseptene. Dette inkluderer type, relasjoner og attributter. Eksportstøtte for spreadsheet er derfor kraftigere enn import-støtten for disse.

Turtle-eksporten fra Topbraid inneholder også ontologien. Dette gir oss muligheten til å importere også denne i andre verktøy. Formatet ser ut til å være en egen dialekt. Det er ikke trivielt å importere Turtle-eksport til et annet verktøy med støtte for dette uten å gjøre en transformasjon.

N-Triple eksporterer alle triplene i prosjektet slik at disse kan importeres i en tripplestoredatabse for



SPARQL-spørringer.

Vi har ikke testet RDF import, men denne ser ryddig ut. Det er uvisst om den vil ha en dialekt som er forenelig med andre klienter. Det ser ikke ut som om ontologien er en del av RDF-eksporten.

## Navnerom

Vi har muligheten til å dele inn konseptene i subkategorier. Vi kan også definere dette opp i ontologien som en egen type.

Vi kan også sette dette som eget attributt i ontologien, men vi må selv vedlikeholde det. Se Figur.



Figur 15: Navnerom

Hvis vi deler det opp på denne måten kan vi bruke det til å snevre inn søket ved å legge navnerommet inn i *has broader*.

### Custom Properties

Navnerom: [Enhet](#)

Figur 16: Navneromattributt

## Søke- og sammenlikningsmuligheter

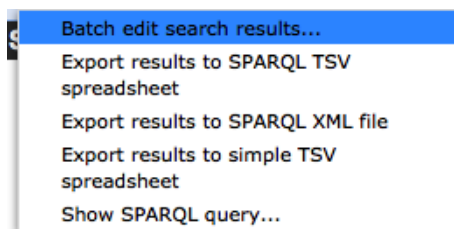
EVN har en avansert søkefunksjon, delvis vist i Figur. I søkefanen lengst til høyre (trykk *Search* dersom denne ikke er synlig) har man mulighet til å spesifisere alle attributter samt type man vil søke på. Dette gjelder også custom attributes. I tillegg til å spesifisere innhold i attributter eller navn, kan man også sette kriterier som equals, contains etc.

Figur 17: Søkefunksjon

I panelet under *Search Concepts* ligger *Search Results*. Ved å trykke på det som kommer opp her, vil begrepet bli valgt i *Concept Hierarchy*.

Man kan også i søk eksportere søkeresultatet til forskjellige spreadsheets.

Man kan ved å bruke *has broader* for eksempel snevre inn søk på fagområder. En feature vi savner er å kunne snevre inn søket basert på Concept-Scheme, altså etat/domene.

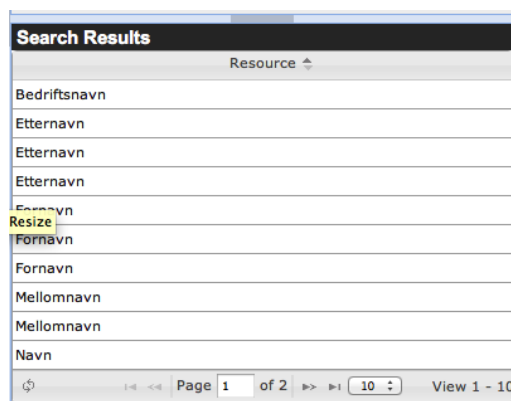


Figur 18: Alternativer for søkeresultat

Søkeresultatet viser også kun preferred label for resultatene. I Figur har vi søkt etter preferred label Navn. Vi må trykke på hver enkelt for å se hvilken instans vi søker etter. Når det er sagt er dette javascript-basert, slik at navigeringen skjer veldig raskt, men her burde navnerom eller path vært vist.

Når det gjelder sammenlikning av konsepter velger man de attributtene man vil ha i søkeresultatet

ved å velge sjekkboksen ved siden av feltet.

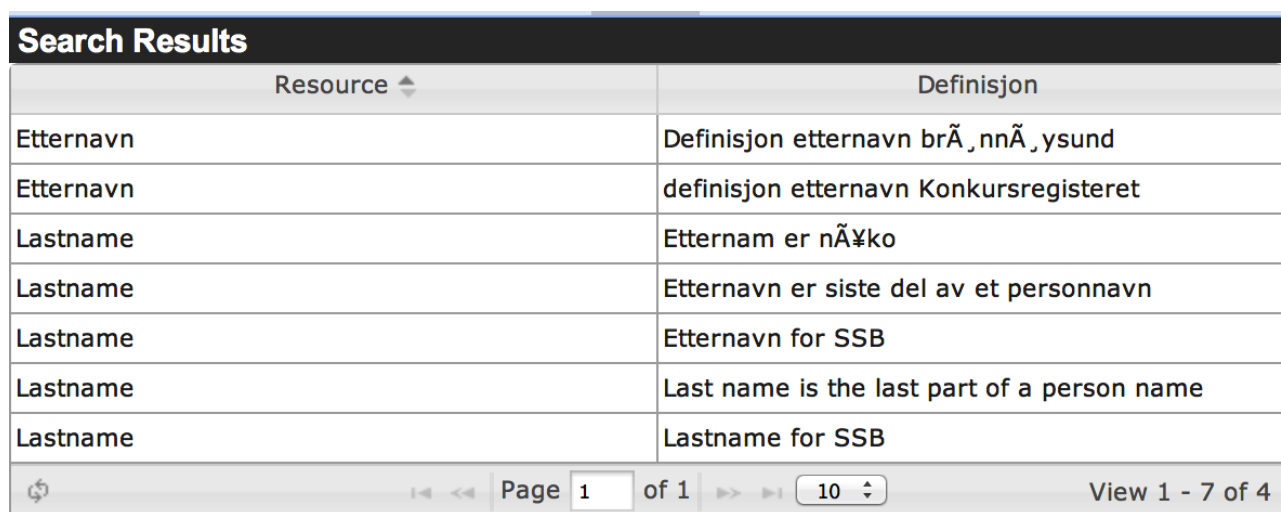


Resource
Bedriftsnavn
Etternavn
Etternavn
Etternavn
Etternavn
Etternavn
Etternavn
Fornavn
Fornavn
Fornavn
Mellomnavn
Mellomnavn
Navn

Figur 19: Søkeresultater

Man kan sammenlikne *Working Copy* og *Production vocabulary* for å se endringer. Ellers kan man også søke opp forskjellige begreper og velge dem hver for seg, eksempelvis i forskjellige faner/browsere, og sammenlikne dem manuelt. Det er litt avhengig av hva som er interessant å sammenlikne. Hvis man for eksempel ønsker å sammenlikne to definisjoner vil dette være en mulighet. Dette er illustrert i Figur.

Vi kan også bruke EVN sitt SPARQL-endpoint for å selv spesifisere data vi ønsker. Endpoint for dette på vår testinstans er: <http://tellu.topquadrant.net/evn/tbl/sparql>. Vi må i tillegg legge på `?query=<query>`.



Resource	Definisjon
Etternavn	Definisjon etternavn brÃ¸nnÃ¸ysund
Etternavn	definisjon etternavn Konkursregisteret
Lastname	Etternam er nÃ¸ko
Lastname	Etternavn er siste del av et personnavn
Lastname	Etternavn for SSB
Lastname	Last name is the last part of a person name
Lastname	Lastname for SSB

Figur 20: Sammenlikning av definisjoner i EVN

## Evaluering av UseCase

### UC1: Vedlikehold av terminologi / vokabular for et statlig organ

Rollene som er satt opp for dette use caset er

- Terminology steward
- Concept coordinator
- Legal responsible in organisation
- Information architect
- Service line owner

Standard roller for Topbraid EVN er som vist i Figur.

Man kan ikke legge til flere roller.

#### Vocabulary User Roles

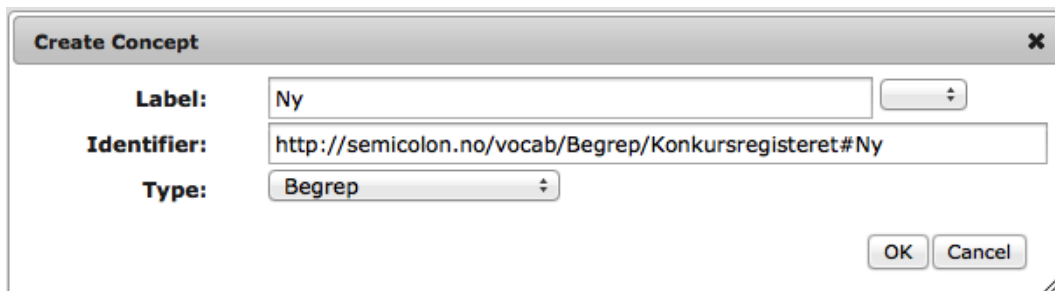


Figur 21: Autorisering i EVN

## Opprette et nytt konsept

Å lage nye konsepter i EVN er trivielt. I *Concept Hierarchy* velg det nivået du skal legge til begrepet i. Trykk på knappen for nytt concept nederst i *Concept Hierarchy*  . Dialogen for nytt concept kommer opp.

Identifiseren må være unik i EVN. Type er klasse hentet fra ontologien vi selv har definert.

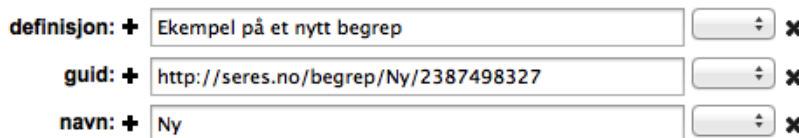


Figur 22: Dialog

for nytt Begrep

Når man trykker ok opprettes konseptet, og en editor kommer opp for å legge til tilleggsinformasjon. Har du valgt *Begrep* som type vil også et sett av *Custom Properties* komme opp. Dette er typisk DIFI-spesifikke attributter. Det er også her man legger til relasjoner. Trykk *Save Changes* nederst for å lagre endringer.

#### Custom Properties



Figur 23: Cusstom

Properties

Conceptet er lagt til.

## Lese konsepter

Dette kravet kan sikkert utdypes. For å lese vokabularet kan vi sette en bruker som *viewer*. Vi setter user2 som *viewer* og logger på med denne brukeren. Trykk på *View Production Vocabulary*.

Oversikten er samme som man har når man editerer, men man har ikke lenger tilgang til editeringsfunksjonene. Hvis du trykker på et Begrep i *Concept Hierarchy* kommer egenskapene til Begrepet opp i informasjonspanelet til høyre.

Publiserte vokabular vil være synlig for alle *Viewers* i systemet. For ikke-publiserte vokabular må *Viewers* tilegnes. Vokabularer vil ikke være synlige for hvermannen.

## Oppdatere konsepter

Vi logger på igjen med user1 som er manager for prosjektet og dermed har skrive-tilgang. Ved å velge concepts i *Concept Hierarchy* kan man velge *edit* nederst til høyre i panelet for detaljer. Før man gjør endringer er det lurt å lage en ny *Working Copy*. Dette gjøres fra prosjektets hovedside.

Trykk *Create Working Copy*.

Når du trykker edit får du mulighet til å endre alle egenskaper til konseptet, med unntak av identifikatoren. Trykk *Save Changes* for å lagre endringene. Hvis du ønsker å publisere endringene gjort på en *Working Copy* må du først committe endringene til produksjon. Hvis en annen rolle enn *Manager* gjør dette må endringene godkjennes av en *Manager* før de blir flettet med produksjon.

### Søke, navigere og sammenlikne konsepter

Når det gjelder navigering tilbyr EVN endel frihet. Vi kan arrangere Concepts hierarkisk. Foreløpig har vi kunn organisert dem som i SERES. Med etatsnavn som høyeste nivå, deretter Begreper, Begrepsrelasjoner og BegrepsrelasjonsTyper. Under disse gruppene ligger selve instansene.

Vi kan også velge å dele inn i flere nivåer dersom det er ønskelig. Som kategorier/navnerom under Begreper etc. Eksempelvis: Person, Organisasjon, Oppgave, Adresse.

Det er ingen spesifikk funksjonalitet for sammenlikning av enkeltkonsepter. Dette må gjøres manuelt i søkt.

Hvis vi ønsker å kunne søke og/eller sammenlikne på tvers av etater må vokabularene være plassert i samme prosjekt. Hver etat representeres med en *Concept Scheme*. En mulighet er å ha etatene i forskjellige prosjekter for så å importere dem i andre etaters prosjekter. Man kan da søke på tvers av etatene selv om de er i forskjellige prosjekter, og lage relasjoner til konsepter definert i andre etater.

Det er en funksjon for å legge til SPARQL-templates. Vi har ikke testet dette, men det vil sansynligvis gjøre det mulig å rendre flere begreper av en gitt kontekst samtidig slik at de kan sammenliknes. Å skrive et slikt template er likevel relativt komplisert.

### Lese konsept og relaterte attributter

Å se enkeltkonsepter gjøres som nevnt ovenfor under *Read Concepts*. Hvis man ønsker å gruppere concepts må man bruke søkefunksjonen. Hvis en relasjon er gjort til et konsept i annen etat vil ikke dette reflekteres i begrepets properties.

Det er ingen funksjonalitet i EVN for å jobbe på flere konsepter samtidig.

### Eksportere til andre verktøy

Vi har ikke testet å importere eksportene fra EVN til andre verktøy, men regner på ingen måte med at dette er en triviell affære. (Turtle fra EVN har en annen dialekt enn PoolParty).

EVN har støtte for en rekke eksport-formater. Disse er vist i Figur nedenfor.

I forsøkene ovenfor for å importere data eksporterte vi først turtle fra EVN for å studere syntaks. Denne er ikke identisk med syntaks for andre turtle-eksporter fra andre verktøy i denne testen, og vi må derfor gjøre en oversettingsjobb før eksporten kan importeres av andre verktøy. Men det er mulig det finnes unntak.

#### Export

[Turtle](#) [N-Triple](#) [RDF/XML](#) [Publish](#) [Hierarchy Spreadsheet](#) [Concept Overview Spreadsheet](#) [SharePoint Import File](#)

Figur 24: Eksportformater

Stylesheet-eksportene fra EVN ser ut til å inneholde all informasjon om en modell. Hvis man importerer dette til et annet verktøy har man mulighet til å mappe verdiene over til sine respektive verdier i det andre verktøyet. Hvis dette verktøyet også støtter å importere relasjoner via stylesheets bør ikke eksporten være noe problem.

### Endre status på konsepter

Det ser ikke ut som om det er noe statusfelt på konseptene. Jeg antar at man derfor ikke kan sette

konseptene til for eksempel *deprecated*. Vi kan legge til attributter som holder styr på gyldighetsperiode, men vi må selv vedlikeholde og passe på at konsepter som ikke lenger er gyldige fjernes eller flyttes.

En mulighet er at man i ontologien lager konseptreet Statusverdi->Kandidat, Til godkjenning og Godkent. Og gir konseptet *Begrep* får et relasjonsattributt med type *Statusverdi* som kan vedlikeholdes av editor.

### Opprettholde konsistens for en gruppe begreper

Hvis man arbeider i en workflow må hele workflow'en godkjennes før den committes til produksjon. EVN har støtte for å legge inn et sett med skos-regler som må valideres før endringer kan lagres.

### UC2: Publisering av terminologi / vokabular

En bruker med rolle *Manager* for et vokabular kan velge å publisere vokabularet. Når vokabularet er publisert kan alle registrerte brukere lese vokabularet. Vokabularet leses da med en read-only versjon av EVN editor.

En *Viewer* vil få de samme rettighetene som en *editor* med hensyn på søkefunksjonalitet.

### UC3: Registrere et konsept med tilhørende relasjoner

Man kan lage relasjoner i ontologien som refererer til andre konsepter. Når man lager instanser av konseptet skriver man inn navnet på konseptet man skal linke til, og velger instansen man skal peke til i en drop-down liste. I denne dropdown listen kommer alle reele alternativer opp, også fra andre synlige vokabular som kan eies av andre etater som vist i Figur.

Dette vil utvilsomt bli noe tungvint etter hvert som antall vokabular skalerer, og ikke minst en kilde til feil i modeller.

The screenshot shows a web form for creating a relationship. It has three main sections: 'fra:', 'til:', and 'ype:'. The 'fra:' field contains the text 'Per'. The 'til:' dropdown menu is open, showing two options: 'Person (http://semicolon.no/vocab/Begrep/Konkursregisteret#Person)' and 'Person (http://semicolon.no/vocab/Begrep/Br??nn??ysundregistre#Person)'. The 'ype:' dropdown menu is also open, showing four options: 'Personidentifikator', 'Personnavn', 'Personstatus', and 'Personstatus'.

Figur 25: Assosiering av

relasjoner

Legg merke til at identifikator kun vises for de konseptene hvor det eksisterer flere med samme navn.

EVN har støtte for å generere *Constraint Violations Report*, *Spell Check* og *Graph Statistics*. Violations genereres basert på et sett med skos-regler satt opp for prosjektet. For eksempel multiplisitet på et attributt. Dette er også den eneste måten å sette multiplisitet på attributter.

For konsepter som har en regel for dette kan man fortsatt legge til flere attributter enn det osm støttes, men man får en feilmelding. Her hadde det vært mer hensiktsmessig å ikke tillate brukeren å legge til flere attributtverdier. Det er jo ikke noe vits i å legge til rette for noe man ikke får gjøre.

Hvis konsept A har en referanse til konsept B, og konseptet B er valgt, vil man ikke ha en referanse tilbake til A. Vi har så langt ikke funnet noen funksjonalitet for billinære referanser. En referanse må eventuelt legges til på begge sider.

### Legge til kommentarer på relaterte konsepter

Hvis et konsept har et relatert konsept kan man trykke på linken til det relaterte konseptet og velge konseptet som er linket til. Hvis man her trykker edit har man mulighet til å endre attributtene. Man

kan legge til et eget attributt *Comment*, eller man kan bruke defaultattributtet *note*. Vi har lagt til et attributt *Kommentar* på import av SSB sine begreper.

## UC4: Traversere og se på konsepter fra et annet statlig organ

### Avgrense en domenemodell for en etat

Man har mulighet til å strukturere konsepter hierarkisk. Vi har gjort dette for SSB sine datatyper. Man kan også lage attributtrelasjoner for å reflektere hvilket navnerom konseptet er plassert under. Vi kan da avgrense begreper til en kontekst eller et fagområde.

Hvis vi bruker funksjonaliteten ovenfor kan vi avgrense søk basert på navnerom. Det kan likevel vise seg problematisk å ytterligere begrense dette søket inn for kun et statlig organ. Det betyr at søk vil gå på tvers av alle domener. Det er mulig man kan gjøre dette ved å legge et property på konseptene som indikerer domene, og sette konstraint på søket til at dette attributtet må være satt riktig.

### UC5: Sammenlikne konsepter

Man kan legge til Templates for SPARQL-spørringer. Man kan velge hvilke attributter som skal være med i søkeresultatet, og på den måten sammenlikne verdier som vist i Figur. Men det er endel attributter som ikke støttes. For eksempel *Top Concept* eller andre data som kan brukes til å skille verdiene i søkeresultatet. Det er også noe merkelig med resultatet. Hvis man trykker på objektene i tabellen skal det objektet velges i EVN. Men dette fungerer kun for noen av objektene.

Serveren har også sitt eget SPARQL-endpoint. Dette kan brukes til å hente de dataene vi er interessert i.

## Andre observasjoner

- Programmet ser ut til å ha støtte for vokabular i flere språk, men det kan se ut som om det kun er et utvalgt knippe språk som støttes, og norsk er ikke blant dem. Det er mulig dette kan konfigureres opp i en installert standalone server. Når vi importerte SSB sitt vokabular tok vi med språk nynorsk med språkkode *nn*. Nynorsk språkkode kom da opp som alternativ på web'en.
- Ingen av relasjonene er bilinære. Hvis bilinære relasjoner ønskes må disse lages og vedlikeholdes manuelt.
- Multiplisitet kan kun settes vha. Skos-regler. Dette er litt tungvint.
- Import av andre vokabular kan ikke revertes. UPDATE Dette støttes visstnok i nyere versjoner. Vi har fått tilbud om nyere versjon.
- Mye av arbeidet med vokabularet må gjøres rett på produksjons-versjonen.
- Hvis man legger til et engelsk property, for eksempel preferred label, blir det dette som vises som standard.

## Konklusjon og inntrykk

EVN er enkel å bruke. Spesielt historikk-funksjonaliteten innad i en workflow er imponerende.

Det er relativt lett å bygge og vedlikeholde en Ontologi. Selv om det hadde vært ønskelig med mer funksjonalitet for å sette multiplisitet etc.

Import og eksport-funksjonaliteten til EVN støtter mange formater, men det burde vært mulig å importere relasjoner og unik ID for konseptene.

Det er enkelt å vedlikeholde og opprette konsepter, men det ser ut som det er vanskelig å snevre inn skopet/kontekst ved søk. Slik at etterhvert som antall vokabularer stiger vil det bli vesentlig

vanskeligere å modellere i verktøyet. Dette gjelder spesielt når man skal lage relasjoner til et Begrep hvor det eksisterer flere med samme navn.

Søkealternativene i EVN er gode, men vesentlige funksjoner som å velge topnivå ser ikke ut til å støttes. Det kan derfor bli vanskelig å søke kun innenfor en etat.

Vi savner også mulighet for å publisere vokabular offentlig, slik at en bruker ikke behøver å logge seg på for å se det.

EVN er en enkel editor som gjør det enkelt å bygge og vedlikeholde begrensede vokabular, men som kanskje ikke er like god på mer avanserte funksjoner. ~~At man ikke har mulighet for å sammenlikne elementer i søk er en mulig show-stopper.~~ Nå viser det seg at man kan det likevel



# Poolparty (PP)

PoolParty er en web-basert løsning for semantiske vokabular ved bruk av åpne standarder (SKOS). Verktøyet har et enkelt grensesnitt som muliggjør drag and drop editering, og som gir god visuell oversikt over modellene.

## Dokumentasjon

**Brukerdokumentasjon:** <https://grips.semantic-web.at/display/public/POOLDOKU/PoolParty+Thesaurus+Server+Documentation>

**Tutorial:** <http://goo.gl/dg8s0>

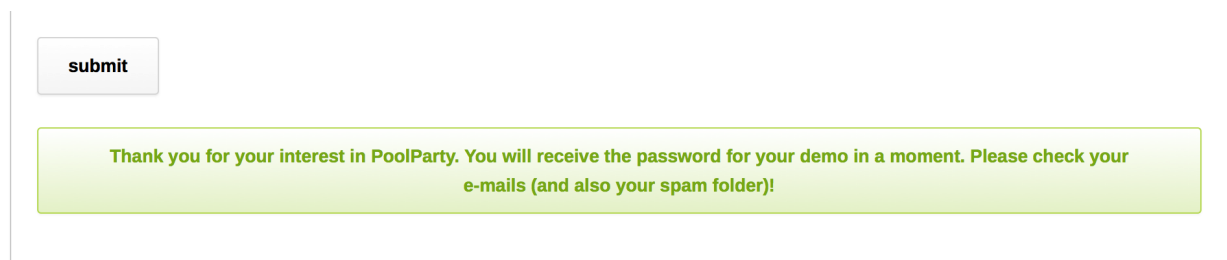
**Demo:** <http://www.poolparty.biz/test-demo/>

**Weblinar:** <http://www.poolparty.biz/portfolio-item/semantic-web-for-developers/>

**Løsninger:** <http://www.poolparty.biz/solutions/>

## Innledende testing

PoolParty har ingen standalone demo-versjon. Vi må sende en forespørsel om å få en testkonto på demo-instans som kjører hos PoolParty. Vi sender en forespørsel om en testkonto vha. Skjema på PP sine sider.



The screenshot shows a web interface with a light gray 'submit' button. Below it is a green rectangular box containing the text: 'Thank you for your interest in PoolParty. You will receive the password for your demo in a moment. Please check your e-mails (and also your spam folder)!'

*Figur 26: Respons angående testkonto*

Etter 2 arbeidsdager og en helg er det fortsatt ikke kommet noen respons per email.

Etter kontak med Poolparty viser det seg at mail med autentiseringsinformasjon har en tendens til å havne i spamfilteret. Jeg har brukt en mailkonto som sletter all spam uten å stille spørsmål.

Når jeg prøver å registrere meg på nytt med tellu-mailen får jeg svar som blir fanget av spam-filteret.

Vi får tilgang til en testkonto som virker veldig åpen. Det ligger mange gamle påbegynte prosjekter, og Custom Schema Types i instansen. Dette betyr at vår data blir delt med, og kan endres av, andre med tilgang til demoinstansen hos PP.

## Erfaringer med leverandør og tekniske utfordringer

Vi har ikke hatt allverdens kontakt med PoolParty. Som nevnt ble automatisk generert informasjon om testkonto slettet av spamfilteret. Da vi sendte en henvendelse til PP om dette fikk vi svar veldig kjappt.

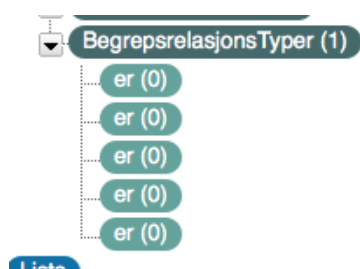
Det kommer til statighet opp en loadbar når man trykker på noe i hierarket til venstre. Dette oppleves veldig forstyrrende. Sannsynligvis fordi den driver og sender AJAX-forespørsler til server. Dette skjer også når du høyreklikker. Dette tar ofte veldig lang tid, for praktisk talt ingenting.

Menyen over headeren har en tendens til å forsvinne. Da det ikke er noen scrollbar på siden er eneste mulighet å hente denne tilbake å bruke piltastene. Tungvint.

Testserveren hang seg opp under testing, og sluttet å svare på forespørsler. Til slutt fikk jeg en 502 Bad Gateway. Forhåpentligvis er dette noe vi slipper unna ved å bruke en instans vi selv har kontroll på. Etter en stund kom den opp igjen. Dette har skjedd flere ganger.

Da vi fikk tilgang til testserver ble denne gitt til <http://open.poolparty.biz/PoolParty/>. Senere, mens vi testet, ble opplegget for dette tydeligvis endret. Addressen ovenfor går nå til en side hvor man velger mellom tre forskjellige test-servere. Dette medførte for det første en del forvirring om hvor testdata vi allerede hadde brukt tok veien, samt at tidligere testdata hadde id'er som pekte til feil URI.

Hvis man linker objekter sammen opprettes en tilbakereferanse hos referenten. Hvis man trykker på denne linken dupliseres begrepet i oversikten. (Se figur nedenfor for 'er' Figur). Duplikatene forsvinner når man refresher siden, men dette er en bug. (Dette skjer ikke hver gang.)



Figur 27: Bug

Ved et tilfelle da jeg forsøkte å importere på conceptnivå, ved å høyreklikke og velge import dupliserte den de konseptene jeg hadde. Og jeg fikk to synkrone hierarkier som tilpasset seg hverandre når jeg endret den ene. Dette er en bug.

En feil liknende den nevnt ovenfor observert ved følgende fremgangsmåte. Hvis man har alle sine begreper som *Top Concepts*. Legger inn et nytt *Top Concept* man vil bruke til å kategorisere et sett begreper. Når man endrer begrepets *Broader Concepts* for å legge det under *Top Concept* vi opprettet for Kategori dupliseres begrepet. Det vil eksistere som *Top Concept* og *Narrower Concept* for kategorien. Begge begrepene har samme ID. Det legges opp til at alle Begreper (i vårt tilfelle) skal ha inndeling i form av *Top Concepts*. Men dette bør være et designvalg vi står fritt til å ta. For å rydde opp i denne feilen må man for hvert begrep trykke *Remove as Top Concept*.

Når man importerer data kommer det opp en statusbar som viser at systemet jobber. Når denne er ferdig kommer det opp en dialog som sier hva som er importert, eventuelt om noe er galt slik at import ikke støttes. Hvis man importerer noe som PP ikke kjenner igjen, men som den heller ikke feiler på kommer det ingen tilbakemelding. Statusbaren forsvinner og det er det.

Konseptenes unike ID er en uri. Hvis man trykker på en URI kommer man til PoolParty-browser for det konseptet man trykker på. PP endret testserver fra open

Oppdatering. Dette ser ut til å fungere lenger. Et eller annet skjedde, en ny testserver ble deployet eller tilsvarende. Hvis man trykker på linken til som er begrepsid'en nå kommer man bare til login-siden. ID for begrepet *Dato* under *Konkursregisteret* har id: <http://enterprise.poolparty.biz/SemicolonTest/Konkursregisteret/Begreper/Dato>. Tidligere var testversjonen open.poolparty.biz, ikke enterprise. Dette ble endret mens vi testet, og det resulterte i at alle ID'ene vi hadde laget pekte feil.

Editorene er, ved relasjoner, basert på at man skriver inn et begrep og så skal det komme forslag til eksisterende Concepts. Hvis navnet er kort, som for eksempel begrepsrelasjonstypen *er*, kommer ikke forslaget opp (Dette gjelder også søk). Og man kan ikke bruke conceptet med mindre man legger til en hidden label for alternativt søk.

## Informasjon om test-instans

Mail fra PoolParty:

Now, login to your PoolParty Thesaurus Server test account and build your own thesaurus:  
<http://open.poolparty.biz/PoolParty/> (user=XXXXXXXXXX / password=XXXXXXXXXX)

**(!OBS! Denne urlen peker nå bare til tre instanser. Bruk enterprise for siste versjon av semicolonprosjektet)**

To test the text mining component of PoolParty, please visit <http://open.poolparty.biz/extractor> - see how it works: <http://goo.gl/XEoMI>

Which use cases in the area of thesaurus management, search, text mining, data integration and/or linked (open) data do you have in mind?

Just let me know if you want to talk with us in person and we will set up a phone/skype/gotomeeting session.

## Rolle, workflow og brukerstøtte

PP har følgende roller:

- PoolParty User – standardrolle som kan tilegnes alle brukere som kan logge på.
- Administrator – gir tilgang til avansert-menyen som har flere avanserte konfigurasjonsmuligheter.
- Superadministrator – tilgang til brukeradministrasjon og admin-dashboard
- Public – Dette er ikke en rolle man tilegner en bruker. Man tilegner rollen til et prosjekt for å indikere at det skal være offentlig lesbart.
- Wiki Editor – gir tilgang prosjektets wiki-view, og kan gjøre endringer via dette. Har ikke tilgang til andre views i prosjektet.

Ut over disse rollene har ikke PP noen støtte for brukere i en workflow. I følge dokumentasjonen foreslås workflow på følgende måte: Konsepter utvikles på en utviklings-instans av PP. Deretter Stages det på en testserver, for så å bli deployet på en offentlig- / produksjonsserver.

Det er ikke lagt opp til noen roller i workflowen ut over de som er nevnt ovenfor.

## Støtte for DIFI-Ontologi

Å jobbe med ontologier er ikke umiddelbart like intuitivt som det var for EVN. PP har enorme mengder metadata per konsept. Konseptene i PP har ikke type eller klasse. De deles inn hierarkisk i følgende orden: *Prosjekt->Concept Scheme->Top Concept->Narrower Concept*.

Man må ikke bruke alle disse. Narrower Concept kan inneholde andre Narrower Concept. Det ser ikke ut til å være noen forskjeller mellom *Top Concept* og *Narrower Concept* ut over type. De har de samme attributtene. Det eneste er at de har forskjellig farge i grensesnittet.

Figur viser Custom Schema som eksisterer i instansen vi bruker. Difihar vi laget selv. SKOS er default. Resten er laget av andre brukere. For å endre/legge til custom attributes velger vi *Advanced->Custom Schema* i menyen til PP. En dialog kommer da opp.



Figur 28: Custom Schema

Legg først til et *New Schema* ved å skrive inn navnet i input-feltet. Trykk Add. Legg deretter til attributtene.

**Schema** – Skjemaet du opprettet

Figur 29: Custom Schema Dialog

**Type** – Label eller URI. Eventuelt concept for referanser. Typene som støttes er *Concept*, *URI*, *Label*, *Literal*, *Integer*, *Date* og *Boolean*.

**Label** – attributtnavnet

URI- styres automatisk, men kan overrides. Det er også mulig å sette multiplisitet for attributter som ikke er relasjoner. Relasjoner må være mange til mange.

Trykk Save. Gå tilbake til tab'en *Create Schema* for å legge på flere attributter. Når du er ferdig trykker du *Close*.

## Versjonering og mulighet for å spore endringer

Figur 30: Forenklet DIFI-Skjema type

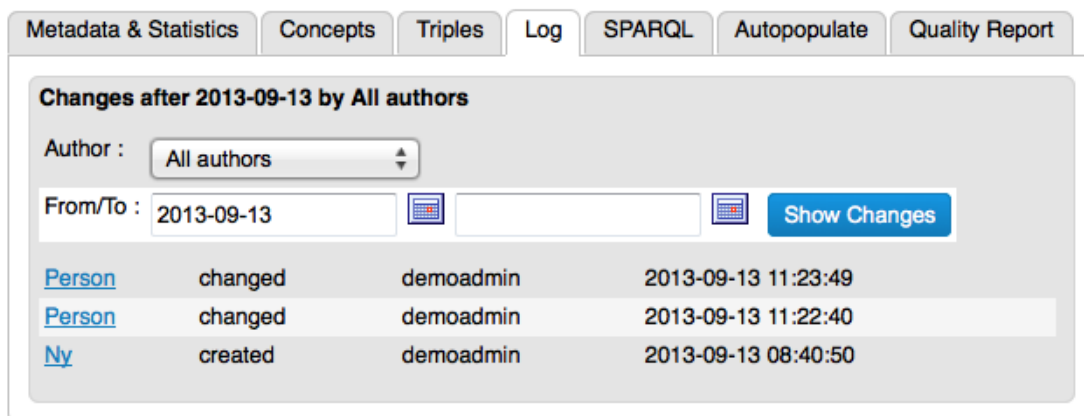
Som nevnt under har PP ingen implisitt støtte for workflows. Versjonering må derfor administreres ved å ha installasjoner på forskjellige servere.

PP har ingen avansert funksjonalitete for å se endringer gjort på modellen. Hvis man velger et konsept og trykker fanen *Metadata* får man opp informasjon om hvem som gjorde siste endring, og når den ble gjort. Men det gis ingen informasjon om hva denne endringen var.

Figur 31: Endringer for et begrep

Feltet *Version* inkrementeres med en for hver endring som blir gjort. Nye konsepter har ikke dette feltet.

Loggen holder oversikt over hvor det er blitt gjort endringer, men ikke hva som ble endret som vist i Figur.

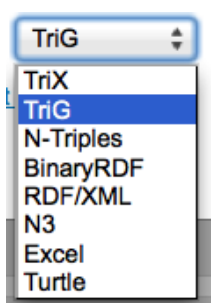


Changes after 2013-09-13 by All authors			
Author :	All authors		
From/To :	2013-09-13		Show Changes
<a href="#">Person</a>	changed	demoadmin	2013-09-13 11:23:49
<a href="#">Person</a>	changed	demoadmin	2013-09-13 11:22:40
<a href="#">Ny</a>	created	demoadmin	2013-09-13 08:40:50

Figur 32: Log

## Import og eksport av Data

PP støtter import fra en rekke formater. De formatene som dukker opp i dropdown listen man kan velge mellom i importdialogen er vist i Figur.



Figur 33: Importformater

Vi tester først å importere en Excel-eksport fra EVN, men denne gir feilmeldingen vist i Figur. Feilmeldingen indikerer at kolonnene i stylesheetet må ha et spesifikt format. Dette er en uting med hensyn på sømløs overføring mellom verktøyene, men det kan også spare oss fra en manuell prosess med å mappe kolonnene over til attributter i PP. Det skal nevnes at PP ikke støtter eksport på dette formatet, og at det derfor kan være vanskelig å se hva disse kolonnene må hete. Noe informasjon om dette gis i dokumentasjonen.



Figur 34: Feilmelding ved import av Excel-fil

Vi tester videre import via Turtle, da turtle er litt kraftigere enn stylesheets. Vi gjør en kjapp test av å importere turtle-data brukt under testing av topbraid. Vi gjør dette på samme måte som ved initiell testing. I.E. Høyreklikke på prosjektet og velge import. På samme måte som tidligere lukker dialogen seg uten tilbakemelding. Vi må sannsynligvis gjøre dette på en annen måte.

For videre forsøk lager vi et begrep i PP i prosjektet som gjør bruk av DIFI schema.

Vi eksporterer dette til turtle og ser hva som er forskjellig fra eksporten vi fikk fra EVN.

Det ser ut til å være et noen forskjeller i syntax. Det ser ut til at alt blir gitt ID med løpenummer, som ikke er spesielt oversiktlig. For eksempel for bedrift i vårt eksempel er identifikatoren <<http://open.poolparty.biz/Testulus/2>> (Testulus er navnet på prosjektet). Id 0 og 1 er henholdsvis Begreper og Konkursregisteret. Vi kan ikke skrive denne ID'en in selv når vi oppretter elementet i PP, men vi kan velge hva den skal være hvis vi oppretter data ved import.

PP er ikke så nøye som EVN i å eksplisitt definere alle attributter og typer som brukes i eksporten. Dette betyr i praksis at definisjonen av ontologien ikke er med i eksporten, slik den er for EVN.

Det at vi ikke har konsepttyper i PP kan bidra til å gjøre ting litt mindre oversiktlig. Dette gjør i praksis at alle *Concept Schemas* vil gjelde for alle konsepter i PP. Hvis vi ønsker å skille prosjekter med forskjellige attributter, slik som SERES gjør med Begrep og Begrepsrelasjon, vil ikke dette la seg gjøre i PP. Alle konsepter må ha alle attributter, men ikke alle attributter behøver å ha verdi.

Man kan legge til *Related Concepts*, men det ser ikke ut som om man får noe kontroll over navngiving eller instans av relasjonen. Alle relasjoner i PP er bilinære (dette kan jukses til fra importen til å kun gjelde en vei, men ikke for elementer som er behandlet i PP). Hvis vi sier at *Personnavn ER Navn*, vil *Navn* ha relasjonen *Navn ER Personnavn*, som ikke er konstant.

Inntrykket er at PP er bedre egnet til å håndtere Data enn metadata.

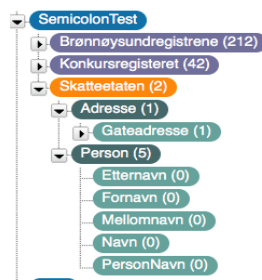
Vi har utvidet schema-typen til Difi til å inneholde Concept type, til og fra.

The screenshot shows a web-based configuration interface for a custom schema type. At the top, there are tabs for 'Details', 'Metadata', 'Documents', 'Linked Data', 'Triples', 'Visualization', and 'Quality Report'. Below these, there are sub-tabs for 'SKOS', 'Online\_Learning\_Relations', and 'Difi'. The 'Difi' tab is selected. The main area displays a list of attributes for a concept type, each with a label and a dropdown menu for selecting a type. The attributes are: 'er', 'synonym', 'omfatter', 'guid', 'gyldigFra', 'navn', 'harRolle', 'gyldigTil', and 'definisjon'. The 'navn' attribute has a dropdown menu with 'no' and 'en' options. The 'definisjon' attribute has a dropdown menu with 'no' and 'en' options.

Figur 35: Custom Schema Type for

DIFI

Vi kan strukturere det på følgende måte som vist i Figur. Ingen av Concept'ene i figuren har fått angitt noen type vi selv har definert. Alle er skos:Concept av type *Concept Scheme*, *Top Concept* og *Narrower Concept*. Og alle har samme attributter for Difi-schema, men ikke alle har det samme attributtene satt. Adresse og Person har ikke verdier for DIFI-Schema Type.



Figur 36: Strukturering

Vi gjør et forsøk ved å lage en parser som kan oversette SERES til turtle for PP, og prøver å gjøre en eksport.

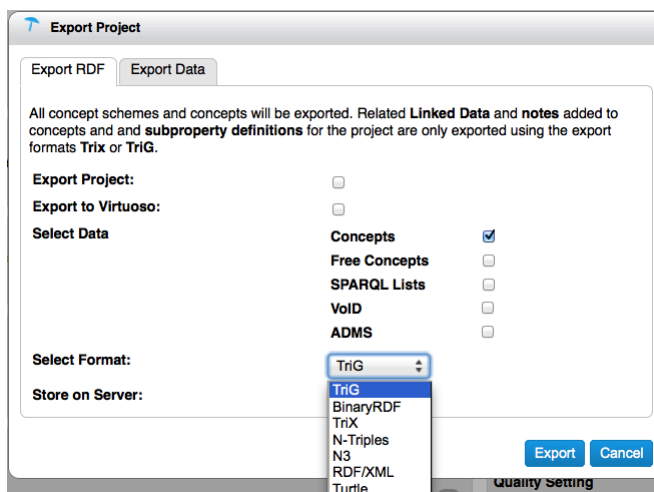
PP gjør bruk av tilbakepekere i sin turtle-eksport. Hvis vi for eksempel har en Begrepsrelasjon mellom to Begreper vil Begrepene ha en referanse tilbake hverandre. Denne blir bare synlig på weben dersom vi tar den med i turtle. På denne måten kan vi lage enveis relasjoner ved import, hvis vi ikke tar med relasjonen for begge Begrepene. Dette er feil i syntaks, og kan ikke vedlikeholdes av PP sin editor.

En annen ting som avanserer transformering mellom SERES-XMI og PP-Turtle er at alle nivåene må ha referanse til alle sine underconcepts. For eksempel i Figur må konseptet *Person* ha referanse til både *Etternavn* og de andre begrepene under. Ellers blir den ikke vist i strukturen på weben. I tillegg må *Etternavn* ha referanse tilbake til *Person*. Hvis ikke begge disse er på plass i Turtle-importen vil ikke strukturen vises på weben.

## Eksportmuligheter

PoolParty sine eksportmuligheter virker noe mer begrenset enn importmulighetene. Da verktøyet støtter import av stylesheets skulle vi også gjerne sett at verktøyet hadde mulighet til å eksportere til dette formatet. Figur viser dialogen for eksport i PP. Vi kan eksportere hele prosjektet, eller et subtre fra et valgt nivå. For et statlig organs domene velger vi typisk det siste.

Fanen *Export Data* lar oss eksportere alle dokumenter og opplastede filer.



Figur 37: Eksportdialog

Formatene *Trix* og *Trig* brukes dersom det er mappet linked data til prosjektet. Det er disse formatene som anbefales å bruke for versjonering.

Dokumentasjonen sier lite om de forskjellige formatene.

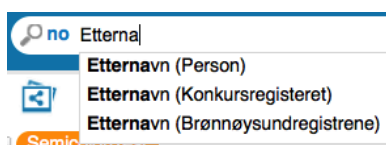
En ting vi har bitt oss merke i er at Turtle eksporten ikke inneholder ontologien. Hvis dette er påkrevd av verktøyet du skal importere til bør du kanskje sjekke ut de andre formatene.



## Navnerom

Strukturen til PP er i høyeste grad lagt opp til at man skal dele inn konseptene i navnerom. Et navnerom vil typisk være et *Top Concept*, men også *Narrower Concepts* kan ha konsepter under seg. Søke- og refereringsmekanikken i PP viser også konseptets *Broader Concept* for å skille konsepter med likt navn fra hverandre.

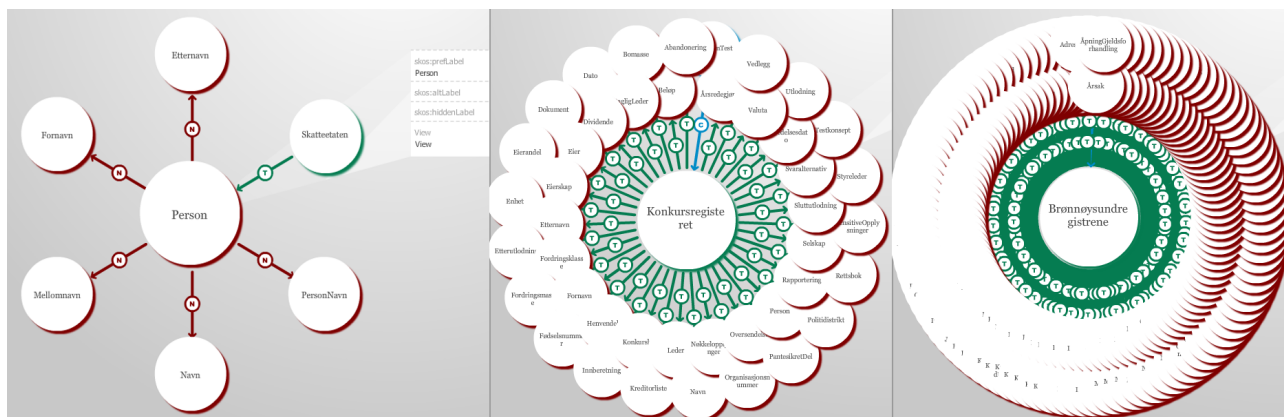
I Figur er *Etternavn* et *Top Concept* i Konkursregisteret og Brønnøysundregistrene. I Skatteetaten er det lagt under *Top Concept Person*. Hvis vi hadde hatt samme strukturen i alle domenene, med *Person* som kategori for *Etternavn* ville vi fått tre like resultater opp i søket. Dette er samme mekanikk som brukes for å relatere to begreper sammen. Det vil derfor oppstå problemer med å velge riktig begrep når man skal opprette relasjoner.



Figur 38: Søkeforslag fra PP

PP har også en fancy navigeringsfeature. Hvis man trykker *Visualization* på et konsept får man opp et vakkert animert navigerbart diagram som viser innholdet i konseptet samt detaljer for konseptet. Dette er en oversiktlig og fin måte å traversere modellene på hvis de er strukturert fornuftig. Hvis man har mange eller alle begrepene på et nivå derimot kan diagrammet derimot bli relativt ubrukkelig som vist i Figur. Figuren viser Visual Browser for Skatteetaten, Konkursregisteret og Brønnøysundregistrene.

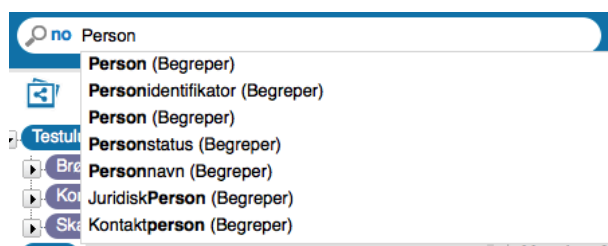
## Søk- og sammenlikningsmuligheter



Figur 39: Visual Browser

Det ser ikke ut som om søkefunksjonaliteten til PoolParty er spesielt avansert. Det eneste vi kan finne på dette fra brukerdokumentasjonen er at man kan søke i et generisk søkefelt øverst til venstre på siden. Som vist i Figur.

Det er ikke noen måte å innsnevre søket basert på attributtverdier, plassering eller eksakt match.



Figur 40: Søkefelt

Ettersom antall vokabularer øker og antall begreper med samme navn øker kan det bli en lang liste å



forholde seg til ved søk.

Merk at Person kommer to steder i listen. Dette er to forskjellige instanser. På grunn av lik struktur i domenene har begge plassering Begreper. Det er derfor ikke mulig å se hvilken som hører til Brønnøysundregisteret og hvilken som hører til konkursregisteret (Uten en annen struktur).

Vi har heller ikke funnet ut hvordan jeg kan bytte fra *no* til for eksempel *en* i søkefeltet. I følge dokumentasjonen vil søk foregå på det som er satt opp som default språk for prosjektet.

Man har mulighet til å gjøre SPARQL forespørsler. Dette vil i så fall kunne gi bedre søk.

I Figur har vi søkt etter definisjonen for alle begreper som har preferredLabel Person (på norsk). Resultatet vises under. Dette hadde vært en fin funksjonalitet dersom man kunne laget maler for søket og lagre disse. Det er ikke tilfellet. Man må skrive inn hele SPARQL-forespørselen hver gang. Det er nok litt i overkant av hva vi kan forvente av hvermannen eller etat-hvermannen.

The screenshot shows a web interface with tabs: Metadata & Statistics, Concepts, Triples, Log, SPARQL, Autopopulate, and Quality Report. The SPARQL tab is active, showing a query endpoint and a query text area. The query is: `PREFIX skos:<http://www.w3.org/2004/02/skos/core#>  
SELECT ?x ?definisjon  
WHERE { ?x skos:prefLabel "Person"@no .  
?x <http://open.poolparty.biz/PoolParty/schema/Difi#definisjon> ?definisjon }`. To the right, there is an 'Add Namespace' section with checkboxes for SKOS, DC, DCTerms, OWL, RDF, and RDFS. Below the query area is a 'Run Query' button. The results are displayed in a table with two columns: 'x' and 'definisjon'. The first row is: `http://open.poolparty.biz/Testulus/Konkursregisteret/Begreper/Person` and `"Personsdefinisjon for konkursregisteret"@no`. The second row is: `http://open.poolparty.biz/Testulus/Brønnøysundregistrene/Begreper/Person` and `"Persondefinisjon for Brønnøysundregistrene"@no`.

x	definisjon
<code>http://open.poolparty.biz/Testulus/Konkursregisteret/Begreper/Person</code>	<code>"Personsdefinisjon for konkursregisteret"@no</code>
<code>http://open.poolparty.biz/Testulus/Brønnøysundregistrene/Begreper/Person</code>	<code>"Persondefinisjon for Brønnøysundregistrene"@no</code>

Figur 41: SPARQL-søk

Hvis man velger *Concepts* tab'en i vist i Figur får man mulighet til å søke litt mer avansert enn med måten nevnt i dokumentasjonen. Her kommer også resultatene opp i en liste.

*Concept Index* kan også brukes for å finne konsepter basert på navn.

The screenshot shows a web interface with tabs: Concept Search and Concept Index. The Concept Search tab is active. There is a 'Filter By' section with a dropdown menu showing 'Any label', 'Preferred label', 'Alternative label', and 'Hidden label'. The search criteria are 'contains' and 'no Person'. A 'Search' button is present. Below the search area, there is a table with the following columns: Concept, Alternative Labels, Broader Concepts, # Narrower Concepts, and Delete. The table contains the following rows:

Concept	Alternative Labels	Broader Concepts	# Narrower Concepts	Delete
<a href="#">JuridiskPerson</a>		<a href="#">Begreper</a>	0	
<a href="#">Kontaktperson</a>		<a href="#">Begreper</a>	0	
<a href="#">Person</a>		<a href="#">Begreper</a>	0	
<a href="#">Person</a>		<a href="#">Begreper</a>	0	
<a href="#">Personidentifikator</a>		<a href="#">Begreper</a>	0	
<a href="#">Personnavn</a>		<a href="#">Begreper</a>	0	
<a href="#">Personstatus</a>		<a href="#">Begreper</a>	0	

Figur 42: Concept Search

Sammenlikning av konsepter kan utføres vha. SPARQL-spørring. Ut over dette er det ingen spesifikk funksjonalitet for å bygge views på dette.

## Evaluering av UseCase

### UC1: Vedlikehold av terminologi / vokabular for et statlig organ

PP har ikke støtte for rolle-indeling som en del av workflow. Rollene i PP sier bare hvilken tilgang man har til selve editoren. Vi kan derfor ikke opprette et rolleregime for å administrere workflow.

#### Opprette et nytt konsept

Konseptene ordnes hierarkisk på samme måte som i EVN. PP vedlikeholder ikke Ontologi på samme måte som topbraid. Man kan lage en *Custom Schematype* som er et sett av attributter. Vi har for eksempel laget schematypen DIFI. Men man kan ikke begrense denne til kun å gjelde for typer av konsepter. Vi har derfor som nevnt ovenfor laget en generell schematype for DIFI som inneholder alle interessante attributter.

For å lage et nytt konsept velger man nivået i modellen der man ønsker å legge konseptet. Hvis vi for eksempel skal legge til et nytt Begrep velger vi for eksempel Konkursregisteret, høyreklikker og velger *Create Narrower Concept* (eller *Create Top Concept* avhengig av struktur).

Skriv inn navnet på konseptet i dialogen som kommer opp.

Konseptet er nå opprettet under begreper. Trykk tab'en Difi for å legge til attributter.

I DIFI *Concept Schema* kan vi legge til navn, guid, definisjon og en gyldig fra dato. Vi kan også legge til relasjoner til andre begreper. Trykk pluss tegnet ved for eksempel *er* og begynn å skrive navnet på referenten. Forslag skal komme opp i liste nedenfor. Velg referent og trykk ok. Merk: hvis referenten har kortere navn enn tre bokstaver vil det ikke komme opp i forslagslisten. I de tilfellene må referenten bruke *Hidden Label* for å gjøres søkbar.

Det ser ut som at all interaksjon skjer på prodsatt vokabular. Det er ingen støtte for workflow nevnt i userguiden. En mulig måte å gjøre dette på er å bruke forskjellige prosjekter, og importere endringer til prod fra "Workspace"-prosjekt. Dokumentasjonen foreslår å bruke forskjellige server-instanser for versjonering.

Det ser ikke ut som man har kontroll over URI'en konseptet får. Dette er styrt av serveren. Vi kan kontrollere dette for importerte konsepter.

#### Lese konsepter

Man velger begreper i hierarki på venstre side, på samme måte som i EVN. Når man velger et konsept kommer detaljer opp til høyre. Dette viewet inneholder tabber for å velge forskjellige konsept schema.

I følge dokumentasjonen skal man kunne sette Usergroups på prosjektet. Hvis man legger til usergroup *Public* skal vokabularet være synlig for alle brukere av PoolParty-instansen, også de som ikke er logget inn. Hvis man går til url'en som er konseptets unike ID får man opp en detaljside om konseptet som viser alle attributter som er satt.

The screenshot shows the PoolParty Linked Data Server interface. At the top, there is a logo for 'poolparty LINKED DATA SERVER' and a search bar. Below the logo is a navigation bar with letters A through Z. The main content area is titled 'Person' and shows a list of properties and their values. The 'VISUAL' tab is selected. The properties listed are:

- http://www.w3.org/2004/02/skos/core#broader**: Begreper
- http://open.poolparty.biz/PoolParty/schema/Difi#guid**: <http://seres.no/guid/Brønnøysundregistrene/Begrep/Person/4926>
- http://purl.org/dc/terms/creator**: demoadmin
- http://www.w3.org/1999/02/22-rdf-syntax-ns#type**: <http://www.w3.org/2004/02/skos/core#Concept>
- http://purl.org/dc/terms/created**: 03 September 2013 11:28 CET
- http://purl.org/dc/terms/contributor**: demoadmin
- http://open.poolparty.biz/PoolParty/schema/Difi#navn**: Person
- http://open.poolparty.biz/PoolParty/schema/Difi#definisjon**: Person er et individ, enkeltmenneske.

At the bottom right of the interface, there is a 'Figur' label.

43: PoolParty Browser

Man kan også bruke *Visual Browser* som vist i Figur.

## Oppdatere konsepter

Oppdatering av konsepter gjøres på samme måte som ved opprettelse av konseptet. Som tidligere nevnt ser det ikke ut som om man kan branche dette ut i egen workflow.

## Søke, navigere og sammenlikne konsepter

PP er bygget opp av mange forskjellige views. Vi har også mulighet til å lage egne, eller integrere søket til PP i egne web-applikasjoner. Det er derimot ingen god måte å snevre inn søkene. Søkene

vil treffe på preferredLabel, alternativeLabel og andre attributter i synkende rekkefølge. Vi skulle gjerne sett at det var mulighet til å snevre inn søk mhp. Domene, navnerom etc.

Navigering i editoren er oversiktlig og intuitivt. Vi har sett hvordan vi kan strukturere vokabularet hierarkisk med navnerom og navigere i treet dette spinner. Vi kan også bruke *Visual Browser* som tilbyr en enkel måte å traversere modellen (dersom den ikke har for mange konsepter på hvert nivå).

Sammenlikning av konsepter kan kun gjøres ved SPARQL-requests. Det er mulig vi kan få litt mer støtte for dette dersom vi integrerer egen front-end mot PP hvor vi selv har kontroll over rendringen av data.

### Lese konsepter og relaterte attributter

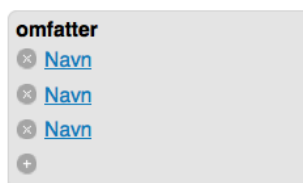
I motsetning til EVN er alle relasjoner (tilsynelatende) bilinære. Det betyr at man kan navigere mellom to konsepter basert på en relasjon mellom dem.

Det kan være vanskelig å se hvilken instans som er referert til i viewene for detaljer for et konsept da de kun representeres ved sitt navn. Hvis flere konsepter heter det samme vet man ikke hvilken instans som er referert til.

### Eksportere til andre verktøy

Dette er ikke testet, men PP støtter en rekke formater for eksport. Desverre støttes ikke stylesheets for eksport. Stylesheets er et greit eksport-format for å kunne dekke i bredden, men det er ikke kan være noe svakt i forhold til å representere relasjoner eller struktur i modellen.

De andre formatene som støttes må i de fleste tilfeller, kanskje alle, innom et mellomledd før vi kan importere dem i andre verktøy. For å importere til SERES må vi lage en egen parser for dette, da SERES kun støtter egendefinert XML. I Figur har vi relatert et Begrep til tre forskjellige instanser med samme navn. Det er umulig å se hvilken som er hvilken.



Figur 44: Relasjoner med samme navn

### Endre status på konsepter

Det er ingen standard støtte for dette i PP, men som med EVN er det fullt mulig å legge dette til som et egendefinert attributt vi selv vedlikeholder.

Det samme gjelder gyldighetsperiode. Dette er noe vi selv legger på, og vi må selv sørge for at begreper som ikke lenger er gyldige eller som blir gyldige er tilgjengelig i offentlig PP-instans.

### Opprettholde konsistens for en gruppe begreper

Versjoneringen foregår ved å eksportere fra en server til en annen. Eksporten som brukes for import kan brukes som versjonsrepresentasjon. Man kan da i ved fremtidige releaser rulle tilbake til denne, etc.

## UC2: Publisering av terminologi / vokabular

### Publisere konsepter

Poolparty tilbyr et property *Usergroups* på prosjektene. Hvis man tar med *public* i dette attributtets verdier skal vokabularet/prosjektet i teorien være tilgjengelig for allmennheten.

The role **Public** is a special role in PoolParty. Assigning the role **Public** to a project indicates that the project is published and the [Linked Data frontend](#), the [Wiki view](#) and also the [SPARQL endpoint](#) are available. If the **Public** role is not assigned to a project, you have to log in to the backend to access the Wiki view and the Linked Data frontend and SPARQL endpoint will not be available.

Når et prosjekt er public kan alle få lesetilgang til det uten å logge på. Tilgang gis via Linked Data Frontend, Wiki view eller SPARQL-endpoint.

Linked Data Frontend er viewet vi kommer til når vi trykker på URI'en som er konseptets ID. Dette er vist i Figur.

PP har flere web-frontends, og muligheter for å lage flere.

Vi har også mulighet til å lese konseptene som rdf ved å legge *.rdf* til URI'en som er id for et konsept. For eksempel vil

<http://enterprise.poolparty.biz/SemicolonTest/Brønnøysundregistrene/Begreper/Personnavn.rdf> resultere i RDF vist i Listing 1.

Frontend for vårt test-prosjekt for semicolon kan ses på

<http://enterprise.poolparty.biz/SemicolonTest.html>.

```
<rdf:RDF>
  <rdf:Description rdf:about="http://enterprise.poolparty.biz/SemicolonTest/Brønnøysundregistrene/Begreper/Personnavn">
    <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
    <dcterms:creator>dmoadmin</dcterms:creator>
    <dcterms:created rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2013-09-23T10:58:18Z</dcterms:created>
    <skos:prefLabel xml:lang="no">Personnavn</skos:prefLabel>
    <skos:topConceptOf rdf:resource="http://enterprise.poolparty.biz/SemicolonTest/Brønnøysundregistrene"/>
    <dcterms:contributor>dmoadmin</dcterms:contributor>
    <guid rdf:resource="http://seres.no/guid/Brønnøysundregistrene/Begrep/Personnavn/4997"/>
    <gyldigFra rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2013-09-23Z</gyldigFra>
    <navn xml:lang="no">Personnavn</navn>
    <definisjon xml:lang="no"> Personnavn består av fornavn og ett enkelt eller dobbelt etternavn og kan i tillegg ha mellomnavn.</definisjon>
    <er rdf:resource="http://enterprise.poolparty.biz/SemicolonTest/Brønnøysundregistrene/Begreper/Navn"/>
    <omfatter rdf:resource="http://enterprise.poolparty.biz/SemicolonTest/Brønnøysundregistrene/Begreper/Etternavn"/>
    <omfatter rdf:resource="http://enterprise.poolparty.biz/SemicolonTest/Brønnøysundregistrene/Begreper/Mellomnavn"/>
    <omfatter rdf:resource="http://enterprise.poolparty.biz/SemicolonTest/Brønnøysundregistrene/Begreper/Fornavn"/>
  </rdf:Description>
</rdf:RDF>
```

Listing 1: RDF for personnavn

## Søke, navigere og sammenlikne konsepter

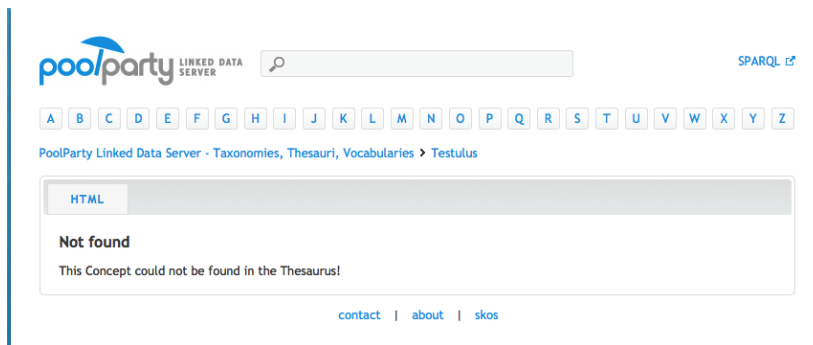
For et prosjekt som er gjort *public* kan brukere som ikke er logget inn traversere eller søke vokabularet ved å bruke Linked Data frontend, Wiki View eller SPARQL endpoint som nevnt ovenfor.

Navigasjon blant konseptene gjøres ved å trykke på linker, avhengig av hvilket view du bruker til å navigere. Linked Data frontend eller wiki har en link-struktur som gjør hierarkiet lett å navigere.

Søkefunksjonen i Linked Data Server ser ikke ut til å være den beste. Når vi søker på navn går den direkte til instansen av Brønnøysundregistrenes konsept Navn, selv om også konkursregisteret har et konsept med samme navn. Søker vi på Etternavn kommer ingenting opp, selv om begge vokabularene har dette konseptet. Indekseringen til dette viewet (bokstavknapper) ser heller ikke ut til å fungere.

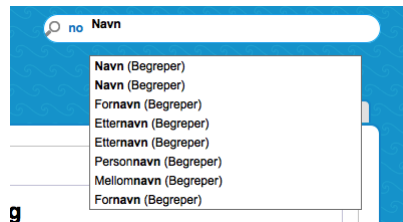
Det kan virke som om frontenden er under konstant endring. Når vi nå (senere) forsøker å søke kommer forslagene opp slik de gjør når man søker i editoren. Hvis man trykker på et av forslagene får man en 404.

Søker man i Wiki'en er resultatet som i de andre editorene.



Figur 45: Indeksert søk

Man får opp forslag mens man skriver, og må velge et av disse for å søke. Hvis strukturen er lik, altså at navn og navn på overordnet concept er likt, blir det umulig å skille forslagene.



Figur 46: Søkeresultater

Hvis man ønsker å sammenlikne konsepter er det så vidt jeg ser to muligheter.

Den mest intuitive er å åpne enten to faner eller to browser-vinduer og søke opp begrepene hver for seg. Deretter stille inn størrelsen på vinduene slik at de kan vises ved siden av hverandre.

Den andre muligheten for litt mer avanserte brukere er å bruke SPARQL til å spørre etter nøyaktig de instansene og informasjonen man er interessert i å sammenlikne. God SPARQL-støtte gir endel frihet m.h.p. Søk og sammenlikning av konsepter. SPARQL view er tilgjengelig for public-brukere.

Ut over metodene nevnt ovenfor er det ingen visuell støtte direkte for å sammenlinke konsepter i PP.

## Lese konsept og relaterte attributter

Hvis man bruker wikien til å søke etter konsepter ser det ikke ut til at attributter fra Custom Schema, for eksempel Difi, kommer med.

Hvis man gjør SPARQL-requests kommer derimot denne informasjonen med hvis den er med i forespørselen.

I Linked Data frontend kan man også se på attributtene.

## Endre en eller et set med attributter

Public-brukere har ikke tilgang til å endre attributter på Linked Data Server.

Public-brukere har muligheter til å legge notater til på wiki-siden.

Public-brukeren har ikke tilgang til standard editor view.

## Export concept to other tools

Public-brukere har tilgang til SPARQL endpoint, som kan eksportere svaret til XML, HTML eller JSON. Etter transformasjon vil dataen kunne importeres til andre verktøy.

I tillegg er RDF-endpoint også åpent for publiserte vokabular. Andre formater som trig kan også hentes over web.

## UC3: Registrering av et konsept med tilhørende relasjoner

Funksjonalitet relatert til å opprette nye konsepter.

## Validere konsistens av graf

I vinduet for detaljer kan man velge en fane *Quality Report*. Man må trykke *Generate* etter endringer er gjort for å vedlikeholde rapporten. Denne viser hva som er inkonsistent med vokabularet vårt. Som dere kan se i Figur til venstre er feilen i våre vokabular at vi har konsepter med samme navn. Grunnen til at vi har det er at vi har to vokabularer som har egen definisjon for flere Begreper. F.Eks. Navn, Etternavn mm.

Details	Metadata	Documents	Linked Data	Triples	Visualization	Quality Report
Last Generated: 13.09.2013 - 10:44						
Hierarchical Cycles (0)						
Non-Disjoint Labels (0)						
Inconsistent Preferred Labels (0)						
No Broaders and not Top Concept (0)						
Omitted or Invalid Language Tags (0)						
Same Label for different concepts (21)						
Relation Clashes (0)						

Figur 47: Quality Report

## Enter comments on related concepts

Det skal være link til related concepts hvis de er opprettet/importert riktig. Hvis man trykker på linken velger man konseptet og kan legge på kommentar, eller endre andre attributter.

I wikien har man mulighet til å legge til *new Note* på konsepter.

## UC4: Traversere og se på konsepter fra et annet statlig organ

Dette er litt avhengig av hvordan vi strukturerer prosjektet. Hvis vi lar hvert statlig organ være et eget prosjekt må vi linke inn de disse prosjektene for at det skal være synlige. Til nå har vi kun testet å ha alle etater i samme prosjekt, og skille dem med *Concept Schema*.

Det er også i PP vanskeligere å snevre inn søk til ikke å vise alle statlige organer, snarere enn å vise dem. SPARQL vil gi oss muligheten til denne innsnevringen.

## UC5: Sammenlikne konsepter

Kun med SPARQL, eller egenlaget frontend.

## Andre observasjoner

- Støtter norsk språk i vokaularet
- Man kan importere flere formater fra fil. Bl. a. RDF/XML og Turtle. Når jeg prøver å importere en turtle-fil jeg har laget med Topbraid tygger den en stund og blir ferdig, uten at konseptene er lastet inn i prosjektet. Det kommer heller ingen feilmelding.
- Laget for firefox, chrome og safari
- Hvis man setter multiplisitet til en for et felt kan man lage en instans per språk støttet av prosjektet. Det er likevel en bug her som gjør at man kun kan importere en verdi totalt. Altså ikke en for hvert språk. Dette ble et problem da vi importerte SSB-eksportene.
- Hvis du prøver å endre multiplisitet på et felt i et custom scheme ved å legge til et likt attributt med forskjellig multiplisitet forsvinner attributtet. Det gis ingen feilmelding eller noe, men attributtet er borte for alltid fra eldre data og fra skjemaet.

## Konklusjon og inntrykk

PoolParty tilbyr en oversiktlig måte å strukturere vokabularet sitt på. Ved å lage egne Custom Schema Typer kan man sortere interessante data i grupper.

Versjoneringen i PP kunne nok vært bedre, med støtte for dette i selve serveren.

Linking av data kan være noe uoversiktlig da Navnerom ikke vises i link.

Verktøyet føles generelt litt halvferdig. Det er ofte funksjoner henger seg opp eller svarer med en HTTP feilkode. Noen av disse feilene er kanskje forskyldt av at vi kjører mot testinstans hos PP.

Når man ser på demoapplikasjonene som tilbys fra PoolParty sine hjemmesider ser det robust og vakkert ut. Men det virker som at en front-end for PP legger opp til mye egen tilpasning.

Støtte for offentlige vokabular.



# Collibra Bussiness Glossary (CBG)

Collibra er en web-basert vokabular-klient. CBG sikter på å vedlikeholde og sammenlikne flere vokabular for en institusjon. CBG har en veldig fleksibel modulbasert arkitektur. Alle komponenter defineres og vedlikeholdes enkeltvis, slik at de kan brukes flere steder.

CBG har et simplifisert grensesnitt som skal være lett å lære for de fleste uansett bakgrunn. Grensesnittet er basert på moderne HTML 5 teknologi. Grensesnittet tilbyr bred støtte for egen tilpasning av språk og utseende.

## Dokumentasjon

Produktdokumentasjon er tilgjengelig på Collibra sine sider [www.collibra.com](http://www.collibra.com).

All annen dokumentasjon har vi fått tilsent i en web-folder:

<https://collibra.app.box.com/s/p0pogjp1gixj26ot33yl>

## Innledende testing

Collibra har satt opp en testserver for oss. Vi får en halvannen times innføring i CBG gjennom en skype-session.

## Erfaringer med leverandør og tekniske utfordringer

Det er ingen nedlastbar testversjon på siden. Man må ta kontakt med Collibra for å få mulighet til å teste. I tillegg til dette må man gjennomgå et (eller flere) innføringskurs i hvordan CBG brukes.

Vi registrerte en bruker for å komme til download-siden. Download-siden inneholder nedlastbare videoer og datashet'et til produktene, men som sagt ikke noen testversjon eller annen software.

Vi sender en mail for å etterspørre en trial versjon som respons på mailen vi fikk da vi registrerte oss for å komme til download-siden (Listing 2).

Hello Benny.

My interest in Business Glossary is based on a project that aims to collect Semantic data from different existing sources, but on the same format, and migrate it in to a common client.

We are testing out different clients for vocabulary / ontology to find the one that best suits our needs. Collibra Business Glossary is one of four candidates which we want to try out. However, I am unable to find a trial version on your web. Do you have a trial version available?

*Listing 2: Mail om testversjon*

Etter nesten en uke har jeg ikke hørt noe, så jeg sender en mail til både [sales@collibra](mailto:sales@collibra) og [contact@collibra](mailto:contact@collibra) (Listing 3).

I underkant av to uker senere har vi fortsatt ikke hørt noe. Vi setter derfor Per Myrseth på saken. Det viser seg at vi blitt plassert i andreprioritet hos Collibra. Og har derfor ikke fått respons. Vi skal nå få både demo og tilgang til testserver.

My interest in Business Glossary is based on a project that aims to collect Semantic data from different existing sources, but on the same format, and migrate it in to a common client.

We are testing out different clients for vocabulary / ontology to find the one that best suits our needs.

Collibra Business Glossary is one of four candidates which we want to try out. However, I am unable to find a trial version on your web. Do you have a trial version available?

### *Listing 3: Mail til salg*

Etter en halvannet times live demo er har vi nå fått satt opp en testserverinstans slik at vi er klare til å teste. Det er Maarten Masschelein som guider oss gjennom innføringen. Testlisensen er en Måned, men skal i følge Per være avtalt å vare ut året.

Maarten har også kontaktet oss senere for å høre hva vi synes og om han kan hjelpe oss med eventuelle problemer. Etter dette har vi hatt en kort session for å klare opp noen spørsmål.

Noen operasjoner krever at man refresher siden etter at man har gjort dem før endringene blir synlige. Dette regner vi med er på grunn av utstrakt bruk av AJAX, men det skal være mulig å oppdatere viewene likevel. Regner med at dette blir utbedret i fremtidige versjoner. Merk, dette avhenger også av hvordan du gjør endringen der du har flere alternativer.

En liknende feil ble observert da vi slettet et *Community* som hadde sub-communities. Community'et ble slettet, og en animasjon fjernet det fra viewet, men sub-community'ene som var under ble ikke fjernet. Da vi forsøkte å slette disse fikk vi en feilmelding om at de ikke ble funnet. De var allerede slettet, men dette ble ikke reflektert visuelt før vi gjorde en refresh.

Det ser ikke ut til å være noen direkte språkstøtte for attributter. De må manuelt dupliseres. For eksempel hvis vi har attributtet Navn må vi også ha attributtet Navn\_en eller tilsvarende. Vi har tatt dette opp med Collibra. En mulighet er å bruke localization, altså flere språk for vokabularet. Men dette vil ikke vise alle språkalternativene på weben, kun for det språket man har valgt.

ID'er er UUID som vedlikeholdes av Collibra, ikke URI'er vi har kontroll på.

Hvis man får en feilmelding / gjør noe feil gir Collibra informasjon om hva du må gjøre for å utbedre feilen. Det vil si den kommer med forslag.

## Informasjon om test-instans

Collibra Data Governance Center

URL: <http://london.collibra.com/DNVKEMA>

Username: [REDACTED]

Password: [REDACTED]

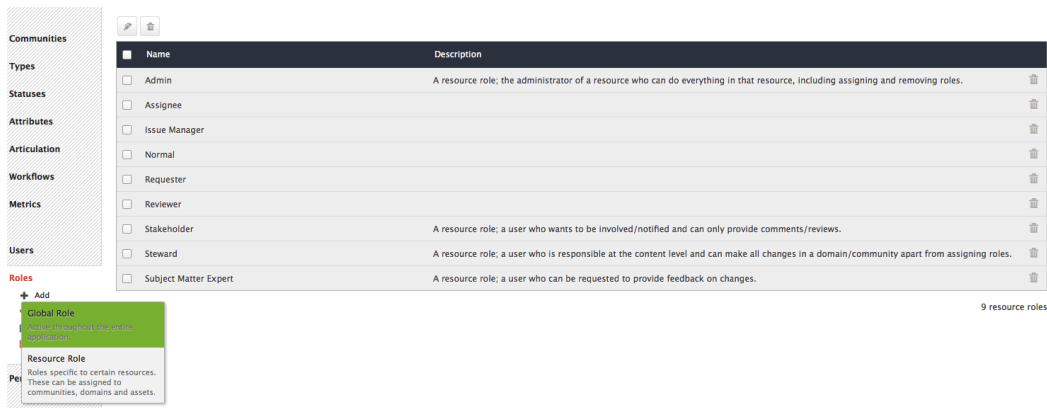
Documentation: <https://collibra.box.com/s/p0pogjp1gixj26ot33yl>

## Roller og brukerstøtte

Som alt annet i Collibra er også roller noe man har tilpasningsstøtte for. Vi kan lage nye roller, gi dem våre navn og bestemme hva de skal gi av tilgang eller muligheter.

Collibra skiller mellom globale roller og roller som har en funksjon for en gitt ressurs. Typisk *Community* eller *Domain*, men også på *Asset* nivå.

I Settings kan man opprette brukere.



Figur 48: Roller

Brukere som opprettes vil får tildelt en mail for å bekrefte konto, men dette kan overstyres. Brukerne kan tildeles brukergrupper. En brukergruppe kan tildeles roller på samme måte som en bruker, slik at man enkelt kan administrere en gruppe brukere i en gitt kontekst.

Figur 49: Dialog for  
brukeroprettelse

En rolle er i utgangspunktet ikke annet enn et navn og en beskrivelse. Funksjonalitet til rollen knyttes på annet nivå. En rolle må som alt annet i CBG godkjennes. Det vil si den er en *Candidate* frem til den har gått gjennom en godkjenningssprosess. Rollene tildeles *Medlemmer*. Dette kan være enten brukere eller brukergrupper.

Når rolle opprettet blir den tildelt et sett med *Permissions*. Her styrer vi hvilke operasjoner de ulike rollene skal ha tilgang til. Her skiller det også på *Resource* og *Global*-nivå.

Name	DataSteward	Flisespikker	Glossary	ReferenceData
Product Rights	✓	✗	✓	✓
Business Semantics Glossary	✗	✗	✓	✗
Reference Data Manager	✗	✗	✗	✓
Data Stewardship Manager	✓	✗	✗	✗
System administration	✗	✗	✗	✗
Workflow	✗	✗	✗	✗
Manage all workflows	✗	✗	✗	✗
Views	✓	✗	✓	✓
Create Views	✓	✗	✓	✓
Create Shared Views	✓	✗	✓	✓
Edit All Views	✗	✗	✗	✗
Compare	✗	✗	✗	✗
Compare	✗	✗	✗	✗

Figur 50: Permission

management

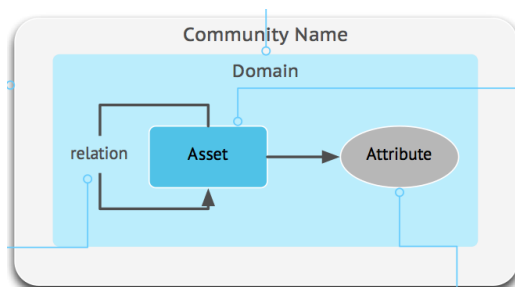
I tillegg til dette kommer administreringen av *Workflows*. Her begynner ting å bli avansert. Spesielt interesserte kan ta seg en kikk her under settings, men foreløpig forholder vi oss i denne testen til workflows som allerede er laget.

Det er tre måter å sende elementer til godkjenning. Den ene er at en med gode nok rettigheter tar en sjefsavgjørelse og oppdaterer status direkte. De to andre er *Approval* og *Simple Approval*. For den

første av disse må flere roller settes opp for å reviewe endringene. Først skal ekspertene reviewe definisjonen. Så får interessentene mulighet til å kommentere og muligens avstemme. Etter dette stemmes det for godkjenning av stewards. Simple approval involverer kun stewards som skal godkjenne endringer. Det blir satt en tidsfrist på godkjennings prosessen. Steward vil få en påminnelse per mail hvis vedkommende bruker for lang tid. Etter dette vil endringene automatisk bli godkjent dersom steward fortsatt ikke gjør noe.

## Støtte for DIFI-Ontologi

Collibra har en litt alternativ struktur når det gjelder konsepter. I bunnen ligger et *Community*, som er et sett av rollene og konteksten som utgjør skopet (Prosjekt). Disse kan organiseres i sub-communities.



Figur 51: Konseptstruktur i CBG

Et Community har flere *Domains*, som igjen inneholder *Assets* (Concept).

En *Asset* er det vi refererer til som et konsept. Vi kan lage en egen Asset-type Begrep.

For å lage ontologien må vi definere attributtene enkeltvis.

Det er tre typer attributter. Disse er Attributt, Relasjon og Kompleks relasjon. Attributt er kun et navn med en verdi. Figur viser dialogen for opprettelse av et attributt. En relasjon i CBG binder to Assets sammen. Disse må forsåvidt være bilinære, men man får mulighet til å definere hva verdien skal være hver vei. For eksempel for relasjonen *Navn omfatter Fornavn*, vil i tillegg kunne gi relasjonen *Fornavn omfattes av Navn*. Se Figur for visuelt eksempel. En kompleks relasjon kan gå mellom flere Assets og ha data (attributter) knyttet til seg. Hvert Asset i relasjonen gis en rolle, for eksempel source/target samt multiplisitet.

Attributtene assignes igjen til grupper som gjelder for de forskjellige Asset-typerne.

Vi lager Custom Attributes for Begrep, slik at vi kan implementere DIFI sin standard. Først må vi lage attributtene enkeltvis. *Settings->Attributes->Add->Attribute Type* (eller *Relation Type*).

Vi får velge et navn, beskrivelse og type for attributtet. Vi biter oss merke i at *Date* eller *URL* ikke er blant alternativene. (Date skal visstnok komme i fremtidige versjoner, men det er ikke bestemt formatet for denne i følge Maarten).

Attribute Type	Description
for an asset.	Text
	Multiple Selection
issue.	Text
	Text
itions that corresp...	Text

Figur 52: Dialog for opprettelse av

attributt

Vi legger til relasjoner for Begrep. Vi tar utgangspunkt i Begrepsrelasjonstypene fra SERES-eksportene. Disse er *omfatter*, *er*, *harRolle* og *synonym*.

Det er foreløpig ingen mulighet til å legge regular expressions på et attributt, men dette skal støttes i

fremtidige versjoner.

Alle feltene i Figur er påkrevd. På samme måte som attributtene ovenfor legger vi relasjonene til i *Begrep Attributes*.



Figur

53: Relasjonstype

Etter av vi har definert alle attributter og relasjoner vi skal bruke må vi knytte disse til Asset-typen vår, Begrep. Trykk *Assign* under *Attributes* i Settings for å lage en gruppe av attributter som knyttes til en Asset-Type. Vi lager en gruppe *Begrep Attributes*.

Figur viser dialogen for å mappe attributter til Asset-typer.

Figur 54: Dialog for attribute assignment

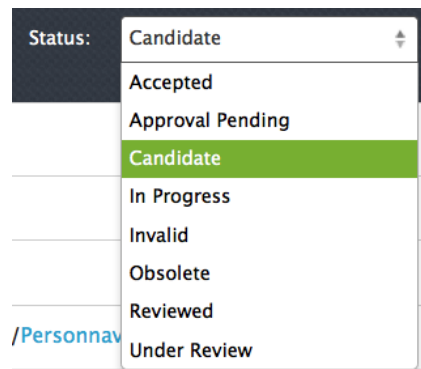
Gruppen *Begrep Attributes* består nå av *Navn*, *Guid*, *Definisjon*, *Gyldig Til*, *Gyldig Fra*, samt relasjonene *er*, *omfatter*, *harRolle* og *synonym*. *Gyldig Til* og *Gyldig Fra* er lagt til som Tekst, da Date ikke støttes.

## Versjonering og mulighet for å spore endringer

Versjonering i CBG foregår ved bruk av Snapshots. Ut over dette er har all data kun en instans som arbeides på. Asset og andre objekter i denne instansen har et statusfelt som indikerer hvor de er i prosessen.

Alle Objekter, både metadata og Asset, som opprettes i CBG har status *Candidate*. Denne statusen kan endres av en Admin, eller som et mål for en omfattende godkjenningsprosess som involverer flere roller.

Denne workflowen godkjenner begreper gruppevis eller enkeltvis. For versjonering bruker CBG et system for Snapshots. Et Snapshot er en instans av eller et subset av et vokabular slik det så ut på et gitt tidspunkt.



Figur 55: Statusfelt

Et Snapshot er i utgangspunktet en instans av et view i CBG. Men til forskjell fra andre views i CBG vil ikke et Snapshot endre seg. Det vil være et statisk bilde av hvordan dette viewet så ut på et gitt tidspunkt i historien. Et view i CBG vil vise et sett data basert på et filter og hvilke attributter som skal være med.

Snapshots kan sammenliknes med hverandre og/eller nyeste revisjon.

CBG holder også til enhver tid oversikt over historikken til termene i vokabularet. Figur viser et utdrag av historikken til begrepet *Eier*. Man kan se hva som er gjort, når det er utført og hvem det er utført av.

☒ Added
 ☒ Changed
 ☒ Removed
 ☐ Other

Name	Definition	Steward	Status	Type
Bomasse			<del>Candidate</del> <u>Accepted</u>	Begrep
DagligLeder	Daglig Leder		<del>Candidate</del> <u>Accepted</u>	Begrep
Dato			<del>Candidate</del> <u>Accepted</u>	Begrep
Eier	<del>Eier-eier-noe</del> <u>Person som er inn...</u>		<del>Candidate</del> <u>Accepted</u>	Begrep
Enhet			<del>Candidate</del> <u>Accepted</u>	Begrep
Etternavn	Siste <del>Ø</del> delen <u>A</u> av <u>N</u> navn		<del>Candidate</del> <u>Accepted</u>	Begrep

Figur 56: Sammenlikning av to Snapshots

En alternativ måte å kjøre versjonering på er å gjøre som for PoolParty at man importerer eksporter fra en utviklings-server til en prod-server.

## Import og eksport av data

Definition  
Person som er innehaver av eiendom, rettighet eller åndsverk.  
**Eier-eier-noe.**

Admin Istrator  
9/30/13 8:52 AM

Simple Approval Process ended (6 days later)

Accepted (Previous status: Candidate)  
9/26/13 1:21 PM

Definition  
Eier eier noe.  
**Eier**

Admin Istrator  
9/19/13 2:22 PM

+ Definition  
Eier

Admin Istrator  
9/19/13 2:20 PM

+ Definisjon  
Eier

Admin Istrator  
9/19/13 2:18 PM

Simple Approval Process started

Admin Istrator  
9/19/13 1:21 PM

+ Gyldig Til  
24.12.2013

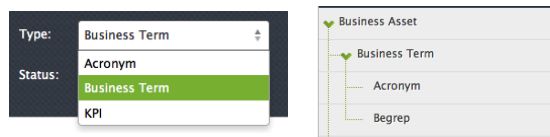
Admin Istrator  
9/19/13 11:25 AM

Figur 57: Activity for begrepet Eier

Som med de andre verktøyene forsøker vi å først lage et lite eksempelprosjekt i CBG, eksportere dette og studere eksporten slik at vi kan replikere syntaks for import fra SERES. Vi starter med å lage et *Community* kalt Semicolon Test og legger til et *Domain* Skatteetaten.

Til tross for at vi har fått litt innføring i Collibra finner vi det litt vanskelig å komme i gang med. Det første vi prøver å gjøre er å legge til et Begrep *Fornavn*. Fornavn blir lagt til som *Business Term*. Jeg har forsøkt å legge til en ny type Begrep, men denne typen kommer ikke opp som forslag når jeg forsøker å instansiere Fornavn.

Etterhvert finner jeg ut hvordan jeg kan konfigurere opp *Glossary* til å akseptere Asset type Begrep. Under *Settings->Types->Assignment* kan man trykke på *Glossary Concept Types* og legge den til.



Neste steg blir å lage Custom Attributes for Begrep, slik at vi kan implementere DIFI sin standard. Først må vi lage attributtene enkeltvis. *Settings->Attributes->Add->Attribute Type* (eller *Relation Type*).

Det er foreløpig ingen mulighet til å legge regular expressions på et attributt, men dette skal støttes i fremtidige versjoner.

Vi instansierer begrepene *Fornavn*, *Etternavn* og *Navn*. Vi legger til referansen *Navn* omfatter *Fornavn* og *Etternavn*. Det er lett å gjøre flere relasjoner i en transaksjon, se Figur.

Figur 58: Legge

til flere begreper av gangen

Instansen for *Navn* blir seende ut som følger:

Legg merke til at vi har linker til *Fornavn* og *Etternavn* i definisjonen. Dette er noe som kommer

opp automatisk dersom termen er lik navnet på et Begrep. I figuren har vi brukt stor bokstav på henholdsvis Fornavn og Etternavn i definisjonen, men dette er ikke nødvendig for at CBG skal kjenne det igjen og linke til det.

**Navn**

Navn

**Guid**

<http://seres.no/guid/Navn/2342333>

**Definisjon**

Navn er en identifikator for en person som består av Fornavn, Mellomnavn og Etternavn

**Gyldig Fra**

17.09.2013

**Omfatter**

Name	Definition	Descriptive Example	Domain
Etternavn			Skatteetaten
Fornavn			Skatteetaten

2 relations *Figur 59: Visning av begrepet*

### Navn

I eksempelet nedenfor vises definisjonen til Personnavn for Brønnysundregistrene. Hvis man trykker på en av linkene i definisjonen kommer det opp forslag til termer som har dette navnet. Trykker man på linken velges begrepet som har dette navnet (**Sweet (-: )**)

**Definisjon**

Personnavn består av fornavn og ett enkelt eller dobbelt etternavn og kan i tillegg ha mellomnavn.

**Gyldig Fra**

18.09.2013

**Omfatter**

- Fornavn Semicolon Test > Brønnysundregistrene
- Fornavn Semicolon Test > Skatteetaten
- Fornavn Semicolon Test > Konkursregisteret

*Figur 60: Link med synonymer*

Vi får følgende oversikt over vårt domene *Skatteetaten*

Hvilke kolonner som skal være med i viewet kan administreres enkelt ved å velge i Settings (Tannhjulet).

Default Public. The default view to show all the assets in this vocabulary. [Reset] [Save] [Settings]

[Hierarchy] [Discover] [Export] [Filter] [Settings]

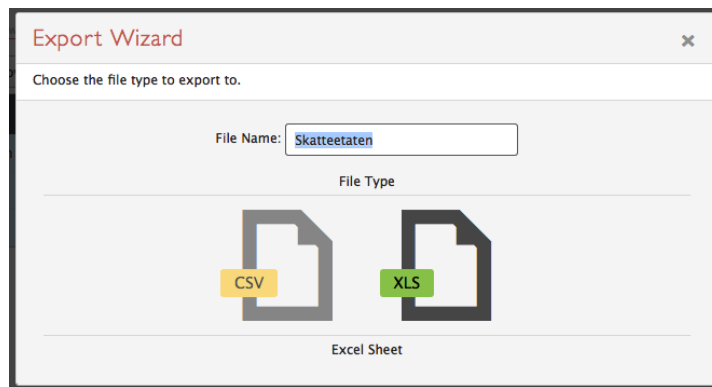
Name	Navn	Guid	Gyldig Fra	omfatter	omfattet av	Steward	Status	Type
<input type="checkbox"/> Etternavn	Etternavn	<a href="http://seres.no/guid/Etternavn/23423432">http://seres.no/guid/Etternavn/23423432</a>	17.09.2013		Navn		Candidate	Begrep
<input type="checkbox"/> Fornavn	Fornavn	<a href="http://seres.no/guid/Fornavn/23423">http://seres.no/guid/Fornavn/23423</a>	17.09.2013		Navn		Candidate	Begrep
<input type="checkbox"/> Navn	Navn	<a href="http://seres.no/guid/Navn/2342333">http://seres.no/guid/Navn/2342333</a>	17.09.2013	Etternavn Fornavn			Candidate	Begrep

*Figur 61:*

### Oversikt testdomene

Vi har fått inntrykk av at import/eksport-mulighetene er litt mer begrenset enn i de andre kandidatene. Med det menes at Collibra støtter litt færre formater, og kun Stylesheets. Det er mulig vi tar feil. Når vi trykker *export to file* kommer følgende dialog opp.

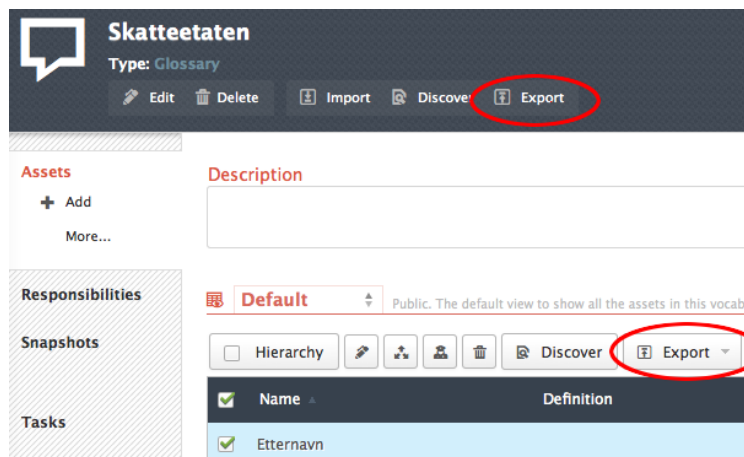




Figur 62: Initiell eksportdialog

I neste dialog får velger vi hvilke felter som skal være med, og hva de skal hete i csv/xls-filen.

Nå viser det seg at det er mulig å eksportere på flere måter. Hvis vi trykker *Export* i menylinjen for domenet kommer dialogen nevnt ovenfor. Trykker vi i listen over Begreper får man opp alternativene Excel eller CSV. Alle synlige attributter blir eksportert. Den siste er mest hensiktsmessig da man slipper å manuelt mappe alt man vil ha med.



Figur 63: Eksport-alternativer

En gjennomgang av eksporten viser et potensielt problem. ID-systemet i Collibra (som ikke er gjengitt i web-viewene) ser ut til å være basert på et 8-4-4-4-12 UUID-system, for eksempel 9a1be84f-61a6-4ab0-abde-4bfe0c7dc7c4. Hvis vi må definere disse kan vi få kollisjoner. Det er ikke gitt at id'ene MÅ være på det formatet. Så når vi importerer fra SERES kan vi muligens sette disse til å være en url eller tilsvarende. Hvis ikke er det mulig vi kan få Collibra til å generere dem ved å la dem være tomme i importen. (Fungerer med url som id).

En annen ting som er verdt å merke seg er at Collibra ulikt de andre kandidatene ikke gjør bruk av en URI som unik id for konseptene.

Nå når vi har ontologien på plass lager vi et script for å importere SERES-exportene.

Vi får ingen kontroll over unik id til konseptene. Hverken ved import eller opprettelse i CBG.

Vi lager en modul i transformeren som konverterer SERES til Excel-filer. Denne inneholder følgende kolonner:

Name	Navn	Guid	Definisjon	Gyldig Fra	Gyldig Til	Synonym	Synonym omfatter	omfattet av	referert av	har Rolle	Referert ro	Status	Type
Utgåttårs	Utgåttårs	http://seres.no/guid/B	18.09.2013					Årsak				Candidate	Begrep

Figur 64: Format for import-stylesheet

For å importere denne i Collibra lager vi et nytt *Domain* manuelt. Dette kan gjøres via importen, men det er enklere og mer oversiktlig å gjøre det manuelt. Velg type *Glossary* og velg *Save*.

Velg det nye domenet.

Trykk *import* og naviger til filen du vil importere og trykk *next*.



Figur 65: Velge fil

Etter dette må du mappe verdiene i excel-dokumentet til attributter i Colibra. Dette må desverre gjøres manuelt.

Trykk *Next*.

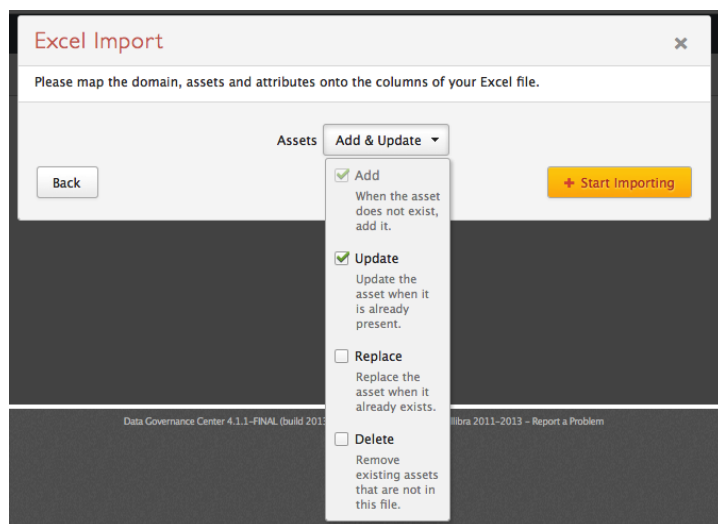


Figur 66: Mappe kolonne til

attributt/relasjon

I neste steg kommer noe som vi tror er litt unikt for Colibra. En dialog ber deg velge hvilke operasjoner importen skal gjøre.

Man har mulighet til å velge Replace og Delete. Dette åpner for at vi lettere kan synkronisere med endringer som har skjedd i eksterne editorer, også med hensyn på slettede begreper.



Figur 67: Import-operasjoner

Når vi importerer stylesheets får vi også mappet relasjonene. Dette er unikt for CBG i denne testen da de andre verktøyene har ikke støttet å binde relasjoner i stylesheets. Det mappes ved å bruke navnet, ikke unik ID. Mappingen blir riktig også når begreper med samme navn eksisterer i andre vokabular. Vi har ikke testet om det er mulig å importere relasjon til begreper i andre vokabular.

Dette betyr i praksis at CBG kan importere data fra alle editorer som støtter stylesheets.

Det ser ikke ut som om det er noen støtte for å importere ontologien til CBG. Denne må være definert på forhånd.

## Eksportmuligheter

Vi har ikke funnet støtte for eksport til noe annet en stylesheets i form av Excel eller CSV. Hva du vil eksportere i disse formatene har du derimot god kontroll på.

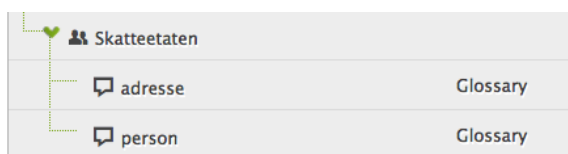
Når du har valgt et domene/vokabular kan du velge å eksportere hele vokabularet ved å velge hvilke kolonner/attributter du vil ha med. Men en kanskje enklere og mer fleksibel måte å gjøre det på er å lage et view med de begrepene du ønsker med de attributtene/relasjonene du vil ha med. Disse viewene kan også gå på tvers av domener. Når du har laget viewet kan du eksportere det direkte til stylesheet uten å måtte mappe kolonner til data.

## Navnerom

CBG har bred støtte for hierarkisk inndeling av konsepter. Ytterst har vi det som kalles *Community*. Et Community kan igjen ha sub-communities. Vi har valgt å ha Prosjektet som Community. Men vi kan også ha hvert statlig organ som eget community. Under dette har vi det som heter Domain som er instansen av vokabularet.

Vi kan strukturere domener i flere nivåer, og bruke dette til å filtrere når vi lager views/snapshots. I stedet for å ha etatene som *Domain* er det i den sammenheng mer hensiktsmessig å la etatene være *Community* eller *sub-communities*. Grunnen til dette er at *Domain* ikke kan ordnes hierarkisk. Hvis vi har et sub-community Skatteetaten kan dette inneholde domene (type glossary) person og adresse. Vi kan ikke dele domenene person og adresse inn ytterligere. Dette kompliserer også importen. For eksempel må vi for SSB som er delt inn i kategorier importere hver kategori enkeltvis i eget stylesheet. Importeringsprosessen er noe tungvinn. Man har et eget felt «Domain» i stylesheetet som sier hvilket domene begrepet skal til, men importen skjer i et eksisterende domene. De andre domenene opprettes fortløpende.

For referanser vil CBG vise to nivåer. Uten å ha funnet dette i dokumentasjonen, eller testet flere alternativer ser det ut til at disse to nivåene er nærmeste Community og domene.



Figur 68: Inndeling skatteetaten

## Søke- og sammenlikningsmuligheter

Enkelt eller dobbelt etternavn og kan i tillegg ha

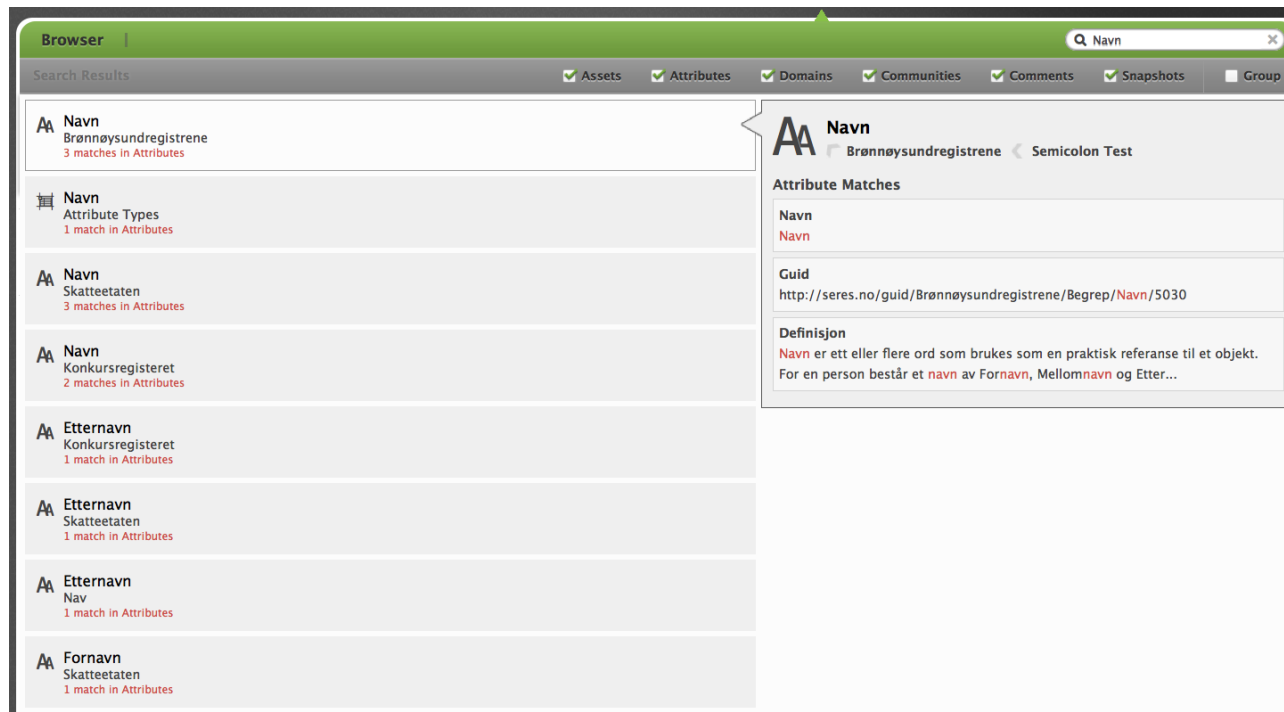
<b>Etternavn</b> Skatteetaten > person
<b>Etternavn</b> Semicolon Test > Nav
<b>Etternavn</b> Semicolon Test > Konkursregisteret
<b>Etternavn</b> Semicolon Test > Brønnøysundregistrene

Figur 69: Søkeforslag

synonymer

Søkefunksjonaliteten i Collibra lander greit på kryssningspunktet mellom kompleksitet og enkelhet. Søket returnerer termer som inneholder (contains) det man søker på. Man kan snevre inn søket etter type. Se Figur nedenfor.

Figuren ovenfor viser resultatet for søket Navn. Resultatene kommer til høyre. Hvis man klikker på et entry her kommer detaljene til venstre. Dette er javascript-basert og går raskt, slik at viewet i seg selv kan brukes for manuell sammenlikning.



Figur 70: Søk i Collibra

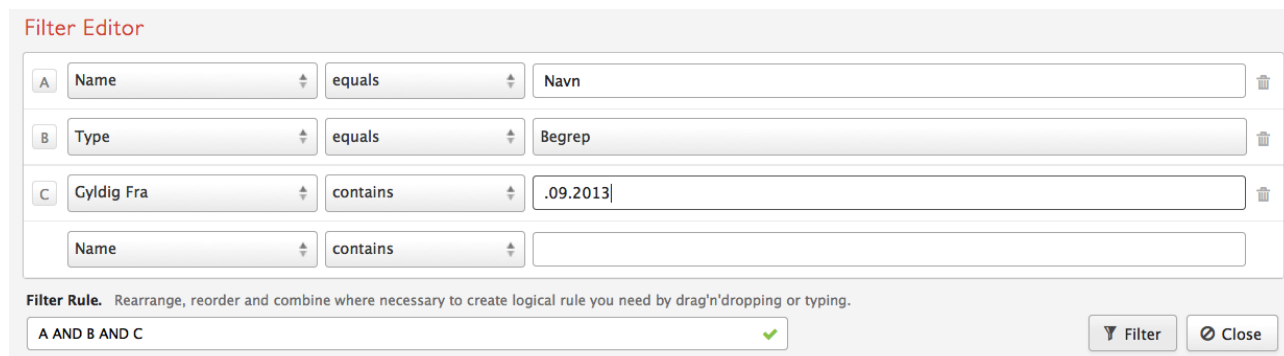
Navigering har vi til en viss grad snakket om. I tillegg til at Collibra ved hierarkisk navigering gjør det oversiktlig å finne begreper er Collibra flink til å legge til linker der det er relevant. I attributtverdier kommer linker opp automatisk dersom de gjenkjennes som Assets.

Når det gjelder sammenlikning av Assets stiller Collibra i en egen klasse. Vi kan customize alle views som viser detaljer for Assets på tvers av domene eller hva som helst.

Vi velger *Community Semicolon Test*. I menyen til Venstre kan vi velge *Business Terms*. Alle Business Terms i Semicolon Test kommer nå opp. (Ideelt sett skulle vi hatt et view som viste alle Begreper, men dette har vi ikke fått til å opprette. Dette krever editering av templatene/html filen. Dette er mulig, men ikke med testversjonen vi har tilgang til, i følge Maarten.)

Vi kan snevre inn begrepene vi ser etter ved å bruke filter-funksjonaliteten. Hvis vi vil se på alle begreper kalt Navn som er gyldige fra september 2013 kan vi kjøre følgende filter:

Et standard view for Business Terms kommer opp med resultatene. Vi kan videre trykke på tannhjulet lengst til Høyre og velge hvilke kolonner vi vil se på.



Figur 71: Filter Editor

Man kan også snevre inn hvilke domener man vil kikke på.



<input type="checkbox"/>	Name	Guid	Definisjon	Status	Domain
<input type="checkbox"/>	Navn	http://seres.no/guid/Brønnøys...	Navn er ett eller flere ord som ...	Candidate	Brønnøysundregistrene
<input type="checkbox"/>	Navn	http://seres.no/guid/Begrep/N...		Candidate	Konkursregisteret
<input type="checkbox"/>	Navn	http://seres.no/guid/Navn/23...	Navn er en identifikator for en ...	Candidate	Skatteetaten

Figur 72: Tilpasset view

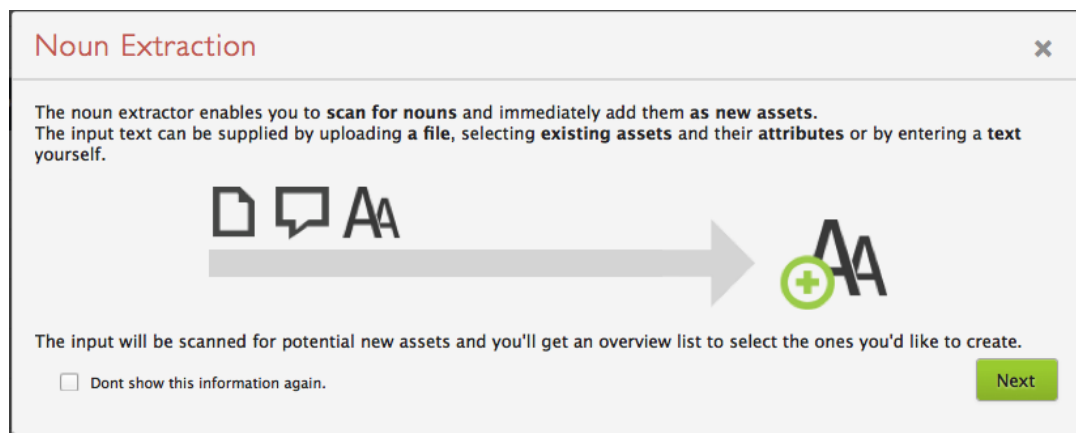
Hvis det kommer opp mange resultater kan man søke på kolonnene for å finne det man leter etter. Viewene kan også lagres for senere bruk.



Figur 73: Søk på kolonner

## Andre features

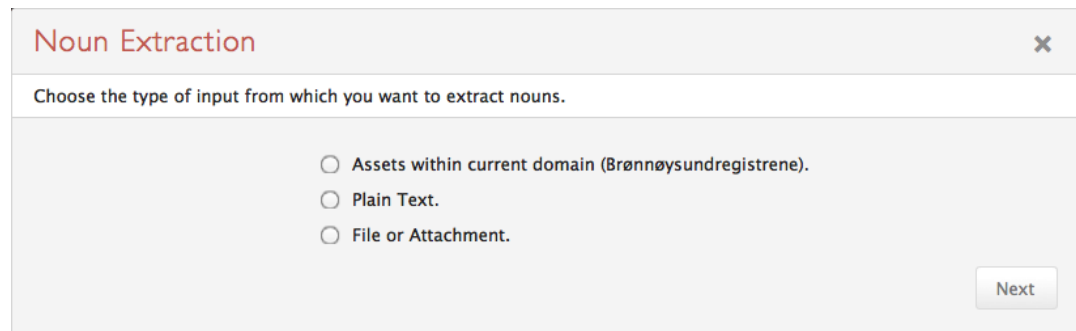
CBG tilbyr en litt unik feature kalt *Discover*. Dette er en funksjon som skal hjelpe deg å bygge vokabular. Hvis vi velger et vokabular ligger knappen *Discover* ved siden av *Approval*. Hvis vi klikker på den kommer det opp en dialog.



Figur 74:

### Discover dialog

Plain text lar oss velge fra en tekst vi skriver eller limer inn. Det samme gjelder File or Attachment. Det øverste alternativet tilbyr å søke gjennom alle attributter i et domene. Vi velger det.



Figur 75:

### Søkeområder

Hva Exclude står for i Figur er uvisst. Resultatet ser ut til å bli det samme uansett hva som er med her. Videre vil CBG gi en oversikt over hvilke ord som eksisterer i vokabularet vårt, og hvor ofte det er brukt. Vi krysser av for de termene vi vil opprette Asset av og trykker *Next*. Dette oppretter de nye termene.

Noun Extraction

×

Select and configure the found nouns before converting them to assets.

<input type="checkbox"/>	# ▾	Noun	Target Domain
<input type="checkbox"/>	9x	Eller	Brønnøysundregistrene ...
<input type="checkbox"/>	8x	Br	Brønnøysundregistrene ...
<input type="checkbox"/>	5x	Av	Brønnøysundregistrene ...
<input type="checkbox"/>	5x	På	Brønnøysundregistrene ...
<input type="checkbox"/>	4x	Enheter	Brønnøysundregistrene ...
<input type="checkbox"/>	4x	Hvor	Brønnøysundregistrene ...
<input type="checkbox"/>	4x	Om	Brønnøysundregistrene ...
<input type="checkbox"/>	4x	Skal	Brønnøysundregistrene ...
<input type="checkbox"/>	4x	Være	Brønnøysundregistrene ...
<input type="checkbox"/>	3x	Det	Brønnøysundregistrene ...
<input type="checkbox"/>	3x	Knyttet	Brønnøysundregistrene ...
<input type="checkbox"/>	3x	Selskap	Brønnøysundregistrene ...
<input type="checkbox"/>	3x	Statuser	Brønnøysundregistrene ...
<input type="checkbox"/>	228x	Guid	Brønnøysundregistrene ...
<input type="checkbox"/>	22x	Er	Brønnøysundregistrene ...
<input type="checkbox"/>	2x	Annet	Brønnøysundregistrene ...
<input type="checkbox"/>	2x	Ansvar	Brønnøysundregistrene ...
<input type="checkbox"/>	2x	Deler	Brønnøysundregistrene ...
<input type="checkbox"/>	2x	Eksempel	Brønnøysundregistrene ...
<input type="checkbox"/>	2x	Etc	Brønnøysundregistrene ...

1

2

3

4

5

6

7

8

>

>>

Back

Next

Figur 76:

*Discover-resultater*

## Evaulering av UseCase

Vi gjør en testing av functionality required fra Use Casene i survey-dokumentet.

### UC1: Vedlikehold av terminologi / vokabular for et statlig organ

Når det gjelder rollene som kreves av use caset stiller Collibra veldig sterkt. Collibra har den kraftigste brukerkonfigurasjonen i denne vurderingen.

Administratorer lager brukere. En Bruker kan legges til i en *User Group* som samler brukere.

I tillegg kan administratorer lage *Roles* og administrere *Persmissions* med skille på *Global* eller *Resource* -nivå. Rollene assignes med *Responsibilities* som kan være følgende: *Admin*, *Asignee*, *Issue Manager*, *Normal*, *Requester*, *Reviewer*, *Stakeholder*, *Steward* eller *Subject Matter Expert*.

## Opprette et nytt konsept

Collibra arbeider i hierarki *Community->Domain->Asset* som tidligere nevnt. Asset er det vi refererer til som Concept. For å lage et nytt konsept velger man Domenet man skal lage Asset'en i, eventuelt opprette et.

I vår implementasjon kan vi velge *Semicolon Test->Konkursregisteret*. På venstre siden kan man trykke *Add* og velge *Begrep* i dropdown listen. *Begrep* er en type vi har laget.

En dialog popper opp for navn på begrepet. Vi kan i denne dialogen velge å legge til et eller flere begreper. Hvis man etter å ha skrevet trykker Tab eller Enter vil det lages en ramme rundt navnet. Se figur nedenfor.

Skriver man bare et navn har man i tillegg valget *Save and View* som bringer oss direkte inn til detaljsiden for Asset'en. På detaljsiden har man mulighet til å legge til de attributtene vi har definert for typen. Denne siden vil være tom før attributter legges til.



Figur 77: Opprette flere Asset av gangen

## Lese konsepter

Collibra har en oversiktlig og ryddig måte å lese konsepter. Hvis vi velger et *Community* får vi opp alle domenene.

Vi kan navigere oss videre for å se på innholdet av de forskjellige vokabularene, samt detaljene for de ulike konseptene.

Name	Domain Type
♥ 👤 Semicolon Test	
🗨 Brønnøysundregistrene	Glossary
🗨 Konkursregisteret	Glossary
🗨 Nav	Glossary
🗨 Skatteetaten	Glossary

Figur 78: Domener

Det er et godt knippe forskjellige muligheter når det kommer til å lese konsepter. Vi kan lage spesifikke views, Snapshots osv.

## Oppdatere konsepter

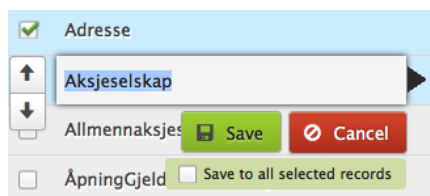
Det er flere måter å oppdatere konsepter på. Man kan oppdatere dem på samme måte som opprettelsen av dem. I views med flere konsepter kan man dobbeltklikke på attributtet man vil endre (eller velge fra en kontekstmeny til høyre). Man kan endre alle egenskaper til en Asset, også typen. For å endre en Asset må brukeren ha en rolle som tillater dette.

Oppdateringene som gjøres på en Asset, eller Domene for den saks skyld, vedlikeholdes i en historikk. Denne historikken viser hva som er endret, verdier før og etter, samt hvem som gjorde endringen og når den ble gjort.

Noen operasjoner støttes også å gjøres for flere Asset av gangen. Hvis man for eksempel ønsker å flytte assets kan man krysse av de relevante Asset'ene og flytte kun dem.

Man har også noe som heter *Rapid Edit*. Her velger man hvilket attributt som skal endres. En dialog gjør det mulig å traversere seg nedover, eller å oppdatere for alle valgte Assets. Man kan sekvensielt endre nedover i listen av valgte elementer, eller krysse av *Save to all selected records*.

Endringer kommer med i historikken til Asset'en.



Figur 79: Rapid edit

## Søke, navigere og sammenlikne konsepter

Vi har gått gjennom søkefunksjonen til CBG i detalj ovenfor. CBG har utvilsomt den mest avanserte funksjonaliteten for sammenlikning blant verktøyene testet i dette dokumentet.

Når det gjelder navigering vil tittelbaren til enhver tid vise hvilket nivå vi er på. Vi kan trykke på denne for å komme til nivå(er) over. Det kan bli litt komplekst å dele inn i nivåer i Collibra.

## Lese konsepter og relaterte attributter

Ja, det kan man. Hvis vi for eksempel velger begrepet navn i Brønnøysundregistrene-domenet får vi følgende oversikt:

Vi kan også bruke views for å vise attributtene for en gruppe av Assets.

Navn

Navn

Guid

<http://seres.no/guid/Brønnøysundregistrene/Begrep/Navn/5030>

Definisjon

Navn er ett eller flere ord som brukes som en praktisk referanse til et objekt. For en person består et navn av Fornavn, Mellomnavn og Etternavn.

Gyldig Fra

18.09.2013

Referert Av

Name	Definition	Descriptive Example	Domain
Personnavn			Brønnøysundregistrene

1 relations

Figur 80: Egenskaper for Asset Navn

## Eksportere til andre verktøy

Man har mulighet til å eksportere domenene via stylesheets, enten CSV eller Excel. Eksportene



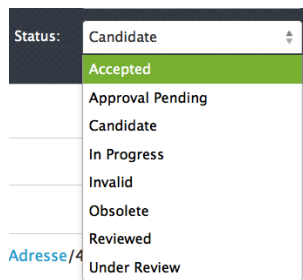
inneholder bare instanser, ikke ontologien til vokabularet. Når man eksporterer kan man enten manuelt gi navn til kolonnene man vil eksportere (og her er det ingen select all) som er en litt tungvinn prosess.

Eller man kan lage et view som vist ovenfor, og eksportere dette. Eksporten vil inneholde de instansene og kolonnene som er i viewet.

### Endre status på konsepter

Jepps! Alt i Collibra har et statusfelt. Typer, Communities, Assets, etc. Alt som opprettes får status *Candidate*. Hvis man har rettigheter kan man ved å trykke *Edit* på Asset'en endre status.

Men statusen er tenkt vedlikeholdt ved en godkjenningsprosess som involverer flere parter.



Figur 81: Statusfelt

### Opprettholde konsistens for en gruppe begreper

Godkjenningsprosessen som legges opp til i Collibra involverer flere roller. Når begreper eller Domener sendes til godkjenning skal de revideres av oppsatte roller, som eventuelt godkjenner Assetene.

Det er to typer godkjenning, vanlig og enkel. Enkel godkjenning trenger bare Steward til å godkjenne. For annen godkjenning skal Stakeholder også være satt opp, og vedkommende må stemme for godkjenning.

Elementene kan godkjennes enkeltvis eller som en gruppe.



Figur 82: Godkjenning

### UC2: Publisering av terminologi / vokabular

I følge Maarten som ga meg innføring i Collibra skal det være mulig å publisere vokabular for offentlig lesing. Altså at man ikke trenger å logge inn som en bruker for å lese vokabularet. Dette er derimot ikke mulig i test-versjonen. Grunnen til dette er, i følge Maarten, at Collibra vil sikre at dataene våre ikke kommer på avveie. Vi får derfor bare stole på at det er mulig.

Når det gjelder publisering av konsepter oppierer Collibra med såkalte Snapshots. Snapshots er en *Versjon* av et vokabular. Det vil si en instans av et vokabular slik den så ut på et gitt tidspunkt. Når alle stakeholdere har godkjent alle begreper de står ansvarlige for kan vi lage en snapshot av et Domene.

En snapshot-konfigurasjon kan sees på som det samme som et view. Etterhvert som vi lager flere

snapshots med samme konfigurasjon kan vi sammenlikne Snapshot'ene, eller sammenlikne snapshots med current data.

Nedenfor vises utdrag av sammenlikning mellom to snapshots for Konkursregisteret.

Vi kan selv velge hva som skal publiseres. Snapshots, Views, etc.

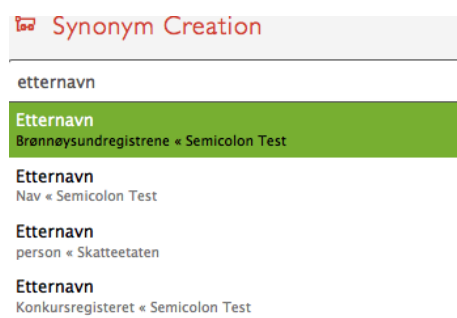
DagligLeder	asfDaglig Leder
Dato	
Eier	Eier_eier noe.
Enhet	
Etternavn	Siste Delen Av Navn

Figur 83:

Sammenlikning av Snapshots

### UC3: Registrere et konsept med tilhørende relasjoner

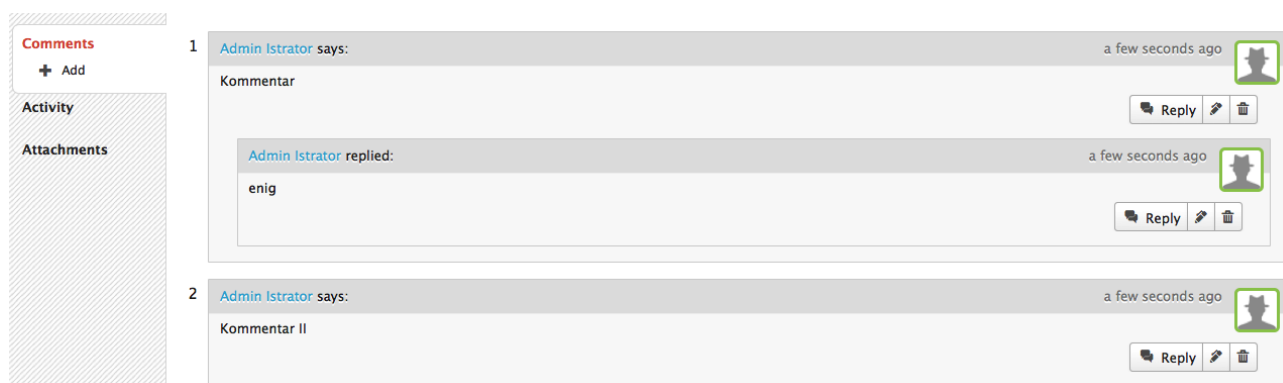
Relasjonene opprettes på samme måte som i andre verktøy. Man skriver inn Begrepet man vil referere til og velger et av alternativene som kommer opp. Vi kan relatere til Begreper på kryss av Communities/Domener.



Figur 84: Forslag

### Legge til kommentarer på relaterte konsepter

Det er en egen Comments funksjonalitet på Asset-nivå.



## UC4: Traversere og se på konsepter fra et annet statlig organ

### Avgrense en domenemodell for en etat

Det er flere måter å gjøre dette på. Ved navigering i hierarkiet vil vi typisk velge en etat av gangen. Vi kan lage et view som viser utvalgte data fra flere etater, eller vi kan bruke filter for å snevre inn til en etat av gangen i et view.

### Konklusjon og inntrykk

Collibra tar litt tid å komme inn i. Når man først kommer inn i oppsettet er strukturen enkel og fleksibel. Det at alt defineres i moduler gjør at man kan gjennbruke ting og dermed spare tid.

Utseendemessig ser CBG veldig moderne ut, med utstrakt bruk av AJAX og HTML 5.

Noen steder blir det tydelig at CBG er et verktøy i konstant utvikling. Under presentasjon og spørsmålsrunde med Collibra var ofte svaret «Comming in future versions».

Når det gjelder sammenlikning av Begreper, og vedlikehold av historikk er CBG utvilsomt det beste verktøyet i dette dokumentet.

Versjoneringen i CBG er litt uoversiktlig. Skulle gjerne sett at man hadde muligheter til å lage egne branches eller tilsvarende, og ikke bare et Snapshot.

Eksport/import-funksjonaliteten til CBG er noe begrenset. Vi kan ikke eksportere eller importere en ontologi. I tillegg må all import gjøres via stylesheets, og mappes manuelt mellom kolonne og attributt.

CBG er en god og relativt robust editor for vokabular som vokser på oss under bruk. Features virker stort sett godt planlagt og utprøvd.

## Annet

I Utdraget fra SERES er det flere definisjoner for Begrepsrelasjonstyper (Se nedenfor).

Vi har attributtet definisjon, samt flere elementer av typen dokumentasjon. Det ser ut som om dokumentasjon lar oss spesifisere i flere språk. Foreløpig tas ikke den med i import-data'ene.

```
<seres.begrep.Begrepsrelasjonstype xmi.id='a1392'
xmi.uuid='http://www.brreg.no/Begrepsrelasjonstype/Semantikk/omfatter'
xmi.label='' navn='omfatter'
guid='http://www.brreg.no/Begrepsrelasjonstype/Semantikk/omfatter'
definisjon='Aggregering mellom to begreper' begrepsrelasjon='a1280 a1277
a1238 a1331 a1307 a1308 a1332' begrepsnivå='a1217'>
  <seres.domeneKjerne.SERESelement.dokumentasjon>
    <seres.domeneKjerne.Dokumentasjon xmi.id='a1394'
      xmi.uuid='0beaa8c245c145f5:2048514a:0000013de3ce2f21:8219'
      xmi.label='' type='beskrivelse' språk='nb' tekst='En omfatter-
      relasjon angir at et begrep (fra) best&#229;r av et annet begrep
      (til). Omvendt kan vi si at begrepet (til) er en del av et det andre
      begrepet (fra).' sereselement='a1392'>
    </seres.domeneKjerne.Dokumentasjon>
  </seres.domeneKjerne.Dokumentasjon xmi.id='a1395'
```

```
xmi.uuid='0beaa8c245c145f5:2048514a:0000013de3ce2f21:821a'  
xmi.label='' type='beskrivelse' språk='nb' tekst='Omfatter baseres  
p&#229; UMLs aggregeringsrelasjon. Defineres mellom to begreper'  
sereselement='a1392'>  
  </seres.domeneKjerne.Dokumentasjon>  
  </seres.domeneKjerne.SERESelement.dokumentasjon>  
</seres.begrep.Begrepsrelasjonstype>
```

Skrive litt om konverteringen mellom SERES og verktøyene?