

NHH



# CLARINO WP 7

# TERMINOLOGY INTEGRATION

CLARINO Consortium meeting

Solstrand, Os, 13 September 2013

Gisle Andersen, Marita Kristiansen, Vemund Olstad





## Towards a national term portal

- In CLARINO, NHH wants to establish a national infrastructure to **harmonize** and **integrate** terminological language resources.
- It is not the aim to develop new terminologies per se, but to provide the technical **architecture** needed for their **integration and accessibility**, and to vouch for their **interoperability** with the CLARIN and ISO/TBX standards.
- **Challenges:**
  - variety of resources
  - variety of formats
  - many resources not updated



# Plethora of resources

Ressurs	Kortnavn	Institusjon	Kontaktperson	E-post	An		
<b>Helsedirektoratets termbase</b>	Hdir	Helsedirektoratet	Jim Yang	<a href="mailto:jim.yang@helsedir.no">jim.yang@helsedir.no</a>	GA		
<b>KB-N</b>	KB-N	NHH	Marita Kristiansen	<a href="mailto:Marita.Kristiansen@nhh.no">Marita.Kristiansen@nhh.no</a>	MK		
<b>NHH Termbase</b>	NHH	NHH	Marita Kristiansen	<a href="mailto:Marita.Kristiansen@nhh.no">Marita.Kristiansen@nhh.no</a>	MK		
<b>NHH Mikroøkonomen</b>	NHH-M	NHH	Marita Kristiansen	<a href="mailto:Marita.Kristiansen@nhh.no">Marita.Kristiansen@nhh.no</a>	MK		
<b>NOT-basen</b>	NOT	UiB/UNI R	Johan Myking/Klaus Johansen		GA		
<b>EØS-basen</b>	EØS	Utenriksdepartementet	Cecilie Ovrum	<a href="mailto:cecilie.ovrum@mfa.no">cecilie.ovrum@mfa.no</a>	GA		
<b>SNORRE</b>	SNORRE	Standard Norge	Kjersti Drøsdaal Vikøren	<a href="mailto:kdv@standard.no">kdv@standard.no</a>	GA		
<b>UHRs termbase</b>	UHR	UHR	Rachel Glasser		MK		
<b>Rådet for Teknisk Terminologi</b>	RTT	Fagbokforlaget (Uni)	Atle Eggereide	<a href="mailto:atle.eggereide@fagbokforlaget.no">atle.eggereide@fagbokforlaget.no</a>	GA		
<b>Forskningsadministrativ terminologi</b>	NFR	Norges forskningsråd	Else Lie		MK		
<b>Termer frå Språkrådets datatermgruppe</b>	DataTerm	Språkrådet	Marianne Åsgård		GA		
<b>Norske oversettelser av termer fra Language Technology World</b>	LTW	Uni Research	Tove Bjørneset	<a href="mailto:tove.bjorneset@uni.no">tove.bjorneset@uni.no</a>	GA		
<b>Språkteknologisk terminologi</b>	SprakTek	Uni Research	Tove Bjørneset	<a href="mailto:tove.bjorneset@uni.no">tove.bjorneset@uni.no</a>	GA		
<b>Kystverket: Begreper</b>	KV	Kystverket			MK		
<b>MeSH (Medical Subject Headings)</b>	MeSH				MK		
<b>Risten.no</b>		Universitetet i Tromsø	Sjur Moshagen		GA		
<b>GeoLeksi</b>		Naturhistorisk museum	Inge Bryhni				
<b>MetLex Meteorologileksikon</b>		Meteorologisk institutt					
<b>Statens kartverk</b>		Statens kartverk					
<b>Helsenett.no</b>		Helsenett					
<b>IKT og helse</b>		Kompetansesenter for IKT i helse- og sosialsektoren					
<b>Karateterminologi</b>		Norges Kyokushin Karate-organisasjon					
<b>Pasienthåndboken: Medisinsk ordbok</b>		NHI					
<b>Ship circular</b>		Uni Research	Tove Bjørneset	<a href="mailto:tove.bjorneset@uni.no">tove.bjorneset@uni.no</a>	MAR	Må kontakte	farkartelzer, rkannet
<b>A dictionary and terminological analysis of merchant ship terms</b>	MS	Marte Salberg	Marte Salberg		MAR	Tilraagnam bruk	ca 90 termparter narzk-enqolrk
<b>Ordlirte fra Å-Å innonzkriptart og affzare</b>		Naroz redorifarbund, <a href="http://www.redori.narozuoibtcmz.narf/dacl04732D08970">http://www.redori.narozuoibtcmz.narf/dacl04732D08970</a>			MAR	Må kontakte	ca 360 uttrykk, mye enqolrk, narzk kart farklarinq
<b>Narzk-enqolrk ordlirte far farkarar av Einar Åurtlid (1971) Orla: Bjarne H. Roenurkaug</b>					MAR	Må kontakte	delvir into qret
<b>Bokkrivnlre av relevante farkeredkap og farkeraktivitet i Narqoz nkanam</b>		Oljedirektoratet			MAR	Må kontakte	
<b>Grønn ordbok</b>		UMB			MK	3	Må kontakte
<b>Cammin</b>		UMB?			MK	2	Må kontakte



# Terminologi.no

- <http://www.terminologi.no/>

The screenshot shows the top part of the website. On the left, there is the NHH logo and the text 'Termportalen' next to a 2x2 grid of icons. On the right, there are links for 'CLARINO | NHH | Kontakt'. Below this is a dark navigation bar with the following items: 'Forsiden' (highlighted), 'Søkeside', 'Om Termportalen', 'Termbasene', 'Om CLARINO', 'Publikasjoner', and two flags (Norwegian and British).

Du er her: [Forsiden](#)

## Termportalen ved NHH

Velkommen til Terminologi.no – en nasjonal portal for terminologi!

I Norge har det i lang tid vært etterlyst et initiativ for å etablere en nasjonal termportal. Dette arbeidet er nå i gang i regi av CLARINO, et nasjonalt prosjekt for forskningsinfrastruktur finansiert av Norges Forskningsråd (NFR).

Termportalen vil gi brukere fri tilgang til norsk terminologi innenfor en lang rekke fagområder, og gjøre det mulig å søke på tvers av termbaser og fagområder.

Termportalen er under utvikling og er i ferd med å fylles med innhold. Arbeidet med Termportalen er organisatorisk tilknyttet Institutt for fagspråk og interkulturell kommunikasjon, Norges Handelshøyskole (NHH), og er et samarbeid mellom NHH, Universitetet i Oslo og Universitetet i Bergen.



# Arbeidet frem til nå

- **Innhenting av data**
  - 8 baser konvertert
  - Nærmer oss avlevering på flere andre
- **Vurdering av innhold / modellering**
  - Ulike eksportformat
  - Ulik struktur på basene (alt fra enkle tabeller til komplekse baser)
  - Veldig varierende innhold
- **Konverteringsrutiner**
  - Skript / stilark for hver enkelt base
  - Alt konverteres til ett standardformat (TBX)
- **Testversjon av Termportalen**
  - Enkel prototype lansert



# Oppgaver for høsten 2013

- **Databaseoppsett**
  - Modelling av relasjonell base
    - Oracle eller Postgres
- **Eksportformat**
  - Oppdatering av TBX
    - Basert på erfaringer fra første runde med konverteringer
  - Baser med ferdigstatus klar for avlevering til CLARIN
    - Koordineres m. UBB
- **Redigeringsapplikasjon**
  - Testversjon klar i løpet av høsten
    - ‘Bare-bones’-versjon med enkelt design
    - Kan forhåpentligvis testes i undervisning
- **Termportalen**
  - Utvide søkefunksjonalitet iht. prosjektbeskrivelse

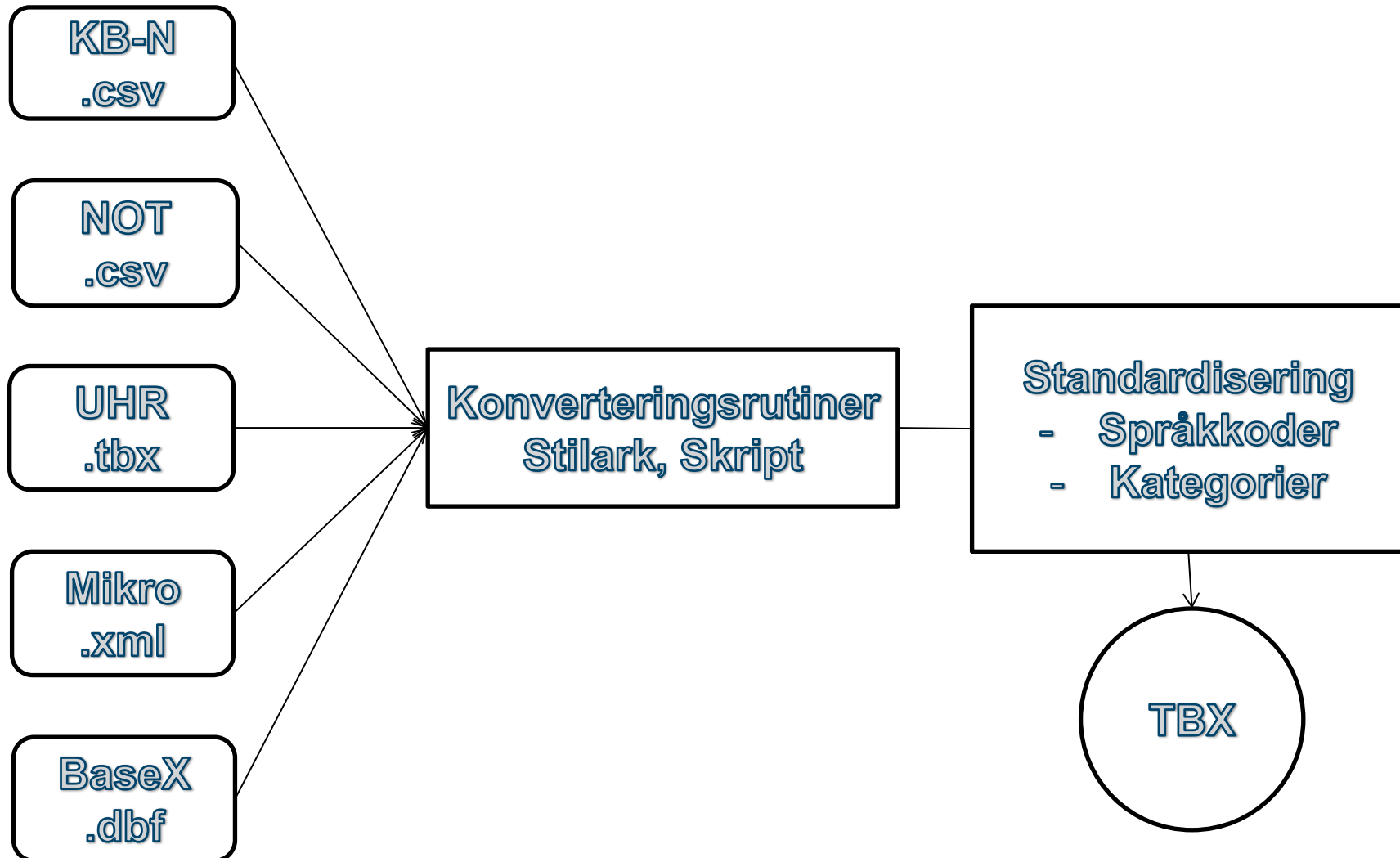


# Status – terminologiske ressurser

Ressursnavn	Fagområde	Termposter	Språk
<b>KB-N</b>	økonomi/administrasjon	8.473	N/E
<b>Mikroøkonomen</b>	mikroøkonomi	798	N/E
<b>NHH Termbase</b>	økonomi/ administrasjon	2859	N/E
<b>UHR</b>	studieadmin. term.	3351	N/E
<b>NOT</b> (Norsk termbank)	oljeterminologi ++	30.525	N/E
<b>RTT</b>	tekniske fagområder	48.311	N/E
<b>LTW</b>	språkteknologi	105	N/E
<b>SprakTek (NorskDok)</b>	språkteknologi	83	N/E
<b>Maritim ordbok</b>	maritime fag	1.226	N/E
<b>Totalt</b>		<b>92.389</b>	



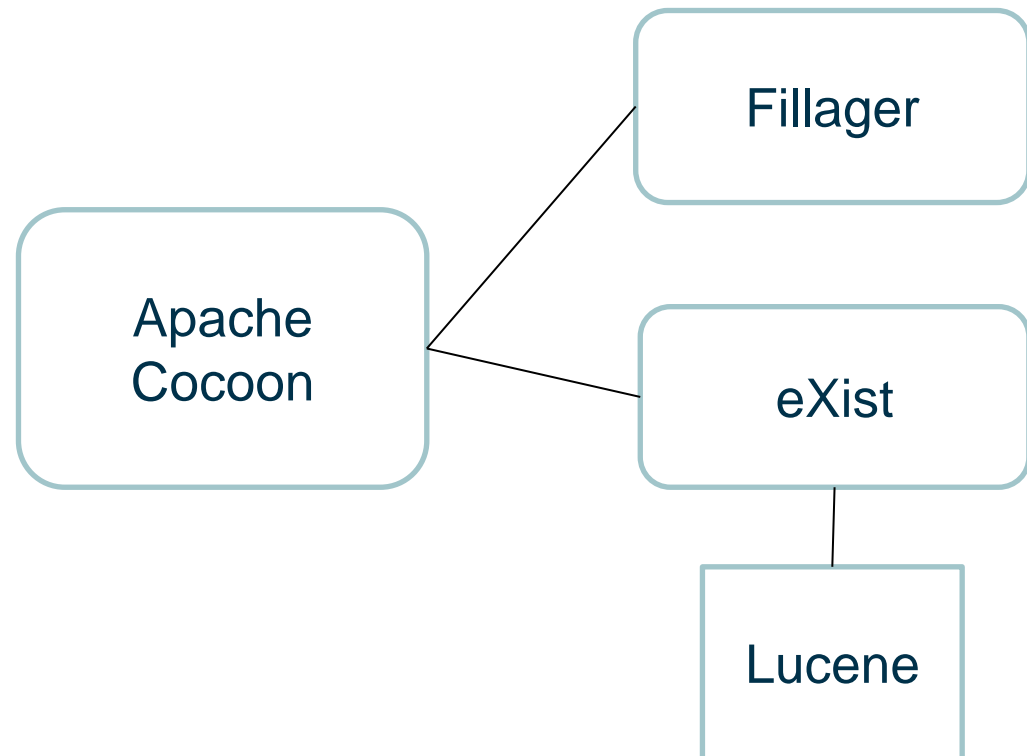
# Konverteringsarbeidet





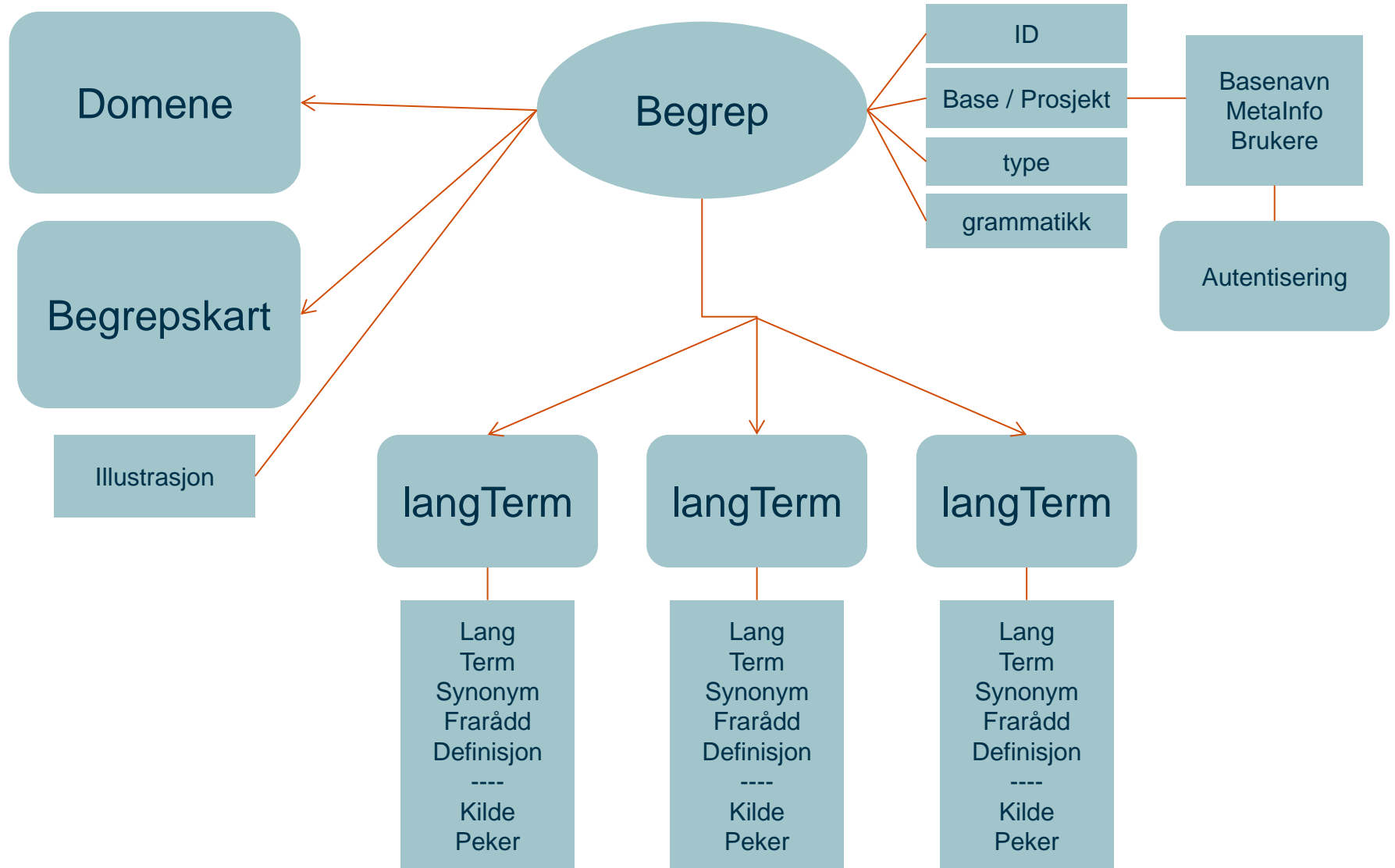


# Demoversjonen - infrastruktur





# Datamodell





# Resource integration

CLARINO | Mikrøøkonomi

## Termportalen

Forsiden
Søkeside
Om Termportalen
Termbasene
Om CLARINO
Publikasjoner

Du er her: [Forsiden](#) > [Termbasene](#) > [Mikrøøkonomen](#) > mikro-535

←
→

**term** sealed-bid auction

**def** auction market in which all bids are made simultaneously in sealed envelopes, the winning bidder being the individual who has submitted the highest bid

**term** auksjon med lukket engangsbud

**takeover bid**

KBN: kbn-3438 (eng) [term]

**current bid price**



# bid

- What the dictionary will tell you

**bid** (subst.) bud, tilbud, anbud, melding, forsøk (Stor engelsk-norsk)

(br.) (am.) offer certain price for something, especially at auction

(br.) (am.) utter greeting or farewell to

**bid**<sup>1</sup> *subst. /bɪd/*

**1** (auksjon) bud  
**EKSEMPEL** • *my highest bid will be £100* mitt høyeste bud blir 100 pund

**2** (handel) tilbud, anbud

**3** (kortspill) bud, melding

**4** forsøk  
**EKSEMPEL** • *the prince made a bid for the throne* prinsen gjorde et forsøk på å bli konge

**5** (hverdagslig, amer.) innbydelse, invitasjon

**UTTRYKK**

**make a bid for something** gjøre alt for å vinne noe, være ute etter noe

**no bid** (kortspill) pass

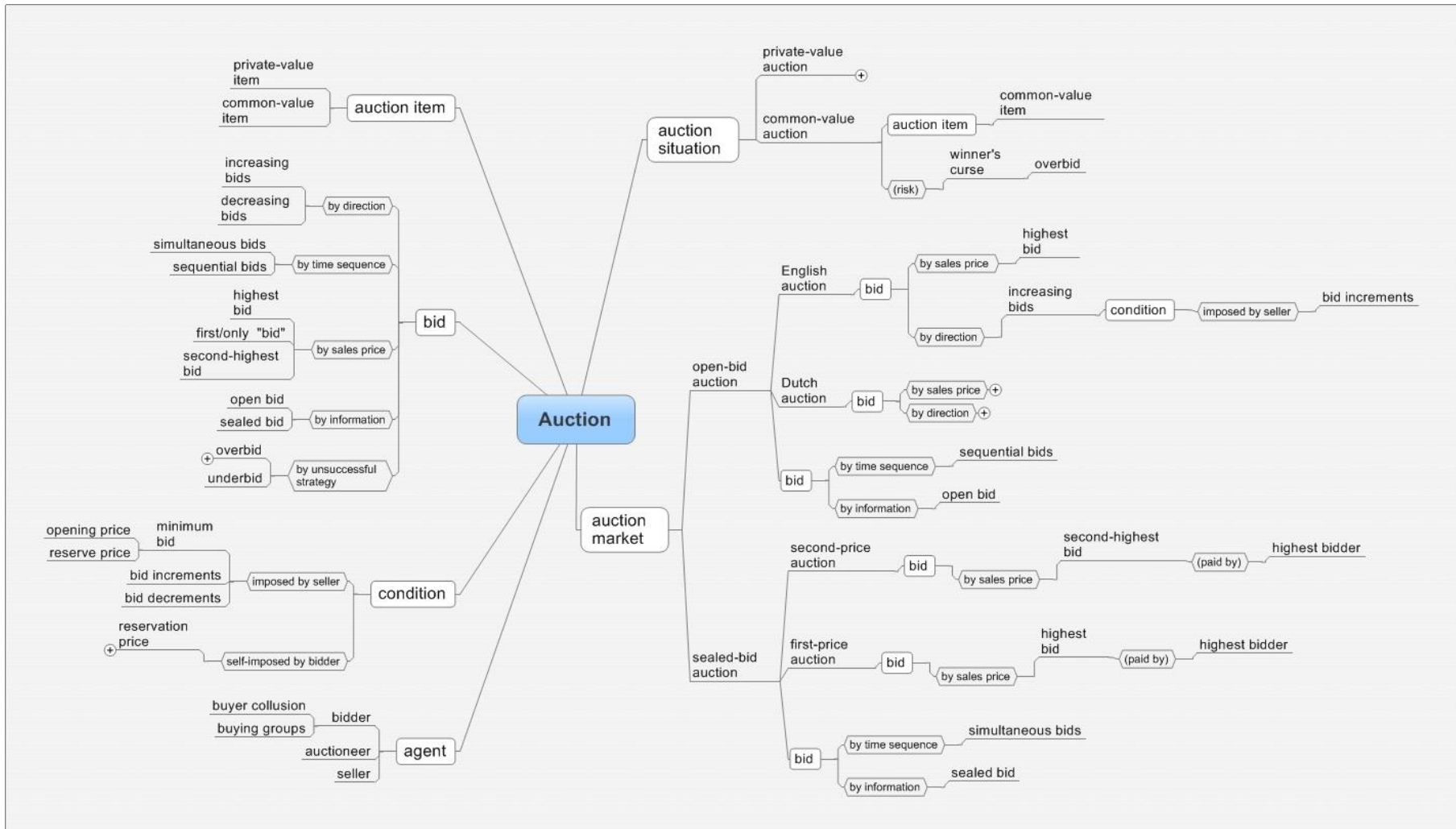
**win the bid** få tilslaget

Publikasjon: [Stor engelsk-norsk](#)

[Send en kommentar til denne artikkelen.](#)

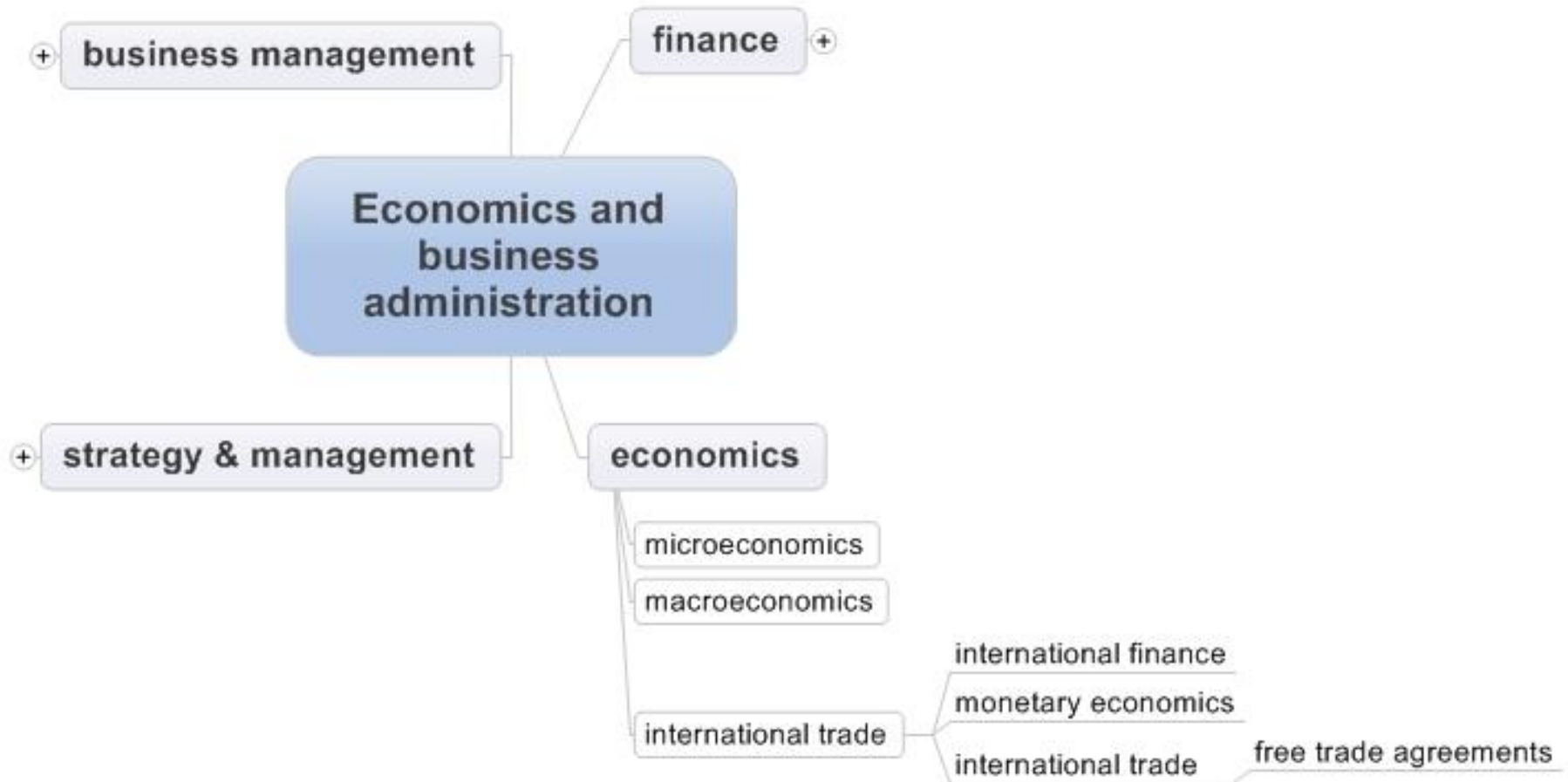


# Conceptually structured data (ontologies)





# Domains and subdomains





# The lexicographical approach

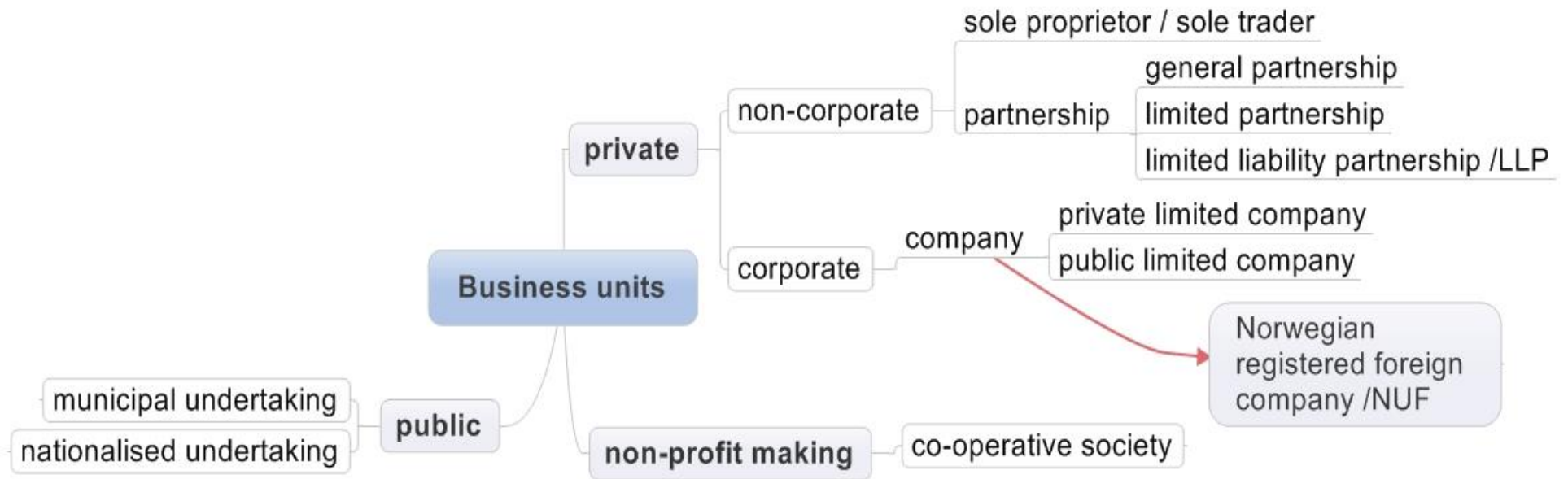
▣ **selskap** (subst.) society, company, companionship (Stor norsk-engelsk)

**selskap** *subst. n*

- 1 (samvær) society, company, companionship
- 2 (festlig sammenkomst) party
- 3 (festdeltakere) company, party
- 4 (økonomisk foretak) company, corporation (*amer.*)
- 5 (gruppe personer) party
- 6 (forening) society, association
- 7 (skuespillertrupp) company
- 8 (samfunnslag) society



# Terminology







# Resource integration and use cases

- CLARIN federated content search
  - Incoming/outgoing
  - Downloading of whole TBX termbases
  - Direct access to terminological information in term posts
- Added value for Terminologi.no
  - Links to corpus search, e.g. Norsk aviskorpus (and vice versa)
  - Morphosyntactic information from Norsk ordbank
  - Decomposition of composita from Oslo-Bergen tagger
  - Parse trees from INESS
- Added value for CLARIN(O) users generally
  - Exploitation of terminology resources for various technology purposes
  - Integration into multilingual/semantic application
  - WordNet, WSD, MT, ontology engineering, etc.



# Challenges

## Technology

- Integrating data into CLARIN repositories
- Integrating other resources into Termportalen
- Visualisation: dynamic (on-the-fly) generation of concept diagrams (3D, zoomable)

## Content

- Domain/subdomain classification of resources lacking this info
- Updating content