

NHH



Terminology methods and tools developed for the national terminology portal *Termportalen* (Norway)

Gisle Andersen, Marita Kristiansen, Vemund Olstad

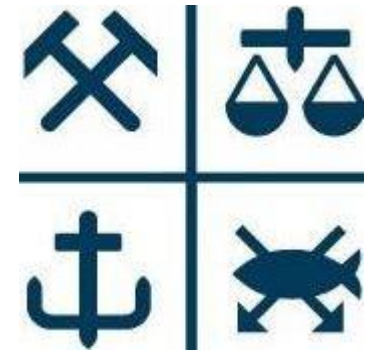
IITF Colloquium – Terminology tools: state of the art and future perspectives; 9 July 2015



Outline of presentation

- Background and status
 - Briefly on motivation, the project and status
- Integration of resources from terminology owners
 - Consolidation of domain information
 - Non-consistent use of data fields
- Terminology management tools
 - Generic data model
 - Terminology CMS System





Termportalen

BACKGROUND & STATUS



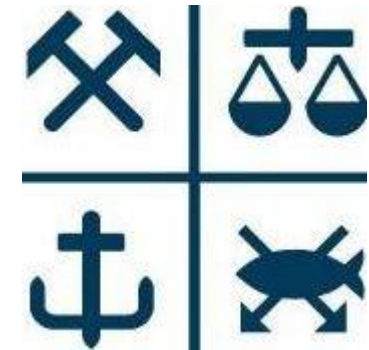
Aims of subproject Termportalen

- Overcoming the **fragmentation problem** in the field of terminology in Norway
- Establish <http://www.terminologi.no/> - a **national infrastructure** whose purpose is to **harmonize** and **integrate** terminological language resources (TLRs) into CLARINO for the benefit of **a variety of end users**
- Provide the technical **architecture** needed for **integration and accessibility** of TLRs and to vouch for their **interoperability** with the CLARIN and ISO/TBX standards
- Thereby taking a national responsibility to develop and maintain Norwegian terminology in line with **national language policy** and the Norwegian higher education act
- part of WP 7 **Terminology Integration** of the RCN project: CLARINO (Common Language Resources and Technology Infrastructure), aiming to make LRs available for research purposes etc.



Availability of TLRs in Termportalen/ CLARINO

- Owners, distributors and end users of TLRs
- Termportalen generally open access
- Searchable and/or downloadable in CLARINO, depending on owner's choice
- Accessibility levels and licence agreements:
 - **Public**: Creative Commons/CC-by licence, neither authentication or authorisation
 - **Academic**: authentication but not necessarily authorisation
 - **Restricted**: authentication and authorisation
- CLARINO scheme for metadata (CMDI) for each resource



Termportalen

INTEGRATION OF RESOURCES FROM TERMINOLOGY OWNERS

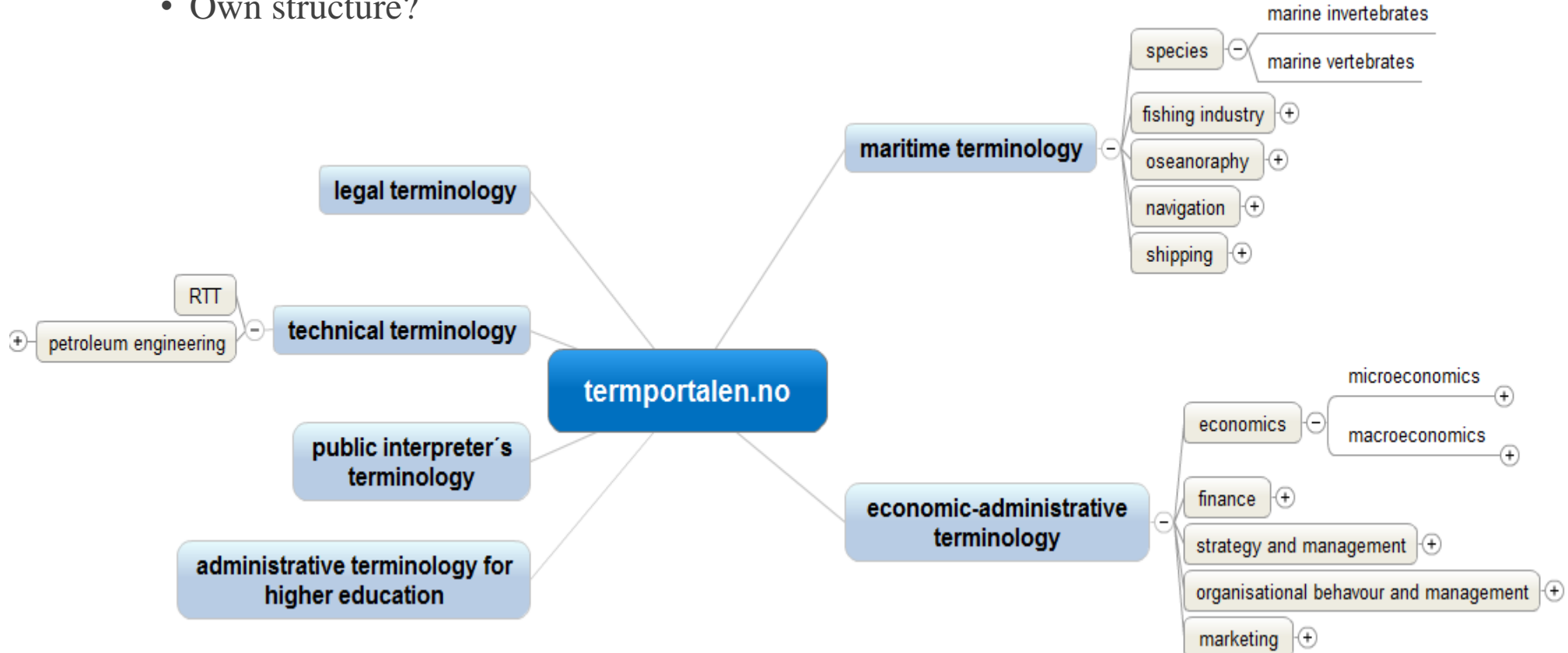


Consolidation of domain information

• How to build an over-arching domain structure

- Possible alternatives:

- Dewey's decimal classification?
- Delineation based on e.g. DanTermBank or others?
- Own structure?





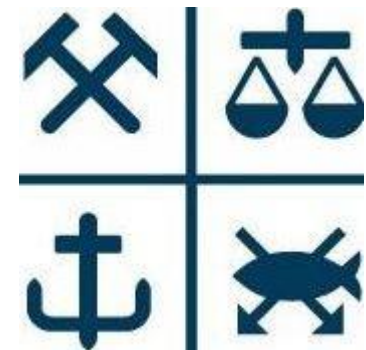
Non-consistent use of data fields (1)

onani				
opium				
opphopning (av væske)				
oppkast				
oppstøt				
skinnebein				
skinnlegg				
Skiveforskyvning (prolaps), skiveutglidning			Przemieszczenie/ wysuniecie dysku	
skjede				
skierdekatarr				
bevegelighet				
bevegelighet, mangel på				
bevisstløs				
hibulebetennelse				



Non-consistent use of data fields (2)

blodkar			
Blodkarsvulst			Bura xiddid-dhiig
Blodkreft/leukemi	Leukemi, blodkreft, felles navn på en gruppe ondartede blodsykdommer. Blodkreftcellene utgår fra celler i benmargen og/eller lymfeknuter.	https://sml.snl.no/leukemi	Kansarka dhiiga
blodlevring			
Blodomløp			Wareega dhiig
abort (spontan-)			Fausse couche, avortement spontané
abort (ta)			avorter
anke v; påanke v	jf. klage bringe en sak til ny behandling og prøving for en høyere instans	Ordbok for samfunn og helse – sosialtermer	
ansvarsfordeling			
kontaktlinser			
Dags-, døg-, månedslinser			Lentilles de (contact / correctrices) de jour / 24h/24 / mensuelles
Harde og myke linser			Lentilles de contact/correctrices



Termportalen

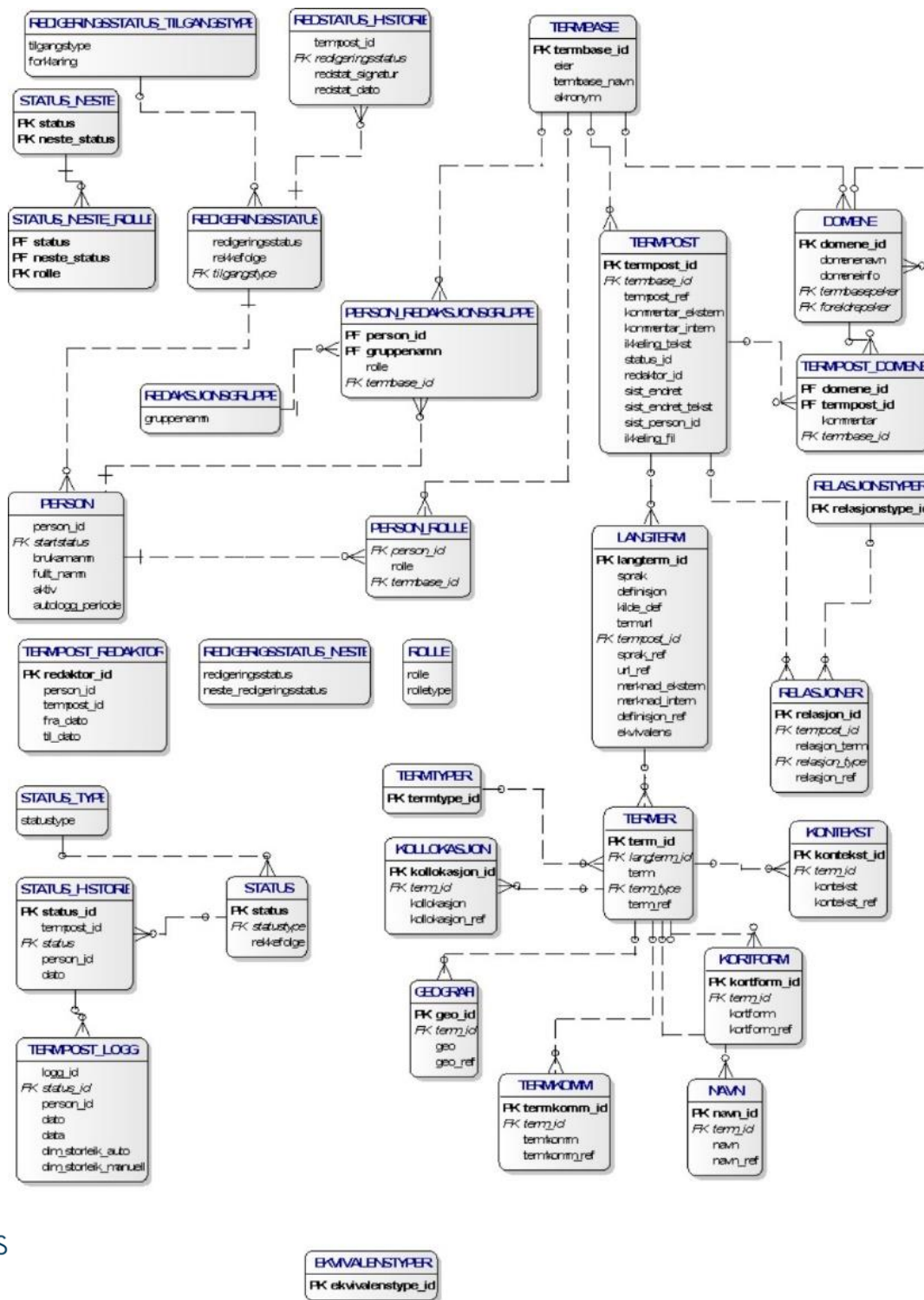
TERMINOLOGY MANAGEMENT TOOLS

Generic data model for interoperable terminology data



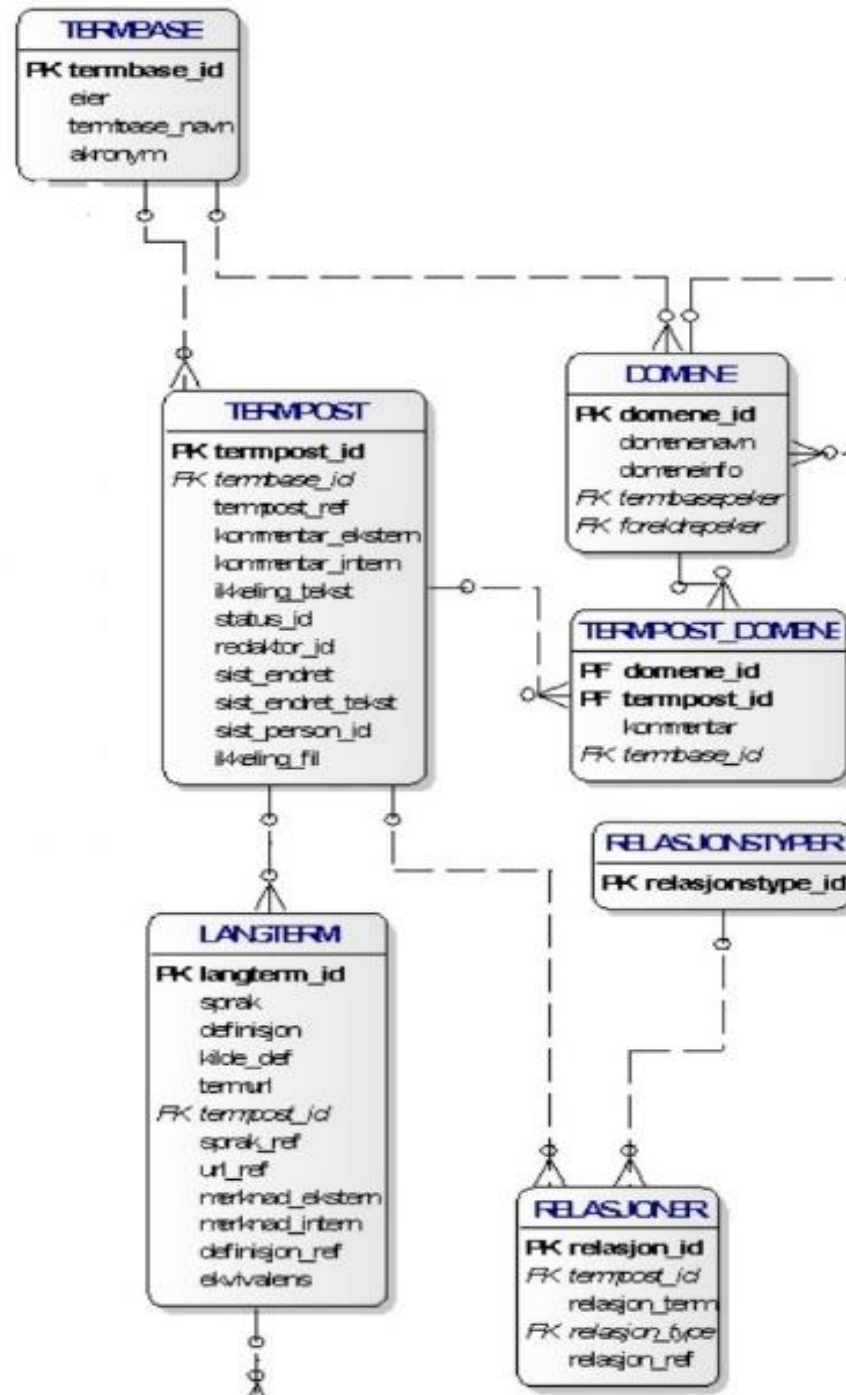
- Compliant with standards like:
 - *ISO 704 Terminology work: principles and methods*
 - *ISO 16642 Computer applications in terminology; TMF*
- Based on reviewing data categories from 'our' bases
 - Special attention given to NOT-basen (Norwegian oil terminology)
- Cooperation with *DanTermBank*, the Danish national terminology project – led by the Danterm Centre/ Copenhagen Business School (CBS)

Data model



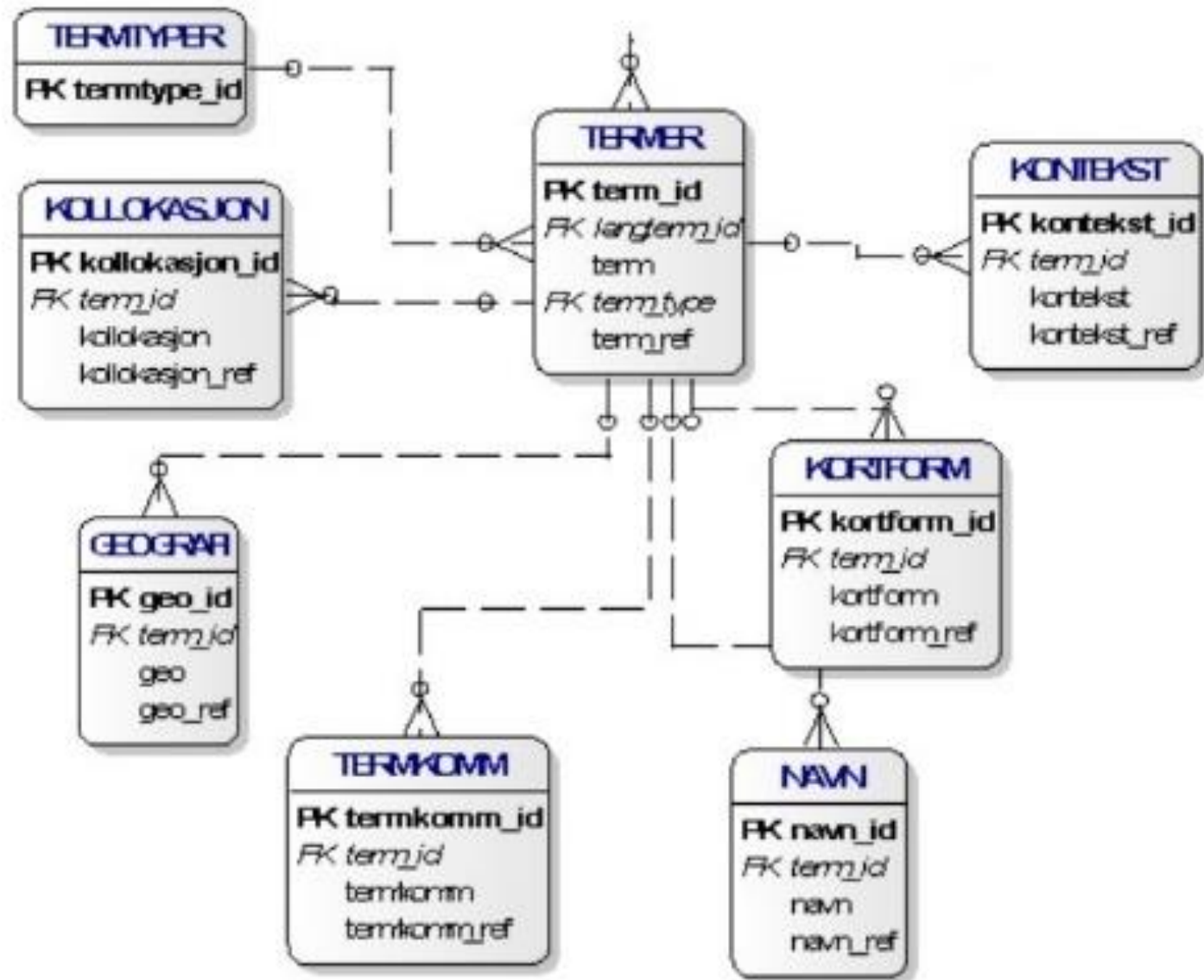


Concept





Term





Terminology CMS System

- Based on existing CMS for lexicography (*Norsk ordbok*)
 - developed by *Unit for Digital Documentation* (EDD, <http://www.edd.uio.no/>), who are responsible for '*Bokmålsordboka*' and '*Nynorskordboka*', the two largest online lexicographic dictionaries in Norway

Termpost (1) x

Søking Objektvisning

Søk Resultatoppsett Søkeoppsett

Antall treff pr. side

Tabellnavn/Søkefelt Verdi

Termpost

- Akronym
- Termbase-id
- Språk
- Definisjon
- Term_Type
- Term finans%
- Kontekst
- Kollokasjon
- Kortform
- Termkomm
- Merknad, intern
- Merknad, eksternt

Hent alle rader Resultatliste Redigér som trær Gruppér/Sortér

TERMKOMM	TERMPOST...	TERM	AKRONYM	TERM_TYPE	DEFINISJON
	20706	finansinntekt	KBN-basen	hvd	
	20834	finanspolitisk spareplan	KBN-basen	hvd	
	21320	finansiell stabilitet	KBN-basen	hvd	
	21321	finansielle sektorbalanser	KBN-basen	hvd	
	21322	finanstransaksjon	KBN-basen	hvd	
	21323	finansiell utsatthet	KBN-basen	hvd	
	21324	finansieringskostnader	KBN-basen	hvd	
	21324	finansieringskostnader	KBN-basen	hvd	
	21325	finansieringsstruktur	KBN-basen	hvd	
	21326	finansiell krise	KBN-basen	syn	
	21408	finansielt foretak	KBN-basen	syn	
	21409	finansielt system	KBN-basen	hvd	
	20457	finansministerium	KBN-basen	hvd	
	20554	finansielt instrument	KBN-basen	hvd	
	23843	finansanalytiker	KBN-basen	hvd	
	24133	finansstrøm	KBN-basen	hvd	
	23174	finansiell buffer	KBN-basen	hvd	
	23190	finansiell post	KBN-basen	hvd	
	23191	finansiell styringsvirksomhet	KBN-basen	hvd	
	22810	finansiell risiko	KBN-basen	hvd	
	23027	finansiell rapportering	KBN-basen	syn	
	22329	finansielle aktiva	KBN-basen	hvd	
	22331	finanstheori	KBN-basen	hvd	
	22510	finansiell rådgivning	KBN-basen	hvd	
	22511	finansiell rådgiver	KBN-basen	hvd	
	23507	finansiell ustabilitet	KBN-basen	hvd	
	23508	finansiell eiendel	KBN-basen	hvd	
	23508	finansiell eiendel	KBN-basen	hvd	
	23509	finansieringsalternativ	KBN-basen	hvd	
	23510	finansieringsproblem	KBN-basen	hvd	
	23379	finansiell boble	KBN-basen	hvd	



Parallel editing view

Termpost (1) x

Søking | Objektvisning

Termpost

- Termpost-referanse
- Termpost-id
- Termbase-id
- Status
 - Redaktør
 - Kommentar, intern
 - Kommentar, ekstern
 - Ikke-lingvistisk informasjon
 - Ikke-lingvistisk informasjon
- V_Termpost_Domene
- Langterm_Rus
- Langterm_Fra
- Langterm_Aao
- TermpostLogg
- Langterm_Pol
- Status
- Langterm_Nbo
- Langterm
- Langterm
- Langterm
- Langterm
- Termbase
- V_Status_Historier

Ny

Lagre

Slett

Frisk opp

Forrige post

Neste post

Admininfo

Redigeringsmodus

Parallellvisning

Parallellvisning

Norsk

Liste over tilknyttede termer

Type	Term	Referanse
hvd	aktivitør	

+

✖

Definisjon...

person som arbeider på en institusjon innen helse- og sosialsektoren med å sette klienter eller pasienter i gang med ulike former for aktiviteterEn aktivitør kan arbeide som assistent for en ergoterapeut.(pol) (fra) (aao)

Merknad, intern

Merknad, ekstern

Engelsk

Fransk

Arabisk

Polsk

Russisk

Liste over tilknyttede termer

Type	Term	Referanse
hvd	معاون معالج وظيفي	

+

✖

Definisjon...

شخص يعمل في مؤسسة صحية أو اجتماعية لمساعدة المرضى لقيام بنشاطات مختلفة وقد يعمل كمساعد للمعالج الوظيفي

Merknad, intern

Merknad, ekstern



Full editing view

Søking
Objektvisning

Termpost

- Termpost-referanse
- Termpost-id
- Termbase-id
- Status
 - Redaktør
 - Kommentar, intern
 - Kommentar, ekstern
 - Ikke-lingvistisk informasjon
 - Ikke-lingvistisk informasjon
- V_Termpost_Domene
- Langterm_Rus
- Langterm_Fra
- Langterm_Aao
- TermpostLogg
- Langterm_Pol
- Status
- Langterm_Nbo
- Langterm
- Langterm
- Langterm
- Termbase
- V_Status_Historier

Ny
Lagre
Slett
Frisk opp
Forrige post
Neste post

Admininfo
Redigeringsmodus
Parallellvisning

Detaljinformasjon

Språk
nob
pol
fra
rus
aao

Definisjon

person som arbeider på en institusjon innen helse- og sosialsektoren med å sette klienter eller pasienter i gang med ulike former for aktiviteterEn aktivitør kan arbeide som assistent for en ergoterapeut.(pol) (fra) (aao)

Definisjon (ref)

Merknad, intern

Merknad, ekstern

Liste over tilknyttede termer

Type	Term	Referanse	UTF-8
hvd	aktivitør		

Kontekster

Kontekst	Kontekst Ref

Kollokasjoner

Kollokasjon	Kollokasjon ref

Kortformer

Kortform	Kortform ref

Kommentarer til termer

Kommentar	Kommentar ref



Administrative info

Ny Lagre Slett Frisk opp Forrige post Neste post

Admininfo Redigeringsmodus Parallellvisning

Admininfo

Termbase
5

Termpost
46889

Termbase
Tolketjenestens termbase

Eier
Tolketjenesten i Bergen

Domenenavn	Domeneinfo	Domene_jd
Sosialomsorg og sosiale tjenester		6

Termpostref

Status
Importert

Endre status

Redaktør

Endre redaktør



Concluding remarks

- Termportalen – a solid framework for methodological and systematic national terminology work:
 - CMS and data model set a **standard** for how to register terminological data
 - **Reduces terminological variation** across local resources
 - **Enables harmonisation of terminology and domains** across organisations, e.g. public bodies
- Extended view: «terminological infrastructure» not only terms and termbases, but also related **technology, research, competence, best practices, language standardisation**, etc. Access to research and technology competence enables e.g. semi-automatic **term extraction** from corpora

- **The combination of terminology work, research and technology development is a prerequisite for a successful development of Termportalen**