



# *Combining data from heterogeneous sources*

Hanne Erdman Thomsen



# Overview

## Harmonization of framework (DB)

- Data fields
- Data structure
- Electronic format

## Harmonization of content

- Classification
- Lists
- Concepts



# *Data fields*

- ISO 12620:1999, *Computer assisted terminology management — Data Categories*
- ISOcat Data Category Registry  
<http://www.isocat.org/>
- DS 2394-1. (1998). Lexical data collections – Description of data categories and data structure – Part 1: Taxonomy for the classification of information types, Danish Standards



# *DS - STANLEX*

## Taxonomic structure

- Administrative information
- Etymological information
- Grammatical information
- Graphical information
- Language
- Phonetic information
- Semantic information
- Structural information
- Usage

Grammatical information	<ul style="list-style-type: none"> <li>● Part of speech</li> <li>● Gender</li> </ul>	
	<ul style="list-style-type: none"> <li>● Information on inflection</li> </ul>	<ul style="list-style-type: none"> <li>● Stem</li> <li>● Paradigm information</li> <li>● Inflected form</li> </ul>
	<ul style="list-style-type: none"> <li>● Word formation</li> </ul>	
	<ul style="list-style-type: none"> <li>● Syntax</li> </ul>	<ul style="list-style-type: none"> <li>● Syntactic frame</li> <li>● Specification of syntactic frame</li> <li>● Specification of auxiliary verb</li> <li>● Syntactic function</li> </ul>

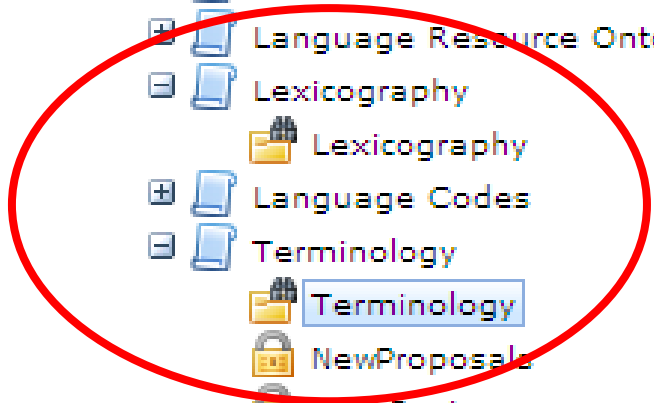


enter keywords here



My Workspace

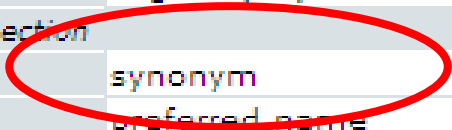
- Public
  - Thematic Views
    - Metadata
    - Morphosyntax
    - Semantic Content Representation
    - Syntax
    - Language Resource Ontology
    - Lexicography
      - Lexicography
    - Language Codes
    - Terminology
      - Terminology
      - NewProposals
      - TBX-Basic
      - TBX-Default\_DCS
  - Multilingual Information Management
  - Lexical Resources



#	Name	Version
329	abbreviated form	1:0
64	abbreviated form for	1:0
331	abbreviation	1:0
1418	abbreviation	1:0
65	abbreviationfor	1:0
334	acronym	1:0
66	acronym for	1:0



Language	English (en)
2.1.1 Name Section	
Name	synonym
Name Status	preferred name
2.1.2 Definition Section	
Definition	Any term or lexeme that represer
Source	ISO12620:1999
2.1.3 Explanation Section	



# *Data structure*

## *Definition*

Language 1

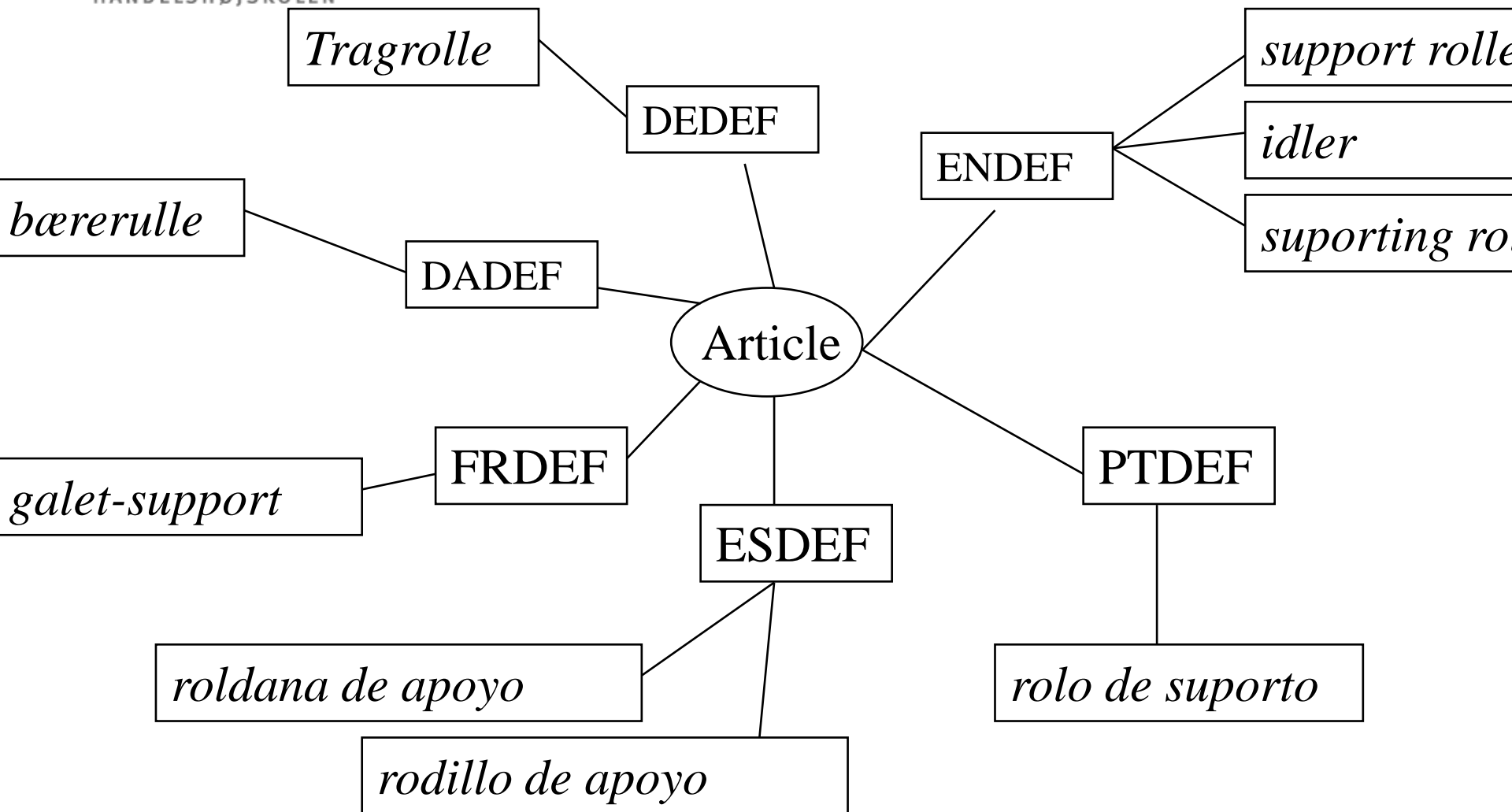
- Term
- Term
- Term

Language 2

- Term
- Term
- Term



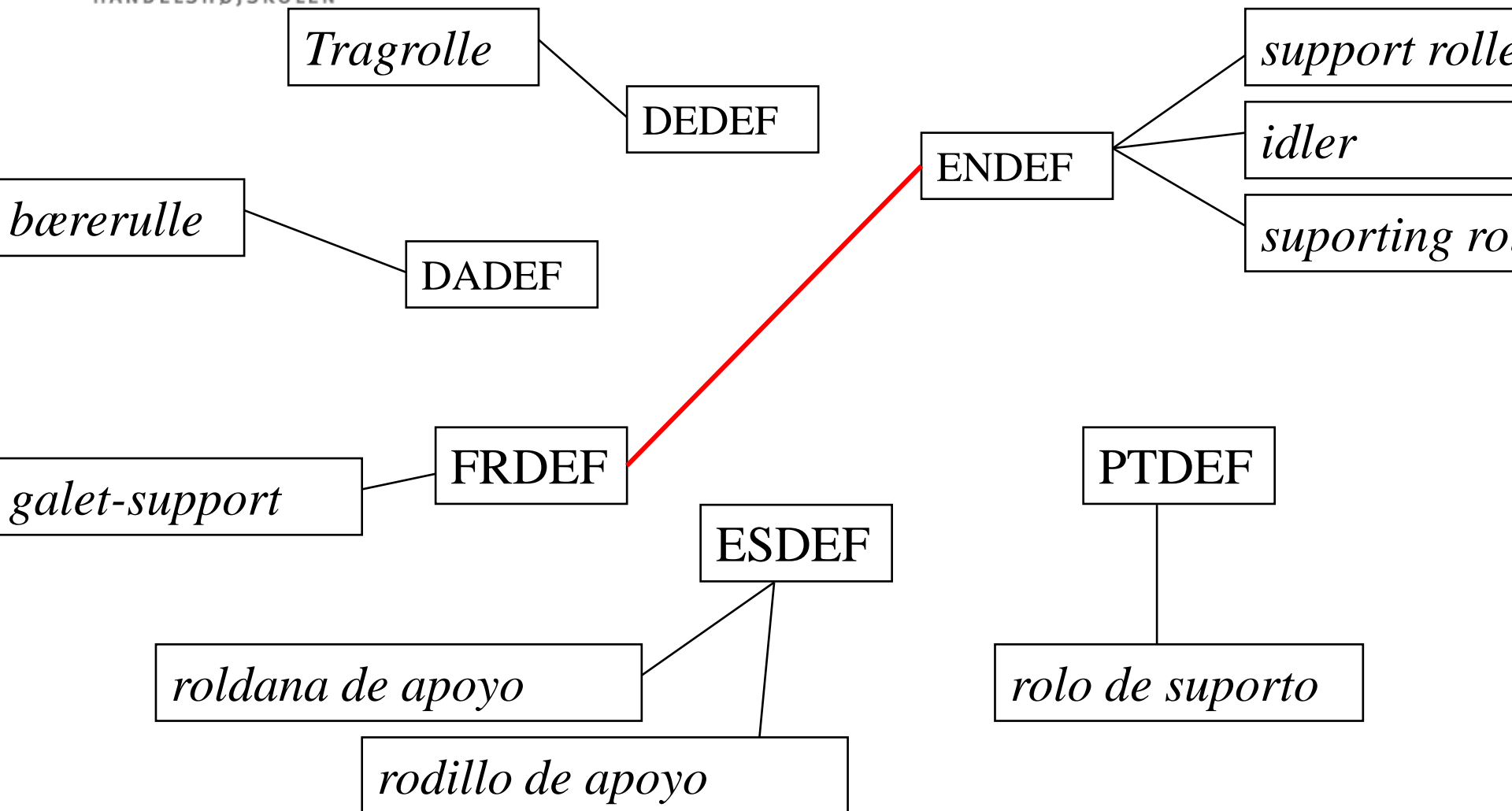
# Data structure







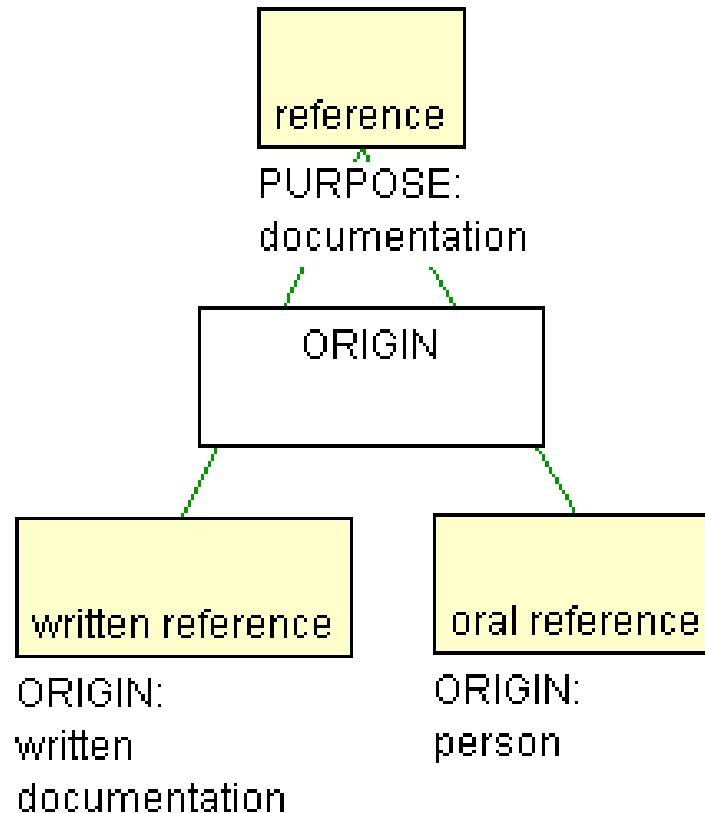
# Adhere to working method



# OLIF entry

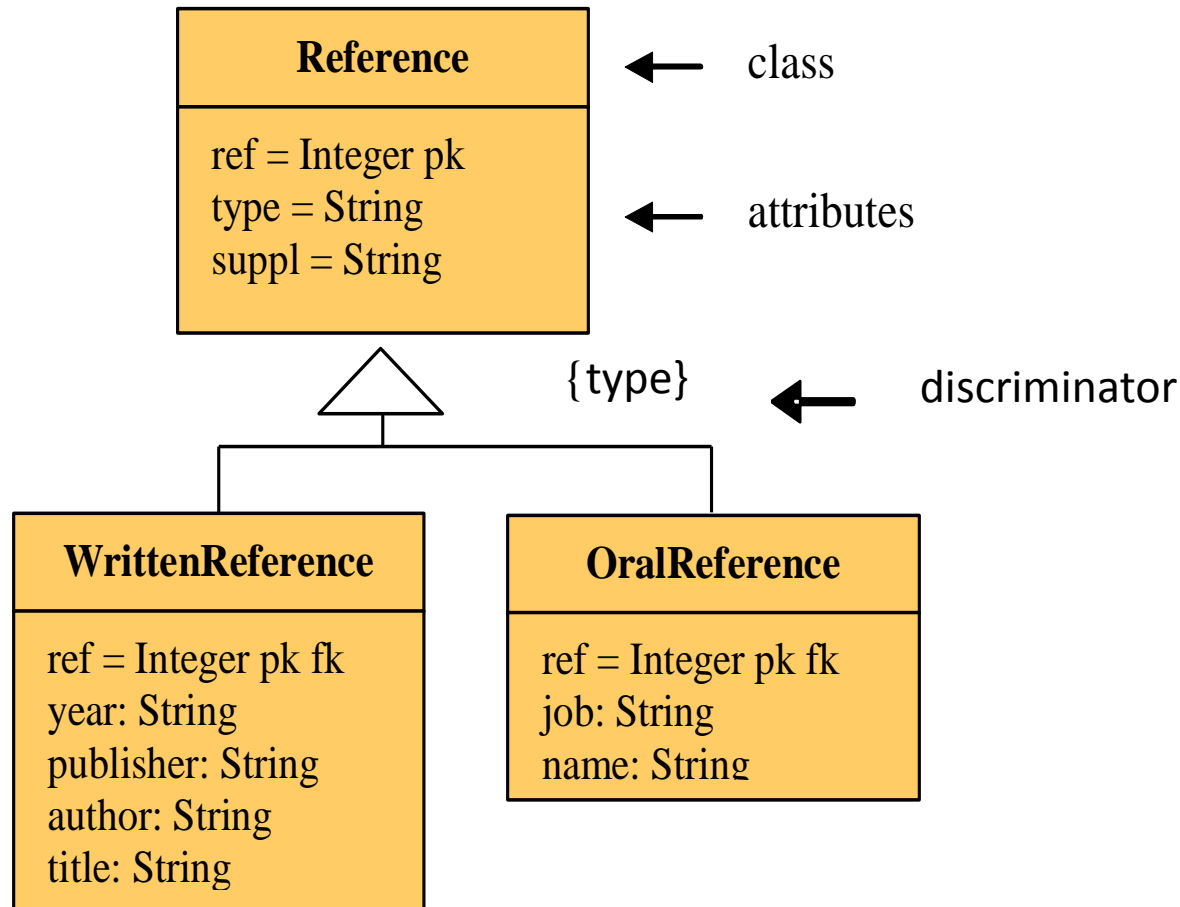
```
<entry>
  <mono>
    <keyDC>
      <canForm>table</canForm>
      <language>en</language>
      <ptOfSpeech>noun</ptOfSpeech>
      <subjField>general</subjField>
    </keyDC>
  <monoDC>
    <monoAdmin>
      <originator>Weber</originator>
    </monoAdmin>
    <monoSem>
      <definition>An arrangement of words, numbers, or signs or combinations of them, as in parallel columns, to exhibit a set of facts or relations in a definite, compact, and comprehensive form.</definition>
    </monoSem>
  </monoDC>
</mono>
<crossRefer>...</crossRefer>
<transfer>...</transfer>
</entry>
```

# *Ontology vs. conceptual data model*



*Extract of an ontology for terminological information*

# *Ontology vs. conceptual data model*



*Conceptual data model for a part of a terminology database*

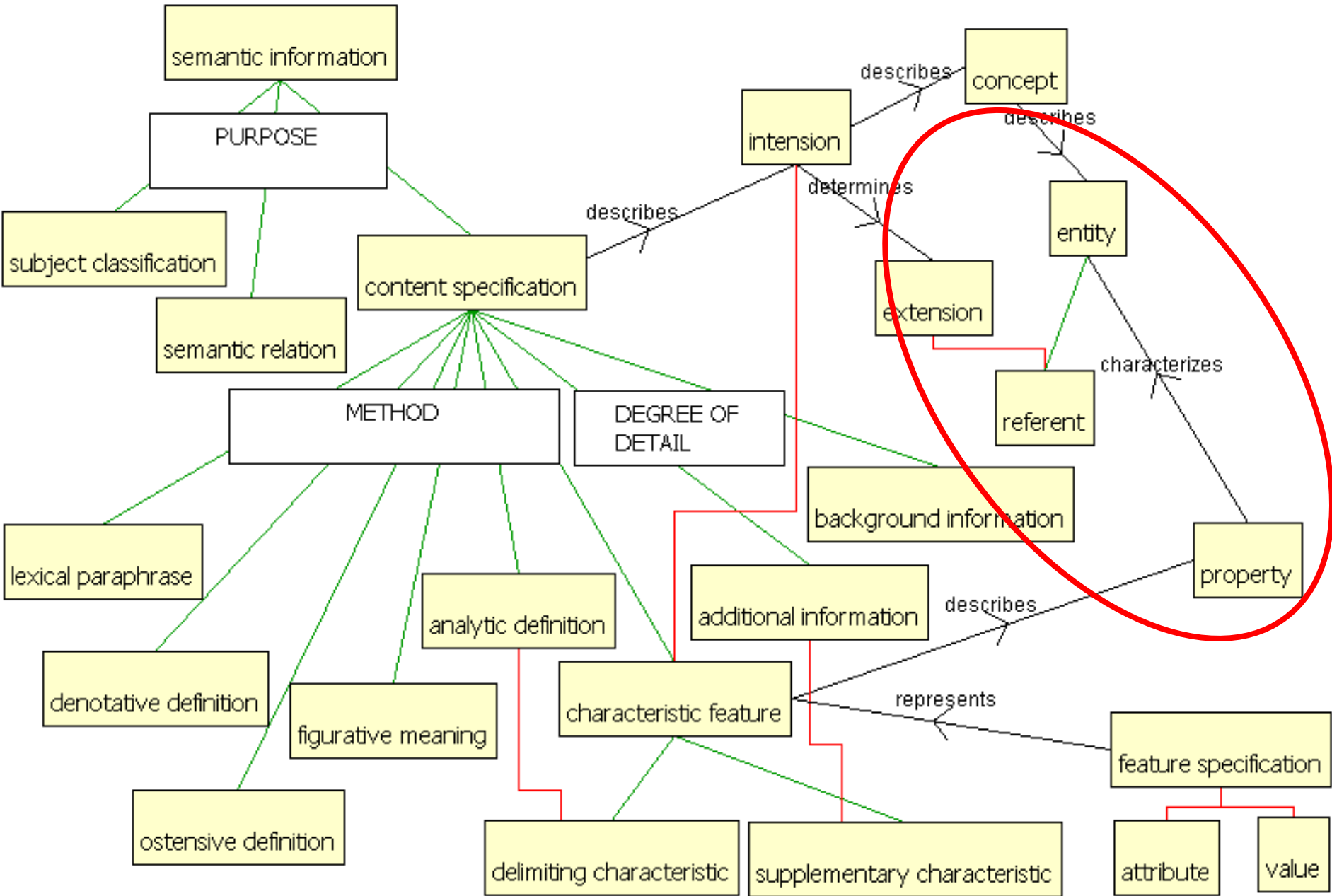
# *Ontology vs. conceptual data model*

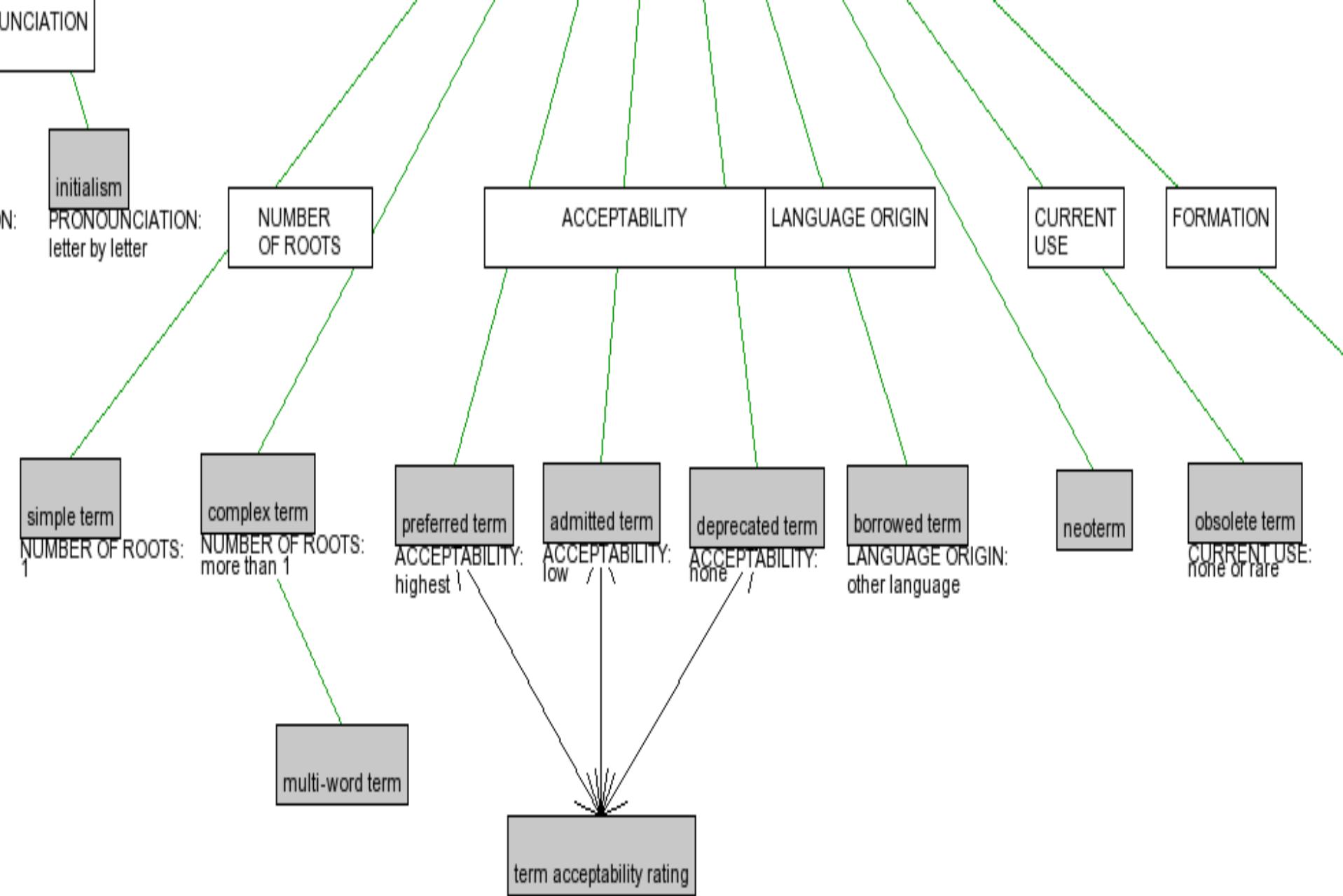
Ontologies (concept models) and conceptual data models have different aims:

- ontologies aim at concept clarification, mutual understanding of concepts and consistent use of terms
- conceptual data models aim at specifying the information types of an IT system and their mutual relationships

# *Ontology vs. conceptual data model*

Ontologies (concept models)	conceptual data models
information about concepts in the form of feature specifications and concept relations (information about meaning)	information about the classes in the form of attributes and associations between the classes  Attributes convey no meaning of class

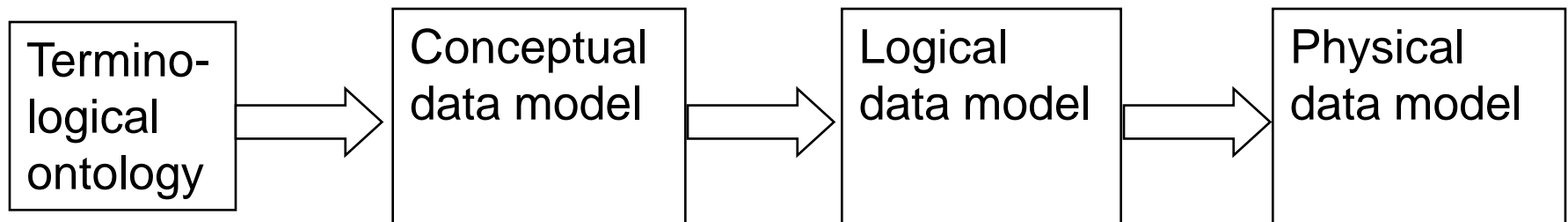






# *Ontology vs. conceptual data model*

Ontologies may form the basis for conceptual data models.



**Recommendation:**

**Always develop a terminological ontology before developing a conceptual data model!**



# *Classification:* *DANTERMklassifikationen*

- Split a large terminology collection into manageable subsets – ideally up to 500 terminological units
- Fast retrieval
- Data interchange
- 2 levels?



- A. Samfundsvidenskab
- B. Kunst og litteratur
- C. Fritid og sport
- D. Stat, offentlig forvaltning, interne relationer
- E. Jura
- F. Økonomi og handel
- G. Naturvidenskab og matematik
- H. Medicin
- I. Teknik, industri, håndværk, forskellige erhverv
- K. Bygningsteknik
- L. Råstofindustri
- M. Datalogi og datateknik
- N. Energiteknik
- O. Maskinteknik
- P. Kommunikationsteknik
- Q. Landbrug
- R. Økologi og miljø
- S. El-energi, elektroteknik, elektronik



## *E. Jura (law)*

E3000 privatret

E3100 personret

E3200 familieret

E3500 arveret

E4000 obligationsret

E4100 obligationsrettens alm. del

E4500 obligationsrettens specielle del



- Was used as basis for subjects in a technical dictionary
- Bodil Nistrup Madsen, Heribert Picht, Gert Engel.



# *Concept merging*

## absorption

- Risktermbanken 68
- IATE 665
- EuroTermBank 42 (-75?)



## absorption sv

FÖRKLARING: En ljus yta reflekterar det mesta ljus som träffar den, medan en mörk yta ...

KÄLLA: Institutet för infologi, Rune Pettersson: Informationsdesign från a till ö | 2008

[Visa hela termposten](#)

## absorption sv

DEFINITION: den process som innebär att ljus och annan strålning omvandlas till värme ...

KÄLLA: Energimyndigheten: Ordlista | 2003

[Visa hela termposten](#)

## absorption sv

DEFINITION: process som åstadkommer energiförlust hos strålning när den passerar ...

KÄLLA: MSB; Myndigheten för samhällsskydd och beredskap: Ordlista | 2003, s. 78–97

[Visa hela termposten](#)

## absorption (1) sv

DEFINITION: process (1) vid vilken ett ämne (1) upptas i det inre av ett annat ämne ...

KÄLLA: Terminologacentrum TNC: Tekniska basord | 1995 [ Ordlistan är under revidering.]

[Visa hela termposten](#)

## absorption (2) sv

DEFINITION: process (1) genom vilken ljudenergi minskas genom energiupptagning hos ...

KÄLLA: Terminologacentrum TNC: Tekniska basord | 1995 [ Ordlistan är under revidering.]

[Visa hela termposten](#)

## absorption (3) sv

DEFINITION: omvandling av strålningsenergi till värme

KÄLLA: Terminologacentrum TNC: Tekniska basord | 1995 [ Ordlistan är under revidering.]

[Visa hela termposten](#)

## absorption (1) sv

DEFINITION: inträngning och upptagning av ett ämne i det inre av ett annat ämne

ÖVRIGA SPRÅK: de, en, fr

KÄLLA: Terminologacentrum TNC: Plan- och byggtmer | 1994

[Visa hela termposten](#)

## Life sciences [COM]

[Full entry](#)

EN	soaking	★★★★ →@
	<a href="#">absorption</a>	★★★★ →@
DE	Durchtränkung	★★★★ →@

## AGRICULTURE, FORESTRY AND FISHERIES [Council]

[Full entry](#)

EN	<a href="#">absorption</a>	★★★★
DE	Übernahme	★★★★

## Technology and technical regulations, Migration [Council]

[Full entry](#)

EN	<a href="#">absorption</a>	★★★★
DE	Wegschlagen	★★★★ →@
	Eindringen	★★★★ →@

## Migration, Technology and technical regulations [Council]

[Full entry](#)

EN	<a href="#">absorption</a>	★★★★
	optical <a href="#">absorption</a>	★★★★
DE	optische Absorption	★★★★
	Absorption	★★★★



RESULTS (42)

**absorption**

EN PL  
natural and applied sciences;

**absorption**

EN LV RU  
social affairs; miscellaneous industries;

**absorption**

EN LV RU  
social affairs; miscellaneous industries;

**absorption**

EN LV  
electronics and electrical engineering; energy;

**absorption**

EN LV  
electronics and electrical engineering; energy;

**absorption**

EN LV  
electronics and electrical engineering; energy;

**absorption**

EN LV

ADVANCED SEARCH

absorption

From English (en) To Any Any subject

- Search also in definitions and explanations  Search in all forms
- Search also in external databases

Exact match

-normative authoriza

SELECTED RESULT

Collection: **Mežtehnikas, mežsaimniecības un kokrūpniecības terminu vārdnīca**  
Subject: wood industry (EuroVoc: 6836)

LV **absorbicija**

EN **absorption**

DE **Absorption**

FR **absorption**

RU **абсорбция**

Discuss this entry...

RESULTS (42)

**absorption**

EN PL  
natural and applied sciences;

**absorption**

EN LV RU  
social affairs; miscellaneous industries;

**absorption**

EN LV RU  
social affairs; miscellaneous industries;

**absorption**

EN LV  
electronics and electrical engineering; energy;

**absorption**

EN LV  
electronics and electrical engineering; energy;

**absorption**

EN LV  
electronics and electrical engineering; energy;

**absorption**

EN LV

ADVANCED SEARCH

absorption

From English (en)

- Search also in definitions and external data
- Search also in external data

SELECTED RESULT

Collection: **Mežtehni**  
Subject: wood indust

LV **absorbciija**

EN **absorption**

DE **Absorption**

FR **absorption**

RU **абсорбция**

Discuss this entry...

RESULTS (42)

**absorption**

EN DE ES ET FR LT LV RU  
wood industry; social affairs; education; electronics and electrical engineering; environment; law; natural and applied sciences; chemistry; miscellaneous industries; energy; transport; mechanical engineering;

**absorption**

EN DE ES ET FR LT LV PL RU  
natural and applied sciences; miscellaneous industries; humanities; mechanical engineering;

**absorption**

EN DE ES ET FR LT LV PL RU  
natural and applied sciences; miscellaneous industries; humanities; mechanical engineering;

**absorption**

EN DE LT LV RU  
social affairs; financing and investment; health; miscellaneous industries; economic analysis;

ADVANCED SEARCH

absorption

From English (en) To A

- Search also in definitions and external data
- Search also in external data

SELECTED RESULT

Entry compounding is done by caution. See original entries.  
Subject: natural and applied s

LV **absorbciija, elastibas pa**

EN **absorption** (of the bed, P

**Definition:** In chemical term structure of another, as disti  
**Definition:** The process in v process always results from radiation into some other for absorbing medium. The absc energy conversion has occur

DE **Absorption** (Fritz.1989: 5

**Definition:** Art der Sorption Masse der absorbierenden S Oberflächenerscheinung wie  
**Definition:** Wellenlängenabl bei welcher Strahlungsenerg Sonnenstrahlung an der Erdc Atmosphäre als Energiequell Aerosolen absorbieren die G Infrarotbereich, wodurch sich Treibhauseffekt begründet. [ durch Ozon ist Ursache für d



# *Concept merging*

- Matching the sense? Potential sense disambiguators:
  - Part of speech
  - Subject field
  - Genus (hypernym)
  - Keywords in definitions
  - Synonyms (if in lex resource)
- Characteristics
- Position in ontology