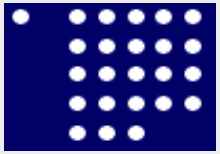
Terminological Data Modeling

Traditional and New Aspects

TKE 2010 – Workshop on ISO/CDB
14th August 2010 – Dublin



Klaus-Dirk Schmitz
Fakultät 03 - ITMK
Fachhochschule Köln
klaus.schmitz@fh-koeln.de

Content

- **General principles of terminology data modeling**
  - Data categories
  - Meta model
  - Modeling principles
- **Practical implementations and improvements**
  - Concept orientation
  - Term autonomy
  - Locale problem
  - (Non-linguistic representations)
- **New aspects and outlook**
- **Conclusion**

Data Modeling

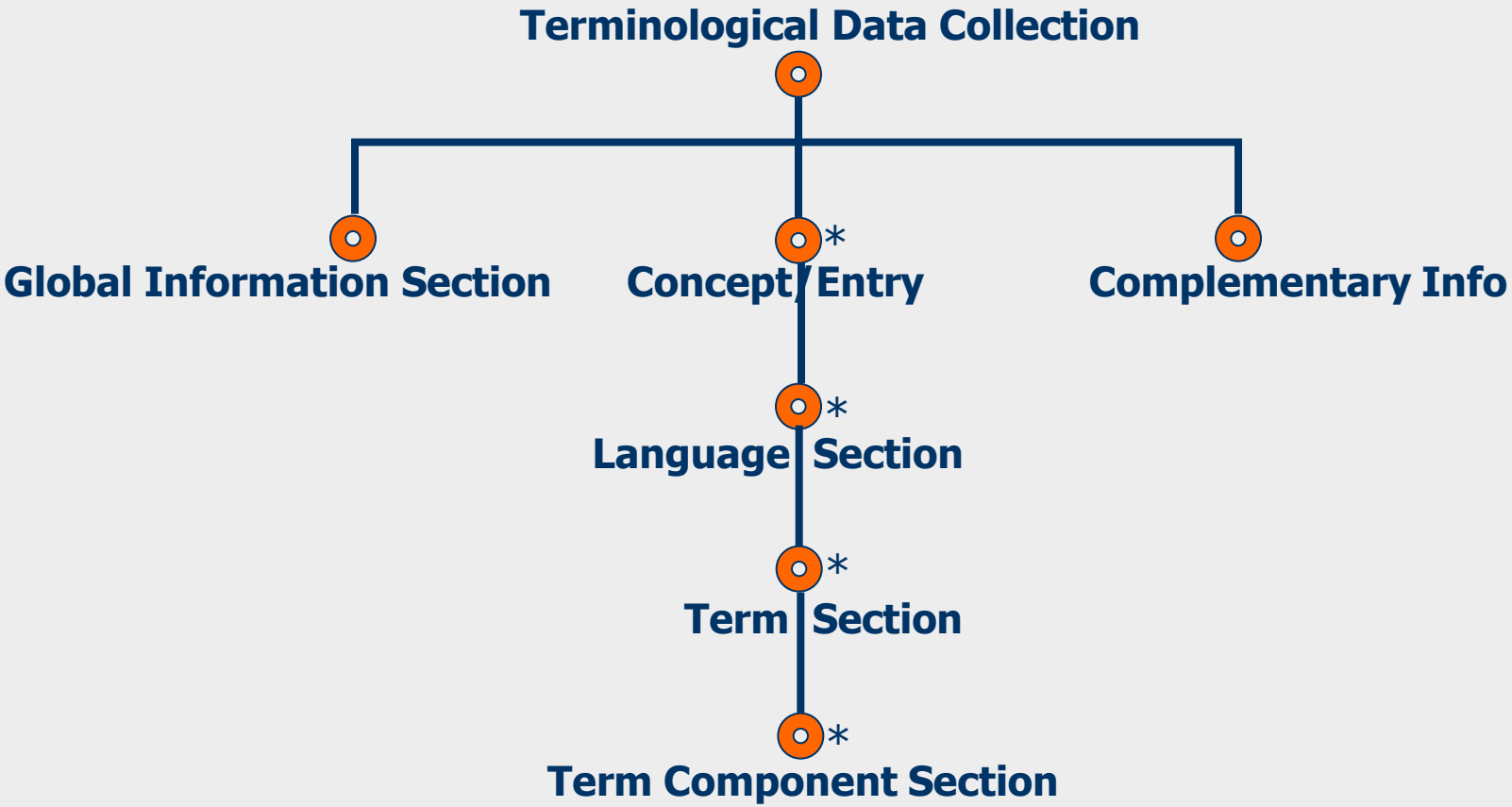
- A data model is an abstract model that describes how data are represented and accessed.
  Data models formally define **data elements** and **relationships among data elements** for a domain of interest.

- Data elements = data categories: ISO 12620 (ISOcat)

- Relationships = meta model: ISO 12200 / ISO 16642 (TMF)

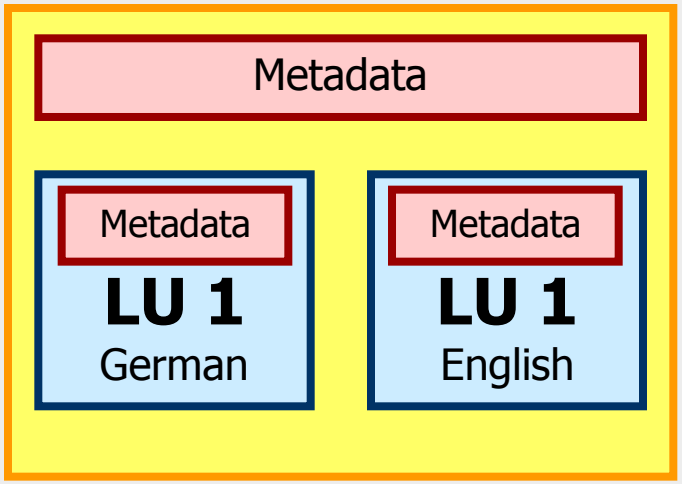
- A **terminological meta model** is an abstraction of data models for terminological applications (data bases, formats etc.) highlighting the common properties of these applications.

Terminological meta model



Argumentation for TMF

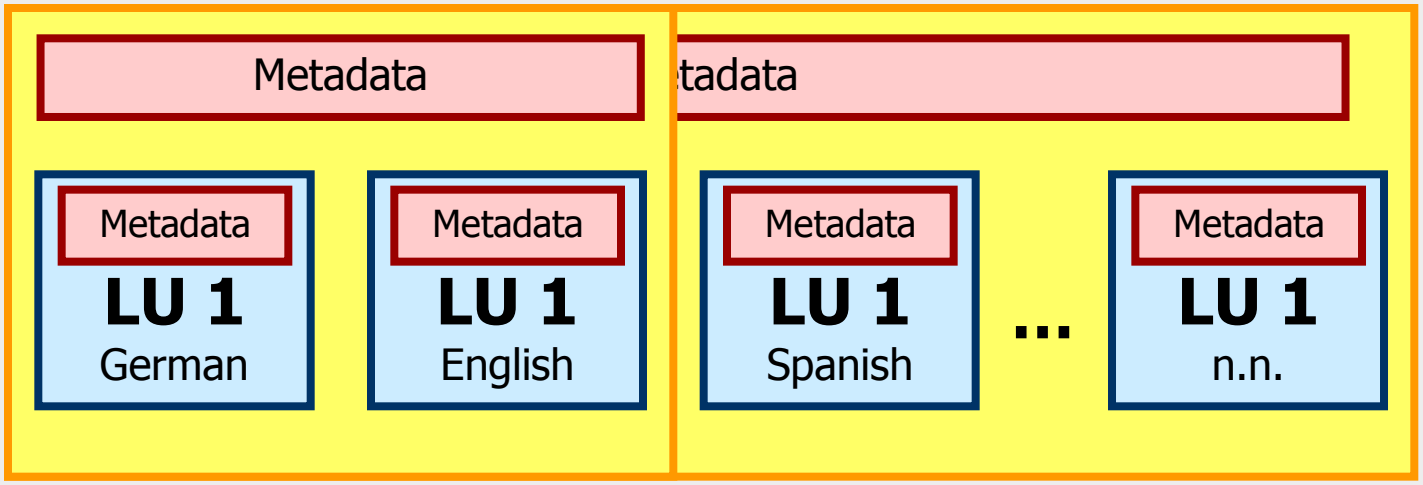
- How does the common **Meta Model** look like?



LU = Linguistic Unit (Term, Phrase, Segment)

Argumentation for TMF

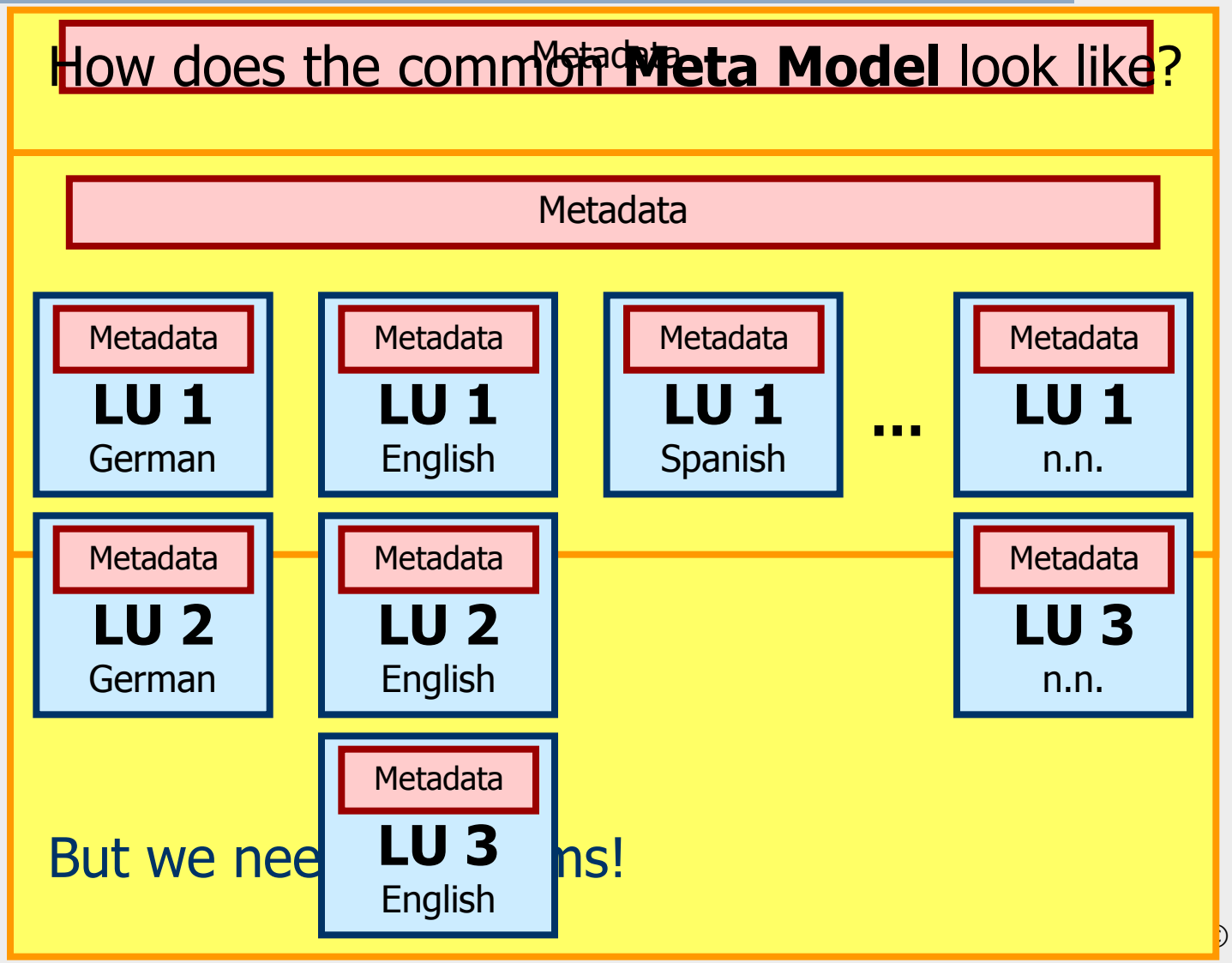
- How does the common **Meta Model** look like?



- But we need multilinguality!

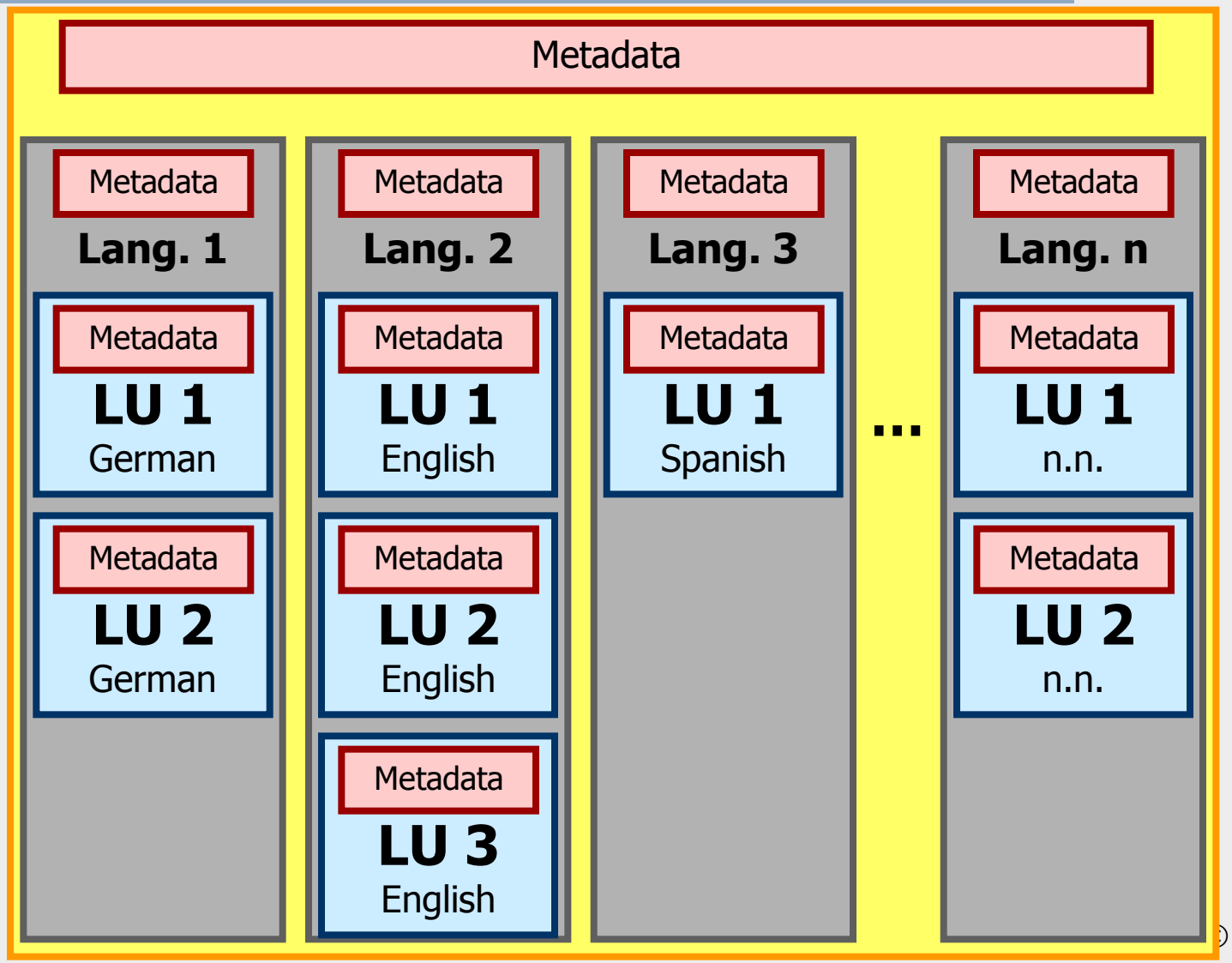
Argumentation for TMF

- How does the common **Meta Model** look like?



- But we nee ms!

Argumentation for TMF



Modeling Principles: Concept Orientation

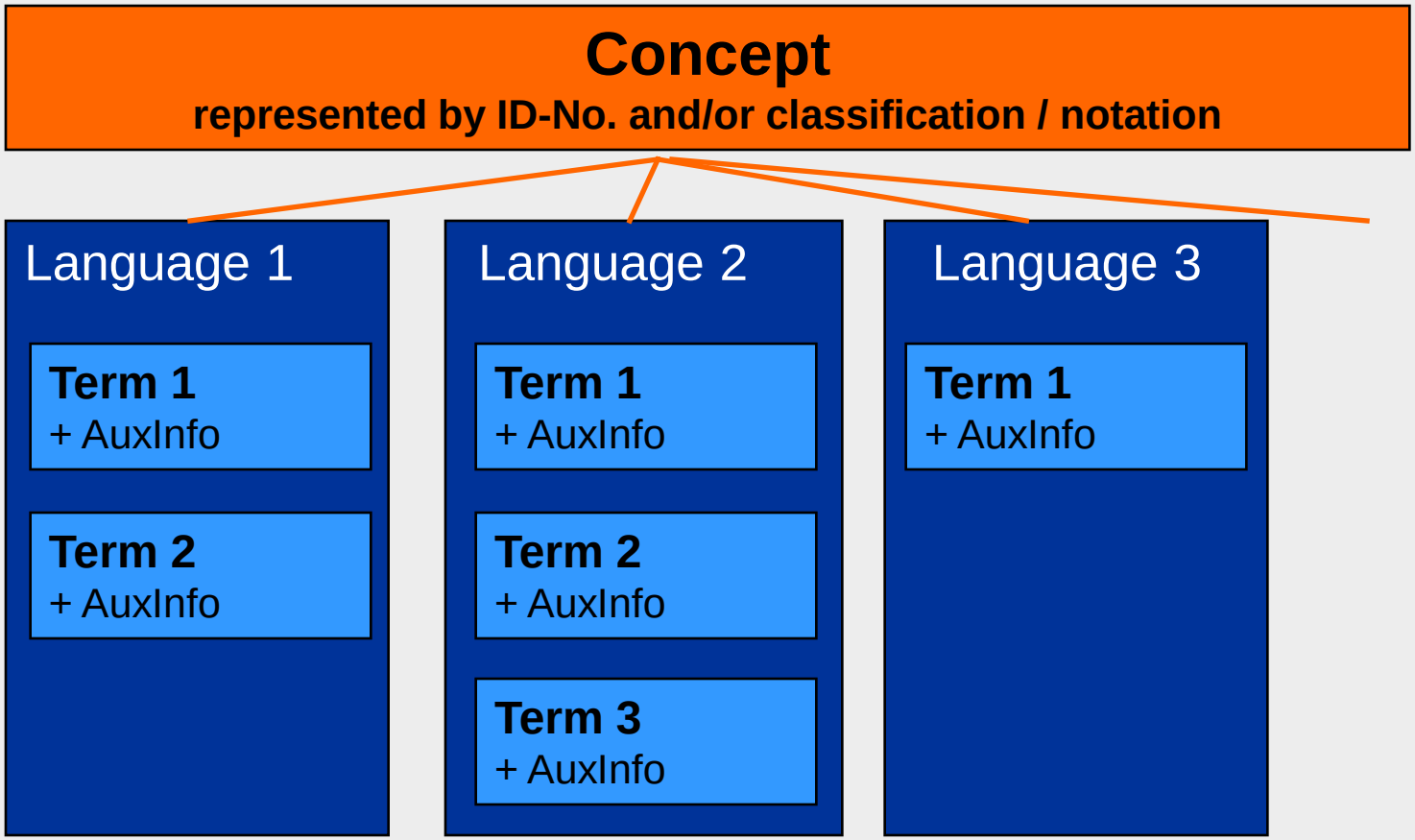
- Terminological data modeling has to follow the principle of **concept orientation**:
  a terminological entry has to contain all terminological data related to one concept (all languages, all terms, all related information).
- The meta model (and ISO 704 + 1087-1) demands this.
- The terminological entry is a logical unit, not a physical one (≠ record).

Modeling Principles: Term Autonomy

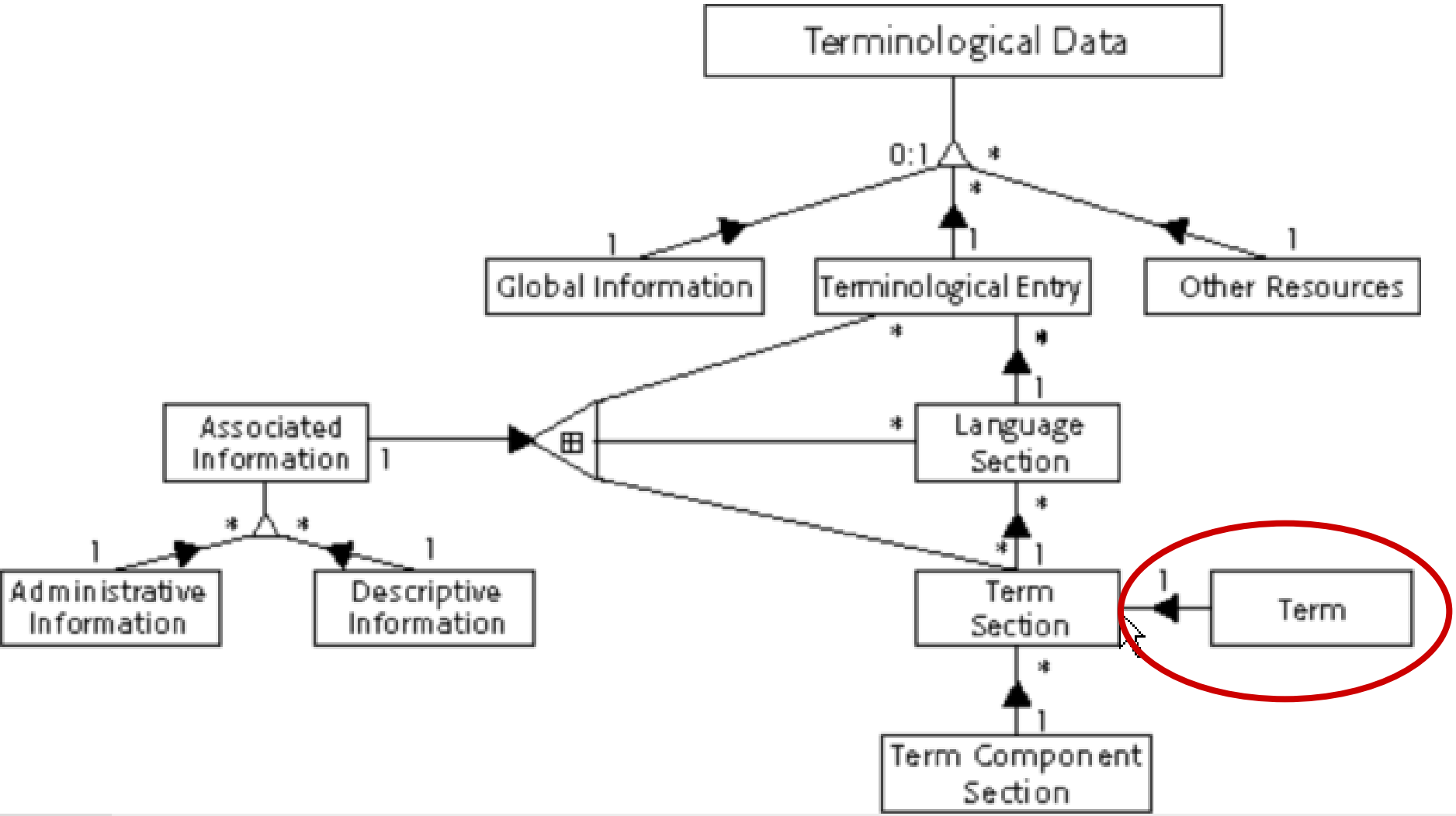
- Terminological data modeling has to follow the principle of **term autonomy**:
  all terms representing the concept (including synonyms, abbreviated forms and orthographic variants) can be retrieved and documented with all necessary term-related data categories.

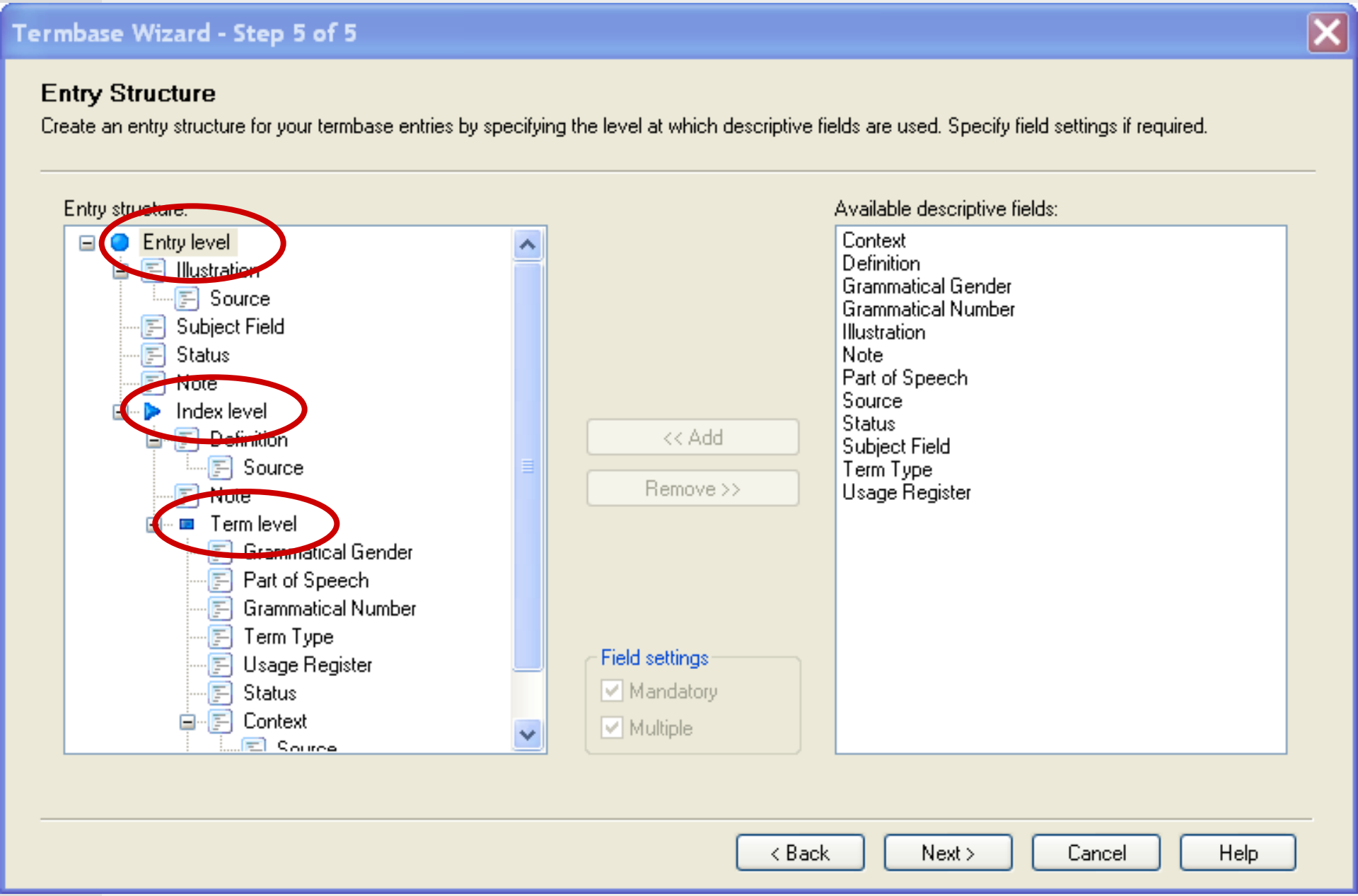
- Terminological meta model: Unlimited number of term sections, but only one term in each section.

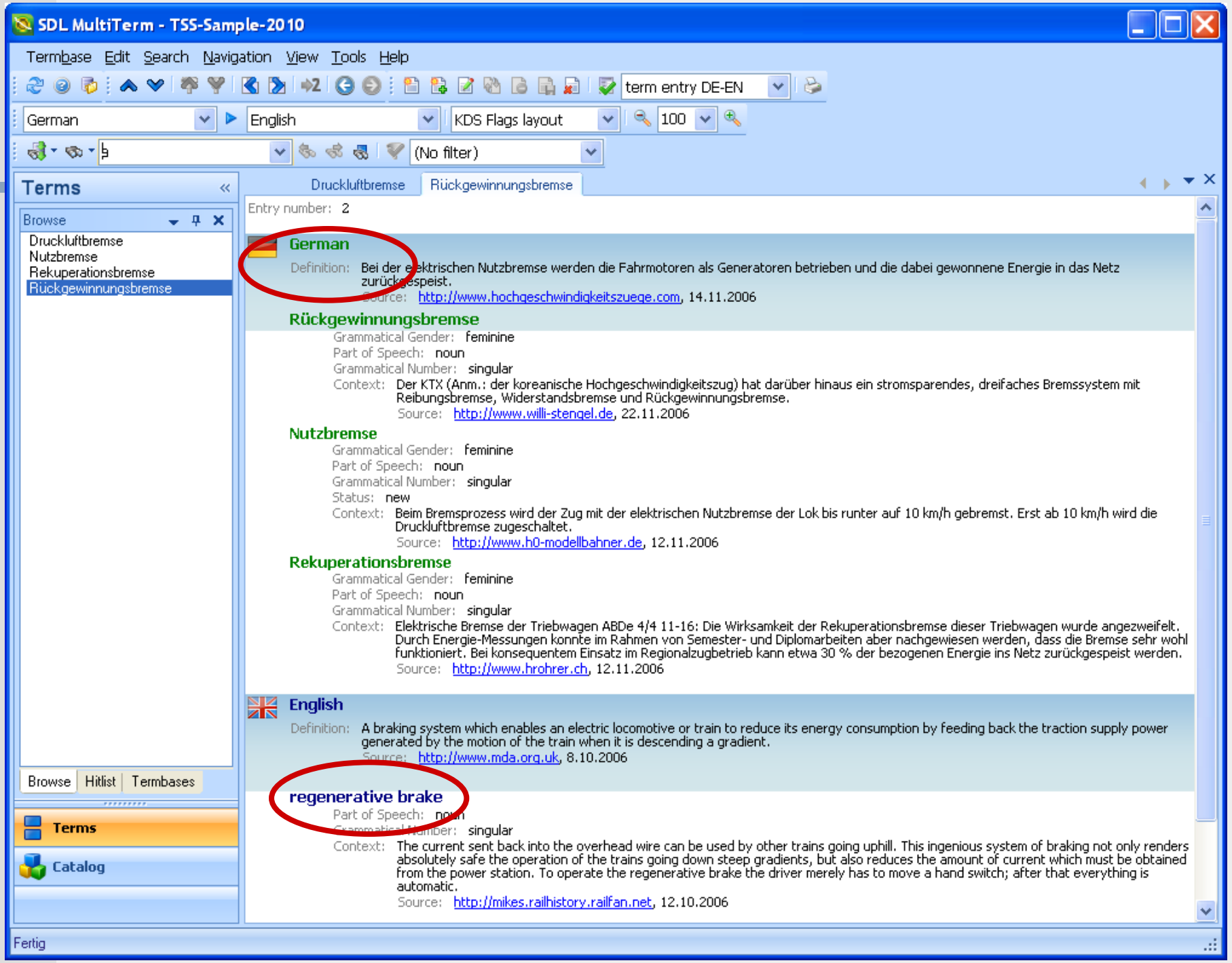
Modeling Principles



Terminological Meta Model (12200)




Termbase Wizard - Step 5 of 5
Entry Structure
Create an entry structure for your termbase entries by specifying the level at which descriptive fields are used. Specify field settings if required.
Entry structure:
Entry level
Illustration
Source
Subject Field
Status
Note
Index level
Definition
Source
Note
Term level
Grammatical Gender
Part of Speech
Grammatical Number
Term Type
Usage Register
Status
Context
Source
<< Add
Remove >>
Field settings
Mandatory
Multiple
Available descriptive fields:
Context
Definition
Grammatical Gender
Grammatical Number
Illustration
Note
Part of Speech
Source
Status
Subject Field
Term Type
Usage Register
< Back
Next >
Cancel
Help


SDL MultiTerm - TSS-Sample-2010
Termbase Edit Search Navigation View Tools Help
term entry DE-EN
German
English
KDS Flags layout
100
(No filter)
Terms
Browse
Druckluftbremse
Nutzbremse
Rekuperationsbremse
Rückgewinnungsbremse
Browse Hitlist Termbases
Terms
Catalog
Druckluftbremse
Rückgewinnungsbremse
Entry number: 2
German
Definition: Bei der elektrischen Nutzbremse werden die Fahrmotoren als Generatoren betrieben und die dabei gewonnene Energie in das Netz zurückgespeist.
Source: http://www.hochgeschwindigkeitszuege.com, 14.11.2006
Rückgewinnungsbremse
Grammatical Gender: feminine
Part of Speech: noun
Grammatical Number: singular
Context: Der KTX (Anm.: der koreanische Hochgeschwindigkeitszug) hat darüber hinaus ein stromsparendes, dreifaches Bremssystem mit Reibungsbremse, Widerstandsbremse und Rückgewinnungsbremse.
Source: http://www.willi-stengel.de, 22.11.2006
Nutzbremse
Grammatical Gender: feminine
Part of Speech: noun
Grammatical Number: singular
Status: new
Context: Beim Bremsprozess wird der Zug mit der elektrischen Nutzbremse der Lok bis runter auf 10 km/h gebremst. Erst ab 10 km/h wird die Druckluftbremse zugeschaltet.
Source: http://www.h0-modellbahner.de, 12.11.2006
Rekuperationsbremse
Grammatical Gender: feminine
Part of Speech: noun
Grammatical Number: singular
Context: Elektrische Bremse der Triebwagen ABDe 4/4 11-16: Die Wirksamkeit der Rekuperationsbremse dieser Triebwagen wurde angezweifelt. Durch Energie-Messungen konnte im Rahmen von Semester- und Diplomarbeiten aber nachgewiesen werden, dass die Bremse sehr wohl funktioniert. Bei konsequentem Einsatz im Regionalzugbetrieb kann etwa 30 % der bezogenen Energie ins Netz zurückgespeist werden.
Source: http://www.hrohrer.ch, 12.11.2006
English
Definition: A braking system which enables an electric locomotive or train to reduce its energy consumption by feeding back the traction supply power generated by the motion of the train when it is descending a gradient.
Source: http://www.mda.org.uk, 8.10.2006
regenerative brake
Part of Speech: noun
Grammatical Number: singular
Context: The current sent back into the overhead wire can be used by other trains going uphill. This ingenious system of braking not only renders absolutely safe the operation of the trains going down steep gradients, but also reduces the amount of current which must be obtained from the power station. To operate the regenerative brake the driver merely has to move a hand switch; after that everything is automatic.
Source: http://mikes.railhistory.railfan.net, 12.10.2006
Fertig

Practical Implications of Terminological Data Modeling

Language vs. locale:

- In computing and software localization, a **locale** is a set of parameters that defines the user's country, language, key board layout, writing system, date format etc.
- A locale identifier consists of at least a language ID and a region/country ID.

Practical Implications of Terminological Data Modeling

Language vs. locale:

- Locale can be used in terminology management to describe that different terms within the "same" language are used in different geographical regions to express the same concept.

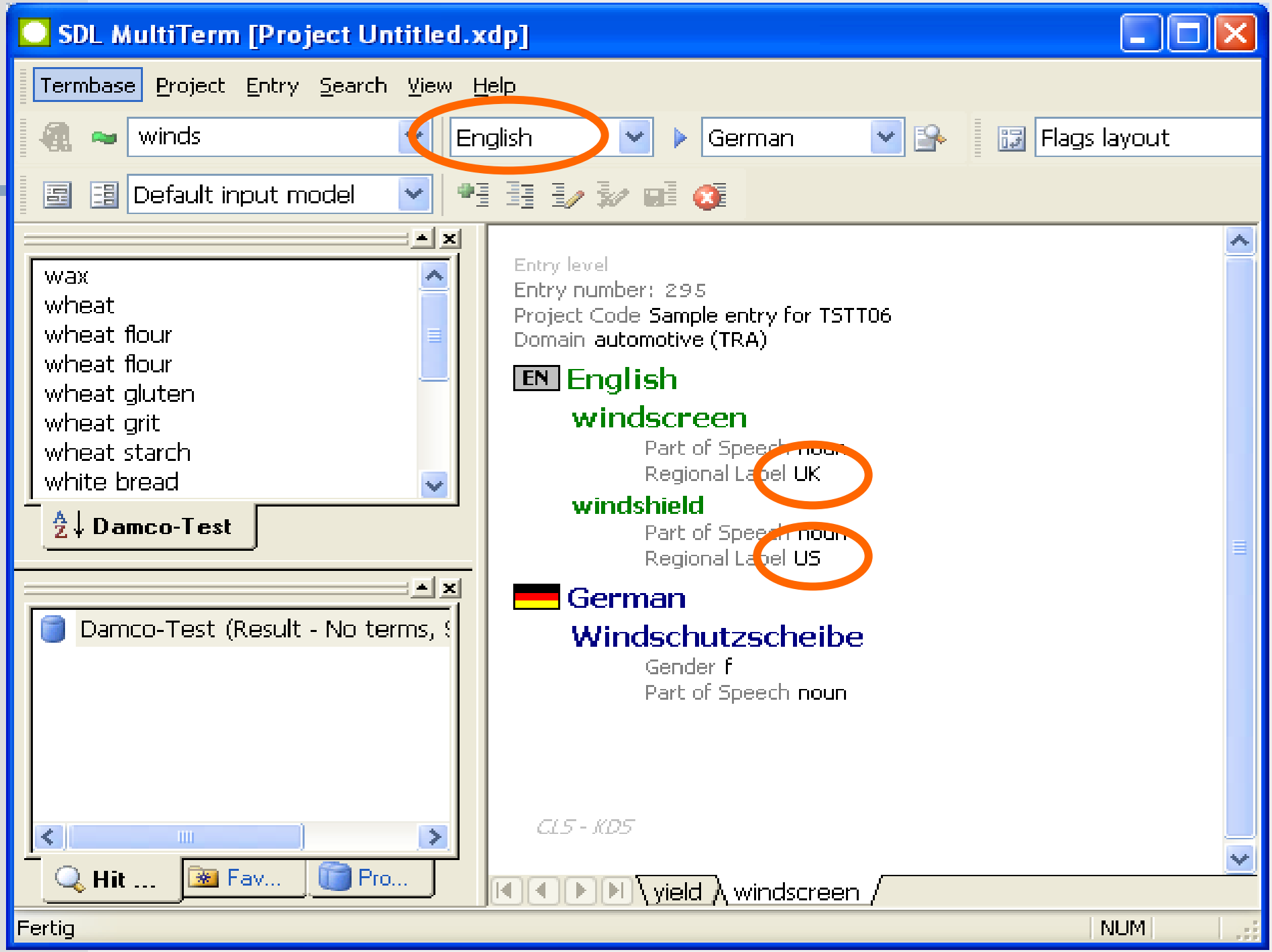
- **Example:**

  English: **windscreen <UK>**
  English: **windshield <US>**

Practical Implications of Terminological Data Modeling

Language vs. locale:

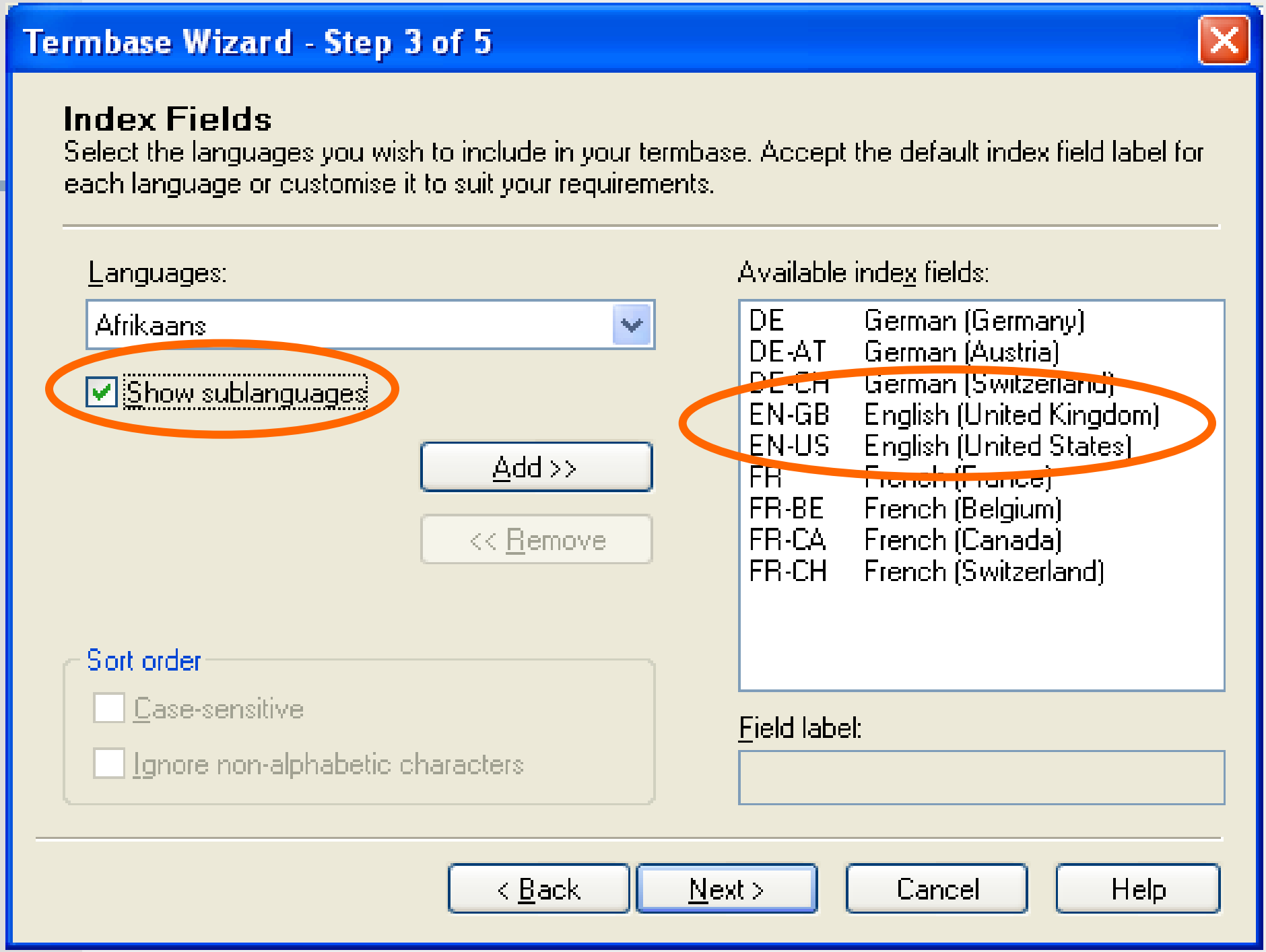
- **Traditional way:**
  - use the language level as in the meta model
  - manage regional variants of the term as synonyms
  - use data category for geographical usage (ISO country code) as attributes to the term

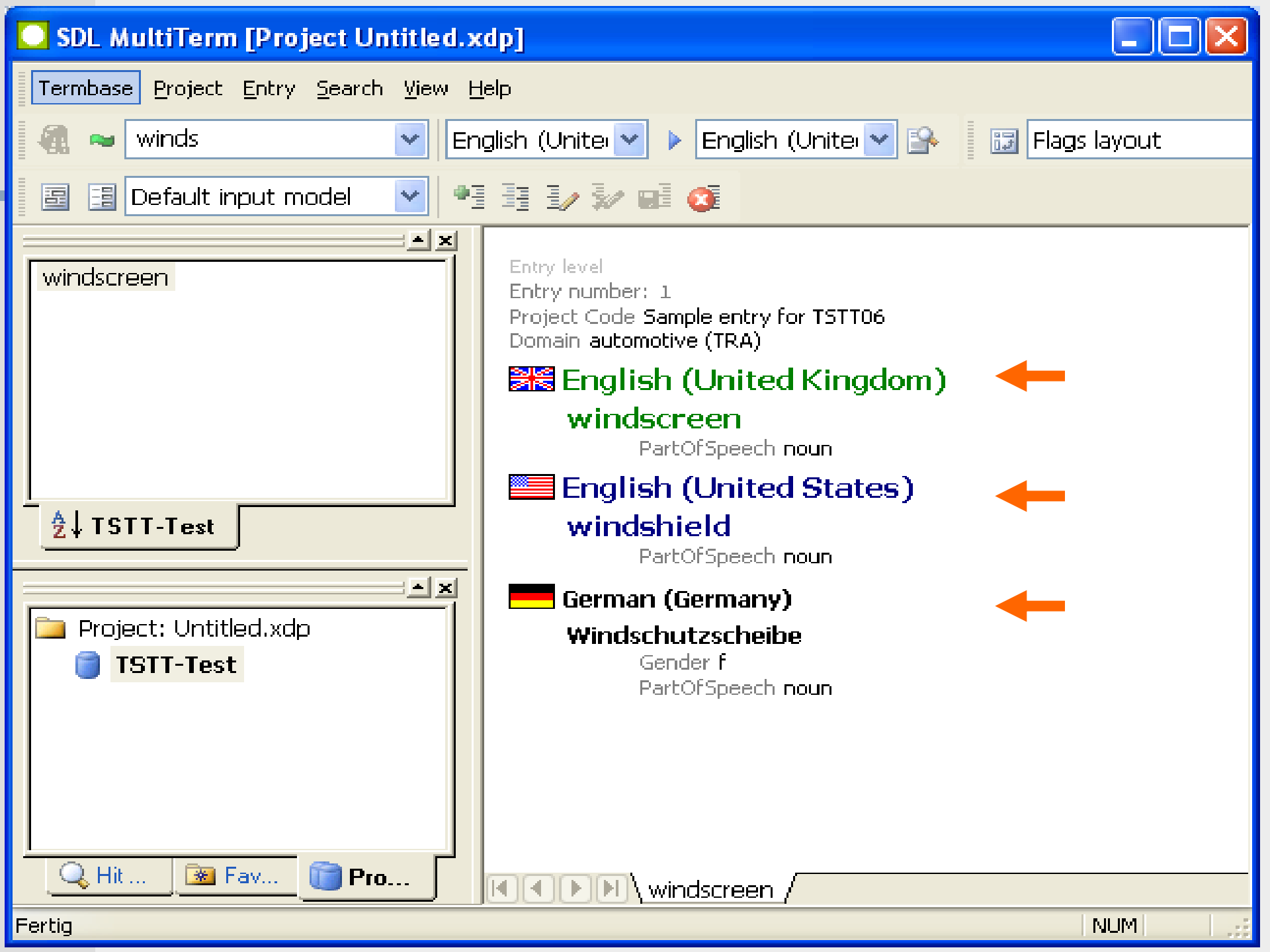

SDL MultiTerm [Project Untitled.xdp]
Termbase
Project
Entry
Search
View
Help
winds
English
German
Flags layout
Default input model
wax
wheat
wheat flour
wheat flour
wheat gluten
wheat grit
wheat starch
white bread
Damco-Test
Damco-Test (Result - No terms,
Hit ...
Fav...
Pro...
Entry level
Entry number: 295
Project Code Sample entry for TSTT06
Domain automotive (TRA)
EN English
windscreen
Part of Speech noun
Regional Label UK
windshield
Part of Speech noun
Regional Label US
German
Windschutzscheibe
Gender f
Part of Speech noun
CLS - KDS
yield
windscreen
Fertig
NUM

Practical Implications of Terminological Data Modeling

Language vs. locale:

- **Other approach** (e.g. as in MultiTerm)**:**
  - index level (=language level) can be specified by a combination of language and country (=locale)
  - terms of a specific locale are managed in separate language/locale section
  - searching for terms is difficult if many terms are the same in all regions (as for US and UK English)
  - but allows for separate definitions / figures for each locale


Termbase Wizard - Step 3 of 5
Index Fields
Select the languages you wish to include in your termbase. Accept the default index field label for each language or customise it to suit your requirements.
Languages:
Afrikaans
Show sublanguages
Add >>
<< Remove
Sort order
Case-sensitive
Ignore non-alphabetic characters
Available index fields:
DE German (Germany)
DE-AT German (Austria)
DE-CH German (Switzerland)
EN-GB English (United Kingdom)
EN-US English (United States)
FR French (France)
FR-BE French (Belgium)
FR-CA French (Canada)
FR-CH French (Switzerland)
Field label:
< Back
Next >
Cancel
Help


SDL MultiTerm [Project Untitled.xdp]
Termbase
Project
Entry
Search
View
Help
winds
English (Unite
English (Unite
Flags layout
Default input model
windscreen
TSTT-Test
Project: Untitled.xdp
TSTT-Test
Hit ...
Fav...
Pro...
Entry level
Entry number: 1
Project Code Sample entry for TSTT06
Domain automotive (TRA)
English (United Kingdom)
windscreen
PartOfSpeech noun
English (United States)
windshield
PartOfSpeech noun
German (Germany)
Windschutzscheibe
Gender f
PartOfSpeech noun
windscreen
Fertig
NUM

SDL MultiTerm [Project Untitled.xdp]

Termbase Project Entry Search View Help

winds | English (Unite | English (Unite | Flags layout

Default input model

plug
windscreen
windscreen

TSTT-Test

Project: Untitled.xdp
TSTT-Test

Hit ... Fav... Pro...

Entry level
Entry number: 4
Project Code Sample entry for TSTT06

English (United Kingdom)
Non-ling. Repres.

plug
PartOfSpeech noun

English (United States)
Non-ling. Repres.

plug
PartOfSpeech noun

French (Switzerland)
Non-ling. Repres.

windscreen | windscreen | plug

Fertig NUM

SDL MultiTerm [Project Untitled.xdp]

Termbase Project Entry Search View Help

winds | English (Unite | English (Unite | Flags layout

Default input model

plug
windscreen
windscreen

TSTT-Test

Project: Untitled.xdp
TSTT-Test

Hit ... Fav... Pro...

French (Switzerland)

Non-ling. Repres.

fiche

Gender f

German (Germany)

Non-ling. Repres.

Stecker

Gender m
PartOfSpeech noun

German (Switzerland)

Non-ling. Repres.

Stecker

windscreen | windscreen | plug

Fertig

NUM

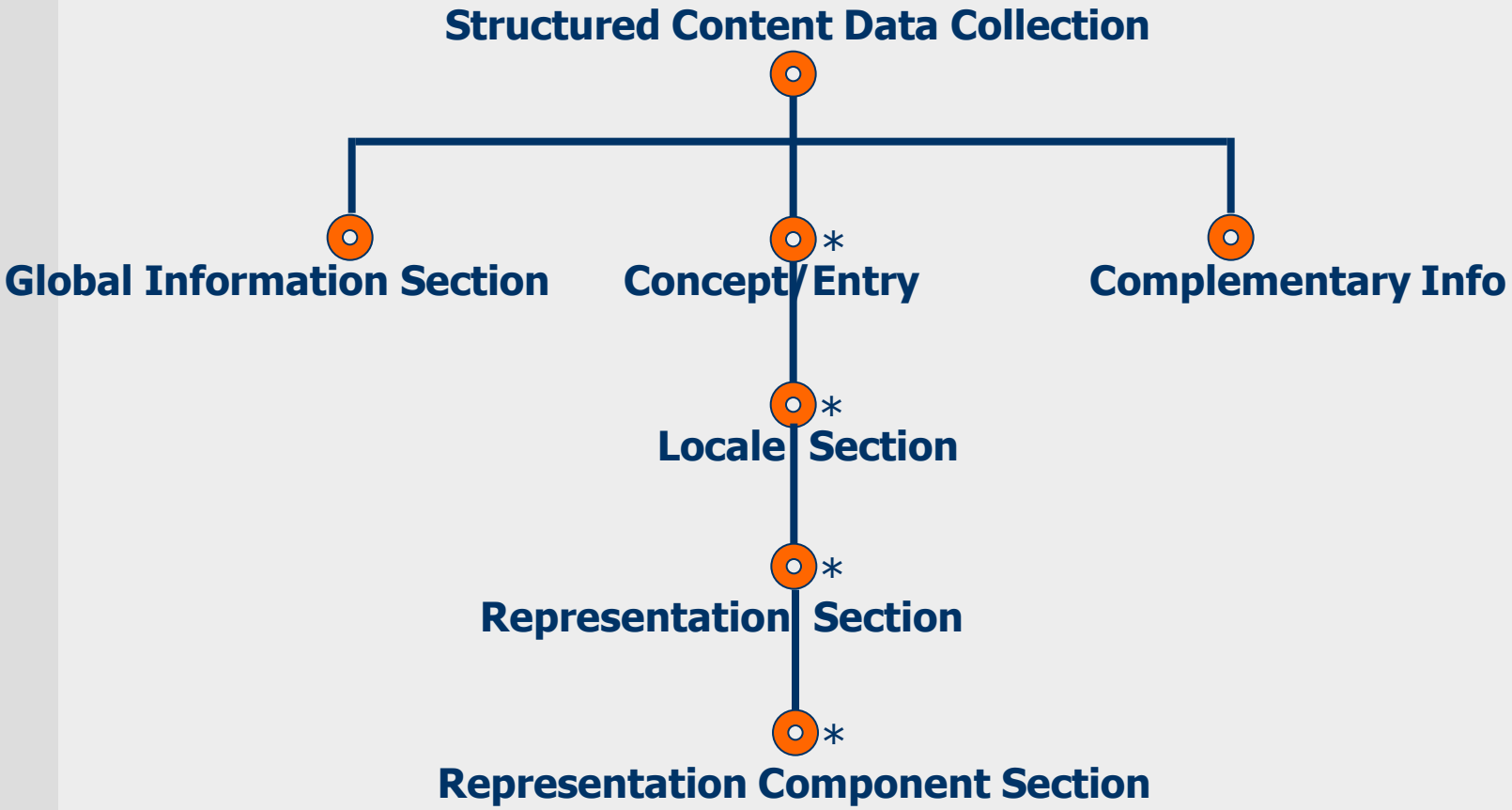
Practical Implications of Terminological Data Modeling

Language vs. locale:

- Such a locale (language+country) solution will also support data modeling for distributed / federated terminological applications.
- Instead of a region, an **organization** or a **company** can be seen as a locale, e.g. ISO, ANSI, BSI as locales
- Application: ISO/CDB for ISO **and** its member bodies

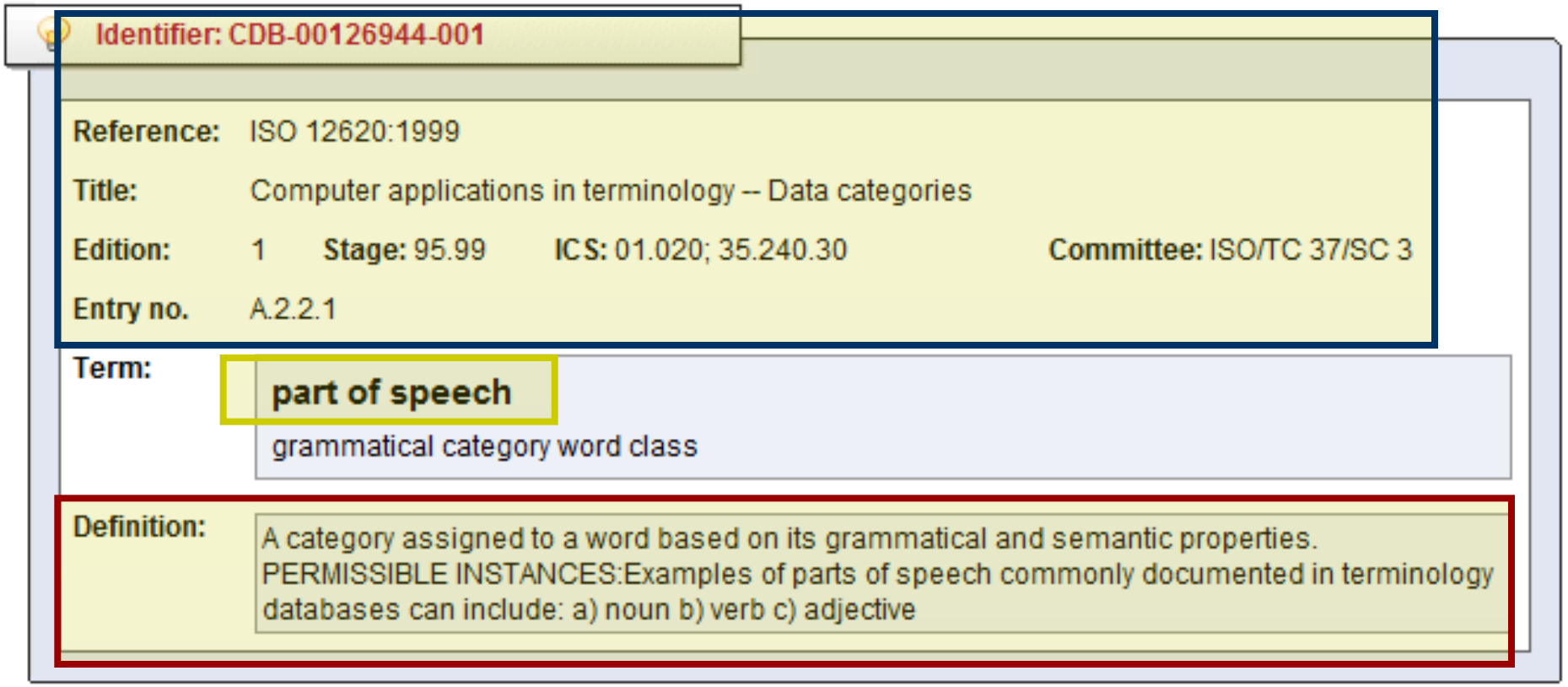
- Do we need a Locale Section instead of the Language Section in our meta model? **What are the consequences?**

Structured Content Meta Model



ISO/CDB

- **Sample entry:** not so many term-related datCats! term autonomy? what is locale-specific?



ISO/CDB

Identifier: CDB-00126954-001

Reference: ISO 12620:1999

Title: Computer applications in terminology -- Data categories

Edition: 1 **Stage:** 95.99 **ICS:** 01.020; 35.240.30 **Committee:** ISO/TC 37/SC 3

Entry no. A.2.2.2

Term: **grammatical gender**

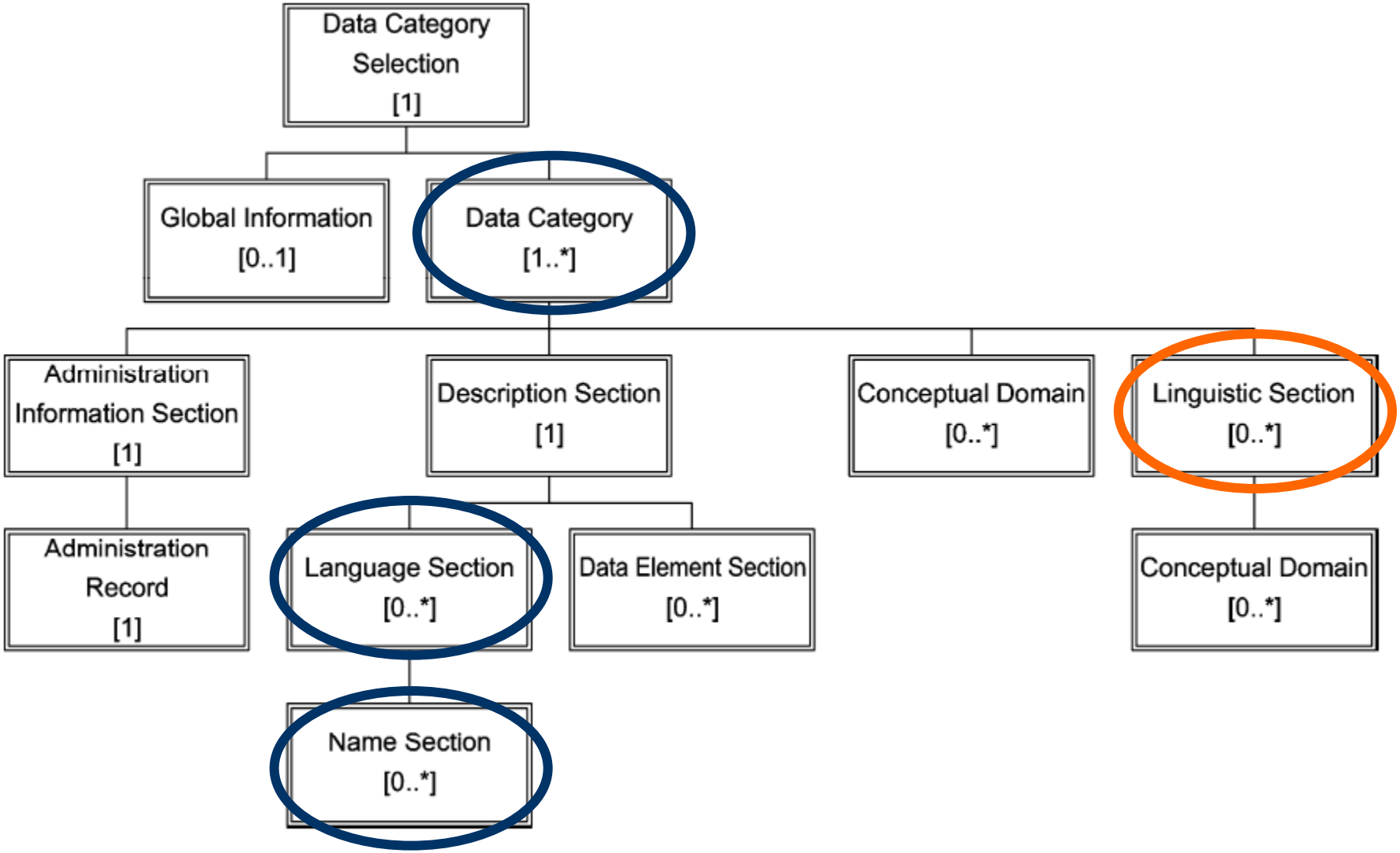
Definition: A grammatical category that indicates grammatical relationships between words in sentences. PERMISSIBLE INSTANCES: Types of grammatical gender commonly documented in terminology databases include: a) masculine b) feminine c) neuter d) other

Example(s):

1. In French, vie (life) is feminine and is used with feminine articles such as la, the feminine pronoun elle, and feminine adjective endings, e.g., une vie longue.

Note: The concept of gender varies from language to language and is not a universal feature of all languages.

ISOcat data model



Conclusion

- ISO standards provide a set of profound principles, methods, and formalisms for terminological data modeling.
- Improvements in the fields shown will make ISO results more suitable for advanced terminological applications.
- The proposed adaptations / modifications will enlarge the scope of applicability to data modeling for other kinds of structured content, particularly for multilingual and multicultural applications.

Thank you for your attention
Questions ?

Prof. Dr. Klaus-Dirk Schmitz

Fachhochschule Köln
Fakultät 03 - ITMK/IIM

Mainzer Str. 5

50678 Köln

klaus.schmitz@fh-koeln.de