

Esther Scheven

ISO 25964 and GND

Contents

GND in general

ISO 25964 in general

Interoperability in general

Data model, ISO 25964-1, and GND

ISO 25964 Part 2

Recommendations for Mapping

Interoperability in GND

How to obtain ISO 25954

Information about GND

Gemeinsame Normdatei – GND (Integrated Authority File)

- Started April 2012; licence: [Creative Commons Zero](#)
- One file: four former files:
 - Gemeinsame Körperschaftsdatei (GKD) (file for corporate bodies)
 - Personennormdatei (PND) (file for person names)
 - Schlagwortnormdatei (SWD) (file for subject headings)
 - Normdatei für die Einheitssachtitel der Musik (EST-Datei) (file for music uniform titles)
- Volume: ca. 10 mio records (ca. 7,5 mio persons; 1,7 mio corporate bodies + conferences; 0,3 mio places; 0,2 mio topical terms; 0,2 mio uniform titles)
- Cooperative file of the German library networks and other users
- Hybride file: thesaurus for topical terms; name authority file for the other entities
- Further information: http://www.dnb.de/EN/Standardisierung/GND/gnd_node.html
www.dnb.de/gnd (in German)

ISO 25964

ISO 25964: Thesauri and interoperability with other vocabularies

- Part 1: Thesauri for information retrieval
- Part 2: Interoperability with other vocabularies

Part 1 contains guidelines for establishing monolingual and multilingual thesauri.

Part 2 deals with mappings between thesauri and other types of vocabularies for information retrieval.

Part 1: published 2011

Part 2: published March 2013

Developed by an international working group 2008-2013

Project leader: Stella Dextre Clarke, Great Britain

Predecessors of ISO 25964

- ISO 2788-1986 Guidelines for the establishment and development of monolingual thesauri
= BS 5723:1987
- ISO 5964-1985 Guidelines for the establishment and development of multilingual thesauri
= BS 6723:1985

BS 8723: Structured vocabularies for information retrieval – Guide

Part 1: Definitions, symbols and abbreviations

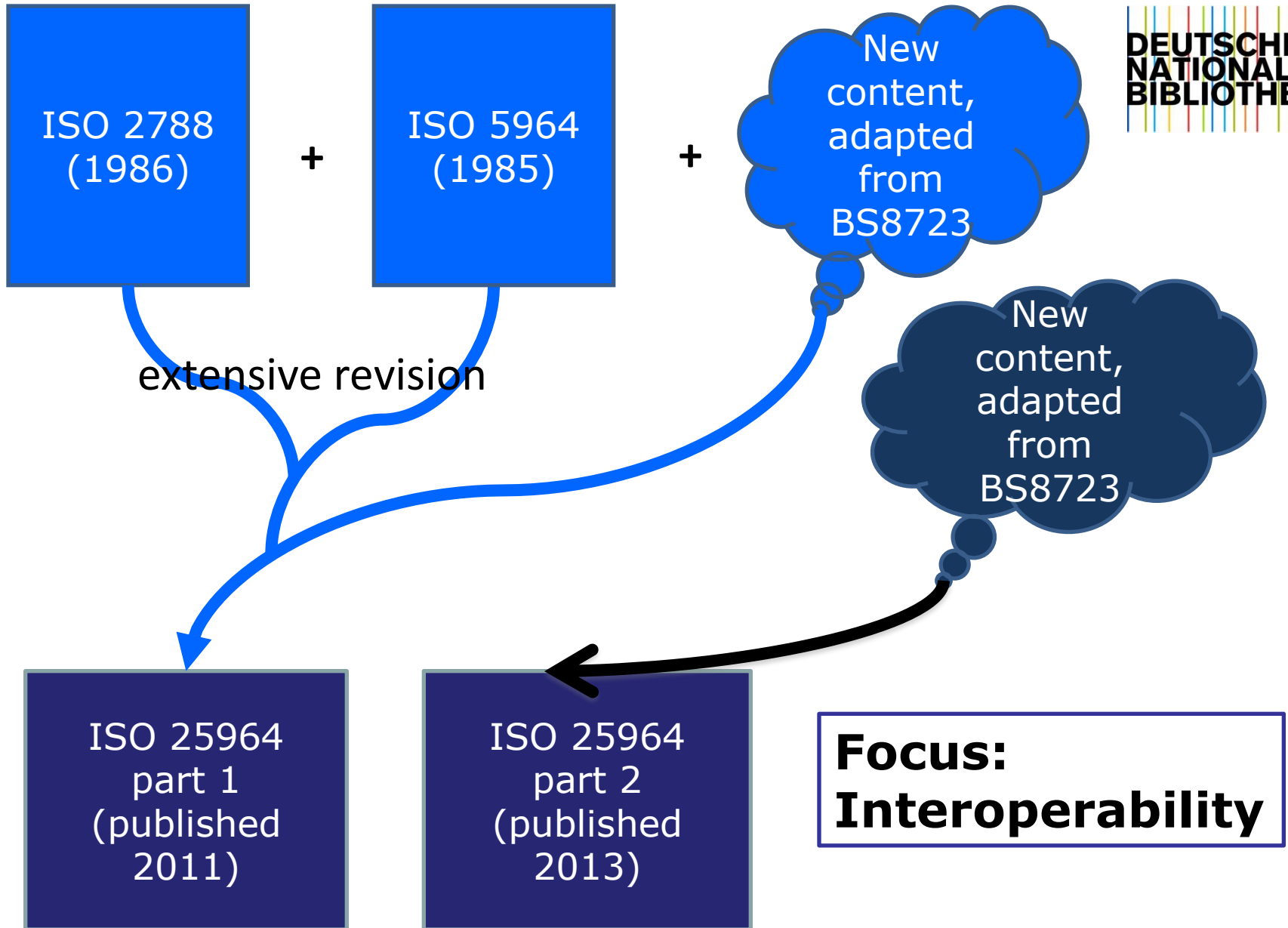
Part 2: Thesauri

Part 3: Vocabularies other than thesauri

Part 4: Interoperability between vocabularies

Part 5: Exchange formats and protocols for interoperability

published 2005-2008



What does „interoperability“ mean?

Definition: The ability for two or more systems to exchange information and to use the information that has been exchanged.

Concerning thesauri and knowledge organisation systems (KOS) there are two ways to achieve interoperability:

- Presenting data in a standard way to enable import and re-use in other systems
- Providing mappings between the terms/concepts of one KOS and those of another

How does ISO 25964 support interoperability?

Part 1:

- Normative guidelines for establishing thesauri
- Normative guidelines how to handle multilinguality
- Data model for creating a thesaurus in UML
(UML: Unified Modeling Language)

Part 2:

- Guidelines for mappings
 - Structural models
 - Specification of mapping types
 - How to handle pre-coordination
 - Mapping to other vocabularies, e.g. classification schemes
 - How to handle mapping data

ISO 25964-1

- Exhaustive definitions
- Concept (ThesaurusConcept) – Term (ThesaurusTerm)
How to create the terms
- Relationships: equivalence, hierarchical, associative for monolingual and multilingual thesauri
- Display of thesauri (alphabetical, hierarchical etc.)
- Data management
- Appendix with examples

Example: GND
Hierarchical display
<http://melvil.dnb.de/gnd>

OB3 [Wissenschaft](#) ▾ ▶
 OB2 [Naturwissenschaften](#) ▾ ▶
 OB1 [Chemie](#)
 [Analytische Chemie](#) ▾ ▶
 [Angewandte Chemie](#) ▾
 [Anorganische Chemie](#) ▾
 [Bauchemie](#)
 Geochemie ▶
 UB1 [Anorganische Geochemie](#)
 UB1 [Biogeochemie](#)
 UB1 [Hydrogeochemie](#) ▲
 UB1 [Isotopengeochemie](#)
 UB1 [Mineralchemie](#) ▲
 UB1 [Organische Geochemie](#)
 UB1 [Umweltgeochemie](#) ▲
 [Gerichtliche Chemie](#)
 [Grüne Chemie](#)
 [Klinische Chemie](#) ▶
 [Kosmochemie](#) ▶
 [Metallorganische Chemie](#)
 [Militärchemie](#)

Data model for thesauri

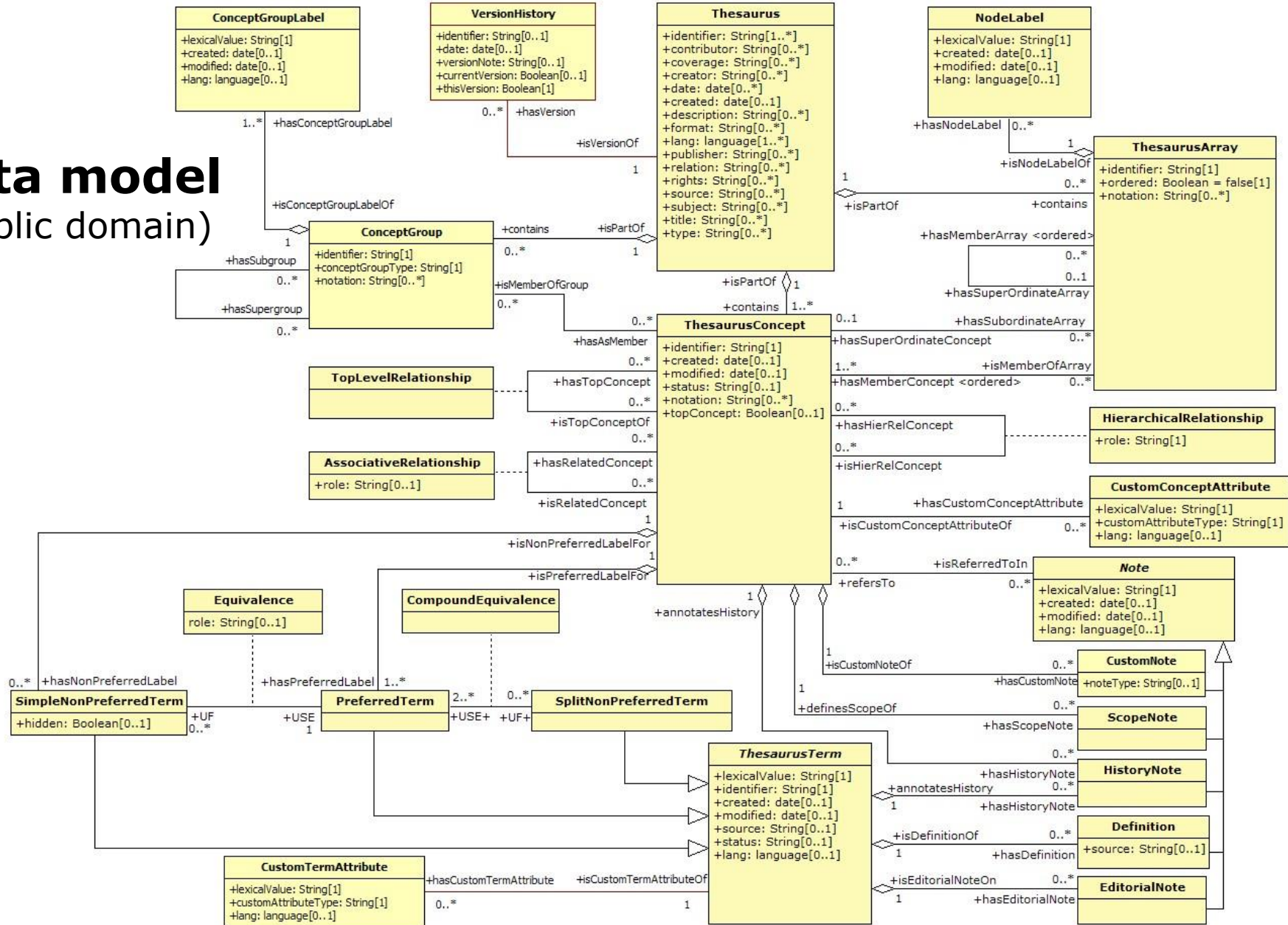
- Mainly aimed at IT people (not everybody has to understand it).
- Can be applied to all types of thesauri.
- Facilitates data exchange.

The model represents

- Concepts of a thesaurus (ThesaurusConcept)
- Use of identifiers (*in GND: every record has an identifier*)
- Traditional hierarchical and associative relationships and user assigned relator types between concepts
- Different kinds of notes like "scope notes", "history notes", "editorial notes", "definitions ", ...
- Preferred terms (descriptors) and variant terms
- Concept arrays and node labels
- Concept groups

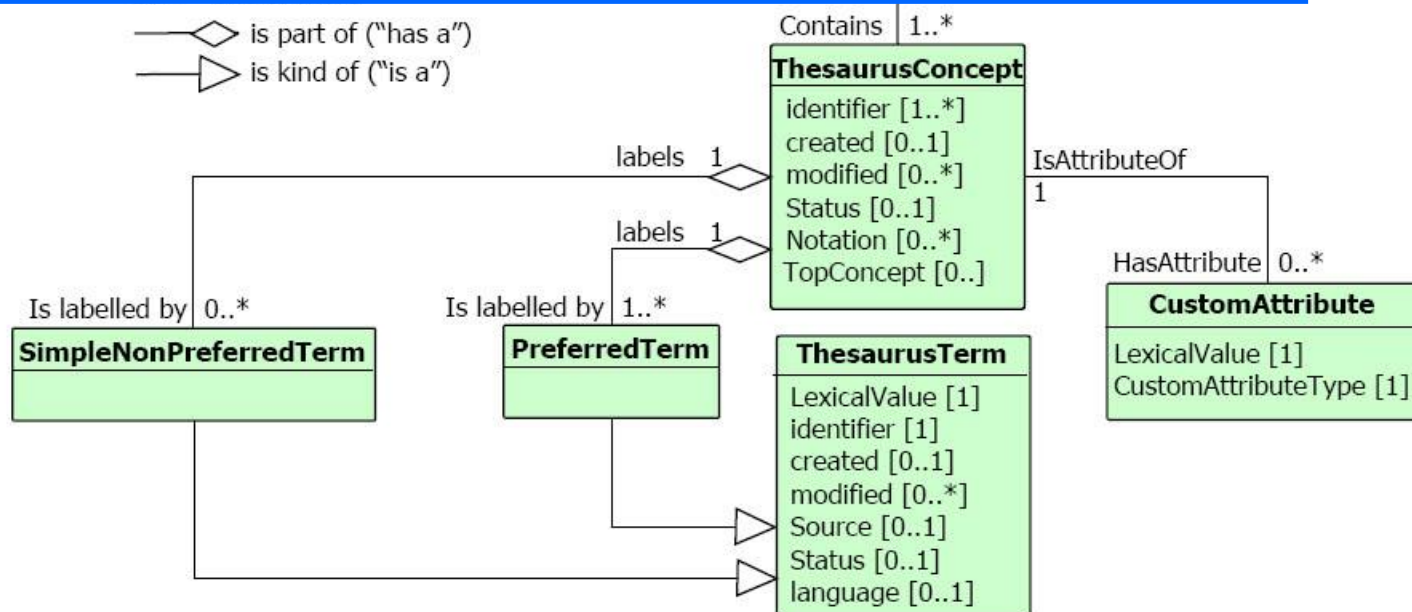
Compatible with SKOS, but with extensions.

Data model (public domain)



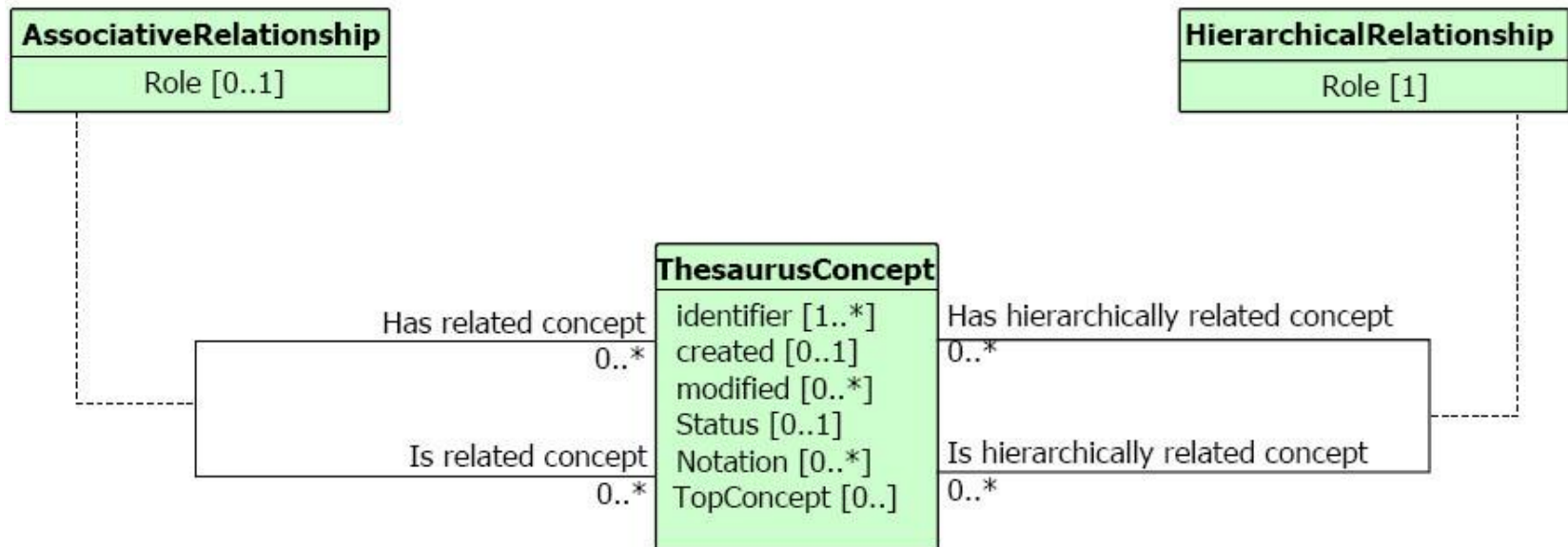
Concept (ThesaurusConcept) and Term (ThesaurusTerm)

- A thesaurus contains concepts (ThesaurusConcept)
- Concepts are represented by terms of natural language (ThesaurusTerm)
- Each concept has one descriptor (PreferredTerm)
- Each concept can have many alternative terms (SimpleNonPreferredTerm); it is possible to specify the type (*realized in GND*)



Relationships between concepts (ThesaurusConcept)

- Hierarchical relationships can be labelled generally (BT/NT) or specifically (BTG/NTG, BTP/NTP, BTI/NTI) *(in GND specifically marked by relator codes: obge, obpa, obin)*
- Associative relationships can be labelled generally (RT) or specifically according to their role (means/purpose, cause/effect, ...) *(in GND specifically marked by relator codes)*



Relationships in GND

Expressed by four-letter relator codes
e.g. general geographical relationship: "geoa"

internal system

006 <http://d-nb.info/gnd/4030449-8>
 150 Kibbuz
 450 Kibbutz
 450 Kibutz
 450 Kibbuzim
 550 !043214762!Gemeinschaftssiedlung\$4obge
 551 !040278085!/srael\$4geoa
 667 Beispiel in RSWK 3. Aufl.
 670 M

Marking the relationships
between entities (Concepts)

Marking the types of variant
terms of one concept
e.g. "abku" for "abbreviation"
or "navo" for "fuller name"

More important for individual names: e.g. building <-> architect

List of relator codes in GND

GND: Liste der Codierungen in Unterfeld \$4¹ – alphabetisch nach Code

Stand: 01. Dezember 2011, überarbeitet: 03. Mai 2012², 13. Juli 2012³

Nr.	Code ^{4 5}	Relation	MARCTerm
1.	abku	Abkuerzung	Abkuerzung
2.	adel	Adelstitel	Adelstitel
3.	adre	Adressat	Adressat
4.	adue	administrative Ueberordnung	Ueberordnung
5.	affi	Affiliation	Affiliation
6.	akad	Akademischer Grad	Akademischer Grad
7.	akti	Taetigkeitsbereich	Taetigkeitsbereich
8.	anla	Anlass	Anlass
9.	anno	Annotator	Annotator
10.	arch	Architekt	Architekt
11.	arra	Arrangeur	Arrangeur
12.	aust	Aussteller	Aussteller
13.	aut1	Verfasserschaft, erste	Verfasserschaft1
14.	auta	Verfasserschaft	Verfasserschaft
15.	autf	Fiktiver Verfasser	Fiktiver Verfasser
16.	autg	Verfasser, zugeschrieben	Zugeschriebener Verfasser
17.	autw	Verfasser, zweifelhaft	Zweifelhafter Verfasser
18.	autz	Verfasser, zitiert	Ziterter Verfasser
19.	bauh	Bauherr	Bauherr
20.	bear	Bearbeiter	Bearbeiter
21.	befr	Besitzer, fruеherer	Fruеherer Besitzer
22.	berc	Beruf, charakteristisch	Charakteristischer Beruf
23.	beru	Beruf	Beruf

8 records found for: **Psychology**

Click on the [number] to display the records indexed with that descriptor in unesdoc/unesbib.

Term: **Attitudes** [435]

Preferred Term

MT 4.10 Psychology

NT Attitude change [321]

NT Intolerance [80]

NT Prejudice [64]

NT Tolerance [416]

Term: **Behaviour** [429]

Preferred Term

MT 4.10 Psychology

UF Human behaviour

NT Emotions [102]

NT Lateral dominance [5]

NT Psychological effects [281]

....NT2 Mental stress [69]

NT Sensorimotor activities [13]

....NT2 Perception [157]

....NT2 Visualization [30]

NT Sexual behaviour [359]

NT Unconscious [9]

....NT2 Sleep [17]

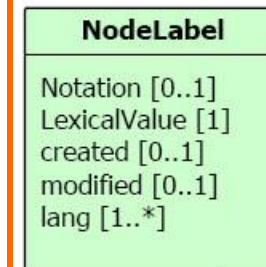
Non Preferred Term

Hierarchical relationships

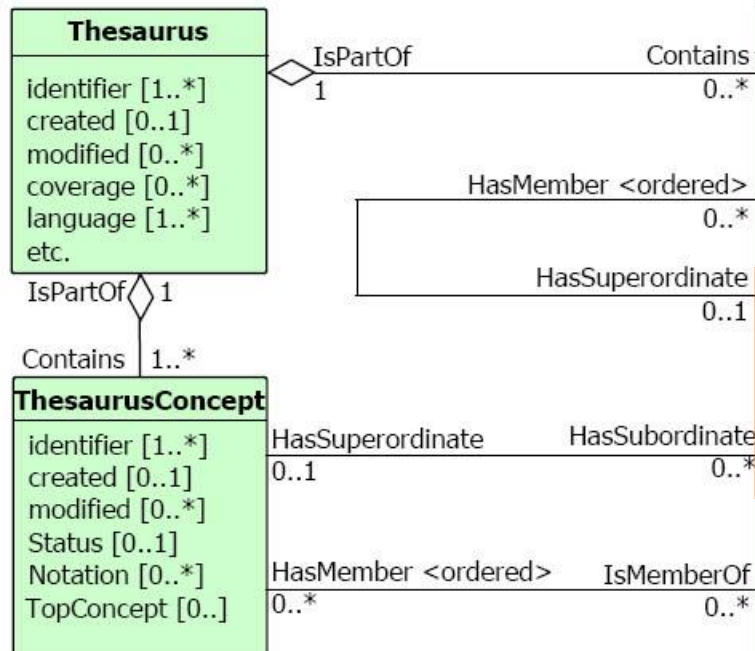
Term: **Cognition** [541]

Thesaurus array and Node label (representation of polyhierarchies)

- A thesaurus array (ThesaurusArray) is a group of sister concepts (concepts with a common broader term).
- A node label represents the criteria by which the thesaurus arrays are differentiated, e.g. (*persons according to age*).



IsLabelledBy 0..1



- Under a node label there may be as many concepts (ThesaurusConcept) as needed. N.B.: This relationship is not BT/NT-relationship.

Example: Thesaurus Array and Node label: Art & Architecture Thesaurus (AAT)

Built Environment (Hierarchy Name)

.... Settlements and Landscapes

..... inhabited places

..... <settlements by function>

In AAT node labels are called *Guide Terms*.
The node labels marked with arrows are
subordinated under the node label
<settlements by function>.

..... <settlements by function: administrative>

..... capital cities

..... cathedral cities

..... colonial cities

Thesaurus
Array

All concepts listed here, e.g. *capital cities*, *municipalities*, *bastides*, or *military towns*, are narrower terms (NT) of the broader term (BT) *inhabited places*. Node labels group these concepts according to different aspects (e.g. according to their function: *administrative* or *military*). The concepts under a node label represent a thesaurus array.

..... <local administrative centers>

..... missions (settlements)

..... municipalities

..... pueblos (settlements)

..... <settlements by function: military>

..... bastides

..... castelli (fortified towns)

..... military towns

..... presidios

Quelle: <http://www.getty.edu/vow/AATHierarchy?find=&logic=AND¬e=&subjectid=300008407>

Polyhierarchy in GND (hierarchical representation)

GND has no "Node Label"

OB4 [Wissenschaft](#) ▾ ▹
 OB3 [Naturwissenschaften](#) ▾ ▹
 OB2 [Chemie](#) ▾
 OB1 [Geochemie](#)
 [Anorganische Geochemie](#)
 [Biogeochemie](#)
 Hydrogeochemie
 [Isotopengeochemie](#)
 [Mineralchemie](#) ▲
 [Organische Geochemie](#) ▹
 [Umweltgeochemie](#) ▲

OB5 [Wissenschaft](#) ▾ ▹
 OB4 [Naturwissenschaften](#) ▾ ▹
 OB3 [Geowissenschaften](#) ▾ ▹
 OB2 [Hydrologie](#) ▾
 OB1 [Hydrochemie](#)
 Hydrogeochemie
 [Meereschemie](#) ▲

Hydrogeochemie

IDN 042254787

NID 4225478-4

BF Grundwasserchemie

OB Geochemie

OB Hydrochemie

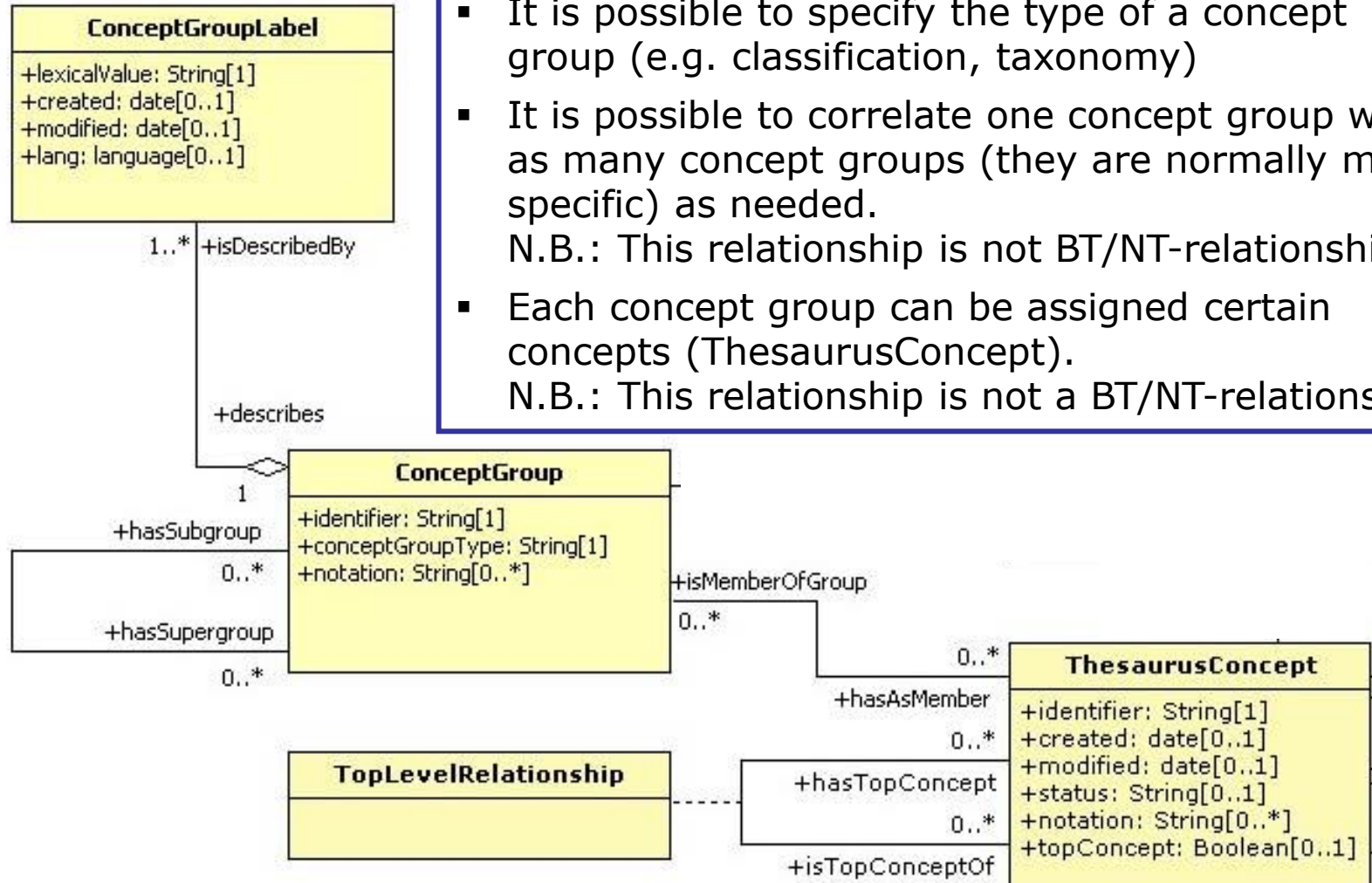
DDC [551.49](#) #2#

OB = BT
 BF = used for
 IDN = identifier
 NID = identifier

Concept group (ConceptGroup)



- Concept groups represent systematical structures like classes of a classification, microthesauri, subject groups, categories, etc.
- It is possible to specify the type of a concept group (e.g. classification, taxonomy)
- It is possible to correlate one concept group with as many concept groups (they are normally more specific) as needed.
N.B.: This relationship is not BT/NT-relationship.
- Each concept group can be assigned certain concepts (ThesaurusConcept).
N.B.: This relationship is not a BT/NT-relationship.



Example „Concept Group“: STW Thesaurus of Economics

Deutsch

ZBW
German National Library
of Economics

- Home
- Alphabetical descriptor list
- A General descriptors
- B Business economics**
- G Geographic names
- N Related subject areas
- P Commodities
- V Economics
- W Economic sectors

Concept groups

B Business economics

B Betriebswirtschaft (german)

narrower

- B.00 Business Economics
- B.01 Management and Business Organization
- B.02 Corporate Finance and Investment Policy
- B.03 Business Accounting and Auditing
- B.04 Human Resources Management
- B.05 Materials Management and
- B.06 Production Management
- B.07 Marketing**
- B.08 Corporate Taxation and Acc
- B.09 Business Information System
- B.10 Operations Research

Persistent Identifier (for bookmarking)

- <http://zbw.eu/stw/thsys/b>

More specific concept groups of the superordinated concept group **B**

Descriptors

- Competition
- Consumer goods marketing
- Country Image
- Gross margin
- International marketing
- Investment goods marketing
- Investor relations
- Key-account management
- Market

Records found for: **Psychology**

Click on the [number] to display the records indexed with that descriptor in unesdoc/unesbib.

[New query](#)

[- Help -](#)

Term: **Attitudes** [435]

MT=Microthesaurus

MT 4.10 Psychology

NT Attitude change [321]

NT Intolerance [80]

NT Prejudice [64]

NT Tolerance [416]

Concept group

Term: **Behaviour** [429]

MT 4.10 Psychology

UF Human behaviour

NT Emotions [102]

NT Lateral dominance [5]

NT Psychological effects [281]

...NT2 Mental stress [69]

NT Sensorimotor activities [13]

...NT2 Perception [157]

...NT2 Visualization [30]

NT Sexual behaviour [359]

NT Unconscious [9]

...NT2 Sleep [17]

1 record found for: **Attitude change**\$

Click on the [number] to display the records indexed with that descriptor in unesdoc/unesbib.

Term: **Attitude change** [321]

Terme français: Changement d'attitude

Término español: Cambio de actitud

Русский термин : Изменение отношения

MT 4.10 Psychology

BT Attitudes [435]

RT Social adaptation [114]

Concept group

Associative relationship

Term: **Cognition** [541]

MT 4.10 Psychology

UF Cognitive processes

Example: topical term Gemeinsame Normdatei (GND)

GND		
Link zu diesem Datensatz	http://d-nb.info/gnd/4185172-9	Preferred Term
Sachbegriff	Thesaurus	Scope Notes
Quelle	M, DIN 1463 Teil 1, RSWK Anlage 6	
Erläuterungen	Definition: Als Formschlagwort für vorliegende Thesauri Verwendungshinweis: Für unstrukturierte Schlagwortlisten benutze s. Schlagwort : f Verzeichnis	
Synonyme	Deskriptorsprache Deskriptor / Verzeichnis	Non preferred Terms
Oberbegriffe	Indexierung <Inhaltserschließung> Verzeichnis	Hierarchical relationship
Thematischer Bezug	Verwandter Begriff: Schlagwortnormdatei	Associative relationship
DDC-Notation	025.49	Concept group
Systematik	6.7 Bibliothek, Information und Dokumentation	

Quelle <http://d-nb.info/gnd/4185172-9>

Part 2: Interoperability with other vocabularies

Scope:

“This part of ISO 25964 deals with thesauri and other types of vocabulary that are commonly used for information retrieval. It describes, compares and contrasts the elements and features of these vocabularies that are implicated when interoperability is needed. It gives recommendations for the establishment and maintenance of mappings between multiple thesauri, or between thesauri and other types of vocabularies.”

Challenges for mapping:
defined as the “process of establishing relationships between the concepts of one vocabulary and those of another”

Part 2: Interoperability with other vocabularies

Contents:

- General guidelines for mappings
 - Definitions, symbols
 - Structural models between vocabularies
 - Guidelines for establishing mappings
 - Impact of the application on the mapping
 - Data management, display, functionality
 - Exchange formats
- Representation of different vocabularies under the aspect of mappings
 - Thesauri
 - Classifications
 - File plans
 - Taxonomies
 - Subject heading schemes
 - Ontologies
 - Terminologies
 - Name authority lists
 - Synonym rings

Part 2: Interoperability with other vocabularies

Normative: guidelines concerning mappings

Informative/descriptive: brief representations of different vocabularies

Why are vocabularies like termbanks or synonym rings included, which are not used for information retrieval?

Because they can provide vital support for searching and indexing by enriching the search terms. The described principles can be applied to them as well.

What about ontologies ? (only heavyweight ontology)

ontology ≠ thesaurus

other elements and other aims

-> no mappings recommended, instead supplementary usage of an ontology

Two different types of vocabularies

structural unity:

The mapped vocabularies have the same structure.

The equivalence of the concepts of such vocabularies is expressed by their identical structural position in the vocabulary. All relationships of the concepts correspond to each other; the concept terms can be automatically linked to pairs.
(e.g. multilingual thesauri of public authorities)

Only real equivalence pairs: = EQ

Two different types of vocabularies *continued*

structural disunity:

The mapped vocabularies do not have the same structure.

Equivalence of concepts has nothing to do with their position in the vocabularies. The mapping process produces either exact equivalence pairs or inexact equivalence pairs.

Different types of equivalences:

exact equivalence: =EQ

inexact equivalence: ~EQ

The concept is broader: BM ("Broader Mapping")

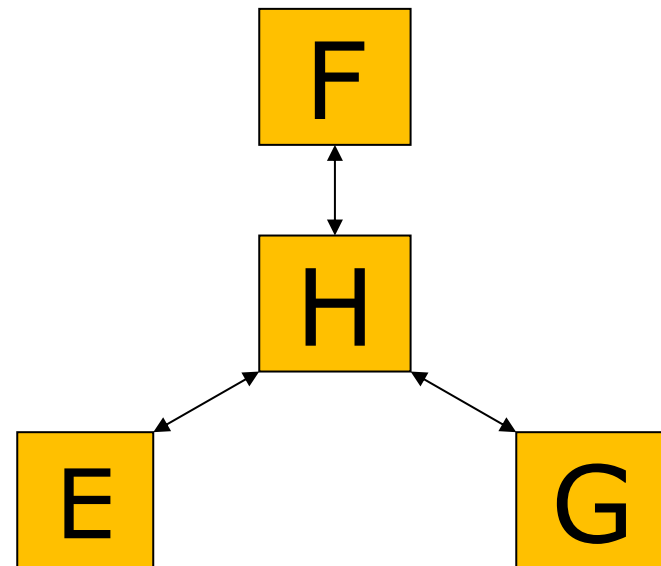
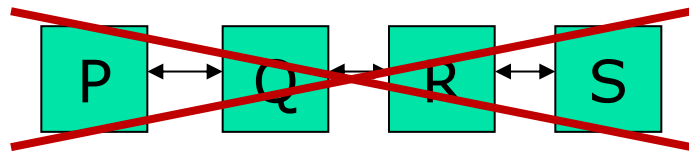
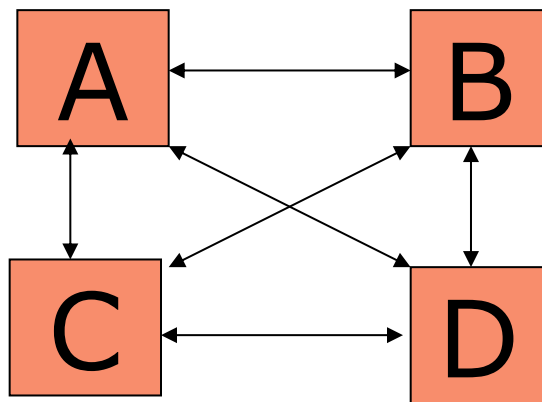
The concept is narrower: NM ("Narrower Mapping")

The concepts are somehow related: RM ("Related Mapping")

1:n-Relationship: - listing of single, narrower terms
- combination of several terms

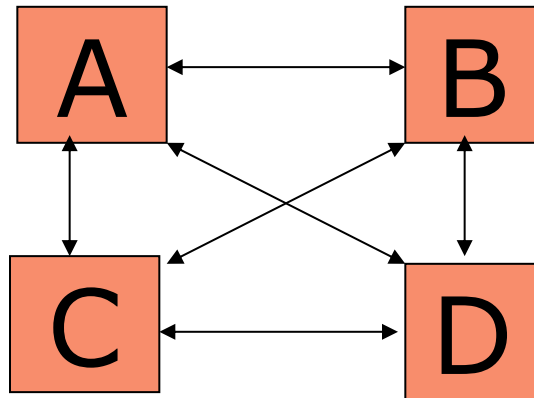
Structural models for mapping

Which of these are recommended?



Structural models for mapping

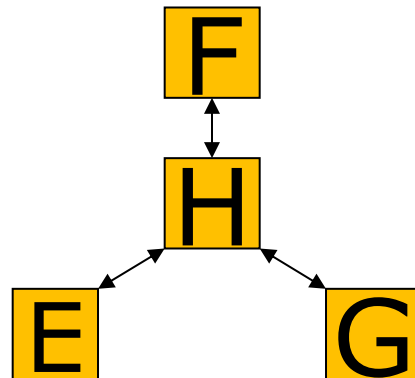
1. All concepts of all vocabularies are tied to each other.



- vocabularies with structural unity: only exact equivalences
 - vocabularies with structural disunity: exact equivalences and many inexact equivalences
- > very difficult and costly; appropriate only if few vocabularies are involved

Structural models for mapping

2. One vocabulary has the role of a hub; the concepts of the different vocabularies are only mapped to the concepts of this vocabulary.



- > appropriate for vocabularies with different structures,
 - especially if many vocabularies have to be mapped.
 - if there is a vocabulary with a dominating position.

Recommended types of mapping

- Equivalence
- Hierarchical
- Associative

Optionally distinguish between

- "exact" and "inexact" equivalence
- compound equivalence, either cumulative or intersecting

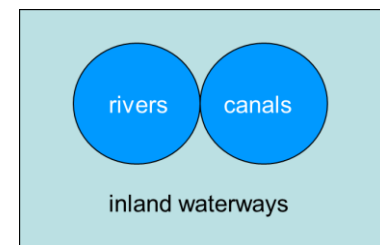
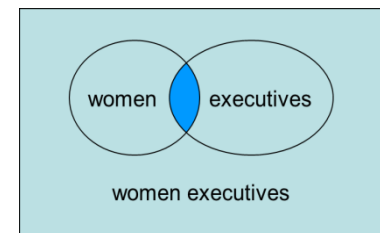
Compound equivalence

Women

executives

EQ Women + Executives

Inland waterways EQ rivers | canals



Precombination / Precoordination

- Typical for classification schemes, subject heading schemes, taxonomies
- > Compound equivalence; can be very complex

Use of Mappings

e.g.

- In the retrieval starting from the classification
- In the retrieval starting from the thesaurus
- > application has impact on the mapping

Interoperability in GND: Name authority file

- Technical integration of name authority files and a thesaurus

-> Same data format used; authority records have the same structure; only one database; relationships between entities are presented using the same method.

Name authority files and thesaurus

Linked by instancial relationships:

Generic term as instancial broader term for an individual name in the sub-file „subject cataloguing“ according specific rules

Bodensee
(Lake Constance)
 BTI See
(Lake)

Schwarze Dame Osterfeld 1988
 BTI Schachklub
(chess club)

BTI: Broader Term,
 instancial
 = obin: relator code
 in GND

Interoperability in GND: Other subject heading schemes

Mappings in the context of MACS (Multilingual Access to Subjects); the mappings are kept in a special system

- GND-LCSH; GND-Rameau
- Only **equivalences** without specifying the rate of accordance (no hierarchical or associative relationships)
- n:1 and 1:n-mappings are possible (compound equivalences)

The data format of GND can only handle 1:1 mappings. It is planned to import all 1:1 mappings into GND (ca. 40.000).

For the 1:n-mappings the data format of GND has to be expanded.

Interoperability in GND: Other thesauri

Mapping between GND / STW Thesaurus of Economics

Mapping between GND / Thesoz Thesaurus for the Social Sciences

- > The type of mapping is specified:
 - equivalence:
 - exact equivalence
 - inexact equivalence: GND-concept is broader
GND-concept is narrower
 - associative

1:n- and n:1-Mappings

Relevance

The concept group of the other thesaurus is carried along.

- > no part of GND; they are kept in a special system

Interoperability in GND: DDC (CrissCross-Project)

- Topical terms and authority records for places get a DDC number
- Application: to explore during retrieval
- Principle of mapping: not systematics of concepts but the literary warrant is important
- 4 degrees of coverage (determinacy)
(determinacy 4 = full conceptual coverage between GND-concept and DDC-class; 1 = only slight conceptual congruency)
- Challenge: subject headings are often only a narrower match (only a subset of a DDC-class)

No information about the type of mapping as EQ, NM, BM or RM

-> A combination of determinacy and type of mapping would improve the retrieval.

Example – Chick-pea

Link zu diesem Datensatz	http://d-nb.info/gnd/4163728-8
Sachbegriff	Kichererbse
Quelle	B 1986
Synonyme	Cicer arietinum
DDC-Notation	583.74 633.37 635.657 641.35657 641.65657
Systematik	24.3 Spezielle Botanik ; 32.4 Gartenbau, Obstbau

Public catalogue
representation

internal system

\$d = determinacy

065 24.3;32.4 (*GND-Systematics*)

083 583.74**\$d2\$st**2007-01-01

083 633.37**\$d2\$st**2007-01-01

083 635.657**\$d3\$st**2007-01-01

083 641.35657**\$d3\$st**2007-01-01

083 641.65657**\$d1\$st**2007-01-01

150 **Kichererbse** (*Preferred Term*)

450 Cicer arietinum (*Synonyma*)

670 B 1986 (*Source; "Scope Note"*)

botany

crop plants

in horticulture

specific food from plant crops

cooking with chick-peas

Chick-pea: DDC classes (determinacy given in brackets)

- **583.74 Fabales (Leguminales) (D2)**
- **633.37 Other legumes** Including: chick-peas, fava beans ... **(D2)**
- **635.657 Chick-peas (D3)**
- **641.35657 Chick-peas , . . . (D3)**
- **641.65657 Chick-peas—Cooking (D1)**

Although 641.65657 deals with cooking of chick-peas the determinacy is low as most literature about chick-peas does not deal with cooking of chick-peas.

ISO 25964 and GND - Evaluation

ISO 25964:

Comprehensive theoretical background for establishing thesauri, illustrated by examples.

Valuable guidelines for creating mappings between different vocabularies.

GND:

Realizes already many recommendations of ISO 25964 concerning the building of a thesaurus and mappings.

-> GND is an innovative tool

How to improve GND according to ISO 25964:

- Broader use of identifiers (not only for the record but also for concept groups or terms)
- For future mappings: the type of mapping (EQ, NM, BM, RM) should be taken into account as well.

How to get ISO 25964

In Germany: both parts ISO 25964
Beuth-Verlag <http://www.beuth.de/de/>

Directly via ISO <http://www.iso.org>

Part 1:

http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=53657 (238,00 CHF)

Part 2:

http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=53658 (196,00 CHF)

Data model: <http://www.niso.org/schemas/iso25964>
(and further general information to ISO 25964)

Information about GND

Homepage of DNB (German, English): www.dnb.de/gnd;
http://www.dnb.de/EN/Standardisierung/GND/gnd_node.html
Public GND-Wiki (only in German): <https://wiki.dnb.de/x/8IYOAw>

Any questions?

Esther Scheven
Deutsche Nationalbibliothek
Inhaltserschließung
Adickesallee 1
D - 60322 Frankfurt am Main
Tel.: +49-69-1525-1547
Fax: +49-69-1525-1555
<mailto:e.scheven@dnb.de>
<http://www.dnb.de>