

# German-Language Content in Biomedical Vocabularies

Stefan Schulz<sup>a</sup>, Josef Ingenerf<sup>b</sup>, Sylvia Thun<sup>c</sup>, Philipp Daumke<sup>d</sup>, Martin Boeker<sup>e</sup>

<sup>a</sup>Institute for Medical Informatics, Statistics and Documentation,  
Medical University of Graz, Austria

<sup>b</sup>Institute for Medical Informatics, University of Lübeck, Germany

<sup>c</sup>University of Niederrhein, Krefeld, Germany

<sup>d</sup>Averbis GmbH, Freiburg, Germany

<sup>e</sup>Institute of Medical Biometry and Medical Informatics,  
University of Freiburg, Germany

stefan.schulz@medunigraz.at

**Abstract.** There are hundreds of English-language biomedical vocabularies, but only a minor part has been translated into other languages. We performed a survey of localized German content in biomedical vocabularies, based on the authors' past experiences, web search, and analysis of the UMLS Metathesaurus. As a result we found German versions for fourteen medical terminologies in a strict sense. ICD 10 (International Classification of Diseases) for disease and OPS (Operationen- und Prozedurenschlüssel) for procedure encoding play a prominent role, as both are used in the German patient classification system. Interestingly they exceed in content and coverage the international sources they had been derived from. German content is also available for ICD-O (oncology), the Medical Subject Headings (MeSH), the Medical Dictionary for Regulatory Activities (MedDRA), the Logical Observation Identifiers Names and Codes (LOINC), the Anatomical Therapeutic Chemical Classification System (ATC), and the International Classification of Functioning, Disability and Health (ICF). From the SNOMED (Systematized Nomenclature of Medicine) family German translations have been created for SNOMED 3, SNOMED CT, and the Wingert Nomenclature. Other sources are RADLEX for radiology terms, the lists of Standard Terms of the European Directorate for the Quality of Medicines & HealthCare (EDQM), the Unified Code for Units of Measure (UCUM), the International Classification of Primary Care (ICPC), and the International Classification of Nursing Practice.

Partly relevant as sources for German medical terminology are fee catalogues for reimbursement like EBM and GOÄ. Medically relevant content is also provided by catalogues for medical devices such as contained in eCI@ss. Latin terms, which still matter in German medical text are available for organisms and anatomy from the NCBI taxonomy and the Terminologia Anatomica, respectively. Lay terms can be found in domain-independent terminology sources like GermaNet. No localised versions exist exclusively biological content such as the OBO Foundry and the NCBO BioPortal ontologies, although German terms occasionally appear as synonyms in an unsystematic way.

The most important German terminology systems are made available by DIMDI, the German Institute for Medical Documentation and Information. A

severe drawback is that the German SNOMED CT translation, which – according to its size (roughly 300,000 concepts) – exceeds by far, all other German medical vocabularies – is outdated and not officially available.

A general observation was that many of the localized versions lag behind the original version and are less rich in synonymous expressions, scope notes and definitions.

**Keywords:** Controlled vocabularies, German language, Medicine

## 1 Introduction

Making best use of biomedical data in health records, literature databases, and on the web requires standardization of structured and unstructured content in order to address a variety of use cases like biomedical research, health statistics, regulatory reporting, document and fact retrieval, decision support and quality management. Systems that provide standardized meaning of domain terms are commonly referred to as vocabularies or terminology systems. More than any other discipline, medicine has put high efforts into their development, from the London Bills of Mortality in the 17<sup>th</sup> century to highly complex systems like SNOMED CT. A large body of research has analysed the different kinds of biomedical vocabularies, their origins, formal foundations, their expressiveness, quality issues, as well as their relations to formal ontologies<sup>1,2</sup>. Numerous vocabularies are available by centralized repositories like the UMLS metathesaurus<sup>3,4</sup> and BioPortal<sup>5,6</sup>.

As a general building principle, vocabularies provide words or multiword terms as basic building blocks, mostly organized in hierarchies where they are organized by broader / narrower meanings. Terms are often enriched by (quasi)synonyms, i.e. other terms or term variants (e.g. acronyms). The criterion for synonym grouping is their exchangeability in the domain of interest. Such groups refer to a language-independent meaning, mostly referred to as concepts. Concepts are also in the centre of translations of terminologies. A translator evaluates the meaning conveyed by the terms assigned to a concept in the source language and attempts to find one or more fitting terms in the target language

Thesaurus-like vocabularies, are mainly descriptive, i.e. they describe the meanings of pre-existing terms and relate them using semantic relations. In contrast, classification systems, a special kind of vocabularies, have a more prescriptive tendency: many of their terms are, on purpose, artificial language expressions, which are not expected to occur in spoken or written language. The reason is that classification systems are intended to assign individual domain entities (such as diseases, procedures) to well-delineated, mutually disjoint classes. The disjointness criterion is important if they are used for statistics or billing.

Finally, a more recent tendency is to provide terminological systems with an ontological foundation using a logic-based framework, which allows for axiomatic description of the entities which fall in a certain class or instantiate a terminology concept. Such axioms describe what is universally true for all members of a class (or all

instances of a concept, or all things denoted by a term). This provides useful reasoning support for knowledge-intensive systems.

From a bird eye's view, biomedical terminologies consist of some or all of the following components:

- **Representational units** (RUs): the primary identifiers of meaning, often called concepts, classes, or categories;
- **Links**: the connections between RUs, such as broader/narrower, is-a, part-of;
- **Codes**: alphanumeric identifiers for a RU;
- **Terms**: words or group of words that describe the meaning of a RU in human language. In case there is more than one term per RU, one of them is generally picked out as the preferred one;
- **Hierarchies**: network of links that constitute a partial order, thus defining trees (single hierarchies) or directed graphs (multiple hierarchies);
- **Axes**: systems of independent, non-overlapping hierarchies;
- **Glossary entries**, also called scope notes: natural language elucidations of the meaning of the RU and its use in a given context;
- **Axioms**: sentences expressed in logic which are always true in the domain, thus providing formal, computable definitions for RUs;
- **Rules**: directives that specify exclusion and inclusion criteria in classifications;
- **Accessibility**: online / offline availability, open access, different licence models;
- **Metadata**: author, copyright, versioning etc.

The usefulness of a terminology system depends on many factors. Due to the fact that all the major terminology systems have been, primarily, built for English, and because English-speaking countries dominate, by far, R & D in biomedical informatics, it has often been neglected that a key requirement for adoption of a vocabulary is the availability of terms and definitions in the local language.

In this paper, we give a survey of biomedical terminologies available in German. We will include both translations of international terminology systems as well as systems that have been built exclusively for German-speaking users. We extend the survey by considering sources which are generally not considered biomedical terminologies, but which are nevertheless useful as a source for medical terms.

## 2 Survey of biomedical vocabularies

In the following we will provide an overview of biomedical vocabularies from which German medical terms are available. Most of them are translations from international terminology systems, where the German terms have been produced as translations.

### 2.1 ICD-10

The International Classification of Diseases (ICD)<sup>7</sup>, currently in its tenth version, classifies not only diseases but also related health conditions into disjoint classes in

single hierarchies, governed by inclusion and exclusion rules. ICD is the flagship of the WHO classifications and has been translated into 42 languages. In Germany there are two versions, one for mortality statistics (the original WHO one with about 14,400 RUs), and a German modification used for reimbursement in the German DRG system, called ICD-10-GM (German modification) with about 12,500 RUs. An important source for additional terminology is the Alphabetical Index, which contains about 77,000 terms each of which is assigned a unique identifier, called Alpha-ID<sup>8</sup>. This is especially important for application areas like drug safety. For example, "Amoxicillin-Allergie" can be communicated using the Alpha-ID "I112547" instead of the much less specific ICD-10 code T88.7 (Unspecified adverse effect of drug or medicament). However, Alpha-ID codes are not assigned to a concept level; hence synonymy and hypernymy between terms are not represented. The eleventh version of ICD is currently in preparations and will be enriched by a concept layer that should allow to replace the Alpha-ID in the future. German-language ICD-10 versions are produced, maintained, and distributed by the German Institute for Medical Documentation and Information (DIMDI)<sup>9</sup>.

## **2.2 OPS**

In Germany, the operations and procedure classification according to § 301 SGB V (OPS)<sup>10</sup> is the official coding system for medical procedures with a focus on surgical interventions. Like ICD it contains the systematic classification and an Alphabetical Index. With about 28.300 classes it is considered as the largest existing procedure encoding system. OPS has grown as an adaptation of the WHO's International Classification of Procedures in Medicine (ICPM). OPS's architecture mirrors ICD's with single hierarchies, disjoint classes, inclusion and exclusion rules. Together with ICD-10 it is used e.g. for the German DRG system, Doctors' Fee Scale within the Statutory Health Insurance Scheme (EBM), quality management, health statistics and is maintained and provided by DIMDI.

## **2.3 ICD-O**

ICD-O ("O" for oncology)<sup>11</sup> is a biaxial classification. It provides terms for tumour morphology and localization, totaling 1,180 entries. It is mostly used in cancer registries. The topography axis is for the topographical codes of the tumor's site. It corresponds to the codes from the C section of malign neoplasms in ICD-10. The morphology axis is for the morphology of the tumor, derived from the M8/9-section of the morphology axis of an older version of SNOMED. The German translation of ICD-O is freely provided by DIMDI and can also be downloaded from the UMLS.

## **2.4 The Medical Subject Headings (MeSH)**

MeSH<sup>12</sup> is the indexing vocabulary of the biomedical literature database MEDLINE. It contains multiple, partly overlapping hierarchies (trees) and is available in 41 languages. The manual assignment of MeSH terms to each new MEDLINE record by the

U.S. National Library of Medicine provides an important added value to this literature database. MeSH is updated annually by the NLM and translated into German by DIMDI. In the current 2013 version of MeSH, it contains 26,853 main headings and 60,067 German entry terms (synonyms). Scope notes are not translated. The German MeSH can be purchased by DIMDI. Due to the limited granularity of thesauri, there is a special problem of quasi-synonymy between descriptor-related terms, called entry terms in MeSH. In order to support the maintenance and translation, an intermediate concept level has been added<sup>13</sup>, which is, however, not visible to the user.

Example (descriptor with D-codes, concepts with M-codes, terms with T-codes):

Descriptor: D003085 "Colic"  
Concepts: M0004742 "Colic" (Preferred),  
M0004741 "Abdominal Cramps", M0439077 "Infantile Colic"  
Terms: T008934 "Colic" (Preferred), ref. to (→) M0004742  
T520657 "Colicky Pain" → M0004742,  
ger0002992 "Kolik" (Preferred) → M0004742  
T008933 "Abdominal Cramps" (Preferred) → M0004742,  
ger0030380 "Abdominelle Krämpfe" (Preferred) → M0004742

Formerly, all quasi-synonymous terms linked as entry terms to one descriptor like "D003085 Colic" could not be differentiated. Term-IDs with prefix "ger" are added as identifier for German synonyms<sup>14</sup>.

## 2.5 MEDDRA

MedDRA<sup>15</sup> (Medical Dictionary for Regulatory Activities) is an international terminology used by authorities and pharmaceutical industry during the regulatory process, from pre-marketing to post-marketing activities, targeting data entry, retrieval, evaluation, and presentation. Additionally, it is used to code adverse events. It encompasses signs, symptoms, diseases, procedures, investigation results and therapeutic indications. MedDRA's German translation consists of 92,000 German terms, grouped by different kinds of synonymy relations and ordered in a single hierarchy with 26 so-called system organ classes at the highest level.

## 2.6 LOINC

LOINC (Logical Observation Identifiers Names and Codes)<sup>16</sup> is a collection of worldwide unique identifiers for medical examinations in general and laboratory tests in particular. It consists of laboratory (clinical chemistry, hematology, serology, microbiology) and clinical investigations and additional content like document types. The LOINC names of some 22,000 frequently used investigation method by medical experts are translated in German and are available at DIMDI.

## 2.7 ATC

ATC (Anatomical Therapeutic Chemical Classification System)<sup>17</sup> is a classification system for drugs<sup>18</sup>, which are grouped according to anatomical entity (A) on which they act, their therapeutic (T) and chemical (C) characteristics. Defined daily doses (DDD) are the assumed average daily maintenance dose for the main indication of each substance in adults. In the German release there are about 7,600 codes in a single hierarchy. However, the same drug may have more than one code, according to its therapeutic use. ATC is used for comparison of prices and for morbidity-oriented risk structure compensation schemes.

## 2.8 ICF

The International Classification of Functioning, Disability and Health is a multiaxial classification, issued by WHO. The German translation from 2005 is available at DIMDI and has close to 1,000 categories<sup>19</sup>.

## 2.9 SNOMED and Wingert Nomenclature

SNOMED (Systematized Nomenclature Of Medicine Clinical Terms), started as pathology-centred terminology (SNOP)<sup>20</sup> and grew into subsequent versions of a multiaxial nomenclature system covering the whole of clinical medicine. Its version 3 provided 12 axes each of which corresponding to a single hierarchy under a clinical category like Chemical, Disease, Function, Morphology etc. In 1984 a German translation was released<sup>21</sup>, later known as Wingert Nomenclature<sup>22</sup>. With the subsequent version SNOMED RT, SNOMED was embedded into a description logics framework. After the fusion with the UK terminology CTC3 it has been promoted as the international terminology standard SNOMED CT by IHTSDO<sup>23</sup>. A new and Wingert – independent German translation was initiated in 2002 by a translation company, totalling an effort of 11.5 person years. It was finished in 2004 with about one million terms for close to 362,000 concepts. However, it has not been officially released.

SNOMED CT is tightly interwoven by semantic relations, from which description logics axioms can be generated. They are mainly used to automatically generate taxonomic links in the production process, so that SNOMED CT exhibits a high degree of multiple parenthood.

## 2.10 RADLEX

RadLex is a vocabulary of standardized radiological terms for structured reporting in medical imaging<sup>24</sup>. It is hierarchically structured and covers 6,240 German terms, out of currently about 30,000 English terms.

## **2.11 EDQM**

The lists of Standard Terms<sup>25</sup> of the European Directorate for the Quality of Medicines & HealthCare cover dosage forms, routes and/or methods of administration, and containers, closures and delivery devices used for medicines for human and for veterinary use. They are translated in 32 world languages. The Standard Terms are used in European Marketing Authorisation applications, the Summary of Product Characteristics (SmPC), labelling and electronic communications.

## **2.12 UCUM**

The Unified Code for Units of Measure is a code system intended to include all units of measures being contemporarily used in international science, engineering, and business. The purpose is to facilitate unambiguous electronic communication of quantities together with their units<sup>26</sup>.

## **2.13 ICPC**

The International Classification of Primary Care (ICPC) was developed by the ICPC Working Party and published in 1987 by WONCA, the World Organisation of Family Doctors<sup>27</sup>. 720 terms have been translated into German and are available via the UMLS Metathesaurus.

## **2.14 ICNP**

The International Classification of Nursing Practice, owned by the International Council of Nurses (ICN), provides a standardized reference for nursing diagnoses, nursing interventions, and nursing results, totalling 2800 concepts in its 2<sup>nd</sup> release. It is a multiaxial terminology system based on seven axes, out of which postcoordinated entries to purpose specific catalogues can be created. Like SNOMED CT, ICNP is increasingly ontology and logic-based. The German version is maintained by the German-Speaking ICNP User Group<sup>28</sup>.

# **3 Special cases**

We will conclude this overview presenting additional sources, which are generally not considered sources for German-language medical terms.

## **3.1 Fee catalogues for procedure-based reimbursement**

In Germany EBM<sup>29</sup> and GOÄ<sup>30</sup> are catalogues for fee schemes used in the ambulatory sector in Germany. Similar catalogues exist in Austria<sup>31</sup> and Switzerland<sup>32</sup>. Formally they mostly resemble a flat classification of disjoint representational units similar to product catalogues. The meanings of the categories are described, in large parts, by

detailed free-text descriptions. However, numerous categories are just described by medical terms, so that we can consider these catalogues, at least partially, as sources for medical terminology. However, there are cases where the names are elliptic, e.g. "Glucose", whereas their exact meaning has to be derived from the context (Glucose measurement in Blood). EBM has about 1,600 categories, GOÄ about 3,000.

In a similar vein, catalogues for medical devices could be used as a source for German terminology. An example is eCI@ss<sup>33</sup>, a comprehensive product categorization system, which also includes biomedical devices. A catalogue for pharmaceutical preparations is provided e.g. by ABDA<sup>34</sup>, which is especially an important source for brand names for drugs.

### 3.2 Sources for Latin terms

In German-speaking countries, Latin terms are more widely used than in most other languages, both biomedical literature and medical documentation Latin terms play a major role in German. Therefore, two large terminologies can be used as sources for Latin terms which are commonly used in German-language biomedical text: **NCBI taxonomy**<sup>35</sup> is a vocabulary for about 10% of all known organisms.

The **Terminologia Anatomica** is a bilingual (English, Latin) nomenclature for anatomy, maintained by two international anatomical societies<sup>36</sup>.

### 3.3 Domain-independent terminology sources

Similar to the English semantic lexicon WordNet<sup>37</sup>, GermaNet<sup>38</sup> is a lexical-semantic network under construction, containing about 111,000 lexical units, grouped by 85,000 concepts (Synsets), linked by about 100,000 relations. Although GermaNet's coverage is not domain-specific, it may be a valuable source of laypersons' medical terms.

## 4 Conclusion

The main sources of German medical terms viz. ICD 10, LOINC and OPS are freely available and they are well supported by extensive synonym lists. In other cases, the German translation covers no or less synonyms than the English original (e.g. MeSH, RadLex), and also misses scope notes and definitions. The German SNOMED CT translation, which – according to its size – exceeds by far, all other German medical vocabularies – is outdated and not officially available. There are sources which are only available in printed format such as the Terminologia Anatomica. It is obvious that only where there is a clear use case for a localised version of an international terminology system, maintenance effort is invested. No localised versions exist for vocabularies that exclusively serve biomedical research, such as the OBO ontologies<sup>39</sup> and the biology sources in the NCBO BioPortal<sup>40</sup>, although German terms can sometimes be found listed as synonyms in an unsystematic way.

## 5 References

1. Freitas F, Schulz S, Moraes E: Survey of current terminologies and ontologies in biology and medicine. *RECIIS - Electronic Journal in Communication, Information and Innovation in Health*, 2009; 3 (1): 7-18.
2. Ingenerf J, Giere W. Concept-oriented standardization and statistics-oriented classification: continuing the classification versus nomenclature controversy. *Methods Inf Med*. 1998 Nov;37(4-5):527-539.
3. Nelson SJ, Powell T, Humphreys LB. The Unified Medical Language System (UMLS) of the National Library of Medicine. *Journal of American Medical Record Association*, (2006) vol. 61, 40-42, 1990.
4. [http://www.nlm.nih.gov/research/umls/knowledge\\_sources/metathesaurus/](http://www.nlm.nih.gov/research/umls/knowledge_sources/metathesaurus/)
5. Whetzel PL; NCBO Team. NCBO Technology: Powering semantically aware applications. *J Biomed Semantics*. 2013 Apr 15;4 Suppl 1:S8.
6. <http://biportal.bioontology.org/>
7. WHO - International Classification of Diseases, 10th Edition. World Health Organization. <http://www.who.int/classifications/apps/icd/icd10online/>
8. <http://www.dimdi.de/static/de/klassi/alpha-id/index.htm>
9. <http://www.dimdi.de>
10. <http://www.dimdi.de/static/de/klassi/ops/>
11. <http://www.dimdi.de/static/de/klassi/icdo3/index.htm>
12. MESH - Medical Subject Headings, <http://www.nlm.nih.gov/mesh/>.
13. Nelson SJ, Johnston D, Humphreys BL. Relationships in Medical Subject Headings (MeSH), chapter 11. In: Bean CA, Green R, editors. *Relationships in the Organization of Knowledge*. New York: Kluwer Academic Publishers; 2001; 171-184, see <http://www.nlm.nih.gov/mesh/meshrels.html>.
14. Nelson SJ, Schopen M, Savage AG, Schulman JL, Arluk N. The MeSH translation maintenance system: structure, interface design, and implementation. *Stud Health Technol Inform*. 2004; 107(Pt 1):67.
15. <http://meddramsso.com/>
16. <http://www.dimdi.de/static/de/klassi/loinc/>
17. <http://www.dimdi.de/static/de/klassi/atcddd/>
18. <http://www.dimdi.de/dynamic/de/klassi/downloadcenter/atcddd/version2013/>
19. <http://www.dimdi.de/static/de/klassi/icf/>
20. Cornet R, de Keizer N. Forty years of SNOMED: a literature review. *BMC Med Inform Decis Mak*. 2008 Oct 27;8 Suppl 1:S2
21. Coté RA, Wingert F. *SNOMED - Systematisierte Nomenklatur der Medizin*. 1984 Springer Verlag Berlin Heidelberg New York
22. Denecke, K., Kohlhof, I., Bernauer, J.: Use of Multiaxial Indexing for IE from Medical Texts. In: Proc. FCTC 2006, Int. Workshop on Foundations of Clinical Terminologies, Timisoara, Romania, April 2006
23. <http://www.ihtsdo.org>

24. Marwede D, Daumke P, Marko K, Lobsien D, Schulz S, Kahn T. RadLex - deutsche Version: ein radiologisches Lexikon zur Indexierung von Bild- und Befunddaten Rofo. 2009 Jan;181(1):38-44. doi:
25. <http://www.edqm.eu/en/standard-terms-590.html>
26. <http://www.dimdi.de/static/de/klassi/ucum/index.htm>
27. <http://www.who.int/classifications/icd/adaptations/icpc2/en/>
28. <http://www.icnp.info/>
29. <http://www.kbv.de/8156.html>
30. <http://www.e-bis.de/goae/defaultFrame.htm>
31. [http://www.bva.at/mediaDB/792066\\_ho.pdf](http://www.bva.at/mediaDB/792066_ho.pdf)
32. <http://www.tarmedsuisse.ch/>
33. <http://www.eclass.de/>
34. <http://www.abda.de/>
35. <http://www.ncbi.nlm.nih.gov/taxonomy>
36. <http://www.unifr.ch/ifaa/Public/EntryPage/HomePublic.html>
37. <http://wordnet.princeton.edu/>
38. <http://www.sfs.uni-tuebingen.de/lsd/introduction.shtml>
39. Smith B, Ashburner M, Rosse C, Bard J, Bug W, Ceusters W, Goldberg LJ, Eilbeck K, Ireland A, Mungall CJ; OBI Consortium, Leontis N, Rocca-Serra P, Ruttenberg A, Sansone SA, Scheuermann RH, Shah N, Whetzel PL, Lewis S. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nat Biotechnol. 2007 Nov;25(11):1251-1255.
40. NCBO BioPortal <http://bioportal.bioontology.org/>