

Standards and Biomedical Terminologies: the CEN TC 251 and ISO TC 215 Categorial Structures. A step towards increased interoperability

Jean M. RODRIGUES^{a,1}, Anand KUMAR^a, Cédric BOUSQUET^a,
Béatrice TROMBERT^a.

^a *University of Saint Etienne, Department of Public Health and Medical Informatics,
St Etienne, France.*

Abstract: Among different biomedical terminologies standardisation strategies the European Standard Body CEN TC 251 followed by the ISO TC 215 have stated that it was not possible to convince the different European or international member states using different national languages to agree on a reference clinical terminology or to standardise a detailed language independent biomedical ontology. Since 1990, they have developed an approach named categorial structure as a step standardising only the terminologies model structure. The methodology and the review of the different existing categorial structures are presented as a step towards increased interoperability between biomedical terminologies thanks to conformity to a minimum set of ontological requirements.

Keywords: Standard; Biomedical terminology; Categorial Structure; Ontology;

Introduction

The standards and the biomedical terminologies have appeared since a long time as two different worlds. Most of developed countries maintain, update and modify their own coding systems for procedures, as well as national adaptations of ICD, in order to manage and to fund their health care delivery. The most significant efforts were done in Australia withACHI (Australian Classification of Health Interventions) or ICD10 AM [1], in Canada with the Canadian Classification of Health Interventions (CCI) [2] developed by the Canadian Institute for Health Information and in France with CCAM (Classification Commune des Actes Médicaux) [3].

Natural language expressions show inconsistencies and ambiguities as assessed by biomedical ontology driven tools [4]. Related knowledge bases consist of multi-hierarchies of concepts organized by subsumption and associative relations. These knowledge representations are named biomedical ontology [5] [6]. The related

¹ Corresponding Author : JM Rodrigues, CHU de St Etienne, SSPIM/CIM42, Chemin de la Marandière, 42 270 Saint Priest en Jarez, France, rodrigues@univ-st-etienne.fr

automated language generation techniques include the linkage of lexicons from different national languages [7]. The most important achievements are GALEN (Generalised Architecture for Languages, Encyclopaedias and Nomenclatures in Medicine) [8], and FMA (Foundational Model of Anatomy) [9].

On the other hand, the standardisation in health informatics started in the US with the HL7 user group. The European Standard Body CEN TC 251 WG2 (Comité Européen de Normalisation Technical Committee 251 Working Group 2) and later the International Standard Organisation ISO TC 215 WG3 elaborated and developed a standard approach for biomedical terminology named Categorical structure. We address the rationale and definition of Categorical structures in part 2 and the review of the existing Categorical structures standards in part 3. We finally discuss the role of this standard approach as a compromise between one single standard for biomedical terminology and ontology based terminologies on the way to a future more complete semantic interoperability.

1. Hypothesis behind the CEN categorical structure approach

1.1. Rationale

As a prerequisite it was clearly stated at the beginning of the CEN standard process that it was not realistic to impose within Europe a standard terminology to health care professionals even as a reference terminology or a pragmatic terminology. The two main supportive arguments were that European countries speak different natural languages and that different health care professionals within the same natural language do not convey the same meaning through a particular terminology.

CEN has considered in the early 90s and ISO in the late 90s it not possible to standardise the clinical knowledge in a single biomedical ontology which could support one reference clinical terminology or different types of clinical terminologies. First the ontology developments were starting and needed consolidation and assessment of feasibility. Second the standards needed to facilitate developments in biomedical terminologies and not prevent the quickly evolving volume of terms used for different goals. Finally clinicians and medical recorders were not convinced that the formal representation of the biomedical terminologies by description logic tools which is an important point in the definition of computer ontology [10] was realistic. For these opposite challenges a third way named Categorical structure was designed and applied to some priority domains of biomedical terminologies.

1.2. Definition.

The CEN Categorical structure was defined within some linguistic variations [11-22], as a minimal set of health care domain constraints to represent a biomedical terminology in a precise health care domain with a precise goal to communicate safely. It is a definition of a minimal semantic structure describing the main properties of the different artefacts used as terminology (controlled vocabularies, nomenclatures, coding systems and classifications): a model of knowledge restricted to 1) a list of semantic categories; 2) the goal of the Categorical structure; 3) the list of semantic links between semantic categories authorised with their associated semantic categories; 4) the minimal constraints allowing the generation and the validation of well formed

terminological phrases. Any biomedical artefact claiming conformance to the standard shall attach with the data sent the Categorial structure of the terminology used. The Categorial structure shall satisfy the 4 constraints but can add more constraints.

2. Methodology of the Categorial structures

2.1. EN 1828 Categorial structure for surgical procedures [11]

We first explain precisely one Categorial structure to make explicit what is a Categorial structure and what it is not. Figure 1 shows the hierarchies of the semantic categories (left tab) and their associated semantic links (right tab).

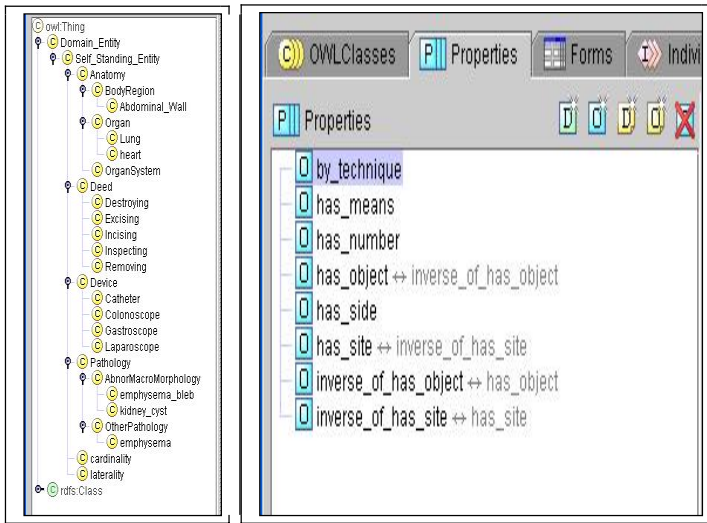


Figure 1. Semantic categories (left tab) and semantic links (right tab)

- 1 The different main semantic categories are Anatomy, Deed, Device and Pathology with 2 qualifiers categories cardinality and laterality.
- 2 The semantic links are *by_technique*, *has_means*, *has_object* (with inverse), *has_site* (with inverse), *has_side* and *has_number*.
 - 2.1 *has_object* is authorised between deed and anatomy or device or pathology
 - 2.2 *has_site* is authorised between device or pathology and anatomy
 - 2.3 *has_means* is authorised between deed and anatomy, device or pathology
 - 2.4 *by_technique* is authorised between deed and deed
- 3 The minimal constraints required
 - 3.1 A deed and *has_object* shall be present
 - 3.2 Anatomy shall always be present either with a *has_object* or with a *has_site*
 - 3.3 Use of pathology shall be restricted to macroscopic lesion and to cases where it allows to differentiate the procedure from procedures using the same deed and the same anatomy;

3.4 When *by technique* is used the deed on the right side of the semantic link must be conform to the rules 3.1, 3.2 and 3.3.

The categorial structure allows ensuring that new terms describing surgical procedures are associated to a formal definition consistent with a common template. For example we evaluated the introduction of six new rubrics. Definitions of these six rubrics were entered in the Protégé editor [23] and classified with the Racer inference engine [24]. Concepts not consistent with the standard were identified during classification and represented as red circled concept nodes in Protégé.

2.2. The Categorial structures [11] to [16] and [18] to [22]

There are six full CEN or ISO standards for Categorial structure: CEN EN 12 264 [13] for the definition of a Categorial structure, CEN EN 1828 [11] for surgical procedures, ISO 1804 [12] for nursing terminologies, CEN EN 12 611 [14] for clinical laboratory, CEN EN 13 940 [15] for continuity of care, and CEN EN 15 521 [16] for human anatomy. On the other hand five Categorial structures are still in a provisional status on medical devices [18] and [20], medicinal products [19], vital signs [21] and patient findings and problems [22].

The standards will be revised in the five coming years by a joint process of two Working Groups from CEN and ISO. This will give the opportunity to harmonise the wording and definitions and to check their applications around the world. Finally the EN 15 521 for terminologies of human anatomy has introduced more extended constraints by prescribing detailed semantic categories and semantic links for the knowledge in the field of anatomy.

3. Results and Discussion

There are at least three different ways to address the standard issues. The first of them can be called a single standard. This is the Family of classifications used since the nineteenth century by the International Classification of Diseases (ICD) and now by the WHO Family of International Classifications (FIC). The same rationale continues to be followed when it comes to revising the ICD into its eleventh revision. However the size of knowledge is increasing so quickly with biomedical research and technology developments that maintenance and update costs are huge and never quality assured. It is an unending work based on divergent individual and domain expert opinions.

The second methodology which has emerged since 15 years can be called the biomedical co-operative approach based on formal ontology and on the biomedical ontology tools coordinated with natural language processing and web based tools [4-9]. It provides a new perspective to semantic interoperability by de-multiplying the workload organised by disseminated social computing. It is complex but separates terminology artefacts from the logic of knowledge and the linguistic characteristics of the different national languages. By associating clinical specialist domain experts, ontology experts and linguists this infrastructure gives the opportunity to assess clinical terminology by logical expertise and as well ontological representation by clinical domain experts working in their national language.

Between these two trends is a third methodology where the international standardisation organisations have the responsibility to support or assist in improving semantic interoperability. Biomedical terminologies need to conform to a minimum set

of ontological requirements. This has been the choice of the CEN European standard body within Technical Committee 251 and ISO Technical Committee 215 for terminology who has proposed the Categorical Structure. The categorical structure proposes a frame for a light ontological organisation to ensure standardisation of the knowledge representation of terminologies. The work began in the 90s with the goal to move from a syntactic or functional semantic interoperability level 1 to a semantic interoperability of level 2 where the recipient is able to understand the meaning of terms used by the sender but cannot process them as SAFELY as he can do with his own terms and meaning (semantic interoperability level 3). Developers of new terminologies or new versions of existing terminologies should conform to this standard in order to increase interoperability. It is limited but supporting the dissemination of knowledge based on a representation for any clinical and biomedical terminologies and a pedagogic step to ontology and semantic interoperability.

With the wider use of CEN categorical structure in the near future, which is imminent given the fact that European standards are mandatory for member states, would provide further insights into the advantages of using such a methodology. Future steps would include studies and trials to compare its clinical and administrative benefits to other methodologies.

Acknowledgements

We wish to thank our partners in the CEN TC251 WG2 and ISO TC 215 WG 3 and the convenors Hendrik Olesen, Ann Harding, Magnus Fogelberg and Chris Chute and terminology and biomedical ontology experts namely Alan Rector from University of Manchester and Barry Smith from the University at Buffalo. Special thanks to our partners in the GALEN program and namely Pieter Zanstra, Egbert van der Haring, Robert Baud, Jeremy Rogers, and the Stanford team for their kindness and the open quality of their website.

References

- [1] National Centre for Classification in Health see <http://www3.fhs.usyd.edu.au>
- [2] Canadian Classification of Health Interventions <http://secure.cihi.ca/cihiweb/dispPage.jsp>
- [3] Trombert-Paviot B, Rodrigues J-M, Rogers J, Baud R, van der Haring E, Rassinoux A-M, Abrial V, Clavel L, Idir H. GALEN: a third-generation terminology tool to support a multipurpose national coding system for surgical procedures. *Int J Med Inform* 2000;58-9:71-85.
- [4] Rodrigues J-M, Rector A, Zanstra P, Baud R, Innes K, Rogers J, Rassinoux A-M, Schulz S, Trombert Paviot B, ten Napel H, Clavel L, van der Haring E, Mateus C. An ontology driven collaborative development for biomedical terminologies: from the French CCAM to the Australian ICHI coding system. *Stud Health Technol Inform*. 2006;124:863-8..
- [5] Gruber T. A translation approach to portable ontology specifications. *Knowledge Acquisition* 1993; 5(2):199-220.
- [6] Smith B, Ceusters W. Towards industrial strength philosophy: how analytical ontology can help medical informatics. *Interdisciplinary Science Reviews* 2003; 28:106-11.
- [7] Baud RH, Rodrigues J-M, Wagner JC, Rassinoux A-M, Lovis C, Rush P, Trombert-Paviot B, Scherrer J-R. Validation of Concept Representation Using Natural Language Generation. In: Masys DR (Ed.). *AMIA Annu Symp Proc*. 1997: 841.
- [8] Rector A. Compositional Models of Medical Concepts: Towards Re-usable Application-Independent Medical Terminologies. In: Barahona P and Christensen JP (Eds.). *Knowledge and Decisions in Health Telematics*. Amsterdam: IOS Press, 1994; pp. 109-14.

- [9] Rosse C, Mejino JVL. 2003. A reference ontology for biomedical informatics: the Foundational Model of Anatomy. *J Biomed Inform.* 36:478-500.
- [10] Semantic Interoperability seminar. Brussels February 2005
- [11] CEN AFNOR EN NF 1828:2002. Health informatics – Categorial Structure for classifications and coding systems of surgical procedures.
- [12] EN ISO -18104, 2003 – Health informatics – Integration of a reference terminology model for nursing
- [13] CEN AFNOR EN NF 12264:2005. Health informatics – Categorial Structure for concept systems.
- [14] EN 1614: 2006, Health Informatics – Representation of dedicated kinds of property in laboratory medicine
- [15] EN 13 940: 2006, Health Informatics – System of concepts to support continuity of care.
- [16] EN 15 521:2007. Health informatics – Categorial Structure for terminologies of human anatomy
- [17] ISO 17115: 2007, Health informatics — Vocabulary for terminological systems
- [18] ENV 12 611: 1996, Medical Informatics – Categorial structure for system of concepts - medical devices
- [19] ENV 12 610:1997 Medical Informatics – Medicinal products identification
- [20] **ISO CD 15225:2000 – Nomenclature – Specification for a nomenclature system for medical devices for the purpose of regulatory data exchange**
- [21] **ENV 13734:2000 Health Informatics – Vital signs information representation**
- [22] **ISO/DTS 22789:2007, Health informatics – Conceptual framework for patient findings and problems in terminologies**
- [23] Noy NF, Fergerson RW, Musen MA. The knowledge model of Protege-2000: Combining interoperability and flexibility. 2nd International Conference on Knowledge Engineering and Knowledge Management (EKAW'2000), Juan-les-Pins, France. 2000.
- [24] Haarslev V and Möller R. Description of the RACER System and its Applications. In Proceedings International Workshop on Description Logics (DL-2001), Stanford, USA, 1.-3;131–141. 2001.