

Ontology-Based Integration of Medical Coding Systems and Electronic Patient Records

Ceusters W^{ab}, Smith B^{bcd}, De Moor G^a

^a RAMIT (Research in Advanced Medical Informatics), University of Ghent, Belgium

^b ECOR (European Centre for Ontological Research), Saarland University, Germany

^c IFOMIS (Institute for Formal Ontology and Medical Information Science), Saarland University, Germany

^d Department of Philosophy, University at Buffalo, NY, USA

1 Abstract

In the last two decades we have witnessed considerable efforts directed towards making electronic healthcare records comparable and interoperable through advances in record architectures and (bio)medical terminologies and coding systems. Deep semantic issues in general, and ontology in particular, have received some interest from the research communities. However, with the exception of work on so-called ‘controlled vocabularies’, ontology has thus far played little role in work on standardization. The prime focus has been rather the rapid population of terminologies at the level of fine detail. In this paper, we argue that more efforts are needed on the side of both research and standardization to ensure that the coding systems used in electronic healthcare records enjoy a semantics that is coherent with the semantics of the record. We propose realist ontology as a method to bring about this coherence by means of a robust system of top-level ontological categories.

2 Introduction

As a result of the MEDIREC project [1], which led through the PROREC initiative [2] to the creation of the European Institute for Health Records [3], and as a result of the activities of CEN/TC251, there is now a Europe-wide agreement that a comprehensive, communicable and secure electronic healthcare record (EHCR) is a prerequisite for the delivery of high quality healthcare [4, 5]. This EHCR should respond to the needs of all involved parties and at the same time allow interconnection with other health telematics applications. Gradually, this European vision has gained international acceptance, leading to standards at ISO level.

Standards are indeed a key issue to achieve interoperability of electronic healthcare systems, and they are required at various levels, including:

1. the healthcare record system architecture (defined in ISO 18308 [6] as: “*the generic structural components from which all electronic healthcare records are built, defined in terms of an information model*”, although what we need are not *information models* but rather a *theory of reality*),
2. the communication mechanisms employed in exchanging data amongst EHCRs (healthcare messages as developed according to the principles described in CR 12587 [7] being a specific example)

and, last but not least,

3. the content itself, i.e. the patient data.

Despite previous and current standardisation efforts, it is still argued that “one of the most significant challenges to implementing electronic health information systems is the lack of

standards for electronic patient medical record information, especially standards around the terminology that expresses clinical documentation” [8]. Coding systems are typical examples.

3 Coding systems

3.1 Purpose of coding systems

The goal of EHCR is to achieve faithful clinical data entry [9, 10], but in such a way as to meet the requirements of communicability for both human and machines. To this end much emphasis has been placed on clinical coding, with the rationale that codes will make it possible to associate precise meanings with the terms used in expressing patient data in a way that can be interpreted by software for further processing for purposes such as statistic analysis, billing, reimbursement, automated decision support, and triggering alerts. To this end patient data has to be described by a physician (or professional encoder) by means of the terms and data-entry-formats available in the coding system to be used. Already in [11] it was argued that for this task is to be carried out adequately one would need:

- 1) a very high level of understanding of the meaning of the patient data (the sources),
- 2) a similarly high level of understanding of the meaning of the entries available in the coding system (the targets),
- 3) at least a certain level of similarity and coherence between sources and corresponding targets, and
- 4) facilities to search the coding system for the targets that match given sources as closely as possible.

For these four conditions to be realizable, however, the EHCR system needs the functionality to provide the clinician (or coding clerk) with assistance in the coding task [12], with the goal, in the long term of transforming the effort of coding free text patient data sources into a completely automatic process [13, 14].

Coding systems were originally designed to annotate data with a specific purpose (such as epidemiology, billing cost containment or insurance claims) in mind. Then, terminologies were developed with a slightly broader focus – as systems primarily designed to stabilize domain jargon. More recently still, we have domain ontologies, which are claimed to provide a greater level of formal rigour than coding systems or terminologies, and in a way that will make them understandable for software applications rather than for humans. Coding systems, terminologies and ontologies exist in many flavours. The Unified Medical Language System (UMLS), designed to “*facilitate the development of computer systems that behave as if they ‘understand’ the meaning of the language of biomedicine and health*” [15], contains over 100 such systems in its MetaThesaurus [16], which comprehends in all some 3 million medical concepts. In spite of this, however, it still cannot be said that complete coverage of the clinical language has been achieved. and this is so even when we focus on some one specific purpose [17]. Even selecting an appropriate coding system need not be an easy task [18, 19], and the same is true *a fortiorissimo* when it comes to building one from scratch [20], or to create mappings between them (table 1).

3.2 Formal methods for coding systems

For some years now it has been commonly accepted that both the development and use of clinical terminology in general and of coding systems in particular should be supported by formal methods. This is a thesis that we strongly support. But we wish no less strongly to

insist that formal methods alone are not enough. (Thus the use of a Description Logic-based system appears, for example, not to have provided any guarantee for the absence of errors in SNOMED-CT [21, 22, 23, 24].)

Source system	Target system(s)	References
ICD-10	ICD-10-AM V1	http://www2.fhs.usyd.edu.au/ncch/whats_new/ICD10toICD10amMappings/ICD-10%20to%20ICD-10-AM%20Mapping%20Tables.htm
ICD-10-AM V2	ICD-10-AM V1	http://www.nzhis.govt.nz/documentation/mapping/mappingdoc.html
ICD-10-AM V3	ICD-10-AM V2	http://www2.fhs.usyd.edu.au/ncch/whats_new/ICD10toICD10amMappings/ICD-10%20to%20ICD-10-AM%20Mapping%20Tables.htm
ICD-10-AM	ICD-9-CM-A	http://www.nzhis.govt.nz/documentation/mapping/mappingfiles.html
ICD-9-CM	All previous versions	http://www.cdc.gov/nchs/data/icd9/icdenv05.pdf
SNOMED-CT	ICD- O3 Map NANDA Nursing Map 2003-2004 NIC Nursing Map 4th Edition NOC Nursing Map 3rd Edition PNDS Nursing Map 2nd Edition HHCC Nursing Map Omaha System Nursing Map: ICD- 10 OPCS 4	http://www.snomed.org/products/content/mappings.html
SNOMED-CT	ICD-9-CM	http://informatics.iupui.edu/aca/undergraduate/hia/jones_presentation.ppt http://www.rivm.nl/who-fic/Colognepapers/cologne52.rtf
SNOMED-CT	LOINC	http://www.snomed.org/clinical/SNOMEDMapping.html
SNOMED-CT	PHIN-NCMT	http://www.cdc.gov/phin/data_models/
LOINC	PHIN-NCMT	http://www.cdc.gov/phin/data_models/
MEDCIN	CPT-4 ICD-9 ICD-10 ICD-O DSM-IV	http://www.medicomp.com/nomen.htm
NCVHS Minimum Data Set	HL 7	http://www.hl7.org/library/committees/govsig/NCVHS%20Mapping%20Report.doc
ICD-10-PCS	ICD-9-CM LOINC	http://www.cms.hhs.gov/providers/pufdownload/icd10.asp? http://www.cms.hhs.gov/paymentsystems/icd9/icd10.asp?
CDC reportable conditions	LOINC	http://www.cdc.gov/nedss/DataModels/dwyer-ii.pdf
ICPC2	ICD-10	http://www.ulb.ac.be/esp/wicc/icpc2.html

Table 1: Initiatives around coding system mappings and conversions

Our principal thesis, here, however, is that a role can and must be played by *realist ontology* in making better biomedical terminologies and coding systems. Realist ontology can not merely help in detecting errors and in ensuring intuitive principles for the creation and maintenance of coding systems of a sort that can help to prevent errors in the future. More importantly still, however, it can also help in ensuring that coding systems and terminologies are compatible with each other. Note that we say ‘*realist ontology*’, in order to distinguish

ontology in our understanding from the various related things [25] which go by this mean in contexts such as formal knowledge representation. It is a realist conception of ontology which underlies statements such as:

The UMLS is an extensive source of biomedical concepts. It also provides a large number of inter-concept relationships and qualifies for a source of semantic spaces in the biomedical domain. However, the organization of knowledge in the UMLS is not principled nor consistent enough for it to qualify as an ontology of the biomedical domain [26]

In the tradition of analytical philosophy, ontology is understood not as a software implementation or as a controlled vocabulary, but rather as ‘*the science of what is, of the kinds and structures of objects, properties, events, processes and relations in every area of reality*’ [27]. Ontology as it concerns us here is a theory of those higher-level categories which structure the biomedical domain, the representation of which needs to be both unified and fully coherent – and as closely allied as possible to the representations used by clinicians in formulating patient data – if terminologies and coding systems are to have the requisite degree and type of interoperability. Ontology in this realist sense is already being used as a method to find inconsistencies in terminologies and clinical knowledge representations [28] such as the Gene Ontology [29] or the UMLS Semantic Network [30]. The method has also proved useful in drawing attention to certain problematic features of the HL7 RIM [31, 32] – which brings us back to the electronic healthcare record.

4 Formal semantics

Biomedical terminologies or coding systems can only be compared amongst each other, or used without loss of information within an EHCR system, if they share a common framework of top-level ontological categories. Often one talks in this connection merely of a shared or common *semantics*, meaning thereby the sort of regimentation that can be ensured through the use of enabling technologies such as RDF(S) [33] and OWL [34] that currently enjoy a wide interest through their association with the Semantic Web project, not to forget systems such as Protégé that are able to cope with them in a user-friendly way [35]. On closer inspection, however, one discovers that the ‘semantics’ which comes with languages like RDF(S) and OWL is restricted to that sort of specification of meaning that can be effected using the formal technique of mathematical *model theory*, which is to say that meanings are specified by associating with the terms and sentences of a language certain abstract set-theoretic structures, taking Alfred Tarski’s ‘semantic’ definition of truth for artificial languages as paradigm [36]. Model theory assumes that the language refers to a ‘world’, and describes the minimal conditions that a world must satisfy in order for a ‘meaning’ (or ‘interpretation’ in the model-theoretic sense) to be assignable to every expression in the language. The idea is to provide an abstract mathematical account of the properties that any such interpretation must have in such a way as to make as few assumptions as possible about its actual nature or intrinsic structure, thereby retaining as much generality as possible. The chief utility of a formal semantic theory is thus not to provide any deep analysis of the nature of the things described by the language. Rather, the power of formal semantics resides at the logical level, above all in providing a technical way to determine when inferences are valid, i.e. when they preserve truth [37].

Model theory is thus metaphysically and ontologically almost completely neutral. Merely to formulate statements in a language such as OWL is far from building an ontology in the sense of ontology that is employed by analytical philosophers, and neither would translating a coding system into OWL turn it into an ontology. Such translation would indeed allow consistent reasoning about the ‘world’ – but only in the model-theoretic sense of ‘world’ that

signifies not the flesh-and-blood reality with which biomedicine is concerned, but rather merely only some highly simplified set-theoretic surrogate. The task of ensuring that the latter somehow corresponds in broad terms to the real world of what happens and is the case is, in the semantics literature, almost never addressed. In our view, however, this task – and indeed the whole detour via semantic models – is in fact superfluous: the job of ontology is not the construction of simplified models; rather, **an ontology should directly correspond to reality itself** in a manner that maximizes descriptive adequacy within the constraints of formal rigour and computational usefulness.

5 Basic Formal Ontology

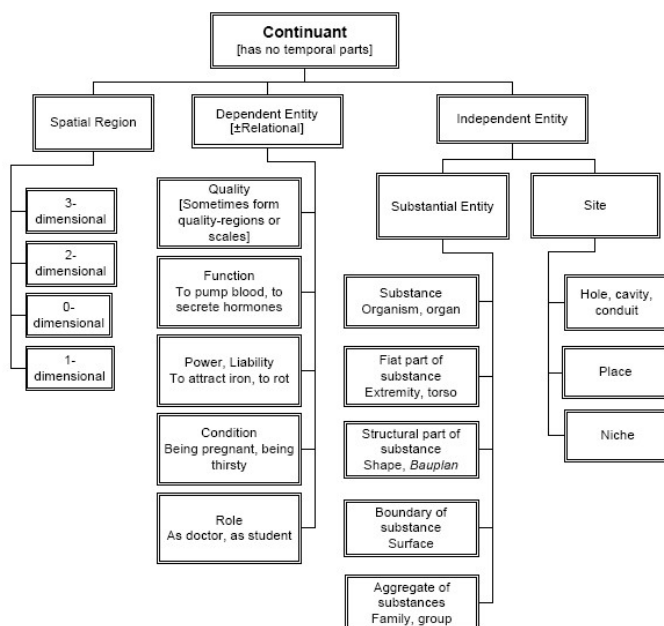
5.1 Main theory

What we need, therefore, is a sound realist ontological theory that relates medical language – formulated in whatever formal or informal language or coding system – to reality. Such a theory is Basic Formal Ontology (BFO) [38], which is one of several closely related ontological theories proposed in the recent literature (for a comparison see [39]).

Basic Formal Ontology consists in a series of sub-ontologies (most properly conceived as a series of *perspectives* or *windows* on reality), the most important of which are:

- SnapBFO, a series of snapshot ontologies (O_{t_i}), indexed by times (see figure)
- SpanBFO a single videoscopic ontology (O_v).

This corresponds to a fundamental division amongst entities between *continuants*, which are entities such as organisms or blood corpuscles which endure self-identically through time, and



occurents, which are processes such as heart bypass surgeries or increases in temperature which are divided into successive temporal phases. Each SnapBFO ontology O_{t_i} is an inventory of all continuant entities existing at the time t_i . Each SpanBFO ontology O_v is an inventory of all processes unfolding through time (standardly within some given temporal interval). Each O_{t_i} is a partition of the totality of objects and their continuant qualities, roles, functions. Each O_v is a partition of the totality of processes. We can say that O_{t_i} and O_v represent complementary perspectives on reality in the sense that (for technical reasons) continuants alone are visible

in the snapshot view, while the occurrent processes in which they are involved are visible only in the span view.

5.2 Granularity

Both SnapBFO and SpanBFO serve as basis for a series of sub-ontologies (windows on reality) at different levels of granularity. Thus the same portion of reality may appear at a plurality of different levels of coarse or fine grain. Masses at one level may be aggregates at another level. What is a tumour at one level is a constellation of cells or molecules at another

level. What counts as a unitary process at one level may be part of a process-continuum at another level. Since we can have no one single ontology which is a window on reality simultaneously at all levels of granularity, each of the ontologies here indicated is partial only: it is a window on just that portion of reality which is visible through the given ontology.

Snap- and SpanBFO acknowledge also sites and settings of continuant and occurrent entities, respectively, which is to say those wider contexts – referred to by expressions such as ‘in the room’, ‘in the lung’, ‘on the table’, ‘the Afghan winter’, ‘Tudor England’, etc. – in which for example organisms exist and processes of walking or fighting take place. Sites and settings are tied in each case to specific physical boundaries or retainers (such as walls, floors, ceilings). Sites are bound portions of space, which can be bound either completely, as in the case of a closed room or an air-bubble inside the human body, or partially, as in the case of a birdcage or nostril. Sites may retain their identity from one instant to the next even though they are projected in succession onto distinct abstract spatial regions – just as objects such as you or me retain their identity from one instant to the next even though they are projected in succession onto distinct aggregates of molecules. Settings are the four-dimensional counterparts of sites (thus they are, very roughly, the Cartesian products of sites with intervals of clock or calendar time) and are illustrated by behavior settings in Roger Barker’s sense (the 5pm train to Long Island, the early morning swim, your meeting with the Dean), with their accompanying physical-behavioural units [40].

Where objects such as cells, organs or planets fall within the natural world of constant causal change, there are also what we might call quasi-objects – such as a surgical team – which belong to the realm of social-political-administrative entities. The latter change in normal circumstances only as a result of administrative acts such as the admission of a new member. Where qualities, similarly, fall within the natural world of constant causal change, there are also quasi-qualities such as a rank or the authority to perform a certain type of medical procedure within a certain hospital. These, too, belong to the realm of social-political-administrative entities and they change (again in normal circumstances) only as a result of administrative acts such as demotion or dismissal from employment. All *roles* are quasi-entities (entities not subject to constant causal-energetic changes).

Processes are dependent entities on the side of occurrents: they exist always only as processes *of* or *in* some one or more independent continuants which are their bearers. Qualities, roles, functions, shapes, dispositions, powers are dependent entities on the side of continuants: they exist always only as the qualities (etc.) of specific independent continuants as their bearers: a smile smiles only in a human face.

Dependent entities, both within the SNAP and within the SPAN ontologies, are divided into relational (for entities dependent on a plurality of entities) and non-relational (for entities dependent on a single entity).

SnapBFO distinguishes in addition to objects also other sorts of independent continuant entities, including aggregates (your family, your teeth), boundaries (the outer surface of your skin), fiat parts (your arms and legs, your skin tissue), and structural parts (the organization of your body). It is in virtue of its structural parts that your body must have some color, temperature and mass, even though the specific color, temperature and mass of your body are not parts of the body but rather dependent continuants which depend upon your body as their substrate. Analogously, SpanBFO distinguishes structural parts of processes, which reflect the difference for example between cyclical and non-cyclical processes, between processes of acceleration and deceleration, and so on.

The BFO framework distinguishes finally (for present purposes) between ontology and epistemology; the former is concerned with representations of reality, the latter with our ways

of gaining knowledge of reality. These ways of gaining knowledge of reality can themselves be subjected to ontological treatment: they are processes of a certain sort, with cognitive agents as their bearers. This fact, however, should not lead us to confuse epistemological issues (pertaining to *what we can know*) with ontological issues (pertaining to *how the world is*).

6 An ontological basis for coding systems and the EHCR

Applying BFO to coding systems and EHCR architectures means in the first place applying it to those entities in reality to which these artifacts of the human intellect refer, such as concrete patients, diseases and therapies. We do this to serve at least one important goal, namely making coding systems coherent, both internally as well as in their relation to the EHCRs in or for which they are used.

It is essential to this endeavour that we give coding systems and EHCRs themselves their appropriate place in reality and that we understand their nature and purposes in terms of a coherent ontological theory. Although coding systems can themselves be viewed as (simple) models of a certain portion of reality, they are in fact as real as the words we speak or write and as the patterns in our brains. We thus use the BFO framework to analyze how coding systems, electronic healthcare records fit into a realist ontology, and thus also how they relate to the patients, physicians, diseases, etc. towards which they are directed.

Already a very superficial analysis of a coding system such as the International Classification of Diseases [41] reveals that this system is not in fact a classification of *diseases* as entities in reality. Rather it is a classification of *statements about disease phenomena which a physician might attribute to a patient*. As an example, the ICD-10 class *B83.9: Helminthiasis, unspecified* does not refer (for example) to a disease caused by a worm belonging to the species *unspecified* which would be some sub-species of *Acanthocephalia* or *Metastrongyilia*. Rather, it refers to a statement (perhaps appearing in some patient record) made by a physician who for whatever reason did not specify the actual type of *Helminth* the patient was suffering from. Neither OWL nor reasoners using models expressed in OWL would complain about making the class *B83.9: Helminthiasis, unspecified* a subclass of *B83: Other helminthiasis*; from the point of view of a coherent ontology, however, such a view is nonsense: it rests precisely on a confusion between ontology and epistemology [42].

A similar confusion can be found in EHCR architectures, model specifications, message specifications or data types for EHCR systems. References to a patient's gender/sex are a typical example. Some specifications, such as the Belgian KMEHR (Kind Messages for Electronic Healthcare Records) [43], refer to it as "administrative sex" (leaving it to the reader of the specification to determine what this might actually mean). The possible specifications of *administrative sex* are then *female, male, unknown, or changed* [44]. *Unknown*, here, does not refer to a new and special type of gender (reflecting some novel scientific discovery); rather it refers to the fact that the actual gender is not documented in the record.

An interpretation along these lines does not work in every case, however. Consider those specifications which refer explicitly to "clinical observations", as is the case for Corbamed-COAS ("Clinical Observations Access Server"), which consists of:

any information that has been captured about a single patient's medical/physical state and relevant context information. This [information] may be derived by instruments such as in the case of images, vital signs, and lab results or it may be derived by a health professional via direct examination of the patient and transcribed [sic]. This term applies

to information that has been captured whether or not it has been reviewed by an appropriate authority to confirm its applicability to the patient record. [45]

When in a EHCR system that claims to follow the COAS specifications the specification “*unknown*” would be registered for gender, then that specification has to be interpreted that an observation has been made with respect to the patient’s gender, and that as a result of that, an unknown kind of gender has been observed. Of course, that is not supposed to be the idea.

7 Conclusions and recommendations

European and international efforts towards standardization of biomedical terminology and electronic healthcare records were focused over the last 15 years primarily on syntax. Semantic standardization was restricted to terminological issues around the semantic triangle paradigm [46] on the one hand and to issues pertaining to knowledge representation (and resting primarily on the application of set-theoretic model theory) on the other hand. Moves in these directions are in indeed required, and the results obtained thus far are of value both for the advance of science and for the concrete use of healthcare telematic applications. We can safely say that the syntactical issues are now resolved and also that the problems relating to biomedical terminology (polysemy, synonymy, cross-mapping of terminologies, ...) are well understood – at least in the community of specialized researchers. Now, however, it is time to solve these problems by using the theories and tools that have been developed so far, and that have been tested under laboratory conditions [28]. This means using the right sort of ontology, i.e. an ontology that is able explicitly and unambiguously to relate coding systems, biomedical terminologies and electronic health care records (including their architecture) to the real world.

To do this properly will require a huge effort, since the relevant existing standards need to be reviewed by experts who are familiar with the appropriate sort of ontological thinking (and this will require some effort in training and education). Even before that stage is reached, however, there is the problem of making all constituent parties – including patients (or at least the organizations that stand up for them), healthcare providers, system developers and decision makers – aware of how deep-seated the existing problems are. Having been overwhelmed by the exaggerated claims on behalf of XLM and similar silver bullets of recent years, that would solve everything, they must be informed about the fact that XML alone isn’t a silver bullet. And for sure, we must also be careful in not giving realist ontology a similar silver bullet status.

In Europe, the CEC’s sixth and seventh Framework Programs might provide good opportunities, and undoubtedly, similar initiatives can be found in the US, in Australia and in the Far East. Collaboration at an international level is in any case required if we want systems developed in different places to be of any value for those that did not contribute to their development.

The message of realist ontology is that, while there are various different views of the world, this world itself is one and unique. It is our belief that it is only through that world that the various different views can be compared and made compatible. To allow clinical data registered in electronic patient records by means of coding (and/or classification) systems to be used for further automated processing, it should be crystal clear whether entities in the coding system refer to diseases or rather to statements made about diseases, or to procedures and observations, rather than statements about procedures or observations. As such, coding systems used in or for electronic healthcare records should be given a precise and formal semantics that is coherent with the semantics of the record as well as with the real world parts that are described by them.

8 References

- [1] Redondo JR, Ceusters W, González JM, Iakovidis I. European Electronic Healthcare Records towards the Future. In: Laires MF, Ladeira MJ, Christensen JP (eds.) *Health in the New Communications Age*, 671-675, IOS Press 1995.
- [2] <http://www.sadiel.es/europa/prorec/>
- [3] <http://www.sadiel.es/Europa/widenet2/default.asp#>
- [4] <http://www.sadiel.es/Europa/widenet2/newsf.htm>
- [5] Waegemann CP. Status Report 2002: Electronic Health Records (<http://www.medrecinst.com/uploadedFiles/MRILibrary/StatusReport.pdf>).
- [6] ISO 18308: Health informatics -- Requirements for an electronic health record architecture. 2002 (<http://www.iso.ch/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=33397>)
- [7] CR12587: CEN Report: Medical Informatics - Methodology for the development of healthcare messages 1996 (<http://www.centc251.org/secret/ENV/12587/12587.htm>)
- [8] Harris MR, Chute CG, Harvell J, White A, Moore T. Toward a National Health Information Infrastructure: A Key Strategy for Improving Quality in Long-Term Care: Executive Summary, May 2003 (<http://aspe.hhs.gov/daltcp/reports/towards.htm>)
- [9] Rector AL, Nowlan WA, Kay S. Foundations for an electronic medical record. *Meth Inform Med* 30: 179-186, 1991.
- [10] Rector AL, Nowlan WA, Kay S, Goble CA, Howkins TJ. A framework for modelling the Electronic Medical Record. *Meth Inform Med* 32: 109-119, 1993.
- [11] Ceusters W, Lovis C, Rector A, Baud R. Natural language processing tools for the computerised patient record: present and future. In P. Waegemann (ed.) *Toward an Electronic Health Record Europe '96 Proceedings*, 294-300, 1996.
- [12] Lovis C, Michel PA, Baud R, Scherrer JR. Use of a semi-automatic conceptual ICD-9 encoding system in an Hospital Environment. In P. Barahona et al. Eds, *Lecture Notes in Artificial Intelligence*, Springer-Verlag, 1995.
- [13] Ceusters W, Jan Lorré, Annelies Harnie, Bert Van Den Bossche. Developing natural language understanding applications for healthcare: a case study on interpreting drug therapy information from discharge summaries. *Proceedings of IMIA-WG6, Medical Concept and Language Representation, Phoenix*, 16-19/12/1999, 124-130.
- [14] Ceusters W. Medical natural language understanding as a supporting technology for data mining in healthcare. In: KJ Cios (ed.) *Medical Data Mining and Knowledge Discovery*, Physica-verlag Heidelberg, New York, 41-67, 2001.
- [15] National Library of Medicine; UMLS fact sheet, updated 7 May 2004 (<http://www.nlm.nih.gov/pubs/factsheets/umls.html>)
- [16] National Library of Medicine; UMLS MetaThesaurus fact sheet, updated 7 May 2004. (<http://www.nlm.nih.gov/pubs/factsheets/umlsmeta.html>)
- [17] Harris MR, Graves JR, Herrick LM, Elkin PL, Chute CG. The content coverage and organizational structure of terminologies: The example of postoperative pain. *Proceedings of the AMIA Annual Symposium 2000*; 335-339.
- [18] SAGOR - Scottish Advisory Group on Read. *Choosing a Coding System for Clinical Data: An approach to option appraisal*. First Edition, 1997.
- [19] Campbell JR, Carpenter P, Sneiderman C, Cohn S, Chute CG, Warren J, and CPRI Work Group on Codes and Structures. Phase II Evaluation of Clinical Coding Schemes: Completeness, Taxonomy, Mapping, Definitions, and Clarity. *J Am Med Inform Assoc*. 1997 May; 4 (3): 238-250.
- [20] Rector AL. Clinical Terminology: Why is it so Hard? *Method Inform Med* 1999; 38: 239-52
- [21] Ceusters W, Smith B. Ontology and Medical Terminology: why Descriptions Logics are not enough. *Proceedings of the conference Towards an Electronic Patient Record (TEPR 2003)*, San Antonio, 10-14 May 2003 (electronic publication).

- [22] Ceusters W, Smith B, Kumar A, Dhaen C. Mistakes in Medical Ontologies: Where Do They Come From and How Can They Be Detected? in Pisanelli DM (ed) "Ontologies in Medicine. Proceedings of the Workshop on Medical Ontologies, Rome October 2003" IOS Press, Studies in Health Technology and Informatics, vol 102, 2004.
- [23] Ceusters W, Smith B, Kumar A, Dhaen C. Ontology-Based Error Detection in SNOMED-CT® Proc. Medinfo 2004.
- [24] Bodenreider O, Smith B, Kumar A, Burgun A. [Investigating subsumption in DL-based terminologies: A case study in SNOMED CT](#). In: Hahn U, Schulz S, Cornet R, editors. Proceedings of the First International Workshop on Formal Biomedical Knowledge Representation (KR-MED 2004); 2004. p. 12-20.
- [25] N. Guarino, P. Giaretta, "Ontologies and Knowledge Bases: Towards a Terminological Clarification". In Towards Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing, N. Mars (ed.), pp 25-32. IOS Press, Amsterdam, 1995.
- [26] Bodenreider O. Medical Ontology Research: A Report to the Board of Scientific Counselors of the Lister Hill National Center for Biomedical Communications. May 17, 2001 (<http://etbsun2.nlm.nih.gov:8000/pubs/pdf/2001-MOR-BoSC.pdf>)
- [27] Smith B. Ontology and Information Systems (draft 2004): <http://ontology.buffalo.edu/smith/articles/ontologies.htm>
- [28] Smith B, Ceusters W. Towards Industrial-Strength Philosophy; How Analytical Ontology Can Help Medical Informatics. Interdisciplinary Science Reviews, 2003, vol 28, no 2, 106-111.
- [29] Barry Smith, Jacob Köhler, Anand Kumar: On the Application of Formal Principles to Life Science Data: a Case Study in the Gene Ontology. In: Erhard Rahm (Ed.): Data Integration in the Life Sciences, First International Workshop, DILS 2004, Leipzig, Germany, March 25-26, 2004, Proceedings. Lecture Notes in Computer Science 2994, Springer 2004, 79-94.
- [30] Schulze-Kremer S, Smith B, Kumar A. Revising the UMLS Semantic Network Medinfo 2004. Proc. Medinfo 2004.
- [31] Lowell V: Actions in Health Care Organizations: An Ontological Analysis in: Proceedings of MedInfo 2004, San Francisco.
- [32] Lowell V, Smith B: Speech Acts and Medical Records: The Ontological Nexus. in: EuroMISE 2004, Prague.
- [33] RDF Semantics. W3C Recommendation 10 February 2004 (<http://www.w3.org/TR/rdf-mt/>)
- [34] OWL Web Ontology Language Semantics and Abstract Syntax. W3C Recommendation 10 February 2004 (<http://www.w3.org/TR/owl-semantics/>)
- [35] Protégé OWL plug-in (<http://protege.stanford.edu/plugins/owl/>)
- [36] Model Theory. Stanford Encyclopedia of Philosophy (<http://plato.stanford.edu/entries/model-theory/>)
- [37] Guha RV, Hayes P. LBase: Semantics for Languages of the Semantic Web. NOT-A-Note 02 Aug 2002 (<http://www.coginst.uwf.edu/~phayes/LBase-from-W3C.html>)
- [38] Smith B. BFO/MedO: Basic Formal Ontology and Medical Ontology. July 2003. (<http://ontology.buffalo.edu/bfo/BFO.htm>)
- [39] Masolo C, Borgo S, Gangemi a, Guarino N, Oltramari A. WonderWeb Deliverable D18: Ontology Library (final) (<http://wonderweb.semanticweb.org/deliverables/documents/D18.pdf>)
- [40] Smith B. Social Objects (<http://ontology.buffalo.edu/socobj.htm>)
- [41] World Health Organisation. ICD-10 - The International Statistical Classification of Diseases and Related Health Problems, tenth revision (<http://www.who.int/whosis/icd10/>).
- [42] Bodenreider O, Smith B, Burgun A. The Ontology-Epistemology Divide: Case Study in Medical Terminology. Submitted to the Third International Conference on Formal Ontology (FOIS) 2004.
- [43] Kmehr-Bis : Kind messages for electronic healthcare record, Belgian implementation standard (<http://www.chu-charleroi.be/kmehr/htm/kmehr.htm>)
- [44] <http://www.chu-charleroi.be/kmehr/xml/CD-SEX.xml>

- [45] 3M, Care Data Systems, Inc., CareFlow/Net, Inc., HBO & Company, Philips Medical Systems, Protocol Systems, Inc. CORBAMED-COAS: clinical observations access server specification. Version 1, April 2001. (<http://www.medcom.dk/picnic/deliverables/01-04-06%20coas%20specs.pdf>)
- [46] Ogden, C.K., & Richards, I.A. (1927). Meaning of meaning. New York: Harcourt, Brace & Company.