IMISE-REPORTS
Herausgegeben von Professor Dr. Markus Löffler

H. Herre, R. Hoehndorf, F. Loebe (Eds.)

OBML 2011
Workshop Proceedings

Berlin, October 6-7, 2011

IMISE-REPORT Nr. 1/2011

UNIVERSITÄT LEIPZIG
Medizinische Fakultät

Impressum

Herausgeber: Universität Leipzig
Medizinische Fakultät
Institut für Medizinische Informatik, Statistik und Epidemiologie (IMISE)
Härtelstraße 16-18, 04107 Leipzig
Prof. Dr. Markus Löffler

Editoren: Heinrich Herre, Robert Hoehndorf, Frank Loebe
Redakteur: Frank Loebe

Kontakt: Telefon: (0341) 97-16100, Fax: (0341) 97-16109
Internet: http://www.imise.uni-leipzig.de

Redaktionsschluss: 25. September 2011

Druck: Inhalt: Universitätsklinikum Leipzig AöR, Bereich 2 - Abteilung Zentrale Vervielfältigung/Formularwesen
Einband: Buch- und Offsetdruckerei Herbert Kirsten

ISSN 1610-7233

Proceedings of the

3rd WORKSHOP OF THE

GI WORKGROUP "ONTOLOGIES IN BIOMEDICINE AND LIFE SCIENCES" (OBML)

Berlin, Germany
October 6-7, 2011

Group Website: https://wiki.imise.uni-leipzig.de/Gruppen/OBML

Organizers

Heinrich Herre	(*coordinator*)	University of Leipzig
Robert Hoehndorf		University of Cambridge, UK
Frank Loebe		University of Leipzig

Local Organizer

Peter Robinson	Charité Berlin

Keynote Speakers

Dietrich Rebholz-Schuhmann	European Bioinformatics Institute, Cambridge, UK
Peter Robinson	Charité Berlin

Program Committee

Robert Hoehndorf	(*chair*)	University of Cambridge, UK
Sören Auer		University of Leipzig
Franz Baader		Technical University Dresden
Martin Boeker		University Medical Center Freiburg
Patryk Burek		University of Leipzig
Fred Freitas		Federal University of Pernambuco, Recife, Brazil
Georgios V. Gkoutos		University of Cambridge, UK
Giancarlo Guizzardi		Federal University of Espirito Santo, Brazil
Heinrich Herre		University of Leipzig
Josef Ingenerf		University of Lübeck
Ludger Jansen		University of Rostock
Janet Kelso		Max Planck Institute for Evolutionary Anthropology, Leipzig
Toralf Kirsten		University of Leipzig
Frank Loebe		University of Leipzig
Axel Ngonga-Ngomo		University of Leipzig
Roberto Poli		University of Trento, Italy
Dietrich Rebholz-Schuhmann		European Bioinformatics Institute, Cambridge, UK
Peter Robinson		Charité Berlin
Michael Schröder		Technical University Dresden
Stefan Schulz		Medical University of Graz, Austria
Luca Toldo		Merck KGaA

Additional Reviewer

Anika Oellrich	European Bioinformatics Institute, Cambridge, UK

Authors

Maurício Almeida	Federal University of Minas Gerais, Brazil
André Andrade	Federal University of Minas Gerais, Brazil
Martin Boeker	University Medical Center Freiburg
Cristine Bonfim	Joaquim Nabuco Foundation, Recife, Brazil
Roberta Fernandes	Federal University of Pernambuco, Recife, Brazil
Fred Freitas	Federal University of Pernambuco, Recife, Brazil
Georgios V. Gkoutos	University of Cambridge, UK
Niels Grewe	University of Rostock
Heinrich Herre	University of Leipzig
Robert Hoehndorf	University of Cambridge, UK
Ludger Jansen	University of Rostock
Oliver Kutz	University of Bremen
Frank Loebe	University of Leipzig
Zulma Medeiros	Federal University of Pernambuco, Recife, Brazil
Anika Oellrich	European Bioinformatics Institute, Cambridge, UK
Djamila Raufie	University Medical Center Freiburg
Dietrich Rebholz-Schuhmann	European Bioinformatics Institute, Cambridge, UK
Norbert Reinsch	Leibniz Institute for Farm Animal Biology, Dummerstorf
Johannes Röhl	University of Rostock
Filipe Santana	Federal University of Pernambuco, Recife, Brazil
Daniel Schober	University Medical Center Freiburg
Stefan Schulz	Medical University of Graz, Austria
Aleksandra Sojic	European School of Molecular Medicine, Milan, Italy
Frank Stumpf	University of Leipzig
Vojtech Svatek	University of Economics, Prague, Czech Republic
Silke Trißl	Leibniz Institute for Farm Animal Biology, Dummerstorf
Ilinca Tudose	University Medical Center Freiburg

Information, reality and epistemology: an ontological take

Maurício Almeida and André Andrade*

School of Information Science – Federal University of *Minas Gerais* – Brazil

ABSTRACT

Medical records are crucial resources for the whole healthcare practice. The amount and complexity of information they bear require the use of automation. In this paper we present a framework to represent information recorded in medical records, drawing on Popper three worlds theory. Then, we test such a framework by using a description of a real clinical case. Finally, we offer recommendations of how data can be properly arranged in order to incorporate assorted representations like ontologies, information models and reasoning rules.

1 INTRODUCTION

The medical record is a complex document employed for several purposes in the healthcare realm. Proper documentation of medical encounters is an important task of a physician's activity. Medical records have a myriad of uses in healthcare processes, such as [11]:

- to support patient care: to remind staff of and communicate information, to help in organizing the care process (e.g. care information used in process coordination, clinical decision making, patient demographics);
- to fulfill external obligations: legal requirements, accreditation, reimbursement regulations (e.g. procedure coding), order documentation (e.g. exams, medication), and events (including adverse events, surgeries, sample collections);
- to support administration: in planning, controlling, and refunding the health care institution's services (e.g. medication and medical materials used, equipment use, procedure coding, diagnostic coding, name of professionals);
- to support quality management: by enabling critical assessment and systematic monitoring of processes (e.g. clinical outcomes);
- to support scientific research: by enabling patient selection and statistical analysis (e.g. possibly relevant clinical information, not yet used in clinical reasoning, according to research protocols);
- to support clinical education: by providing information for critical review and case examples (e.g. contextual information about consultation setting).

* To whom correspondence must be addressed.

As a consequence of those multiple uses, medical information is a mix of facts, impressions, measurements, rules, and knowledge recording. A classification of kinds of information is required for automatic processing by computers, as well for system interoperability.

Formal ontology allows robust reasoning, but restricts representation to reality entities. Non-realist information (called here "epistemological information") [5] comprises some representations of symptoms, since there is no way to ascertain the truth value of these assertions: "Neither signs nor symptoms form a natural kind, but are rather composite classes – fiat collections of bodily features delineated by certain socially established cognitive practices on the parts of clinicians and patients" [20].

In order to integrate realist and epistemological oriented information, one must clearly define what such kinds of information mean in medical records, why they are important and which sort of automatic operations they should support. Moreover, a clear separation between, on one side, the entities in reality, and on the other side, the information about them, makes easier the understanding of medical records, allowing different logical operations and the use for different purposes.

The goal of the present paper is to explore better approaches to represent information registered in medical records, taking advantage of the best characteristics of well-known techniques. In seeking such goal, we rely on philosophical grounds in order to create a framework of analysis. Then, we test the framework against a sample of medical records distinguishing within it: i) references concerning real entities; ii) reference concerning epistemological entities; iii) other kinds of information contained in the record that are relevant to the clinical practice. Finally, we offer a proposal of a general arrangement encompassing all those representations into information systems.

2 METHODOLOGY

In order to reach better possibilities of medical record representation, we need to organize the kinds of information they enclose. We here take advantage of well-known techniques for dealing with medical information, like ontologies and information models. The methodology is composed by the following steps.

First, we develop a framework of analysis, which draws on inputs from philosophy, particularly, from Karl Popper´s

three worlds and its usefulness in health information science [3]. We also consider recent researches on the medical ontologies, namely, the Basic Formal Ontology (BFO) [10] the Ontology of General Medical Science (OGMS) [20] and the Information Artifact Ontology (IAO) [12]. Despite empirical evidences suggesting the feasibility of such approaches, different views can be found in the literature [13, 17]. In addition, we take into account other significant advances in binding realist ontologies and information models.
As second step, we test such framework over a complex clinical history developed by The New England Journal of Medicine. We choose that source for didactic reasons, benefiting from the journal´s academic focus, which summarizes clinically useful information. Everything represented in the summary is important to physicians and therefore all entities are considered in our scope of computer processable information. The record was analyzed with the aim of identifying underlying propositions.
In order to identify propositions, a domain expert transcribed the records in sentential fragments that make sense for him. The domain expert was asked to identify the reason for recording those entities and the information that is being conveyed by the representation. The transcription draws upon principles of logic and controlled languages [8, 9], which allowed identification of entities recorded in natural language, outside the particular context in which the event took place [23]. In addition, on the classification side, we use the rationale underpinning OGMS. On the logical side, we took in account that some natural language parts of speech do not have room in logical statements. Even though this is a well-known fact, for example with respect to an adverb, the very same one may be relevant to characterize a clinical situation.
Finally, we took apart the found according to their suitability to each approach. Thus, we organize the information of the medical record in four kinds, which are so employed to recommend both a data arrangement and a scenario of collaboration among different representations.

3 FRAMEWORK OF ANALYSIS

There is no consensus about the best way to represent the myriad of situations that occurs during a medical encounter. A useful approach relating reality, cognition and representations was proposed by Popper in his theory of three worlds [15]. Those worlds are described as follows:

World 1: the physical world;
World 2: the world of mental states;
World 3: the world of contents of thought.

Within those three worlds, objects are real on their own and each one can modify each other. One example is the learning of a new language, which is a modification of World 2 (the process of learning) by a World 3 entity (language itself). Popper's theories receive critics [3], but also favorable claims in which it is considered a useful model to understand epistemic information [1].Accordingly, one can find additions and improvements of the Popper´s views, which propose additional sub-divisions into the original layers [4, 14]. A complete discussion of such a theory is, though, beyond the goals of this paper.
Then, we propose a framework of analysis as depicted in Fig. 1, which was created to organize information according the best possibility of representation.

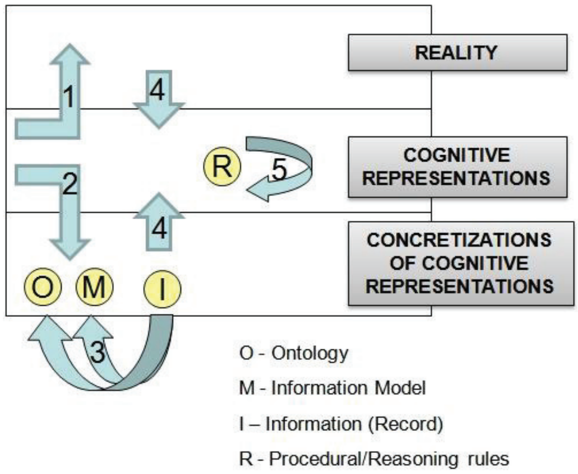

Figure 1 – Framework use for analysis

In this framework, everything begins at level of cognitive representations when a physician observes the reality at the patient side (arrow 1). Each of these entities are filtered by cognition and represented by 2 artifacts (arrow 2). Ontological entities (entity O) are analyzed according to strict philosophical tenets, and are based on reality itself rather than mental representations of a physician. Examples of ontological entities are cells, anatomical features and chemical substances. Information model entities (entity M) stand for cognitive representations of reality, and may include entities without a referent in reality. Examples of these include "severity" of pain and a "feeling well" sensation. Then, the physician creates a record (entity I) to register those representations according to his practical and theoretical knowledge (arrow 3). Constantly, other physicians can interpret records and reality (arrow 4), resulting in new cognitive representations. Finally, the physicians involved in healthcare make judgments, and process past and current information. Some of this processing of information (arrow 5) follows medical training rules, which determine the likelihood of a diagnosis, the correct interpretation of a exam result, to mention but a few. The representation of this process of reasoning is also required for care continuation, a complementary part of the record (entity R). Examples of this include rules to interpret lab data, as hemoglobin level < 12 g/dl means "low hemoglobin level"; and relevant negative information such as "lack of bowel alteration during episodes".

Our approach to the model is rather more pragmatic – our goal here is to establish a methodology to distinguish real entities from epistemological entities represented as information entities in the World 3. Within the framework created we recognize at least four kinds of information to be separated according to their suitability for information systems:

- Data that represent aspects of the reality;
- Data that represent useful constructs for the medical practice not empirically verifiable;
- Data that represent observations about the reality, not reality itself;
- Data that represent observations about the physician understanding of the clinical situation, not about reality.

Under this model, we still contend to the fact that neither representations of the reality nor representations of thought processes are interpreted in the same way by two people. However, allowing manipulation of the World 3 entities is fundamental for the development of new features in medical systems, such as decision support, inferences and information classification, and discovery.

4 TEST OF THE FRAMEWORK

We here make a preliminary test of the framework by analyzing individual information entities contained in medical records. Figure 2 depicts a small extract of the clinical case available at http://www.nejm.org/multimedia/interactive-medical-case [19]. Once we obtain a sentential fragment from a domain expert evaluation, we thus isolate what could be represented in realism-based ontologies following the rationale of OGMS, BFO and AIO. After that, we arrange other information according to kinds mentioned in section 3. The final results systematize the information contained in a medical record in keeping with the information system that it is suitable for.

> "An 88-year-old woman presented to the emergency room with confusion. She began having transient episodes of confusion, dizziness, tremors and anxiety a year earlier. These episodes were unpredictable, lasting for minutes and then abating spontaneously, and had been increasing in frequency since they began. The patient felt well between episodes and reported no abnormal sensation, change in weight, or relation of symptoms to meals, fasting or physical activity."

Figure 2 – Extract of the medical history

In what follows, we present samples of data obtained from the medical record and classified according to kinds proposed in section 3. Fig. 3, 4, 5 and 6 depict such samples.

Data representing aspects of the reality
Physician (BFO Role)
Woman (BFO - Object)
88 years-old (BFO Quality)
Patient report (AOI Information Content Entity)
Confusion, dizziness, tremor (OGMS symptom)
Duration of episodes (BFO temporal region)
Time between episodes (BFO temporal region)
Change in weight (OGMS symptom)
Aspirin (BFO continuant)
Aspirin taken daily (AIO rule)
Physical exam finding of that encounter (OGMS Physical examination finding)
Glucose (BFO Continuant)
Diagnosis of hypoglycemia (OGMS diagnosis)
Insulinoma (BFO continuant)

Figure 3 – Data sample: realist bias

Data that represent useful constructs for the medical practice
… transient episodes of confusion, dizziness, tremors, and anxiety a year earlier (each episode being correlated as caused by a single entity)
No abnormal sensation
… episodes are unpredictable
Confusion
General: well appearing
Chest: clear to auscultation
Abdomen: soft and nontender

Figure 4 – Data sample: data not empirically verifiable

Data that represents observations about the reality
Frequency of episodes
Increase in the frequency of episodes
36º of temperature
76 beats per minute
114/60 mmHg
Glucose concentration
Aspirin dosage

Figure 5 – data sample: observation of the reality

Data that represents observations about the physicians understanding
Insulinoma causing hypoglycemia
Relation symptoms vs. meals

Figure 6 – Data sample: observations of one´s understanding, not reality itself

This data classification was based on both the levels of representation provided in section 2 and the explanation provided in section 3. From the empirical assessment by physicians, the categories suggested from Fig. 3 to Fig. 6

were created. The relation between the proposed framework and the organization of data from medical records can be summarized as follows:

- "Data representing aspects of reality" (Fig. 3) were mapped from processes (1) and (2) to entities (O) (Fig.1);
- "Data that represent useful constructs for the medical practice" (Fig.4) were mapped from the process (1) and (2) to entities (M) (Fig.1);
- "Data that represents observations about the reality" (Fig. 5) were mapped from process (3) to entities (I) (Fig.1);
- "Data that represents observations about the physicians understanding" (Fig. 6) were from processes (4) mapped to entities (R) (Fig.1).

According to the scenario developed so far, we propose a data arrangement to deal properly with all these kinds of data. The data could then be processed by the suitable system and the equivalent representation. The arrangement of data and a scenario of collaboration different systems are depicted in Fig.7:

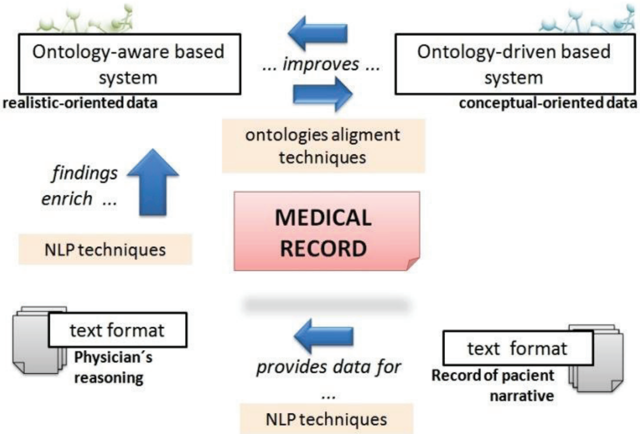

Figure 7 – Final arrangement of medical data

5 DISCUSSION

Medical practice is still heavily grounded in the study of signs and symptoms, which are interpreted by a physician in the search for a diagnosis about a clinical situation. Medical reasoning is a sum of different cognitive practices including induction, abduction and deduction [16]. As a written representation of a medical encounter, a medical record is closely related to medical reasoning practices and how a physician understands pathological process at that time. However, interpretation by computers and semantic interoperability [2] require explicit and shared definition of terms, in order that they can be manipulated without information loss. Our framework attempts make it possible by making clearer the distinctions between reality, medical understanding and the recording of it, while maintaining the medical record as the main data source.

The first important distinction is the separation between information about reality and reality itself. In the context of medical practice, common mistakes have been found when there is no separation of information [6] [22]. For example, a "cancelled surgery" is not a "surgery" which never existed, but a "plan" for a "surgery" which had the content modified". This strategy can be used to allow proper representation of non-existing entities in DL-logic, the most currently embraced family of logic, while maintaining consistency and coherence with realist ontologies tenets [21]. Also, we follow [18] in the description of clinical situations for a useful way of expressing modality, temporal status of symptoms, signs or diseases, and for representing the "subject of care" construct, which is very important for care processes reasons.

Even though one can consider such distinction trivial, we observe that medical systems are currently being developed world-wide without consider what "sort of data" is suitable to what "sort of system". We believe that the commitment to standards addresses only a small piece of the interoperability problem. Our belief is based on the observation that such interoperability problem is yet critical today, despite several standards that have been proposed throughout the years. Our arrangement is an attempt to explore other possibilities of representation for different kinds of systems, considering the real collaboration among them (Fig.7).

With this arrangement proposal, we suggest that it is useful to distinguish what should and what shouldn't be rigorously represented as ontological entities. Our framework suggests that, while the medication (and the analyzed sample) and chemical entities (blood and glucose) are real entities, the results are in fact information about them (data items). For example, the blood glucose measurement refers to the glucose blood concentration at the exact moment of blood sample collection. It is, therefore, empirically verifiable. However, the value of the measurement does not refer to the existence of the enzyme in the real world, and the same entity in reality may be described using different measurement units, laboratory methods and confidence intervals. Besides, the information is analyzed using a sequence of pre-established thinking rules, according to clinical training. For example, in the clinical case present in section 4, the value 40 mg/dl is below normal values (80 mg/dl) and, therefore, suggests the diagnosis of hypoglycemia. It will be interpreted according to a reasoning rule, not according to the structure of reality itself, and is suitable to procedural operations instead of pure logical reasoning. The same reasoning principle holds for other kinds of rates (beats/minute, mg/kg). It is important to emphasize that these rules of interpretation (normality levels) are also based on historical events which had an almost arbitrary definition of normality [23], and may change at any time.

A pragmatic look at the medical record and the categories of our analysis has shown some aspects that current medical systems solve reasonably well using current relational databases, such as medication dosage and laboratory analysis results. Computerized Patient Order Entry (CPOE) systems are relatively widespread and have successfully replaced free-text orders, though actual improvements in healthcare processes haven't yet come to full extent [7].

Finally, one can argue about examples presented (section 4). For example, the entity "confusion" is exhibited both in data with realist bias (Fig. 3) and data not empirically verifiable (Fig. 4). However, the former represents a condition as reported by a patient; the latter represents a physician´s perception of a patient condition. Also, in Fig. 3, one can claim that diagnosis is not exactly entity pertaining to reality. However, in our framework based in OGMS, a diagnosis is taken as a data record.

6 FINAL REMARKS

In this paper, we present a distinction that includes the representation of entities referring to the physician's or the patient's understanding of a situation. Otherwise said, they represent reality as seen and interpreted by a human being, therefore not objective statements. In the focused record, we found many instances where this distinction is beneficial. We argue that, by separating entities as proposed, one is able to safely talk about the clinical case without harming the interoperability of the medical record. We also suggest an arrangement able to encompass these proposed kinds of data and related systems.

The presented approach is an attempt towards the clarification of critical aspects of data categories. Certainly, it needs more progress in order to have direct impact on the interoperability issue. As future work, we intend to create clear rules to divide kinds of information in a semi-automatic fashion. Then, it will be possible to test our approach against a greater sample. In seeking this, we aim to explore the best characteristics of different systems and data representations.

REFERENCES

1. Abbott, R. Subjectivity as a concern for information science: a Popperian perspective. *Journal of Information Science*, *30* (2). 95-106.
2. Almeida, M.B., Souza, R.R. and Fonseca, F. Semantics in the Semantic Web: a critical evaluation. *Knowledge Organization Journal*.
3. Bawden, D. The three worlds of health information. *Journal of Information Science*, *28* (1). 51-62.
4. Bhaskar, R. *A Realist Theory of Science*. Harvester Press, Sussex, 1978.
5. Bodenreider, O., Smith, B. and Burgun, A., The Ontology-Epistemology Divide:A Case Study in Medical Terminology. in *3rd Conference on Formal Ontology in Information Systems*, (Turin, 2004).
6. Brinkman, R.R., Courtot, M., Derom, D., Fostel, J.M., He, Y., Lord, P., Malone, J., Parkinson, H., Peters, B., Rocca-Serra, P., Ruttenberg, A., Sansone, S.A., Soldatova, L.N., Stoeckert, C.J., Jr., Turner, J.A. and Zheng, J. Modeling biomedical experimental processes with OBI. *J Biomed Semantics*, *1 Suppl 1* (22). S7.
7. Eslami, S., de Keizer, N.F. and Abu-Hanna, A. The impact of computerized physician medication order entry in hospitalized patients--a systematic review. *Int J Med Inform*, *77* (6). 365-376.
8. Fuchs, N.E., Hofler, S., Kaljurand, K., Rinaldi, F. and Schneider, G. Attempto controlled english: A knowledge representation language readable by humans and machines. *Reasoning Web*, *3564*. 213-250.
9. Fuchs, N.E., Schwertel, U. and Torge, S. A Natural Language Front-End to Automatic Verification and Validation of Specifications, LMU München, 1999.
10. Grenon, P., Smith, B. and Goldberg, L. Biodynamic ontology: applying BFO in the biomedical domain. *Stud Health Technol Inform*, *102*. 20-38.
11. Haux, R., Knaup, P. and Leiner, F. On educating about medical data management - The other side of the electronic health record. *Methods of Information in Medicine*, *46* (1). 74-79.
12. IAO, 2011. Information Artifact Ontology. http://code.google.com/p/information-artifact-ontology/
13. Merrill, G.H. Ontological realism: Methodology or misdirection? *Applied Ontology*, *5*. 79-108.
14. Niiniluoto, I. *Critical scientific realism*. Oxford University Press, New York, 1999.
15. Popper, K.R. and Eccles, J.C. *The Self and Its Brain: An Argument for Interactionism*. Routledge, 1977.
16. Pottier, P. and Planchon, B. Description of the mental processes occurring during clinical reasoning. *Revue De Medecine Interne*, *32* (6). 383-390.
17. Rector, A. Knowledge Driven Software and "Fractal Tailoring": Ontologies in development environments for clinical systems *Proceeding of the 2010 conference on Formal Ontology in Information Systems: Proceedings of the Sixth International Conference (FOIS 2010)*, IOS Press, Toronto, Canada, 2010.
18. Rector, A.L. and Brandt, S. Why Do It the Hard Way? The Case for an Expressive Description Log-

ic for SNOMED. *Journal of the American Medical Informatics Association*, *15* (6). 744-751.
19. Ross, J.J., Vaidya, A. and Kaiser, U.B. Lying Low, New England Journal of Medicine, 2011.
20. Scheuermann, R.H., Ceusters, W. and Smith, B., Toward an Ontological Treatment of Disease and Diagnosis. in *2009 AMIA Summit on Translational Bioinformatics*, (San Francisco, CA, 2009), 116-120.
21. Schulz, S., Brochhausen, M. and Hoehndorf, R., Higgs bosons, mars missions, and unicorn delusions: How to deal with terms of dubious reference in scientific ontologies (forthcoming). in *ICBO 2011*, (Buffalo, USA, 2011).
22. Schulz, S., Schober, D., Daniel, C. and Jaulent, M.C. Bridging the semantics gap between terminologies, ontologies, and information models. *Stud Health Technol Inform*, *160* (Pt 2). 1000-1004.
23. Vickers, A.J., Basch, E. and Kattan, M.W. Against diagnosis. *Annals of Internal Medicine*, *149* (3). 200-203.