A realism-based analysis of the OpenEHR Entry model

André Queiroz Andrade
Escola de Ciência da Informação-
Universidade Federal de Minas Gerais
Belo Horizonte - Brazil
CEP 31270-901
andradeaq@ufmg.br

Mauricio Barcellos Almeida
Escola de Ciência da Informação-
Universidade Federal de Minas Gerais
Belo Horizonte - Brazil
CEP 31270-901
mba@eci.ufmg.br

ABSTRACT
We describe a proposal of improvement for clinical models of the OpenEHR standard according to the realism-based assumptions. In order to reach such proposal, we analyze OpenEHR entry model and realism-based ontologies created specifically for Medicine. Thus, we check our approach in a test bed of real medical records. Finally, we suggest a new taxonomy with the aim to improve the entry model, and offer conclusions from our underway research.

Categories and Subject Descriptors
H.1.1 [Information systems]: Models and principles – system and information theories; J.3 [Computing Applications]: Life and Medical Sciences — Medical information systems;

General Terms
Design, Standardization.

Keywords
Knowledge representation, ontologies, clinical models.

1. INTRODUCTION

System interoperability – the ability of communication between systems without human intervention – requires shared semantics of the terms used in both systems. In pursuit of such shared semantics, two approaches are worth to be mentioned: ontologies and clinical models. We here use the term "semantics" according to the description of [1].

The use of ontologies for semantic representation has been extensively studied in scientific fields like the biology and medicine [22]. Indeed, the development and wide adoption of a realism-based instance for ontology creation allows for an explicit and stable vocabulary definition with the aim of promoting communication without ambiguities [33]. In this paper, we take the "ontological realism" as a methodology for ontology design – said "realism-based ontologies" – grounded on the philosophical realism principles.

Within the realm of standards, models underlying initiatives like Health Level Seven International (HL7) and Open Electronic Health Records (OpenEHR) try to ensure semantic interoperability by providing templates to represent health records. Those templates consist of a set of common variables suitable to record clinical data according to a unified fashion. The OpenEHR standard, for example, aims to define a generic information model and specific clinical models for semantic interoperability in the medical domain. It is based on the separation of a reference model, containing domain invariant classes; and an archetype model, containing clinical specific information [4].

MIXHS'11, October 28, 2011, Glasgow, Scotland, UK.

Models of the OpenEHR standard[1] have practical orientation and are grounded in an ontology of healthcare information [5]. On the other hand, ontologies developed according to realism-based methodologies constraint the use of certain terms common in clinical practice in favor of a scientific orientation [28]. While realism-based ontologies seems not able to record all clinical data [9, 15] the OpenEHR entry model lacks the ontological soundness required for interoperability[33].

In the present paper, we describe an analysis and a proposal to represent information contained in medical records that considers both clinical models and ontologies. Such approach is reached by the evaluation of OpenEHR entry model according to realism-based ontological guidelines. We test such proposal in the scope of an underway research project in which real health records has been evaluated. The partial results suggest improvements in taxonomy and definitions underpinning the OpenEHR models.

2. ONTOLOGY AND CLINICAL MODELS

The OpenEHR architecture embodies years of research from worldwide medical projects and standards with the aim to support requirements of clinical electronic healthcare records (EHR). Another approach to organize medical data is provided by a set of initiatives based on principles of philosophical realism. Indeed, realism has been shown capable of representing diseases, disorders [24] and symptoms [34], as evidenced by the *Ontology for General Medical Science* (OGMS). This section provides an overview of both OpenEHR initiative and realism-based ontologies.

2.1 The realism-based ontologies

The term "realism" in Philosophy is widely used and controversial [13]. Taken as a methodology, ontological realism is widely used in biomedicine [3, 10], and its generic tenets are the following: i) there is a real world; ii) the reality in which we live in is part of this world; iii) we are capable of knowing the world and reality, even that in an approximate manner[17].

One of the presuppositions of the ontological realism is the theory of universals, which states that reality contains particulars and universals. Particulars are entities described by the observation of the real world, e.g. a clinic or a laboratory. Universals represent what is common among equivalent particulars, what is invariant in reality [30, 31], e.g. the characteristic of having a head that is common to every human being. Since ontological realism proposes that science is the best manner to describe reality, universals are those entities chosen to formulate scientific theories.

According to the ontological realism, the unrestricted creation of classes to represent every possible entity of the world leads to

[1] http://www.openehr.org/home.html
[2] http://www.loa-cnr.it/DOLCE.html
[3] http://www.ontologyportal.org/
[4] http://code.google.com/p/ogms/
[5] Some classes in this site have a *draft* status. However, due to its

inconsistencies. Classes are human creations – e.g. human beings that are men and who likes swimming – and may be interpreted in different ways [[17]. To avoid that, the realism-based methodology restricts the possible classes to those defined by the scientific community as the best approximation of reality. However, the accurately distinction between universals and classes is not always trivial. While universals are grouped by what they are, classes are grouped by how they are [33].

The realism-based methodology uses upper-level ontology to organize universals with a top-down approach. Examples of upper-level ontologies are the Basic Formal Ontology (BFO) [10], Descriptive Ontology for Linguistic and Cognitive Engineering DOLCE[2], the Standard Upper Merged Ontology (SUMO)[3], to mention but a few. In the BFO, adopted in the ontological realism-based initiatives, we can find structuring divisions made by generic universals called continuants and occurrents. This division is based on the notion of SPAN and SNAP [10]. SPAN entities, called occurrent or perdurants, are universals that possess a determined beginning and end, and encompass process (e.g. "the life of an organism") and spatiotemporal regions (e.g. "the eighties"). SNAP entities, also called continuants or endurants, are universals that maintain their identities through time (e.g. a "human being"). Continuants may be dependent (e.g. "the color of an object"), independent (e.g. "a table") or spatial regions (e.g. a "point"). The use of the same upper-level ontology as starting point to create domain ontologies increases the chance that its universals are compatible and, therefore, the chance that they are amenable to integration.

Due to the abovementioned rationale, wide and generally successful use, and the presence of anecdotal evidence, we take the ontological realism as a good reference for ontology creation. However, its scientific status has been questioned in [14], with prompt and sometimes passionate responses for and against the methodology [6]. The proper discussion on merits and problems of realist ontologies are outside the scope of this paper, and the reader should also refer to the discussion in [15, 16, 33] for different positions.

2.2 The OpenEHR ontology of information

The modeling and ontological foundations of OpenEHR are a consequence of several previous efforts to improve the structure and communication capabilities of Electronic Health Records. Probably due to such origins, an important principle of that standard's architecture is the separation between an ontology of reality and an ontology of information [4]. The ontology of information encompasses the information model and the domain content model (clinical archetypes). The ontology of reality is composed of process descriptions, terminologies and medical classifications like International Classifications Disease (ICD) and Logical Observation Identifiers Names and Codes (LOINC).

Even though the artifacts of the so-called ontology of reality are mediated by models, there is no part of the standard concerning direct representations of reality. Indeed, the OpenEHR do not represents the reality, but records about the reality. For example, considering a measure of blood pressure: an ontology of reality would be about the existence of a blood pressure and the diseases characterized by abnormalities in the pressure; the ontology of information is about a process of measuring it, a data scale, measuring procedures and parameters during the measurement.

[2] http://www.loa-cnr.it/DOLCE.html

[3] http://www.ontologyportal.org/

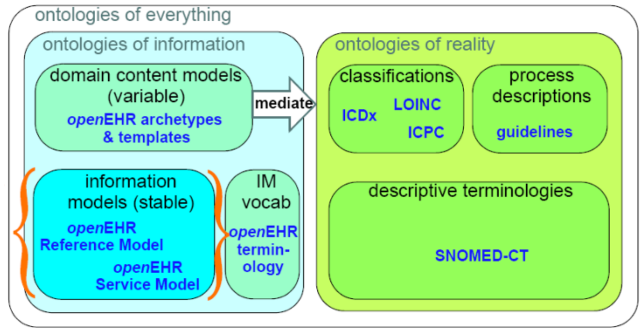

Figure 1 – Ontological landscape adapted from [4]

Due to its goal to deliver healthcare data using information systems, the model is focused on clinical practices and in how records are created during health encounters. Figure 2 depicts such situation:

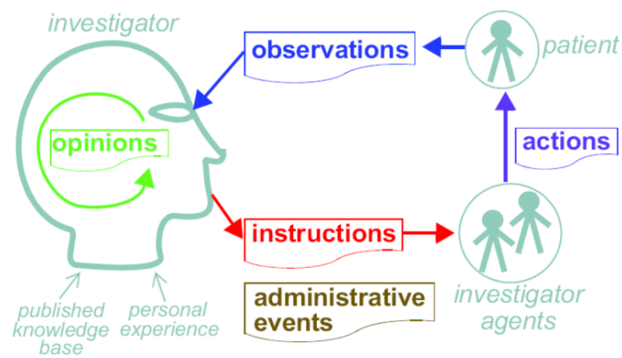

Figure 2 – Information types created by the investigator [5]

3. METHODS

Our methodological steps were focused on the possibilities of instantiating the OpenEHR classes in realism-based ontologies, and then propose improvements to the latter. Our analysis of the OpenEHR standard is based on the documents available at [19]. Since the analysis is focused on the OpenEHR Entry Entity Model, we omit data pertaining to the Reference Model and patient demographic information.

The realism-based ontologies considered here are the Ontology for General Medical Science (OGMS)[24] and the Artifact Information Ontology (AIO)[12, 23], both of them grounded on the Basic Formal Ontology (BFO)[10].

In seeking for clinical entities relevant to medical practice, we employ real medical records as guiding examples. The Ethics Committee of a Brazilian public medical institution released eighty medical records containing information registered since the nineties. In order to reach the goal of this paper, the improvement proposal, we evaluated fifteen of them. Each of those medical records are composed by sections, namely: i) identification; ii) clinical evaluation; iii) medical prescription; iv) exams; and v) annexes. Arranged in accordance with those sections, we meet several documents that allow one following the clinical evaluation of certain patient. For example, there are: demographic information, patients anamnesis, molecular genotyping reports and serological tests (exams section), to mention but a few.

After being properly de-identified, a record of a complete internal medicine encounter was transcribed according to the relevant information which it could provide. Such representation consisting of clinical model improved by realism-based ontological decisions [11].

4. ANALYSIS

Here, we describe briefly the entities and taxonomic relations in both of the initiatives, namely, Open EHR and OGMS. The Open ER entry type's model can be seen at Figure 3:

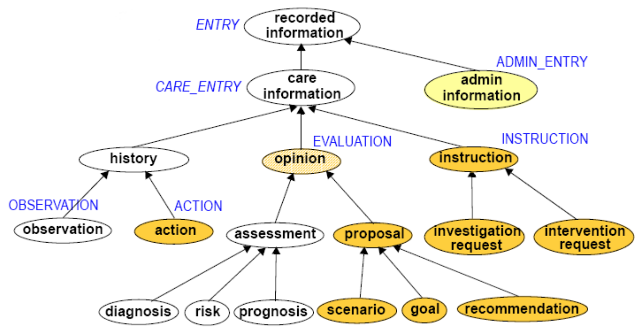

Figure 3 – The OpenEHR entry types model [4]

The entities pertaining to the model depicted in Figure 3 are information records, as one can see observing the top-level class. Follows a brief description of each entity according to [4, 5]:

- Recorded information: what is recorded in the health record is a selective choice of notes about real events, situations, etc., intended to be interpreted by other professionals.
- Observation: information created by an act of observation, measurement, questioning, or testing of the patient or related substance (e.g. tissue, urine, blood), or taken by the patient himself (e.g. measuring temperature), or other information realized by the investigator as relevant to characterize the patient;
- Opinion: inferences of the investigator using the personal and scientific knowledge about what the observations mean, and what is possible to do about them (e.g. diagnoses, assessments, plans, and goals);
- Instruction: opinion-based instructions detailed so as to be executable by other agents (people or machines), in order to effect a desired intervention (e.g. obtaining a sample for further investigation, as in a biopsy);
- Action: a record of intervention actions that have occurred due to instructions;
- Proposal: opinions about what should be done by the investigator;
- Assessments: opinions about what is happening in the patient organism;
- Diagnosis: a group of observed signs and symptoms, which designates them in the understanding of the investigator, as being a known phenomenon;
- Risk: a quantified assessment which provides the basis for investigation (e.g. the family history);
- Prognosis: the assessment outcome related to a current diagnosis;
- Scenarios: a clinical thought based on what-if statements;
- Goals: a clinical thought consisting of statements about a desired state for a patient;
- Plan: a clinical thought consisting of statements about how to reach goals.

The arrows linking entities in Figure 3 seem to be is-a relations, because they are defined as relations between classes and subclasses [5]. As the high-level entities are not formally defined, it is not clear, what kind of is-a relations is considered[29]. The taxonomy of Open EHR Entry types model represented as a hierarchy seems like that:

Recorded information
Care information
History
Observation
Action
Opinion
Assessment
Diagnosis
Risk
Prognosis
Proposal
Scenario
Goal
Recommendation
Instruction
Investigation request
Intervention request

The OGMS, formerly called the Clinical Phenotype Ontology, aims to address issues regarding diseases, signs, symptoms, and findings. Several terms of OGMS hang from BFO and from AIO. Figure 4 presents a fragment of OGMS showing records, thus suitable for comparison with types of recorded information of OpenEHR:

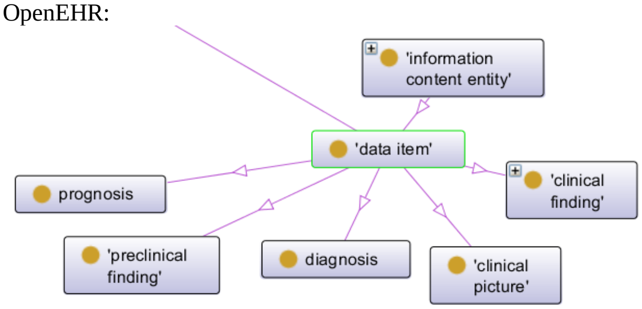

Figure 4 – Fragment of OGMS[4]

The arrows in Figure 4 also would represent is-a relations between particulars. Such is-a relation between A and B are formally defined as: A is-a B = for all x: if x, t, if x instance_of A, then x instance_of B at t. This definition includes a temporal aspect, but is able to deliver only a necessary condition, not a sufficient one. For example, it permits represent *leukocytes in a blood bag is_a leukocytes*, while *leukocytes in a blood bag* are not a special kind of *leukocytes* We consider that those is-a relations stand for relations between defined classes – and not between universals – because the former allows handling of information relevant for clinical investigations. It worth to mention that defined class here has the sense of non-natural classes.

Follows a brief description of entities of OGMS according to [16], some of them present only in IAO:

- Data item: a data item is an information content entity that is intended to be a truthful statement about something (e.g., measurement precision or other systematic errors) and is acquired by a method which reliably tends to produce truthful statements;
- Clinical picture: a representation of the clinically significant bodily components or bodily processes of a human being that is inferred from the totality of relevant clinical findings;

[4] http://code.google.com/p/ogms/

- Clinical history: a series of statements representing health-relevant qualities of a patient and of a patient's family.
- Image finding: a representation of an image that supports an inference to an assertion about some quality of a patient;
- Laboratory finding: a representation of a quality of a specimen that is the output of a laboratory test and that can support an inference to an assertion about some quality of the patient;
- Physical examination finding: not defined
- Preclinical finding: a representation of a quality of a patient that is recorded by a clinician because such quality is hypothesized to be of clinical significance, and refers to qualities obtaining in the patient prior to their becoming detectable in a clinical history taking or physical examination.
- Diagnosis: the representation of a conclusion of an interpretive process that has as input a clinical picture of a given patient and as output an assertion (diagnostic statement) to the effect that the patient has a disease of such and such a type.
- Prognosis: a hypothesis about the course of a disease.
- Rule: a rule is an executable which guides, defines, restricts actions (only in AIO);
- Plan specification: a directive information entity that when concretized it is realized in a process in which the bearer tries to achieve the objectives, in part by taking the actions specified. (only in AIO);
- Study design: a study design is a plan specification comprised of protocols that are executed as part of an investigation (only in AIO);
- Objective specification: (only in AIO);
- Report: a document assembled by an author for the purpose of providing information for the audience.

A fragment of OGMS represented as a hierarchy seems like that (the sign "(…)" indicates that there are subclasses not described at this point.):

```
Thing
  Disposition
    Disease
Entity
    Generically dependent continuant
      Information content entity
        Data item
          Clinical finding
            Clinical history
            Imaging finding
            Laboratory finding
            Physical examination finding
          Clinical picture
          Diagnosis
          Preclinical finding
          Prognosis
        Patient symptom report
      Normal value
    Specifically dependent continuant
      Quality
        Configuration (...)
        Manifestation of disease (...)
        Phenotype (...)
        Syndrome
      Realizable Entity
        Disposition
          Disease (...)
            Homeostasis (...)
            Predisposition to a disease type X  (...)
    Independent continuant
      Material entity
          Disorder
            Epigenetic disorder
            Genetic disorder
          Pathological anatomical structure
          Pathological formation
    Occurrent
      Processual entity
        Bodily process
        Pathological bodily process
        Clinical history taking
        Convalescence
        Disease course (...)
        Etiological process
        Healthcare process
          Healthcare encounter
        Inflammation process
        Laboratory test
        Life course
        Physical examination
        Treatment
  Sign
    Vital sign
  Symptom
Pain
```

We have now at our disposal instruments for evaluating clinical models of the OpenEHR standard. Table 1 presents a mapping between entities of models and realism-based ontologies.

Table 1. *Ad-hoc* correspondences

Entry model		Realism-based ontology
Recorded information	Is-a	Data item
Care information	Is-a	Data item
History	Is-a	Clinical history
Observation	Is-a	Image finding + lab. Finding + Clinical finding + physical examination
Action	Is-a	Report
Opinion	Is-a	Clinical finding
Assessment	Is-a	Clinical picture
Diagnosis	Is-a	Diagnosis
Prognosis	Is-a	Prognosis
Proposal	Is-a	Objective specification
Scenario	Is-a	Objective specification
Goal	Is-a	Objective specification
Recommendation	Is-a	Objective specification
Instruction	Is-a	Plan specification (AIO
Intervention request	Is-a	Plan specification (AIO) AND (is_about some Treatment)
Investigation request	Is-a	Plan specification (AIO) AND (is_about some Laboratory_test)

In comparing entities of the OpenEHR entry model and the OGMS is possible to achieve an *ad-hoc* correspondence between them. That correspondence is said "*ad-hoc*" because OpenEHR not seems to share the soundness and rigor of the realism-based ontologies, perhaps, because its practical bias.

It is worth to mention that some comparisons are not so direct. For example, while "history" in the OpenEHR entry is a broad class, encompassing information about the past of the patient, clinical observation and actions performed by the patient,"clinical history" in the OGMS is not so comprehensive. Thus, the correspondence between them seems to be just an approximation. Likewise, the recording of actions is nothing but a special kind of report that is about some occurrent which had the patient as a participant. In such cases, the category is itself unnecessary and could be replaced by a restriction axiom such as:

Report AND

(is_about some (Occurrent

AND (has_participant some Subject_of_record)))

5. RESULTS

In this section we present partial results of the underway research mentioned in section 3. According to the analysis carried out in sections 4 and 5, we evaluate real records to check the feasibility of our proposal. The Figure 6 depicts an extract of real medical record among the fifteen being evaluated at this phase.

QP: Chest pain and abdominal pain.
HMA: Six months ago, she felt severe precordial pain in addition to nausea and dyspnea. She attempted medical care in the Hospital X, where received isordil + AAS 300mg. Enzimes: CKT 262 CKMB 30. She was not aware of previous pathologies. It was prescribed: Captopril, HCTZ e AAS.
Last month, she felt severe pain again and sought for medical care in a different place. Then, it was prescribed: Losartan, AAS, Sinvastatina e Nebilet.
She sought for medical care in other occasions because of the precordial pain. In addition to the medicine mentioned, she uses Metoprolol - 100 mg 12/12 h.
She reports diffuse and intermittent abdominal pain, which becomes worst in case of stress. It is not related with bowel movement alterations. She also reports rare burning epigastric pain that improves with water drinking.

Figure 5 – Sample extract of a original medical record

Within our investigation, we identify four kinds of information that are present in medical records[2], namely:

- Aspects that represent entities in reality, for example: "chest pain", "metoprolol", "moment of re-incidence of pain (one month ago)";
- Aspects that represent useful constructs for medical practice not empirically verifiable, for example; "severe (precordial) heavy pressure", "rare burning (epigastric pain)";
- Aspects that represent observations about reality (not reality itself), for example: "CKMB 30", "left ventricle ejection fraction: 68%";
- Aspects that represent observations about the physician understanding of the clinical situation (not about reality), for example: "previous consultations and prescriptions", "not related to bowel movement alterations".

It is beyond the scope of this article to examine all of abovementioned aspects. However, for sure, a more sound ontological structure is required in order to properly represent the world of clinical practices. Even that, at a first glance, seems that realism oriented ontologies do not address the real world expectations, they are able to register particulars that in the long term could be turned into universals and vice-versa. This movement can create the basis for the better understanding of clinical entities.

We now present our proposal of taxonomy for recorded information, a proposal to improve the OpenEHR. The terms of OpenEHR models instantiate those of OGMS and AIO, according to defined in section 4. The relations also are considered is-a relations as defined in the section 4. The original OpenEHR terms are not changed, but other terms (in italics) were included for the sake of creating a more understandable taxonomy. The OpeEHR class "recommendation" is not considered because it is not even informally defined; "risk" is not included by inexistence of a direct correspondence in the realism-based ontology.

- Recorded information
  - Care information
    - *Clinical finding*
      - History
      - Observation
      - Opinion
      - Proposal
    - Action
    - Assessment *(Clinical Picture)*
      - Diagnosis
      - Prognosis
- *Directive information entity*
  - *Objective specification*
    - Goal
    - Scenario
  - *Plan specification*
    - Instruction
      - Instruction request
      - Instruction intervention

Figure 6: proposal of an improved OpenEHR entry model

6. DISCUSSION

Understanding what is an "entry" in the OpenEHR is required to interpreting the relation between the "observation" class and the "care_entry" class. An "entry" is defined as "a logically a single 'clinical statement', and may be a single short narrative phrase, but may also contain a significant amount of data, e.g. an entire microbiology result, a psychiatric examination note, a complex medication order"[4]. By comparison with the AIO entities, the "entry" is somewhat equivalent to the "information content entity", which is "an entity that is generically dependent on some artifact and stands in relation of aboutness to some entity". This aboutness relation to some artifact allows proper understanding of the distinction between "care_entry" and "admin_entry", being the former about an entity of the direct care situation, and the latter an entity related to the administrative situation. However, while this distinction is clear to the human reader, a discharge situation is as much about an encounter as a medication order, and there is very little ontological separation between both. A convention was drafted since the professionals responsible for these tasks are different, but that requires seeing these processes as occurrents, having as participants persons with different roles.

Closer looks at the "observation" class suggest that this class represents information about the patient reality, in order to properly characterize the patient. However, the definition makes it unclear whether the information is about an entity on the reality side, referencing a patient feature, or about the process of making an observation or measurement. Further exploration of what is considered an observation is required to understand it.

At the OpenEHR´s Clinical Knowledge Manager [18] we are able to find several examples of Observation classes such as "Body Temperature", "Blood Pressure" and "Story or History"[5]. While the first two refer to measurements processes, the third actually refers to another information entity. This is made clearer by an example, taken from a real medical encounter, as that transcribed to Figures 6 and 7. The "Story or History" class takes another archetype as "Detail" metadata. As we see in the example, the story is about a pain symptom. At the same time, it is explicit stated in the "Symptom" archetype definition that it is "for recording symptoms and information about the symptoms. This archetype allows a statement to be made that the patient does not have this symptom (without any other information) or more detailed statements as required" [18]. At this point, the distinction between reality and information becomes blurred, since the information is about the symptom and no longer about the observation of it. At the same time it can be used to deny its existence, becoming ultimately a statement about an observation and about an entity in reality which never existed.

Even that comparisons and assignments like those described in table 1 (see section 4) seems to be reasonable, they give rise to inconsistent subsumptions if retained the entry model structure. For example, in the entry model, *diagnosis* is-a *assessment* and *prognosis* is-a *assessment*, and we connect *assessment* to *preclinical findings* through a subsumption relation. On the other hand, in the ontology, *diagnosis* and *prognosis* are siblings and do not share subsumption relation between them, which they would have shared if maintained the entry model structure (see Figure 7).

Entry model	**Ontology**
…	…
Assessment	Clinical Picture
Diagnosis	Diagnosis
Prognosis	Prognosis
…	…

Figure 7 - Fragments displaying different ontological decisions

Such situation can be seen in other cases examined, for example: in the entry model, observation is-a history and action is-a history; in the ontology, image finding + lab.finding are siblings and do not share subsumption relation. This fact indicates that different ontological decisions are needed in order to improve OpenEHR models.

The large body of work done in realist ontologies has shown benefit in strict adherence to some principles to ensure robustness, consensus and reliability. At the same time, those principles impose limits about what can be expressed using ontologies, requiring other ways to assert probability, uncertainty, opinions and understanding of the patient condition. We believe, however, that the formalism required in building realism-based ontologies is not a hidrance to a better view of clinical world. Although medicine has many unknown conditions and unidentified diseases, we can work in establishing well-founded basis for what is known at this moment. The separation of information and reality is a useful approach in maintaining the benefits while allowing expression of terms without reference [25], negative statements [26], and other terms that aren't a perfect fit to realist ontologies commonly found in healthcare standards [27].

It is worth to mention that OGMS is also an on-going initiative and probably it will better define some terms. This is the case, for example of risks, which have been discussed under the labels "rates". Our proposal is a step toward seeking for robust models to represent both scientific and clinical data. As future work, we intend to provide a set of more formal relations instead of the ad-hoc ones presented in section 4. We also intend to explore other ontologies gathered in the OBO Foundry [32] and their relation to the OpenEHR information model. Among the approved and candidate ontologies with direct clinical relevance, we can list the Infectious Disease Ontology [8], which describes the pathogens, clinical features and other relevant entities like vaccines, and the Human Phenotype Ontology [20], which describes common phenotypical abnormalities, such as atrial septal defect. We can also mention the Foundational Model of Anatomy [21], which describes anatomical entities and the new Adverse Event Reporting Ontology [7], which describes quality related events in healthcare institutions. Finally, we intend to determine limits between what can be expressed through ontologies and what requires information model for representation [2], seeking a compromise both robust and useful in real-world healthcare systems.

7. SUMMARY AND CONCLUSIONS

So far, we present initiatives to represent entities in the realm of medicine. Then, we make a comparison of their definitions and try to reach a more sound structure to represent data contained in medical records.

Realism has been shown capable of representing entities within biomedicine. The existence of diseases defined as a disposition [24] is well known by medical science. Likewise, symptoms can be seen as body characteristics that a patient experiences. In this case, we represent the body alteration considering its scientific description. On the other hand, the diagnosis itself is not a patient attribute, but rather a conclusion of an interpretive process that has as input a clinical picture.

In our analysis we have shown some inconsistencies in the separation between information and reality in the OpenEHR standard. They are most likely a consequence of the way clinicians see their everyday practice. The clinical encounter is a highly complex situation requiring decision making under incomplete information, with great consequences for the patient. Therefore, a record must contain information about what is there on the side of the patient, about what seems to be there, about the physician's take on the whole situation and about what is planned to be done. It is a document made by physicians to be read primarily by physicians, and the natural language ambiguities are

[5] Some classes in this site have a *draft* status. However, due to its relative stability over time and great importance in a medical record, we consider them in our analysis as valid classes.

easily solved by the similar training and mind models. That is not true when this information is to be processed by computers.

By classifying entities firstly according to their existence in reality, and then according to the event they are about, we hope to have reached our main contribution, the first step towards mapping from the model to other ontologies. We said "first step" because we believe that mapping ontologies is an arduous manual process. In an ideal setting, such preliminary proposal has to be maintained and curated, incorporating new entities as required by practice while retaining capability to make inferences. Indeed, we hope to improve the proposed ad-hoc taxonomy in future works. In such opportunity, some additional questions will be addressed: though grounded in the realist methodology, the ad-hoc mapping was based solely on the natural language description of the OpenEHR classes, since their semantic relations are not strictly subsumption relations; also, technical requirements were here disregarded (such as backwards compatibility), which may require equally practical solutions in order to be useful.

8. ACKNOWLEDGMENTS

We thank the Hemominas Foundation, Govern of State of *Minas Gerais*, Brazil (http://www.hemominas.mg.gov.br) for allowing access to medical records. We also thank the anonymous reviewers for the useful insights.

9. REFERENCES

1. Almeida, M.B., Souza, R.R. and Fonseca, F. Semantics in the Semantic Web: a critical evaluation. *Knowledge Organization Journal*.

2. Andrade, A.Q. and Almeida, M.B., Realist representation of the medical practice: an ontological and epistemological analysis (forthcoming). in *Proceedings of the 4th Ontobras*, (Gramado, Brazil, 2011).

3. Baker, P.G., Goble, C.A., Bechhofer, S., Paton, N.W., Stevens, R. and Brass, A. An ontology for bioinformatics applications. *Bioinformatics*, *15* (6). 510-520.

4. Beale, T. and Heard, S. Architecture Overview, OpenEHR Foundation, 2007, 87.

5. Beale, T. and Heard, S. An ontology-based model of clinical information. *Stud Health Technol Inform*, *129* (Pt 1). 760-764.

6. Brochhausen, M., Burgun, A., Ceusters, W., Hasman, A., Leong, T.Y., Musen, M., Oliveira, J.L., Peleg, M., Rector, A. and Schulz, S. Discussion of "biomedical ontologies: toward scientific debate". *Methods Inf Med*, *50* (3). 217-236.

7. Ceusters, W., Capolupo, M., de Moor, G., Devlies, J. and Smith, B. An Evolutionary Approach to Realism-based Adverse Event Representations. *Methods of Information in Medicine*, *50* (1). 62-73.

8. Cowell, L.G. and Smith, B. Infectious Disease Ontology in Sintchenko, V. ed. *Infectious Disease Informatics*, Springer, 2010, 373-395.

9. Dumontier, M. and Hoehndorf, R. Realism for scientific ontologies *Proceeding of the 2010 conference on Formal Ontology in Information Systems: Proceedings of the Sixth International Conference (FOIS 2010),* IOS Press, Toronto, Canada, 2010.

10. Grenon, P., Smith, B. and Goldberg, L. Biodynamic ontology: applying BFO in the biomedical domain *From D. M. Pisanelli (ed.), Ontologies in Medicine, Amsterdam: IOS Press, 2004, 20–38*, 2004.

11. Grenon, P., Smith, B. and Goldberg, L. Biodynamic ontology: applying BFO in the biomedical domain. *Stud Health Technol Inform*, *102*. 20-38.

12. IAO, 2011. Information Artifact Ontology. http://code.google.com/p/information-artifact-ontology/

13. MacLeod, M.C., 2005. Universals. http://www.iep.utm.edu/universa/

14. Maojo, V., Crespo, J., Garcia-Remesal, M., de la Iglesia, D., Perez-Rey, D. and Kulikowski, C. Biomedical ontologies: toward scientific debate. *Methods Inf Med*, *50* (3). 203-216.

15. Merrill, G.H. Ontological realism: Methodology or misdirection? *Applied Ontology*, *5*. 79-108.

16. Merrill, G.H. Realism and reference ontologies:Considerations, reflections and problems. *Applied Ontology*, *5*. 189–221.

17. Munn, K. and Smith, B. (eds.). *Applied Ontology. An Introduction*. Ontos Verlag, Frankfurt/Paris/Lancaster/New Brunswick, 2008.

18. OpenEHR Foundation, 2011. Clinical Knowledge Manager. http://www.openehr.org/knowledge/

19. OpenEHR Foundation, 2007. OpenEHR. http://www.openehr.org

20. Robinson, P.N. and Mundlos, S. The human phenotype ontology. *Clin Genet*, *77* (6). 525-534.

21. Rosse, C. and Mejino, J.L.V. A reference ontology for biomedical informatics: the Foundational Model of Anatomy. *Journal of Biomedical Informatics*, *36* (6). 478-500.

22. Rubin, D.L., Shah, N.H. and Noy, N.F. Biomedical ontologies: a functional perspective. *Briefings in Bioinformatics*, *9* (1). 75-90.

23. Ruttenberg, A., From Basic Formal Ontology to the Information Artifact Ontology. in *ICBO: International Conference on Biomedical Ontology*, (2009).

24. 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.

25. 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).

26. Schulz, S. and Karlsson, D., Records and situations. Integrating contextual aspects in clinical ontologies (forthcoming). in *Bioontologies 2011*, (2011).

27. 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.

28. Schulz, S., Stenzhorn, H., Boeker, M. and Smith, B. Strengths and limitations of formal ontologies in the

biomedical domain. *RECIIS Rev Electron Comun Inf Inov Saude*, *3* (1). 31-45.

29. Schwarz, U. and Smith, B. Ontological relations. in Munn, K. and Smith, B. eds. *Applied Ontology. An Introduction*, Ontos Verlag, Frankfurt/Paris/Lancaster/New Brunswick, 2008, 220-234.

30. Smith, B., Beyond concepts: Ontology as reality representation. in *FOIS 2004*, (Turin, , 2004), IOS Press, 73-84.

31. Smith, B. From concepts to clinical reality: An essay on the benchmarking of biomedical terminologies. *Journal of Biomedical Informatics*, *39* (3). 288-298.

32. Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., Goldberg, L.J., Eilbeck, K., Ireland, A., Mungall, C.J., Leontis, N., Rocca-Serra, P., Ruttenberg, A., Sansone, S.A., Scheuermann, R.H., Shah, N., Whetzel, P.L., Lewis, S. and Consortium, O.B.I. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature Biotechnology*, *25* (11). 1251-1255.

33. Smith, B. and Ceusters, W. Ontological realism: A methodology for coordinated evolution of scientific ontologies. *Applied Ontology*, *5*. 139–188.

34. Smith, B., Ceusters, W., Goldberg, L.J. and Ohrbach, R.K. Towards an Ontology of Pain and of Pain-Related Phenomena, 2009.