Revisiting ontological foundations of the OpenEHR Entry Model

André Q Andrade[1,2*], Maurício B Almeida[3] and Stefan Schulz[2,4]
[1]Information Science Graduate Program, School of Information Science,
Federal University of Minas Gerais, Belo Horizonte, Brazil
[2]Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Austria
[3]Theory and Management Department, School of Information Science,
Federal University of Minas Gerais, Belo Horizonte, Brazil
[4]Institute of Medical Biometry and Medical Informatics,
University Medical Center, Freiburg, Germany

ABSTRACT

Both information models and ontologies promise advantages in promoting interoperability. Recent research has shown the benefits of rigorous modelling in assuring large-scale consistency.

In previous work we have demonstrated the feasibility of remodelling the OpenEHR information ontology using realist ontologies, such as IAO. We here present an OWL version of the care-entry model, showing that many terms contained in clinical archetypes refer to reality rather than to information.

Even though not covering the domain of the information model (which deals with record structure, data types, *etc*.) we have shown that the harmonization of the OpenEHR standard with realist ontologies is feasible. While useful to resolve ambiguities contained in archetype metadata definition, the proposed merged ontology also revealed several modelling inconsistencies on published archetypes. We have also demonstrated that ambiguity in relations and ontological commitment can be improved by rigorous ontological definitions.

1 INTRODUCTION

The lack of interoperability is recognized in medical informatics communities as one of the main obstacles for the full use of healthcare information systems. This issue has led to the creation of standards development organizations with the purpose to build consensus by proposing commonly agreed message types, terms and architectural patterns. Within the realm of standards, models underlying initiatives like Health Level Seven International (HL7) and Open Electronic Health Records (OpenEHR) try to ensure interoperability by defining basic templates to represent information in health records. Those templates consist of a set of common information and clinical variables that faithfully represent health record information. The OpenEHR standard, e.g., defines

(1) a generic information model, the *reference model* with domain-invariant classes to be instantiated by
(2) specific clinical models, which support semantic interoperability, which are called *archetypes*, containing specific clinical information (Beale & Heard, 2007).

Ontologies, based on the study of reality, are an alternative solution to interoperability issues. With communication standards they share the goal of unifying meaning across different communities, maintaining common (machine-processable) interpretations. The difference is that ontology-based models are based on formal logic and are, to different degrees, influenced by philosophical methodologies.

The practical and pragmatic orientation of the OpenEHR standard[1], which has been described as grounded in an *ontology of healthcare information* (Beale & Heard, 2007), closely follows medical documentation routines. In contrast, ontologies developed according to realism-based methodologies constrain the use of some common terms in clinical practice in favour of a scientific orientation (Schulz et al., 2009). While realism-based ontologies were challenged for not being able to record all kinds of clinical data (Dumontier & Hoehndorf, 2010; Merrill, 2010), the OpenEHR entry model was found to lack the ontological soundness required for interoperability (Smith & Ceusters, 2010).

In previous work (Andrade & Almeida, 2011) we have argued that the basic ontological foundation of OpenEHR archetypes could be better represented by realist ontologies, such as the Information Artifact Ontology (IAO, 2011) and OGMS, the Ontology for General Medical Science (Scheuermann et al., 2009), both based on the BFO, the Basic Formal Ontology (Grenon et al., 2004). We now present an extension of this work, in order to demonstrate the feasibility of representing the OpenEHR care entry information using a formal language within the framework of realist ontology. We then discuss practical and modelling advantages of this approach.

* To whom correspondence should be addressed: andradeaq@ufmg.br

[1]http://www.openehr.org/home.html

This paper is structured as follows: in a brief introduction we describe the OpenEHR two-level modelling approach for clinical data representation; we then compare it with the methodology of ontological realism. Finally, we propose an ontological representation of the entry types, and show practical results obtained and discuss findings.

2 INFORMATION MODELS AND ONTOLOGIES

Required for the development of any information system, modelling is one of the most important aspects of software engineering. However, despite the number of best practices and research developed to this subject, there is not yet an undisputed best way of representing a domain. In a complex domain such as healthcare, it is not surprising that many approaches have been used throughout the years.

The modelling 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 the OpenEHR architecture is the separation between an ontology of reality and an ontology of information (Beale & Heard, 2007). The ontology of information encompasses the information model and the domain content model (including the clinical archetypes). All entities in the ontology of information are information artefacts (terms, documents, images, hypotheses, orders, and so on) and not real clinical entities. The view on information artefacts as immaterial but nevertheless ontologically relevant entities is gradually substituting the view of an ontology-epistemology divide (Bodenreider et al., 2004) which had emerged at a time when realist ontologies were ignoring the existence of information entities.

The ontology of reality represents clinical and (patho-) physiological processes, body parts, chemicals, procedures, etc. As OpenEHR makes no distinction between terminologies, medical classifications and realist ontologies, this category encompasses the International Classifications of Diseases (ICD), Logical Observation Identifiers Names and Codes (LOINC), as well as most parts of SNOMED CT.

The information model is a rather complex and detailed architecture of a generic EHR. It encompasses both definition of records and documents (e.g. classes such as *Folder*, *Composition*, *Section* and *Entry*), and of the basic functions of software systems such as *Data structure*, *Data type*, *Access*, *Version,* etc.). The care entry model "define the semantics of all the 'hard' information in the record" (Beale & Heard, 2007), and represents information recorded during a medical encounter. Figure 1 shows a graphical representation of the ontology leading to the care entry model.

Following a quite different principle, realist ontologies are based on the philosophical study of reality. The term "realism" in Philosophy is widely used and controversial (MacLeod, 2005). Taken as a methodology, ontological realism is widely used in biomedicine (Grenon et al., 2004), 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 if in an approximate way (Munn & Smith, 2008).

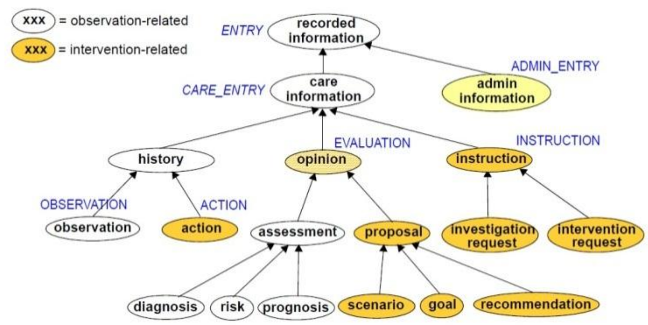

Fig. 1. The OpenEHR Ontology of Recorded Information(Beale & Heard, 2007).

Our goal in this paper is not to analyse the whole standard but only the care entry structure, due to its similarity and overlap with ontologically described entities. For this purposes, the realist approach brings two main advantages. The first is the clear separation between information entities and real entities, which are related by the relation *isAbout* – e.g. the drawing of a horse is about a real horse, or a shadow on a radiological image is about some anatomical structure. That prevents inadvertent incorrect inferences of common language statements, such as "patient blood pressure is an observation, and all observations are created by healthcare professionals, therefore the patient blood pressure is created by healthcare professional". The second advantage concerns the possibility of re-use of several previously developed ontologies adopting the principles of the OBO Foundry (Smith et al., 2007). Those ontologies follow the same upper-level ontology, the Basic Formal Ontology (Grenon et al., 2004). This re-use promotes consensus and orthogonality between ontologies, which increase robustness required for large-scale systems, such as Electronic Health Records.

3 ONTOLOGICAL REPRESENTATION

To demonstrate the compatibility between the OpenEHR model and clinically oriented realist ontologies, we have created an OpenEHR information ontology according to the OGMS guidelines. To properly place each class, we took into consideration the natural language description and basic rationale used to develop the entry types. In such rationale, the history classes ("observation" and "action") represent statements about the past events of the individual subject of record. This includes the description of currently observed characteristics, based on the fact that their appearance necessarily precedes the observation. The "evaluation" classes represent current assessment by the attending health professional, including "diagnosis" and "prognosis", as well as the representation of the imagined future, like "goals" and "scenarios". "Instructions" represent future events that should take place as prescribed by the health professional.

The proposed merged ontology can be seen in figures 2 and 3. While the rationale is quite different, the IAO proved capable of faithfully representing the meaning of each information type. *Observation* is a *Data item* resulting from the medical encounter, being a description of an entity, usually, the patient. By classifying the other classes according to their intended outcome, we merged the *Proposal* classes under *Objective specification* and *Instruction* classes under *Plan specification*. Finally, *Action* was represented as a special type of report, since it necessarily describes a process that has the patient as participant.

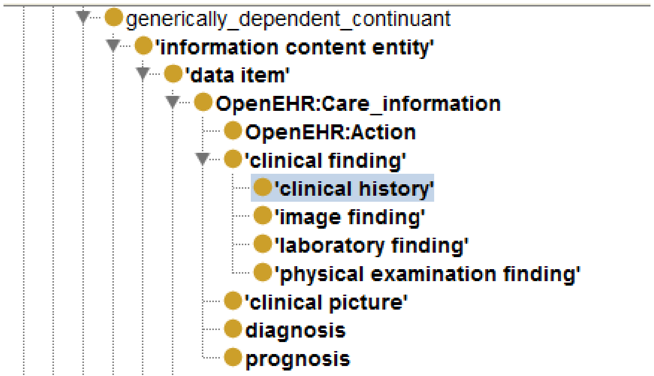

Fig. 2. The OpenEHR Care Information branch

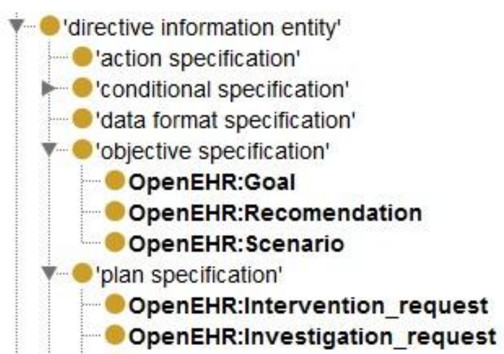

Fig. 3. The OpenEHR entry directive information branch

Using this new ontology, we proceeded to analyse one archetype of each main branch (observation, action, evaluation and instruction) contained in the OpenEHR Clinical Knowledge Manager (OpenEHR Foundation, 2010). Our criteria for selection were based on published status and frequency in medical records. The findings are summarized in the next section.

3.1 Observation Archetypes

We analysed the most commonly used example of an observation archetype, *viz.* the *Blood Pressure Archetype*, due to its stability and simplicity. Some examples are shown in Table 1. It was modelled as information about blood pressure, which is understood as the quality of a portion of blood that exerts pressure toward an artery. For didactical purposes, the ontological aspects of blood pressure were simplified (Goldfain et al., 2011; Kumar & Smith, 2003). Likewise, systolic and diastolic blood pressure measurements were easily modelled and their information status allows unambiguous assignment of data values and types. Some challenges arose while modelling reference to the measurement procedures, rather than to the pressure itself (e.g. the representation of the position of the patient while being measured - sitting, upright - and the size of the cuff used for measurement). This required explicit representation of the measurement as a process having the patient as participant. Also, patient positions could be adequately represented as qualities of the patient, who is a participant of the measurement process.

However, several epistemic entities were not successfully modelled, as they are not properly representable in realist ontologies. As an example, consider the metadata "confounding factors", defined as "Comment on and record other incidental factors that may be contributing to the blood pressure measurement. (…), level of anxiety or 'white coat syndrome'; pain or fever; changes in atmospheric pressure etc." Events such as pain and changes in atmospheric pressure have little or nothing in common that could map them to one category in an ontology. E.g. a confounding factor can be a process, a disposition, or a quality. Whether such "non-ontological" classes – characterized as "defined classes" by (Smith et al., 2006) – belong in an ontology at all, is contentious. However they can represented by logical definitions in an OWL model (Schulz et al., 2011).

Finally, some attributes that are specific to medical practice, such as "Diastolic endpoint", have unclear representation in real ontologies. Defined in the Blood Pressure Archetype as a metadata allowing the user to "record which Korotkoff sound[2] is used for determining diastolic pressure using auscultative method", this attribute depends on training and individual interpretation to be defined, and lack ontological status (Andrade & Almeida, 2011). This is clearly shown by the lack of rigor in the distinction between the 4th and 5th sounds, which refer to perceptive capabilities of the actor, defined as (our emphasis) "phase IV, sounds become **muffled and softer**; and phase V, sounds disappear completely. The fifth phase is thus recorded as the last **audible** sound" (Pickering et al., 2005). However, such attributes can also be represented by logical definitions, and should be subject of further investigation.

[2] Sounds heard during measurement of blood pressure.

Class name	*openEHR-EHR-OBSERVATION.blood_pressure.v1*
Elucidation	The local measurement of arterial blood pressure which is a surrogate for arterial pressure in the systemic circulation. Most commonly, use of the term 'blood pressure' refers to measurement of brachial artery pressure in the upper arm.
Axiomatization	**is_about**some *arterial_blood_pressure* **is_output_of** some *blood_pressure_measurement_process* **has_part** some *blood_pressure_measurement_datum*
Superclass	*'physical examination finding'*
Class name	*Systolic_blood_pressure_data*
Elucidation	Represents the Systolic attribute, defined as Peak systemic arterial blood pressure - measured in systolic or contraction phase of the heart cycle.
Axiomatization	**is_output_of** some *Systolic_blood_pressure_measuring_process*
Superclass	*blood_pressure_measurement_datum*
Class name	*human_position_configuration*
Elucidation	Represents the Position attribute, defined as The position of the subject at the time of measurement.
Axiomatization	**is_quality_of** some *human_being*
Superclass	*configuration* (subClass of *quality*)
Class name	*blood_pressure_measurement_process*
Elucidation	Represents the actual process of measurement that will result in the blood pressure observation. It is not directly referred in the OpenEHR archetype. Extended and modified from the Vital Sign Ontology.
Axiomatization	**has_participant** some (*human_being* and (**has_quality** some *human_position_configuration*) and (**has_quality** some *sleep_status*) and (**has_role** some *patient_role*) and (**has_part** some *artery_wall*))
Superclass	*'planned process'*

Table 1: Mapping the Blood Pressure Archetype to OWL

Overall, the ontological interpretation of the blood pressure archetype revealed definitions that could only be understood by a domain specialist. It also made clear that the archetype is information about the patient, the examination procedures, the examination artefacts used in the procedure (real entities) and the health professional's interpretation of the process (information entities). As such, it requires additional modelling of reality, e.g. roles and parts, in order to express the real entities represented in the information artefacts.

3.2 Action Archetypes

We examined the *Medication Action Archetype*, which represents one of the most commonly described healthcare interventions. Its precise reconstruction was not straightforward, as it included states that contradict the existence of a process, e.g.*Cancelled* or *Postponed* states. In other words, a cancelled process is not a kind of process, since the process never actually took place. Therefore, a different treatment is required, as only plans about medication administration processes can be cancelled or postponed, not the processes themselves (Raufie et al., 2011; Schulz & Karlsson, 2011).

Furthermore, because this archetype includes medication-specific information, it is not clear what kind of relation between the action and the medication holds. Since this template has information such as *Name of medication* as well as *Reason for ceasing the medication*, an explicit definition of those relations in the archetype is required before further modelling.

3.3 Evaluation Archetypes

For this analysis, we used two archetypes. The first is a publicly published evaluation archetype, the *Clinical Synopsis Archetype*. While extremely simple, this class proved conformant to its information status, being defined as "narrative summary or overview about a patient, specifically from the perspective of a healthcare provider, and with or without associated interpretations." This definition suggests that a class such as OGMS *Clinical Picture* perfectly describes its meaning, though not as an overarching concept to diagnosis or objective specifications.

The second one is an archetype that demonstrates the classification as *Objective specification*, the *Goal Archetype*. It conformed to modelling, requiring specification of the objective intended, along with the time where it is expected and the standard that will evaluate its success. Overall, this class appears to be adequately represented.

3.4 Instruction Archetypes

For this analysis of this archetype category we evaluated the medication order, being one of the most common instructions. It is defined as "an order or instruction created by a clinician which specifies which medication to take, when, for how long etc." It is directly related to the medication action archetype by an item called *Medication activity* (See section 3.2). This was suitable for ontological representation since it can be shown that both actions and instructions are about processes. While the actions are kinds of processes, the instructions are plans that specify a process type which may be or not instantiated in the future. Not only the ontological representation faithfully describes the entities about medication, but also makes clearer the distinction between a suspended plan (the medication will not be administered to the patient) and a suspended administration (the administration process started but was stopped before completion).

4 CONCLUSION

Though not addressing those aspects of the information model that deal with record structure, data types, *etc.* we have presented examples which demonstrate that the harmonization of the OpenEHR standard by representing archetypes as realist ontologies is feasible. While useful to detect and fix ambiguities in archetype metadata definitions, the merged ontology also revealed several modelling inconsistencies on published archetypes. We have also shown that ambiguity in relations and ontological commitment can be improved by providing rigorous ontological definitions. Future work should focus on standard ways of ontologically representing epistemic and interpretative clinical information, together with their linkage to reality entities. Also, the precise logical formulation of value constraints, data types and cardinality requires further studies, to guarantee universal interpretability of these kinds of representation.

ACKNOWLEDGEMENTS

This work is supported by the Hemominas Foundation (Belo Horizonte, Brazil). AQA was being financed by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Brazil) – Programa de Doutorado no País com Estágio no Exterior, process number 2380-11-0, during the writing of this paper.

REFERENCES

Andrade, A.Q. and Almeida, M.B. (2011) A realism-based analysis of the OpenEHR entry model. *Proceedings of the MIXHS'11*. ACM, Glasgow, Scotland, UK, pp. 51-58.

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

Beale, T. and Heard, S. (2007) Architecture Overview. OpenEHR Foundation, pp. 87.

Beale, T. and Heard, S. (2007) An ontology-based model of clinical information, *Stud Health Technol Inform*, **129**, 760-764.

Bodenreider, O., Smith, B. and Burgun, A. (2004) The Ontology-Epistemology Divide:A Case Study in Medical Terminology. In Varzi, A. and Vieu, L. (eds), *3rd Conference on Formal Ontology in Information Systems (FOIS 2004)*. Turin.

Dumontier, M. and Hoehndorf, R. (2010) Realism for scientific ontologies. *Proceedings of the Sixth International Conference on Formal Ontology in Information Systems (FOIS 2010)*. IOS Press, Toronto, Canada, pp. 387-399.

Goldfain, A., Smith, B., Arabandi, S., Brochhausen, M. and Hogan, W.R. (2011) Vital Sign Ontology. In Shah, N., *et al.* (eds), *Proceedings of the BioOntologies SIG at ISMB/ECCB 2011*. Vienna, Austria.

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

IAO (2011) Information Artifact Ontology.(last accessed in January 30th 2011).<Available at http://code.google.com/p/information-artifact-ontology/>.

Kumar, A. and Smith, B. (2003) The ontology of blood pressure: A case study in creating ontology partitions in biomedicine.<Available at http://ontology.buffalo.edu/medo/BPO.pdf>.

MacLeod, M.C. (2005) Universals.(last accessed in June 28th 2011).<Available at http://www.iep.utm.edu/universa/>.

Merrill, G.H. (2010) Ontological realism: Methodology or misdirection?, *Applied Ontology*, **5**, 79-108.

Munn, K. and Smith, B. (2008) Applied Ontology. An Introduction. Ontos Verlag, Frankfurt/Paris/Lancaster/New Brunswick, pp. 342.

OpenEHR Foundation (2010) Clinical Knowledge Manager.(last accessed in January 30th 2012).<Available at http://openehr.org/knowledge/>.

Pickering, T.G., Hall, J.E., Appel, L.J., Falkner, B.E., Graves, J., Hill, M.N., Jones, D.W., Kurtz, T., Sheps, S.G. and Roccella, E.J. (2005) Recommendations for blood pressure measurement in humans and experimental animals: Part 1: blood pressure measurement in humans, *Hypertension*, **45**, 142-161.

Raufie, D., Schulz, S., Schober, D., Jansen, L. and Boeker, M. (2011) Redesigning an Ontology Design Pattern for Realist Ontologies. *OBML*. Berlin, Germany.

Scheuermann, R.H., Ceusters, W. and Smith, B. (2009) Toward an Ontological Treatment of Disease and Diagnosis. *2009 AMIA Summit on Translational Bioinformatics*. San Francisco, CA, pp. 116-120.

Schulz, S., Brochhausen, M. and Hoehndorf, R. (2011) Higgs bosons, mars missions, and unicorn delusions: How to deal with terms of dubious reference in scientific ontologies. *Proceedings of the 2nd International Conference on Biomedical Ontology. 2011*. CEUR-WS, Buffalo, NY, USA, pp. 183-190.

Schulz, S. and Karlsson, D. (2011) Records and situations. Integrating contextual aspects in clinical ontologies. In Shah, N., *et al.* (eds), *Proceedings of the BioOntologies SIG at ISMB/ECCB 2011*. Vienna, Austria, pp. 49-52.

Schulz, S., Stenzhorn, H., Boeker, M. and Smith, B. (2009) Strengths and limitations of formal ontologies in the biomedical domain, *RECIIS Rev Electron Comun Inf Inov Saude*, **3**, 31-45.

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. (2007) The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration, *Nature Biotechnology*, **25**, 1251-1255.

Smith, B. and Ceusters, W. (2010) Ontological realism: A methodology for coordinated evolution of scientific ontologies, *Applied Ontology*, **5**, 139–188.

Smith, B., Kusnierczyk, W., Schober, D. and Ceusters, W. (2006) Towards a reference terminology for ontology research and development in the biomedical domain. *KR-MED 2006*. pp. 57-66.