ICBO: International Conference on Biomedical Ontology
July 28-30, 2011 · Buffalo, NY, USA

Information Models and Ontologies for Representing the Electronic Health Record

Daniel Karlsson[1], Martin Berzell[2], Stefan Schulz[3]

[1]Department of Biomedical Engineering, Linköping University, Sweden
[2]Division of Health and Society, Department of Medical and Health Sciences, Linköping University, Sweden
[3]Institute of Medical Informatics, Statistics and Documentation, Medical University Graz, Austria

1 Introduction

The term "information model" is often used in biomedical informatics to refer to an aggregation of information items used to represent (what a health professional knows about) the health status of the patient, as well as planned, scheduled or carried out health care activities. As an example of this use, the International Standard ISO 12967:2009 defines the term "information object" as "information held by the system about entities of the real world, including the ODP (Open Distributed Processing) system itself, is represented in an information specification in terms of information objects, their relationships and behaviour" [1]. Thus, information models seem to closely relate to ontologies, for example as defined by Spear [2]. The aim of this paper is to add to the clarification of the relationship between information models and ontologies and to propose and discuss demarcations of these two kinds of representation.

2 Background

The representations and communication of facts, beliefs, and opinions in the field of health care, and therefore the scope of the Electronic Health Record, clearly exceeds the simple instantiation of mind-independent representational units such as provided by typical biomedical ontologies. The EHR includes observations, opinions, instructions (plans), proposals, requests, etc. Ontologies catering to these needs must necessarily go beyond the realm of the mind-independent, as the reference to entities in such discourses do not necessarily imply the existence of those entities. For instance, a record entry on a "denied tonsillectomy", initially does not refer to any tonsillectomy in clinical reality. Nevertheless, such a proposition should be adequately represented by referring to some artifact which includes a representational unit "tonsillectomy".

Beale and Heard propose a Clinical Investigator Record Ontology [3] distinguishing entries in the record of various types, including:

- if something has been done or if something is yet to be realized (or not), for example distinguishing a plan from a performed activity,
- if something concerns the state of the patient or if something concerns health-care activities, for example distinguishing the blood pressure of a patient from the observation thereof, and
- if something is an assessment or opinion or if something is (more or less) directly observed, for example distinguishing the observation of a blood pressure of 150/90 from a diagnosis of stage 1 hypertension.

From this Clinical Investigator Record Ontology, the *open*EHR foundation has constructed a reference (information) model and, in a larger cooperation, a framework for representing specialisations of that reference model known as archetypes [4]. Archetypes consist of constraints on how the reference model may be instantiated in addition to the constraints of the reference model. The demarcation of the reference model and its specialisations through archetypes, called the two-level modelling approach, is based on the stability of the respective models [5]. The reference model consists of a core representation which is assumed to be stable over time and across organisations while archetypes are used to build representations where such assumptions cannot be made. The *open*EHR foundation also keeps a repository of clinical and demographic archetypes [6].

```
ELEMENT[at0008] occurrences matches {0..1} matches { -- Position
    value matches {
        DV_CODED_TEXT matches {
            defining_code matches {
                [local::
                at1000, -- Standing
                at1001, -- Sitting
                at1002, -- Reclining
                at1003, -- Lying
                at1014; -- Lying with tilt to left
                at1001] -- assumed value
            }
        }
    }
}
```

Figure 1. Part of an openEHR Blood Pressure archetype specification

However, not all classes in the *open*EHR reference model are clear cut. For example, the **OBSERVATION** class represents, on the one hand, the quality which is observed in the patient, and on the other hand, the performing of the observation activity, thus causing an overlap with the **ACTION** class, which represents the performing of activities. This is however consistent with another principle of *open*EHR: that each single archetype should contain everything necessary to be clinically relevant for specified use cases. In order to assess a result of an observation it is important to have knowledge of how the observation was performed.

Traditionally, the burden of formally representing clinical practice and the health status of the patient has been divided upon a number of technical solutions. Certain parts of this domain have been represented using ontology-language solutions using different kinds of Description Logic (DL), for example the one in which SNOMED CT is represented [7], while other parts of this domain have been represented using database schemas, UML diagrams or, as described above, archetypes. This division into different representational artifacts can be seen as a "divide-and-conquer" approach to dealing with the inherent complexity of medicine and clinical practice. This division is however not clear and there is no consensus on how to make this division, although attempts has been made [8, 9].

This lack of consensus can be exemplified by looking at the (now classical) blood pressure archetype and the SNOMED CT representation of the same entity. The archetype includes a separate slot for stating the state of the patient at the time of observation (see figure 1), for example whether the patient was standing, sitting or lying down. SNOMED CT gives (almost) equal opportunities for representation with codes like:

> **163034007|Standing blood pressure (observable entity)|, 163035008|Sitting blood pressure (observable entity)|, and 163033001|Lying blood pressure (observable entity)|.**

The selection of representation in each specific case is up to the clinical modeller and although emerging guidelines exist [8] our own experiences tells us that such guidelines are hard to apply. Also, we have yet not found evidence of any systematic useage of such guidelines.

3 Representational Requirements of the EHR

The requirements for representing the EHR can be divided into two separate cases: runtime requirements, and design-time requirements. In runtime, the amount of data stored is typically very large and the computational complexity of individual inferences performed must be kept very low for example when doing inferencing on a population scale. In design time, when for example building archetypes, the size of the models is moderate at most. Due to the uniqueness of archetypes, ontologies can generally be partitioned, for example when doing validation through satisfiability checking only the reference model and any specialised archetypes need to be reasoned over. Thus, inferences of greater complexity may be allowed in design time.

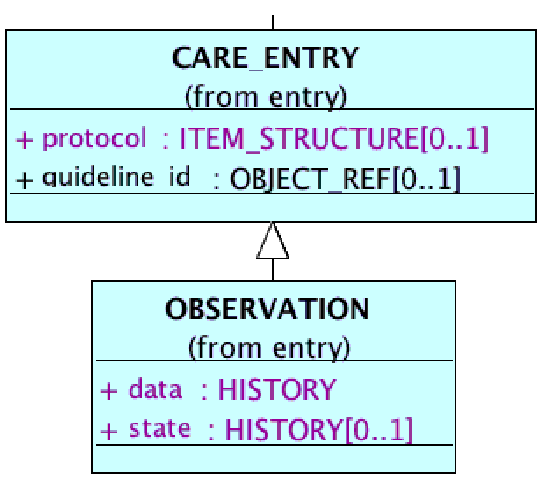
Earlier work has shown that it is feasible to represent information model schemas such as the *open*EHR reference model [10, 11]. Analysing the ontologies representing the reference model, the concept constructors used are concept conjunction (⊓), existential- (∃), value- (∀) and number restrictions, and nominals. Value restrictions are used for attribute data type specifications in archetypes and cardinality- and occurrence constraints in archetypes are represented using existential- and number restrictions. As shown by Baader et al., subsumption testing is intractable with respect to general TBoxes [12, 13]. Number restrictions, when added to conjunction and existential restrictions, lead to ExpTime--completeness. Archetypes used to specialize reference model classes further add qualified number restrictions [14] to the list of required concept constructors.

Some parts of the *open*EHR reference model will require special attention, specifically what is called the openEHR Archetype Profile which includes some classes with special semantics. For example, the reference model class **DV_QUANTITY** includes the attribute unit to represent unit of measurement. This attribute should be consistent with an Archetype Profile constraint on kind of quantity, for example that a meter unit is consistent with a length kind of quantity.

As an example of an *open*EHR reference model class, the **OBSERVATION** class is used to represent observations of the patient's health status, see figure 2. The **OBSERVATION** class inherits from the **CARE_ENTRY** class and adds two attributes: data, to represent the results of the observation and state, to represent the state of the patient at the time of the observation. From the **CARE_ENTRY** class the inherited attribute protocol is used to represent the observation procedure. The **CARE_ENTRY** class in turn inherits the attribute subject to distinguish which patient this aggregate of information concerns.

The **data**, **state** and **protocol** parts may be further specified in archetypes. Figure 3 shows part of an observation archetype from the *open*EHR archetype repository expressed in a DL syntax. The **data** part represents the results of the observation, in this case the blood pressure values.

Additionally, opinions, plans, and proposals are significant parts of the health record. However, as noted by Rector and Brandt as well as Schulz et al., representations of plans require a more expressive description logic than the EL profile used by most large scale DL-based ontologies including SNOMED CT [15, 16].




```
OBSERVATION ≡  CARE_ENTRY ⊓
               ∀ data . HISTORY ⊓
               = 1 date ⊓
               = state . HISTORY ⊓
               ≤ 1 state
```

Figure 2. The *open*EHR OBSERVATION class in UML and DL

```
        Blood-pressure ≡ OBSERVATION ⊓
                           ∀ data . ( HISTORY ⊓
                                      ...
                                      ≤ 1 items . Systolic ⊓
                                      ≤ 1 items . Diastolic ⊓
                                      ...
                                      ) ⊓
                           =1 data
                           ...
        Systolic         ≡ ( ELEMENT ⊓
                           ∀ value .( C DV QUANTITY ...) ⊓
                           =1 value )
      Diastolic          ≡ ( ELEMENT ⊓
                           ∀ value .( C DV QUANTITY ...) ⊓
                           =1 value )
```

Figure 3. Part of an *open*EHR OBSERVATION archetype in DL

Recently, the IHTSDO has presented a style guide for representing observable entities and procedures in laboratory medicine in SNOMED CT [17] influenced by, among other things, the OBO phenotypic quality ontology (PATO) [18]. Both rely on the EL profile for representation: PATO for representing qualities and the SNOMED CT observables model for representing qualities as well as how those qualities are observed and represented, indicating that both qualities and observation procedures may be represented using tractable DL.

4 Discussion

When choosing means of representation for a specific use case, there is always the trade-off between expressivity and computational complexity to consider. Major use cases such as exclusions (negation) and plans (value restriction) as well as common information model constructs (for example value and qualified number restrictions) cannot be faithfully represented using DLs with polynomial time subsumption testing [12, 13], and must, at least in runtime be ruled out. However, evidence guiding the choice of representation are still relatively sparse. For example partitioning of ontologies for specific use cases might allow the practical use of logics with non-tractable complexity.

Then, what could be those guiding principles for the choice of representation? From the perspective of Bodenreider et al., the division between the representation of "biomedical reality" and the representation "how this reality is perceived" (observed) can be understood as the distinction between ontology and epistemology [19]. But the observation of qualities in biomedical reality is in itself a part of biomedical reality, although a part distinct from the quality being observed. So what is from one perspective thought of as belonging to the domain of epistemology may in some other perspective belong to the domain of ontology. Also, as shown by the observables example above, both the ontology of qualities as well as epistemological aspects of qualities (for example observation procedures) may be represented using tractable ontology languages.

If using expressive DLs in runtime for reasoning on typical information models is not possible, there still might be a place for such logics in design time. An alternative could be to encode both ontologies and information models in DLs from which several (computationally tractable) linearisations are generated in order to address different reasoning needs (for example satisfiability checking). Checking validity and conformance to a reference model are examples of design time use cases where satisfiability checking may be applied.

A research program for testing the hypothesis that expressive DL may be used for typical design time EHR representation tasks such as information modelling and terminology binding would include:

– tools for automatically translating *open*EHR archetypes to an OWL representation, including existing terminology bindings, allowing the testing of different approximations to the archetype semantics,

- translation of the full set of archetypes in the *open*EHR public repository, and
- running typical design-time inferencing tasks.

The demarcation between ontology- and information model representations cannot be easily found in the nature of the entities represented, but rather in the complexity of the kinds of inference needed in a specific use case. As noted by Berzell, the context and the intended use of the EHR will be most important when choosing what methods, principles and technologies to use [20]. This is something that will have to be decided in close cooperation between expertise in both healthcare, ontology and informatics.

Acknowledgements

This work was supported by the DebugIT project of the EU 7th Framework Program grant agreement ICT-2007.5.2-217139.

References

1. ISO. *ISO 12967-2:2009: Health informatics - Service architecture – Part 2: Information viewpoint*. International Organization for Standardization, Geneva, Switzerland, 2009.
2. A.D. Spear. Ontology for the twenty first century: An introduction with recommendations. Saarbrucken*, Germany: Institute for Formal Ontology and Medical Information Science*, 2006.
3. Thomas Beale and Sam Heard. An ontology-based model of clinical information. *Studies in Health Technology and Informatics*, 129(Pt 1):760–764, 2007. PMID: 17911819.
4. *open*EHR. The openEHR reference model. http://www.openehr.org/svn/ specification/TRUNK/publishing/roadmap.html, 2010.
5. T. Beale. Archetypes: Constraint-based domain models for future-proof information systems. In *OOPSLA 2002 workshop on behavioural semantics*. Citeseer, 2002.
6. *open*EHR. *open*EHR Clinical Knowledge Manager. http://www.openehr.org/ knowledge/.
7. Stefan Schulz, Boontawee Suntisrivaraporn, Franz Baader, and Martin Boeker. SNOMED reaching its adolescence: ontologists' and logicians' health check. *International Journal of Medical Informatics*, 78 Suppl 1:S86–94, April 2009. PMID: 18789754.
8. HL7 terminfo project. http://www.hl7.org/special/committees/terminfo/ index.cfm.
9. D. Markwell, L. Sato, and E. Cheetham. Representing clinical information using SNOMED clinical terms with different structural information models. In *KR-MED 2008*, page 72, 2008.
10. C. Martínez-Costa, M. Menárguez-Tortosa, J. T Fernández-Breis, and J. A Maldonado. A model-driven approach for representing clinical archetypes for semantic web environments. *Journal of biomedical informatics*, 42(1):150–164, 2009.
11. Roland Hedayat. Semantic web technologies in the quest for compatible distributed health records. Master thesis, Uppsala University, 2010.
12. F. Baader, S. Brandt, and C. Lutz. Pushing the EL envelope. In *International Joint Conference on Artificial Intelligence*, 19, page 364, 2005.
13. F. Baader, S. Brandt, and C. Lutz. Pushing the EL envelope further. In *Proceedings of the OWLED 2008 DC Workshop on OWL: Experiences and Directions*, page 22, 2008.
14. O. Kilic, V. Bicer, and A. Dogac. Mapping Archetypes to OWL. *Middle East Technical University, Ankara, Türkei*, 2006.
15. A.L. Rector and S. Brandt. Why do it the hard way? The case for an expressive description logic for SNOMED. *Journal of the American Medical Informatics Association*, 15(6):744, 2008.
16. Stefan Schulz, Daniel Schober, Christel Daniel, and Marie-Christine Jaulent. Bridging the semantics gap between terminologies, ontologies, and information models. *Studies in Health Technology and Informatics*, 160(Pt 2):1000–1004, 2010. PMID: 20841834.
17. IHTSDO. SNOMED CT® style guide: Observable entities and evaluation procedures (Laboratory), June 2010.
18. C. Mungall, G. Gkoutos, N. Washington, and S. Lewis. Representing phenotypes in OWL. *OWL: Experiences and Directions (OWLED 2007)*, Innsbruk, Austria, 2007.
19. O. Bodenreider, B. Smith, and A. Burgun. The ontology-epistemology divide: A case study in medical terminology. In *Formal ontology in information systems: Proc Third International Conference (FOIS-2004)*, 185-195. IOS Press.
20. Martin Berzell. *Electronic Healthcare Ontologies: Philosophy, the real world and IT structures*. PhD thesis, Linköping University, 2010.