

Clinical Ontologies Interfacing the Real World

Stefan Schulz

(Department of Medical Informatics, University Medical Center Freiburg, Germany
stschulz@uni-freiburg.de)

Holger Stenzhorn

(Department of Medical Informatics, University Medical Center Freiburg, Germany
and
Institute for Formal Ontology and Medical Information Science, Saarbrücken, Germany
holger.stenzhorn@uniklinik-freiburg.de)

Martin Boeker

(Department of Medical Informatics, University Medical Center Freiburg, Germany
martin.boeker@uniklinik-freiburg.de)

Rüdiger Klar

(Department of Medical Informatics, University Medical Center Freiburg, Germany
klar@mi.ukl.uni-freiburg.de)

Barry Smith

(Department of Philosophy, University of Buffalo, State University of New York, USA
and
Institute for Formal Ontology and Medical Information Science, Saarbrücken, Germany
phismith@buffalo.edu)

Abstract: The desideratum of semantic interoperability has been intensively discussed in medical informatics circles in recent years. Originally it was assumed by many that this issue could be addressed simply by insisting on the application of shared clinical terminologies. More recently however the use of the term ‘ontology’ has been steadily growing. We here address the issue of the degree to which the use of ontologies represents any real advance on the road to semantic interoperability.

Keywords: Clinical Ontologies, Knowledge Representation

Categories: I.2.4, SD J.3

1 Introduction

The desideratum of semantic interoperability has been intensively discussed in medical informatics circles over the past decade [Rossi-Mori 98, Ingenerf 01, Garde, 07]. Consider for example the evolution of the Unified Medical Language System

(UMLS)¹, of the Open Biomedical Ontologies (OBO)², of the HL7 Common Document Architecture (CDA) [Dolin 06], of openEHR [Kalra 05], or of SNOMED CT³.

Originally the issue of semantic interoperability was supposed to be addressed mainly by applying clinical terminologies. More recently, however, we have seen steady growth in usage of the term ‘ontology’. The issue addressed here is: does this constitute any real advance or advantage? There is indeed good reason to cast at least some doubt on the claims made on ontology’s behalf: Too many recent publications, calls for research proposals and project descriptions have embodied what are in our view (and thus sometimes already proven themselves to be) insupportable expectations. It is thus understandable that some have been tempted to see in ontology just one more new and flashy buzzword. To begin, we must first ask where the proper difference lies between terminology on one side and ontology on the other. Since neither one of those two terms has been unambiguously defined so far, we will adhere in the context of this paper to the following definitions:

- Terminology: A set of terms representing the system of concepts of a particular subject field. [ISO 00]
- Ontology: The study of what there is. Formal ontologies are theories that attempt to give precise formulations of the types of entities in reality, of their properties and of the relations between them. [Quine 1948]

2 Delimiting The Concept of Ontology

2.1 Contrasting Ontologies with Terminologies

We summarize our position on clinical ontologies and terminologies as follows:

Terminologies are term centered. They relate the senses or meanings of linguistic entities. Classes of (quasi-)synonymous terms are commonly referred to as ‘concepts’. In many terminology systems (often called thesauri or semantic lexicons), concepts are furthermore related by informal semantic relations which are often closely related to natural language predicates⁴. In medical informatics, this language-centered view characterizes the UMLS legacy. In spite of its well-known shortcomings, the UMLS can be seen as a robust and highly useful platform for the retrieval of terms belonging to of heterogeneous, context-dependent, informal terminology systems.

Ontologies are intended to describe a portion of reality that exists independently of human language. Their constituent nodes are (entity) *types*, not concepts. Types (often also referred to as ‘categories’, ‘kinds’ or ‘universals’) are well suited to hierarchically order the particular entities (patients, lesions, surgical procedures) which exist on the side of reality. The existence of certain types and the basic structure of ontological principles are subject to major philosophical disputes. However, at any given stage in the development of science, there is a consensus core of scientific understanding of reality, and it is this (on our view) which should serve as starting point

¹ <http://www.nlm.nih.gov/research/umls>

² <http://obofoundry.org>

³ http://www.snomed.org/snomedct/documents/january_2007_release.pdf

⁴ In the context of this paper we do not demand the existence of inter-concept relations as a necessary criterion for terminologies.

in developing science-based ontologies. Examples of statements belonging to this consensus core are: that humans are vertebrates, that cells contain membranes, that adenosine diphosphate is phosphorylated in mitochondria, that the retina contains photosensors. However, types of the sort with which we are concerned in medicine are elucidated not only by such observation-based descriptions of nature, but also often result from a prescriptive, defintory process: appendectomy (type) is *defined* as the “*surgical removal of an appendix*”, and hepatitis (type) is *defined* as an “*inflammation of liver tissue*”. (Such axioms may be ignored but cannot be falsified.) Types in an ontology apply to classes of entities in the real world (also called individuals or instances) since classes indeed include collections of the entities instantiating a given type. The main construction tenet for ontologies is the taxonomic principle: a type *S* is a subtype of a type *T* if and only if all instances of *S* are also instances of *T*.

2.2 Ontology Constraints

Based on understanding ontologies as representations of types of entities and of relations between them, we can rule out some common misconceptions that often obscure the sharp difference between ontologies, terminologies and other artifacts such as representations of contextual knowledge and information models for data acquisition.

As a fundamental principle, *all* properties of a given type in any ontology are true for *all* instances of this type. Thus all instances of appendectomy are performed on some instance of appendix; all instances of water molecules contain oxygen and hydrogen. This restricts the ability of an ontology to express seemingly obvious assertions such as “*hands have thumbs*” or “*aspirin alleviates headache*”, because there are hands without thumbs and not all aspirin tablets are used to alleviate any headache.

A further restriction is that probabilistic assertions, which are of tremendous importance for everyday clinical reasoning, cannot be expressed in an ontology in a simple way. For example, if a prevalence of 1% is ascribed to lung cancer then this is not a property inhering in any instance of lung cancer. It is rather a factual statement about some given population with respect to the occurrence of this disease. However, the common consensus in science is in many areas based on probabilistic theories which describe results in terms of probabilistic states, processes and events. So is the assessment of risks (of signs, symptoms, and therapies for specific diseases) commonplace in medical practice. (e.g., arterial hypertension is considered a risk for stroke). Unsatisfactorily, the related entity types and relations cannot be straightforwardly represented in formal ontologies following the principles described above. A possible solution is to introduce probabilistic dispositions [Jansen, 07] into the ontology, i.e., dispositions to do something (under certain circumstances) with a certain probability. Such dispositions are related to events by the relation of *realization*. They are special kinds of dependent entities, in that they need not be realized in order to exist. E.g., “risk for stroke” could be represented in such a way.

These fundamental constraints are corollaries of the fact that all assertions of relations between types in ontologies should be of the basic form of universal statements: “*for all instances of type T there is some...*”. We could, of course, consider types and instances as two different ranges for our quantifiers. Then, however, we would have to accept some higher-order logic, which would cause problems for machine reasoning since such logics are known not to be computable in all circumstances using cur-

rent algorithms. By contrast, languages from the family of Description Logics [Baader et al. 03] are computable and therefore frequently used in the ontology development.

2.3 Epistemological Classification Criteria

Classes are the basic building blocks for clinical classification systems as the ICD [ICD 07], which, for the time being, provide the most significant support of semantic interoperability of clinical data. It has repeatedly been observed that medical classification systems (even claiming to classify entities in reality) are distinguished from ontologies because of their use of “*un-ontological*” classification criteria (i.e., such as to represent the knowledge independent reality of the entities) but rather “*epistemological*” (i.e., to represent the knowledge one has about these entities) [Bodenreider et al. 04]. Thus the current ICD makes a classificatory distinction between cases of tuberculosis diagnosed by bacterial culture and those diagnosed by histology. But, a particular disease is not different in nature only because of a different diagnostic method.

Epistemological issues are nevertheless crucial for medical documentation. Diagnostic statements tend to be error-prone with vital decisions often based upon brittle evidence. Necessary or desirable information may simply be missing. So a place must exist to encode the information one actually has in the practically available form. An ontology is, however, not the right place for this. Classes such as “*unspecified tumor stage*” or “*infection of unknown origin*” do not stand for more specific subclasses. They just manifest lack of adequate knowledge, mixing up “*what is*” with “*what we know*”. Such knowledge is important in the clinical scenario but requires additional means to represent contextual knowledge in encoding specific clinical instances.

3 Clinical Ontologies in Practice

All of the questions addressed above arise, to different degrees, in cases where formal domain ontologies are expected – in the framework of clinical research projects but also in routine documentation – to improve data acquisition, standardization, interoperability, as well as data analysis. We report on experiences within the projects ACGT (Advancing Clinico-Genomic Trials on Cancer)⁵ and @neurist (Integrated Biomedical Informatics for the Management of Cerebral Aneurysms)⁶, in which customized ontologies are required and are being currently developed.

Both ACGT and @neurIST aim at setting up integrated information technology infrastructures by implementing common software platforms to improve disease management through a more efficient processing and presentation of knowledge and data. ACGT focuses on nephroblastoma and breast cancer, basing its work on a master ontology for cancer supporting the facility to create clinical report forms automatically to support clinical trial research in cancer genomics. @neurIST is concerned with acquiring and estimating the risk of intracranial aneurysms and subarachnoid hemorrhage based on multimodal data. The goal of the developed platforms is the integration of data from various sources and disciplines within the projects (e.g., clinical studies, genomic research and patient management). These data are highly

⁵ <http://www.eu-acgt.org>

⁶ <http://www.aneurist.org>

fragmented and heterogeneous in regard to format, scale and their particular content reflecting the projects' specific sub-domains.

Consequently, it is a big challenge to design ontologies that acknowledge this broad scope and are capable of integrating available data. Semantic interoperability here means that all data collected for each individual patient, for each experiment, or of each literature abstract considered relevant for the domain, should point to nodes in a domain ontology. One issue that particularly complicates this task is the multitude of entity types and the different scales of spatial and temporal granularity (i.e., medical, biomolecular or epidemiological entities from single cell division to human life). Further, the ontologies have to integrate various levels of description in the available data (e.g., literature, clinical databases, imaging databases and terminologies).

4 Interfaces of Ontologies

Lessons learned from ACGT and @neurIST have shown that the shortcomings and problems described in section 2 can be alleviated by clearly defining the interfaces between the ontologies and other artifacts in the semantic interoperation environment.

4.1 The Interface between Clinical Ontologies and Terminologies

Ontologies, in a strict sense, are domain descriptions that are independent of human language, so they need not incorporate any lexical or term information at all. The fact that – for practical reasons – they commonly employ human-readable names is not a contradiction to this claim. These names may, but do not have to, coincide with actual domain terms. They do so because ontologies need to be maintained by humans, and are often used by humans in expressing their results without any intervention of a machine. Terms and descriptions in ontologies should be precise, unambiguous and self-explanatory (which is often not the case with typically used domain terms).

The interface between domain ontologies on the one hand and domain term lists on the other hand is characterized by a many-to-many relationship: Several terms may be connected to one ontology class due to the phenomena of synonymy and cross-language translation, and polysemy has to be accounted for to link an ambiguous term to more than one ontology class. For instance, the natural language terms “*mamma carcinoma*”, “*breast cancer*” and “*Brustkrebs*” are linked to the same node in the ACGT ontology, whereas “*ulcer*” points to two different ontology nodes, *viz.* first, the process of ulceration and then, the pathological structure (the *result* of this process).

4.2 The Interface between Clinical Ontologies and Upper Ontologies

In our application contexts the project-specific ontologies are rooted in upper ontologies. According to the Standard Upper Ontology Working Group⁷ upper ontologies provide generic categories or types suited to address a broad range of domain areas at a high level in a way which can support integration of the underlying data. These upper-level types provide a highly general structure which in turn helps the consistent representation of the entity types in the associated domains. Following this, the use of an upper ontology is intended to improve the development of the actual domain on-

⁷ <http://suo.ieee.org>

tologies by providing a consistent and sound top-level framework. Whereas the ACGT ontology uses the Basic Formal Ontology (BFO) [Grenon et al. 04] as its upper level, the @neurIST ontology employs the Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) [Gangemi et al. 02]. For a comparison of the two upper level ontologies see [Mascardi et al. 07].

4.3 The Interface between Clinical Ontologies and Non-Ontological Knowledge

As we said before, the domain representations generated by large research projects must extend what can be expressed by ontologies. This extension is here referred to as “*non-ontological knowledge*”. Typical examples are assertions such as “*A treats B*” and “*C is a risk for D*”. This knowledge is still at the terminological level (rather than at the level of instance data), but it is not knowledge which should properly be included in an ontology because it does not express what holds universally of the given types. There are, in principle, two different ways to represent such knowledge:

- The first follows the representational scheme of the UMLS Metathesaurus. Non-ontological knowledge is represented in a thesaurus-like terminology, often based on *concept–relation–concept* triplets. Thus it does not support logic-based reasoning about *classes*, as in logic-based representations, but is anyhow available for informal reasoning about *concepts*, e.g. related concept search by graph transversal.
- The second solution has been applied in the @neurIST ontology and consists in a parallel system of non-ontological, reified classes (e.g. the class “*suspected risk factor for aneurism rupture*”) thus inserting epistemological categories into the taxonomy. These categories are irrelevant for the correct ontological entity description but needed for the specific retrieval requirements of users in the project. From the representation perspective, however, this difference is ignored. So “*hypertension*” is a subclass of the above class just as it is a subtype of “*cardiovascular disorder*”.

4.4 The Interface between Clinical Ontologies and Information Models (and the World)

Information models are templates for the acquisition of clinical data which enable semantic interoperability in the scope of the given information model but not between different information models. They can be based, e.g., on openEHR archetypes [Beale et al. 01, Kalra et al. 05], and are built in such a way as to involve reference to ontologies but they are not by themselves ontologies. In an information model we encode what we know about concrete instances in a certain situation and under certain circumstances. Besides offering a template for the facts to be reported, the models may further include the conditions of measurement, the certainty of an assertion, or other contextual factors. This is why information models are necessary and the simple instantiation of ontology classes is not sufficient in clinical documentation. Ontologies provide the types for the particular instances to be recorded in an information model. This relation has recently received increased attention in the context of openEHR Archetypes, HL7 Version 3 Clinical Document Architecture, and SNOMED CT [Recator et al. 06], and has been further discussed by including experiences from large scale implementation attempts such as the UK Connecting for Health project.

4.5 The Interface between Clinical Ontologies and the Ontology Engineer

The actual interface between clinical ontologies and their developers is an ontology editing environment that ideally offers supports ontology development and maintenance by a graphical user interface and releases the ontology engineer from the need to access and edit the actual source code, e.g. the Web Ontology Language (OWL)⁸. Most editors allow users to further describe ontology nodes both with textual information and logical definitions. The latter can be used by terminological reasoners to enable automated checking of the structural and (to some extent) semantic correctness of the ontology. Both ACGT and @neurIST use open-source software, *viz.* Protégé⁹ (currently the most widely-used ontology editor) together with the reasoner Pellet¹⁰.

4.6 The Interface between Clinical Ontologies and the Application Builder

Application builders need a way to programmatically access the content and structure of an ontology in order to create software systems that refer to this ontology, as in the ACGT and @neurIST projects. Therefore generic application programming interfaces (API) have been developed that can be used by application builders for example to take a given entity type from the ontology and link it with a multilingual terminology system. Another example is the development of an easy-to-use retrieval interface for the ontology content, since it turned out that ontology editors such as Protégé are too complex and therefore less suited for application builders.

5 Conclusion

Ontologies are important informatics resources for large multicentric clinical research projects because they foster semantic interoperability. They offer a stable, language-independent vocabulary that helps standardize and explain the meaning of domain terms. However, ontologies are often mixed up with terminologies, thesauri, and representations of contingent or probabilistic domain knowledge, as well as with database-centered information models which serve recording of instance data. This often creates exaggerated expectations on the part of the users of an ontology. We argue in the above that in order to minimize such expectations we should clearly delimit the scope of ontologies from that of other knowledge and information resources. To this end we defend the introduction of clearly defined interfaces between ontologies and other supporting artifacts. Only after their success being proven in clinical research projects, formal ontologies can be expected to be seriously used to enable semantic interoperability in the clinical routine, as well.

Acknowledgements

We thank our colleagues and ontology co-developers: Susanne Hanser (@neurIST) and Mathias Brochhausen and Cristian Cocos (ACGT). Work on this paper has been carried out cooperatively within the @neurIST and the ACGT integrated projects funded by the European Commission (IST-027703, IST-026996).

⁸ <http://www.w3.org/TR/owl-features>

⁹ <http://protege.stanford.edu>

¹⁰ <http://pellet.owldl.com>

References

- [Baader et al. 03] Baader, F., Calvanese, D., McGuinness, D., Nardi, D., Patel-Schneider, P.: “The Description Logic Handbook: Theory, Implementation and Applications”; Cambridge University Press, Cambridge, United Kingdom (2003)
- [Beale et al. 01] Beale, T., Goodchild, A., Heard, S.: “EHR Design Principles”; London, United Kingdom (2001)
- [Bodenreider et al. 04] Bodenreider, O., Smith, B., Burgun, A.: “The Ontology-Epistemology Divide: A Case Study in Medical Terminology, Proc. FOIS-2006, Torino, Italy (2004)
- [Dolin et al. 06] Dolin, R., Alschuler, L., Boyer, S., Beebe, C. Behlen, F., Biron, P., Shabo Shvo, A.: “HL7 Clinical Document Architecture, Release 2”; J Am Med Inform Assoc, 13, 1 (2006) 30-9.
- [Gangemi et al. 02] Gangemi, A., Guarino, N., Masolo, C., Oltramari, A., Schneider, L.: “Sweetening Ontologies with DOLCE”; Proc. EKAW-2002, Sigüenza, Spain (2002)
- [Garde et al. 07] Garde, S., Knaup, P., Hovenga, E., Herd, S.: “Towards Semantic Interoperability for Electronic Health Records”; Method Inf Med, 46, 3 (2007) 332-343
- [Grenon et al. 04] Grenon, P., Smith, B., Goldberg, L.: “Biodynamic Ontology: Applying BFO in the Biomedical Domain”; in Pisanelli, D. (ed.): “Ontologies in Medicine”, IOS Press, Amsterdam, Netherlands (2004).
- [Ingenerf et al. 01] Ingenerf, J., Reiner, J., Seik, B.: “Standardized Terminological Services Enabling Semantic Interoperability between Distributed and Heterogeneous Systems”; Int J Med Inform, 64, 2-3 (2001) 223-40.
- [ISO 00] International Organization for Standardization: “ISO 1087-1: Terminology work – Vocabulary – Part 1: Theory and applications”, Geneva, Switzerland (2000)
- [Jansen 07] Jansen, L. On Ascribing Dispositions. In Bruno Gnassounou and Max Kistler, editors, Dispositions and Causal Powers, pages 161–177. Ashgate, Aldershot, 2007.
- [Kalra et al. 05] Kalra, D., Beale, T., Heard, S.: “The openEHR Foundation”; Stud Health Technol Inform, 115 (2005) 153-73
- [Mascardi et al. 07] Mascardi, V., Cordi, V., Rosso, P.: “A Comparison of Upper Ontologies”; Technical Report DISI-TR-06-2, Genova, Italy (2007)
- [Quine 1948] On What There Is. Review of Metaphysics. (1948)
- [Rector et al. 99] Rector, A., Zanstra, P., Solomon, W., Rogers, J., Baud, R.: “Reconciling Users Needs and Formal Requirements: Issues in Developing Re-Usable Ontology for Medicine”; IEEE Transactions on Information Technol in BioMedicine, 2, 4 (1999) 229-242
- [Rector et al. 06] Rector, A., Qamar, R., Marley, T.: “Binding Ontologies and Coding Systems to Electronic Health Records and Messages”; Proc. KR-MED-2006, Baltimore, USA (2006)
- [Rossi-Mori et al. 98] Rossi-Mori, A., Consorti, F.: “Exploiting the Terminological Approach from CEN/TC251 and GALEN to Support Semantic Interoperability of Healthcare Record Systems”; Int J Med Inform, 48, 1-3 (1998) 111-124
- [Uschold and King 96] Uschold, M., King, M.: “Ontologies: Principles, Methods, and Applications”; Knowledge Eng. Rev., 11, 2, (1996) 93-155