

## Ontological Theory for Ontological Engineering: Biomedical Systems Information Integration

James M. Fielding<sup>1,2</sup>, Jonathan Simon<sup>1</sup>, Werner Ceusters<sup>2</sup>, Barry Smith<sup>1,3</sup>

<sup>1</sup>Institute for Formal Ontology and Medical Information Sciences, Härtelstrasse 16-18, 04107 Leipzig, Germany

<sup>2</sup>Language and Computing nv., Maaltecenter Blok A, Derbystraat 79, 9051 Sint-Denijs-Westrem, Belgium

<sup>3</sup>Department of Philosophy, University of Buffalo, Buffalo, U.S.A.

### Abstract

Software application ontologies have the potential to become the keystone in state-of-the-art information management techniques. It is expected that these ontologies will support the sort of reasoning power required to navigate large and complex terminologies correctly and efficiently. Yet, there is one problem in particular that continues to stand in our way. As these terminological structures increase in size and complexity, and the drive to integrate them inevitably swells, it is clear that the level of consistency required for such navigation will become correspondingly difficult to maintain. While descriptive semantic representations are certainly a necessary component to any adequate ontology-based system, so long as ontology engineers rely solely on semantic information, without a sound ontological theory informing their modeling decisions, this goal will surely remain out of reach. In this paper we describe how Language and Computing nv (L&C), along with The Institute for Formal Ontology and Medical Information Sciences (IFOMIS), are working towards developing and implementing just such a theory, combining the open software architecture of L&C's LinkSuite<sup>TM</sup> with the philosophical rigor of IFOMIS's Basic Formal Ontology. In this way we aim to move beyond the more or less simple controlled vocabularies that have dominated the industry to date.

### 1. Introduction

The central hypothesis behind the collaboration of Language and Computing nv (L&C) and The Institute for Formal Ontology and Medical Information Science (IFOMIS) is that the methodology and conceptual rigor of a philosophically inspired formal ontology will greatly advance the reasoning capacities of application ontologies. To this end we have submitted L&C's ontology, LinKBase®, which has been

designed in part to integrate and reason across various external databases simultaneously, to the conceptual demands of IFOMIS's Basic Formal Ontology (BFO).

With this project we aim to move beyond the level of the controlled vocabularies that have dominated the industry to date, typically only loosely formalized schemas for which the necessary tools for formal analysis have still to be applied. With attention to sound ontological theory we aim to yield a sound application ontology with the ability to support advanced reasoning applications correctly and efficiently across various external databases simultaneously.

Our general procedure has been the implementation of a meta-ontological definition space in which the definitions of ontological elements and relations within LinKBase® are standardized in a framework of first-order logic. By 'standardization' we mean providing a system of axioms that may be used to reason across the entire domain and thereby provide a solid basis for a consistent standard of data integration and quality control. In this paper we first describe how implementing this standardization has already led to an improvement in the LinKBase® structure by providing a greater degree of *internal* consistency. We then describe how this resulting coherence has increased our ability to utilize LinKBase® as a 'base ontology,' or translation hub, for the purposes of coupling external databases with a marked increase in *external* coherence and expressive capabilities. In the discussion that follows, we demonstrate how this offers a genuine advance over other application ontologies that have not yet submitted themselves to the demands of philosophical scrutiny, and suggest some further avenues of possible research that such an approach has offered us.

## 2. LinKBase® and Basic Formal Ontology

### 2.1 LinKBase®

LinKBase® is a biomedical domain ontology that has been designed in part to integrate large terminological systems and databases. Upon these databases the ontology management system (OMS), LinKFactory®, runs applications designed for natural language processing and unstructured text information retrieval. The base ontology itself, before the integration of external database information, contains over one million language-independent concepts of a medical and general-purpose variety. In turn, these concepts are linked in a semantic network via some 480 different “link-types,” used to express varying sorts of relationships. Both concepts and links are language independent, and are in turn cross-referenced to approximately 3,000,000 terms in various natural languages and vocabularies. In this way, LinKBase® provides a central hub ontology with fixed structured definitions upon which external medical terminologies and databases, such as UMLS®, SNOMED®, and the Gene Ontology (GO), may be grafted (Flett, Dos Santos, and Ceusters 2002).

At times, however, this task turns out to be an exceedingly difficult endeavor. To a large extent, these difficulties lay in the fact that the various terminologies and databases that are to be integrated are often internally and mutually inconsistent. Yet, as all these terminologies must essentially speak about the same reality, we believe there is a common thread that binds them. Our approach is based on the idea that it is possible to integrate these various structures on the basis of a sound understanding of those basic categorical distinctions that are common to them all.

### 2.2 Basic Formal Ontology

IFOMIS's Basic Formal Ontology (BFO) is a philosophically inspired top-level ontology (Grenon and Smith, forthcoming) that provides a coherent and unified understanding of these basic ontological elements and relations that are fundamental to our reality, exactly those elements that are required to successfully integrate the diverse domain specific terminologies that lay scattered throughout the biomedical informatics community. BFO is currently being implemented as an open-source backbone ontology for LinKBase®. BFO provides a common foundation upon which we may map external ontologies, terminologies, and databases onto LinKBase® in a manner that is designed to facilitate successful merger of these structures, as well as to provide a useful guide for the future development of

algorithm-based cross-ontology navigation (Monteyne and Flanagan 2003)

### 2.3 Ontological Distinctions

We begin by reviewing a small number of the fundamental ontological distinctions that form the basis of our philosophical methodology (cf. <http://ontology.buffalo.edu>). These distinctions will figure in the case studies cited below.

#### 2.3.1 Universals and Particulars

As realist philosophers in the Aristotelian tradition we have come to distinguish between universals (also called kinds, species, or types) and particulars (individuals, instances, or tokens). An example of a universal would be the species “Malaria” that a doctor studies in medical school, or the general function “to boost insulin production.” An example of a particular would be *this* malaria present in *this* blood sample, or the function of *this* gene to boost insulin production in *these* beta cells in your pancreas.

#### 2.3.2 Endurants and Occurrents

From each of these distinctions (i.e. universals and particulars) we can further derive two sorts of entities: endurants and occurrents. These two sorts of entities relate differently to time. Endurants are those entities, which, as the name implies, *endure* through time; they are wholly present at each moment of their existence. Examples of endurants are those mesoscopic objects common to the world of tables and chairs, people, operating rooms, cells, and chromosomes. All of these kinds of entities, and all of their parts, maintain their full identity from one moment to the next.

Occurrents, on the other hand, are those sorts of entities that are never fully present at any one given moment in time, but instead *unfold* themselves in successive phases, or temporal parts. Entities that occur are processes or events such as a morning run, a surgery session, or cellularization.

In a parallel fashion, where your arm is a part of you and your hand is a part of your arm, your youth is a part of the processes which is your life and your first day at school is a part of your youth. It is important to note at this point that parthood never crosses these boundaries - parts of endurants are always endurants and parts of occurrents always themselves occur.

#### 2.3.3 Dependent and Independent

Some entities have the ability to exist without the ontological support of other entities; these are entities such as people, tables, or molecules. These sorts of entities we call independent. On the other hand, there are entities, every bit as real, but that require the existence of these first sort of entities for their own

existence, like a morning run needs a runner, or a viral infection is dependent on the virus and the organism infected.

All occurrent entities require an independent entity in which to inhere; in other words, there is no process without a substance to bear it. But within the category of enduring entities there are both dependent and independent entities, such as the function of an organ that depends on the existence of that organ.

## 2.4 Meta-Ontological Theory

Inevitably, when working across various sub-domains within an area of scientific research, as application ontologies are often required to do, difficulties in accounting for apparently incommensurable perspectives on reality will arise. It is our belief therefore, that above and beyond the ontological theories that BFO provides, *meta-ontological* theories are also required. The Theory of Granular Partitions (Bittner and Smith 2003), being developed by the researches at IFOMIS, aims to provide a meta-ontological theory of sufficient sophistication to allow us to move between various sub-domains or levels of exactitude consistently and in this way account for some of the diverse perspectives on reality that may arise in our integration attempt.

It is often the case in ontological engineering that modelers may be tempted to conflate divergent perspectives on reality within a single hierarchy (Smith and Rosse, forthcoming). When “carving-up” reality, modelers must beware not to converge categorically diverse entities in a unnaturally unified “slice.” Without sufficient formal definitions for structuring the ontology itself, these systems are error prone when one attempts to reason across these diverse levels uniformly. This is particularly the case where an ontology makes use of multiple inheritance hierarchies. The users of these systems often lack clear criteria, or *differentiae*, by which to distinguish those categories that properly belong to one level of granularity, as species and genus do, and one may easily be mistaken for the other.

With the Theory of Granular Partitions, we aim to provide a formal system for reverse engineering these sorts of modeling errors. Through the addition of meta-ontological, or structural information, we may isolate and select particular hierarchies within an ontological structure, hierarchies that conform to one level of granularity with clearly defined *differentiae*, while keeping other structures separated. Highlighting one hierarchy, and keeping another in the background, allows one to reason smoothly in one level of granularity while leaving another related but ontologically variant level in the background. Such a

system allows us to address an application ontology that may confuse ontological distinctions, not through altering the structure in question, but by adding information about how various tree structures relate *outside of ontological relations*. While not making any claims about the particular object in question we can describe the structure of the ontological commitments of the various models without explicitly forbidding either.

## 3. General Procedure

### 3.1 Standardization

As ontologies and terminologies expand and as the drive to integrate them increases, it is natural that semantic consistency will become increasingly difficult to maintain. The root of this difficulty is typically the ambiguities and inconsistencies that result from the lack of a standard unified framework for understanding those basic relations that structure our reality. The BFO top-level ontology provides for this with a solid set of standardized, first-order definitions for ontological elements, definitions that may subsequently be exploited by advanced reasoning applications, including those designed for natural language understanding.

In natural language, an identical term may be used to describe very different ontological structures. It is essential therefore, when dealing with these terms in an application ontology that we have a means by which to disambiguate these varied meanings so that our ontology does not in turn mirror this semantic ambiguity. For example, the term “dislocation” may at once be used to describe a dislocated *structure* (an enduring entity) as well as the dislocation *process* (an occurring entity). It is natural therefore, that ontology modelers may be tempted to model the term “dislocation” as both a structure and a process. As we stated above however, parts of endurents are never parts of occurrents and if our modeling does not reflect this distinction, erroneous reasoning is bound to follow. When the dislocation took place as a process, for example, would be indistinguishable from the span of time in which the dislocated structure was present, from its initial occurrence to the time when the bones are set back into place.

By disambiguating the ontological structures underlying informal definitions of insufficient precision, these ontologically inspired formalizations aid in the passage of domain knowledge between users and software agents, and thus improve coherence and adaptability in and between ontologies as well.

The consequent standardization reflects an implementation of philosophical rigor along two dimensions. Initially, it establishes internal consistency

within the base ontology on the basis of precise, philosophically informed conceptual analyses of the elements involved, be they objects, processes, functions, or roles. Ontologies such as LinKBase® (as well as SNOMED and GO) may be viewed as object languages with a certain “surface structure.” These surface structures generally consist, particularly in the case of application ontologies, of networks of concepts joined in binary relations. For the most part however, these relations and concepts are initially given only in natural language and as such, they remain primarily defined on the basis of semantic information. As we have seen, the grammatical form that lies at the base of this information often leads ontology modelers into error due to the various ambiguities inherent in the natural language that expresses such concepts. Thus, the project of defining a common “deep structure” to which every ontological concept, relation, and axiom, may be mapped requires sound conceptual analysis of the elements involved. The standardization effort gives us a tool by which we may identify and repair internal inconsistencies and ambiguities in LinKBase® itself.

The second dimension of rigor requires the use of the standard first-order logical language in which the concepts of BFO are defined and axiomatized. In this way the rigor of the BFO classification system may be imported into an external ontology from the outside via the LinKFactory® ontology management system. This importation is *meta*-ontological, in the sense that changes are not made directly within the external ontology itself; rather, their place in the BFO re-articulated base ontology, in this case LinKBase®, is marked via an external mapping algorithm in a way that provides the degree of consistency required to navigate between the various third-party ontologies such as GO and SNOMED.

The analysis runs as follows:

1. For every ontological element C, the definition consists in a mapping to a pair: <the class named by C, the extension of the class named by C>
2. For every ontological relation R(X,Y), the definition consists in a mapping to a logical formula of the following form: For all x such that x is in the extension of the class named by ‘X’, there is a y such that y is an element in the extension of the class named by ‘Y’, and R\*(x,y). (Where R\* is a relation in the formal language of BFO, for example part-of)
3. Axioms, which are essentially instantiated relations, are defined by a mapping similar to the definition of relation presented above, differing only in that the variables are replaced by specific concepts within the ontology.

In the remainder of this article we discuss the accomplishment of two goals. First, we witness ways in which the philosophical insights afforded by this standardization have allowed us to disambiguate the LinKBase® ontology itself. Second, we discuss the way that the BFO standardization has assisted in our ontology integration effort. We take as our test case L&C’s integration of SNOMED-RT® and the Gene Ontology (GO).

#### 4. Conceptual Analysis and Problems of Internal Consistency.

While the structural difficulties we will present here could certainly be amended were their respective authors to adopt a similar approach based on formal ontological distinctions such as those presented above, our intension was not explicitly to change or remodel SNOMED-RT® or GO - each of these terminologies have their own aims and specific advantages. Rather, it is essential that if we should attempt to integrate these ontologies that we do indeed focus on *integration*, and not assimilation, in order that we do not do away with their strengths along with their weaknesses. Though naturally, this integration attempt requires a certain degree of mutual consistency. Through the use of the BFO defined structure we have been able to map these databases to the LinKBase® ontology simply by adding structural information and not by altering the ontologies in question. This will be discussed further in section 5.

##### 4.1 SNOMED-RT® and the “Parthood” Relation

Identically named concepts often have very different denotations, as we saw above in the case of the naturally vague term “dislocation”; however, this ambiguity is equally problematic in the use of relations as well. The degree of internal consistency required to apply the BFO standardization accurately to an ontology requires that these relations be disambiguated. One common variety of disagreement within a taxonomic system centers on divergent uses of the relation “parthood.” In the initial release of SNOMED-RT®, for example, the concept “amputation of toe” (ID# 57836005/P1-19430) is a special case of the concept “amputation of foot.” (ID# 70638009/P1-19434). But while the toe certainly is a part of the foot, we must recognize that an amputation of the toe is not an amputation of the foot. The former ought to be represented either as a *part of* an amputation of the foot, or alternatively, as an amputation of *part of* the foot. Depending on the context, these are two very different sorts of things.

SNOMED-RT® here confuses the two types of entities discussed above in section 2.2, endurants and

occurs. It confuses that element of parthood associated with the foot, an entity that exists wholly through time, with that parthood associated with an amputation, an event unfolding in temporal parts. It is for reasons such as these that these two dimensions of parthood must be kept apart.

#### 4.2 Objects and Processes in GO

GO is divided into three disjoint hierarchies: the *cellular component*, *biological processes*, and *molecular function* ontologies. The first, equivalent to that of anatomy in the medical domain, is an ontology of endurants. It allows users to access the physical structure with which a gene or gene product is associated. A biological process, on the other hand, is defined in GO as “a phenomenon marked by changes that lead to a particular result, mediated by one or more gene products.” This ontology, as it is apparent, is a hierarchy of occurrents.

There is however some confusion over the role of the molecular function hierarchy (Smith, Williams, and Schulze-Kremer 2003). While GO defines molecular function as “the action characteristic of a gene product,” it is clear that functions do not occur as actions, but rather *endure*; the function of a gene or gene product exists wholly throughout its existence, present at all times even if that function fails to be realized, as in the case of mutant genes. These mutants retain their function, as, for example, “signal transducer activity” remains the function of the EPO\_HUMAN protein even though it is incapable performing the “signal transduction” process.

Molecular functions and biological processes are obviously closely related. The function “signal transducer activity” certainly *involves* performing “signal transduction” in some sense; yet, in GO this relationship is vague and confused. The authors of GO have attempted to clarify this relationship, stating, “a biological process is accomplished via one or more ordered assemblies of molecular functions,” (Gene Ontology general documentation. Cf. <http://www.gene-ontology.org/doc/GO.doc.html>) in order to suggest that the relation is one of agency. Here, functions *initiate* biological processes, but this would further seem to suggest that they share in a relation of parthood.

At the same time however, GO’s authors insist, correctly in our view, that parthood only holds between entities of the same hierarchy. Yet, so long as the associated relations continue to conflate distinct ontological categories within the ontology itself, internal confusions and limitations to reasoning will remain.

#### 5. Applying External Consistency and Mapping Ontological Elements

The Mapping Databases onto Knowledge Systems tool (or MaDBoKS) is an extension of the LinKFactory® OMS that administers and generates mappings from external databases onto LinKBase®. This mapping mediates the data contained in the external database in a manner that “virtually” expands the hub ontology, leaving the structure of the foreign ontology untouched. The mapping tool can map column as well as cell record data in such a way as to carry over relationships into the ontology.

This mapping is split into two broad phases: a conceptual analysis (as discussed above) and the physical mapping phase. In the first phase, the model of the external database is mapped to the ontology semi-automatically to a degree dictated by the structure of the foreign ontology itself. After the structure of the database has been analyzed and mirrored within the LinKBase® ontology using conceptual tools such as those that BFO provides, a generic mechanism translates database data and relations to the LinKBase®-BFO structured ontology. The result of this process is the creation of an XML mapping definition file. In the second phase, described in this section, the external database is physically linked to the OMS. This phase allows users to query the ontology along with the newly integrated databases as if it were a part of the ontology. This means that all the relations mapped to the ontology can be localized in the ontology. All the relations a given concept has with other concepts, whether they originate in the LinKBase® ontology or the original external database, are retrieved automatically from the database using the mapping information defined in the first phase.

The MaDBoKS system meets our requirement that the ontology itself does not change upon coupling or decoupling of the database. In this manner the OMS is able to navigate the difficulties within an external database using the BFO standardization as an effective translation mechanism between semantically heterogeneous databases. High-level queries to the OMS are conversely translated between database queries and LinKBase® queries. Effectively, the results of these queries may then be processed and presented in such a way that all results are presented to the user in a homogenous manner, without the users awareness that several sources are being interrogated at once (Deray and Verheyden 2003).

Unfortunately, there are many obstacles that make fully automated mapping difficult. In a large ontology, a specific name may map to several properly distinct concepts. In a small ontology, a specific name may not map to any of the concepts. The final decision of

mapping any specific database item to an ontological entity remains the user's responsibility, dictated by conceptual difficulty and demand. It is at this stage that philosophical scrutiny proves its worth. Issues of knowledge representation cannot always be reduced to technical procedures and developing a well-formed ontology reveals itself to be equally a matter of craft as science. In what follows we explain a few cases where philosophical scrutiny has brought added perspicuity to our treatment of databases in our integration attempt.

### 5.1 Cross-Mapping SNOMED-RT® to LinKBase®

The philosophically inspired LinKBase® incorporates not only the notion of “part”, but also “proper part,” a distinction long utilized by philosophers in the field of mereology. Such a distinction allows us to build an accurate representation in which both conceptions of “amputation of foot” discussed earlier are recognized as distinct and their relation to each other can be mapped. The distinction rests on the philosophical notion of parthood, whereby every entity is a *part* of itself, but no entity is a *proper part* of itself (Smith 1996). Essentially, “proper part” corresponds to the set of all those things we usually associate with parthood, whereas “part” corresponds to the set of all proper parts *plus the entity itself included*. In other words, where the toe is considered a “part” and a “proper part” of the foot, the foot is considered simply a “part” of itself. In LinKBase®, this notion of parthood is captured by the concept “structure,” both the toe and the foot itself are subsumed by the concept “foot structures.”

This configuration is then mapped to the SNOMED-RT® ontology, where “foot structure” (any part of the foot including the foot itself) is related to “amputation of foot structure”, which subsumes two further concepts “complete amputation of foot” and “partial amputation of foot”. Here the SNOMED-RT® concept “amputation of foot” is linked via the IS-A relation to the former and the concept “amputation of toe” is linked to the latter. Both are linked via IS-A relations to “amputation of foot structure.”

In this way we maintain a hierarchical structure that subsumes both the toe and the foot without reducing either one to the other and thus allowing each to be related to different, and possibly incommensurable concepts without the problematic inconsistencies that may be derived through inherited criteria as we saw above.

### 5.2 Mapping GO to LinKBase®

During the conceptual analysis phase, we carefully investigated the top-layer concepts of the three GO sub-domains that act as our gateway between the

LinKBase® concepts and GO terms. We identified the more general concepts of GO in LinKBase® and created new concepts if they were not adequately recognized. In this way we are able to map GO's molecular function hierarchy to both of the other hierarchies via the BFO inspired axioms at the top-level concepts.

If we return the EPO\_HUMAN protein example from earlier, we see now that LinKBase® is able to incorporate this example and model the relations with a greater degree of clarity, essentially mirroring the BFO defined structure. The connection between a GO protein and its activity in LinKBase® is captured by a “has-function” relation, and the connection between an activity and its corresponding processes is captured by what we call the relation of “realization.” The first reflects the relation between a substance and its function, and the later, that between a function and its actualization without resorting to the whole/part relation properly left exclusive to each hierarchy.

In this manner not only is GO consistently mapped to LinKBase®, but the expressiveness of GO itself has been expanded without any major alterations required in its core structure (Vershelde et al., forthcoming).

## 6. Conclusions

It is for us no surprise that after having reviewed 35 systems for their ontology mapping capacities, Kalfoglou and Schorlemmer conclude: “[O]ntology mapping nowadays faces some of the same challenges we were facing ten years ago when the ontology field was in its infancy. We still do not understand completely the issues involved.” (Kalfoglou and Schorlemmer 2003). This is because many researchers in the field forget that ontology is not a discipline that stood in its infancy just ten years ago, but rather, almost 2400 years ago, since the seminal works of Aristotle. For millennia, when we have encountered difficulties understanding reality, we have turned to philosophers for solutions. Why should we not do likewise today? The return to a realist philosophical foundation means a return to those foundations that reflect over two millennia of ontological research, but this in no way requires that we abandon our pragmatic perspective. In his *Physics*, Aristotle writes, “When the objects of an inquiry, in any department, have principles, conditions, or elements, it is through acquaintance with these that knowledge, that is to say *scientific knowledge*, is attained,” and we would do well to keep such words in mind today when we seek to design an adequate inventory of ontological elements for database integration and navigation.

In general, the difficulties that we are presented with in this project are not those of isolated instances,

but rather, they illustrate a general pattern, present in one form or another in near all existing application ontologies. The 'ad hoc' character of many biomedical ontologies, the main cause of the so-called "Tower of Babel" problem of interoperability, is not without a history. These features have developed because ontology engineers were forced during the initial stages, moving from printed dictionaries and nomenclatures to digital systems, to make a series of uninformed decisions about complex ontological issues, indeed, the very same issues that philosophers have been pondering for millennia. To date, the importance of philosophical scrutiny has been obscured by the temptation to seek immediate solutions to apparently localized problems. In this way, the forest has been lost for the trees, and the larger problems of integration have thus appeared unsolvable.

The BFO-driven restructuring of LinKBase® is still in its infancy, yet we already have examples demonstrating increased adaptability through the application of philosophical knowledge and techniques. We have discussed here some examples in which changes were made leading to an enhanced internal consistency, allowing the level of access necessary for a general database translation hub. If early successes like those presented here are any indicator, we have great reason to expect that the thoroughgoing integration of BFO and LinKBase®, of which the above results are merely preliminary groundwork, will greatly enhance the capacity of LinKBase® to effect direct integration between foreign databases such as SNOMED-RT® and GO, raising ontology-based information management up to its true potential.

### Acknowledgements

We are grateful to the work and helpful comments provided by Mariana Dos Santos and Jean-Luc Verschelde. This work has been supported by the Language and Computing Research Division and the Wolfgang Paul Program of the Alexander von Humboldt Foundation.

### References

Bittner, T, Smith, B. A Theory of Granular Partitions. *Foundations of Geographic Information Science*, Duckham, M, Goodchild, M, and Worboys, M, eds., London: Taylor & Francis Books, 2003: 117-151.

Deray T. & Verheyden P., Towards a semantic integration of medical relational databases by using ontologies: a case study. R. Meersman, Z. Tari et al., (eds.), in *On the Move to Meaningful Internet Systems 2003: OTM 2003 Workshops*, LNCS 2889, page 137 - 150, 2003. Springer Verlag.

Flett A, Dos Santos M, Ceuster W. Some Ontology Engineering Processes and their Supporting Technologies. Siguença, Spain, October 2002. EKAW 2002.

Grenon P, Smith B. SNAP and SPAN: Towards dynamic special ontology. Forthcoming.

Kalfoglou, Y, Schorlemmer, M. Ontology Mapping: The State of the Art. *The Knowledge Engineering Review* 18(1), 2003.

Monteyne F, Fanagan J. Formal Ontology: The Foundation for Natural Language Processing. January 2003.

<http://www.landcglobal.com>

Smith, B, Rosse, C. The Role of Foundational Relations in the Alignment of Biomedical Ontologies. Forthcoming.

Smith B, Williams J, Schulze-Kremer S. The Ontology of the Gene Ontology. *Proceedings of the AMIA Symposium 2003*, forthcoming.

Smith B. Mereotopology: a theory of parts and boundaries. *Data & Knowledge Engineering* 1996; 20: 287-303.

Verschelde J.L., Dos Santos M, Deray T, Smith B, Ceusters W. Ontology-assisted Database Integration to Support Natural Language Processing and Biomedical Data-mining. *Journal of Integrative Bioinformatics*, forthcoming.