

# Referent Tracking for Corporate Memories

Werner CEUSTERS and Barry SMITH

Ontology Research Group  
New York State Center of Excellence in Bioinformatics and Life Sciences  
701 Ellicott Street  
Buffalo NY, 14203  
USA  
phone: +1 716.881.8971  
email: ceusters@buffalo.edu

Department of Philosophy  
University at Buffalo  
135 Park Hall  
Buffalo, NY 14260  
phone: +1 716.645.2444  
email: phismith@buffalo.edu

## Abstract

For corporate memory and enterprise ontology systems to be maximally useful, they must be freed from certain barriers placed around them by traditional knowledge management paradigms. This means, above all, that they must mirror more faithfully those portions of reality which are salient to the workings of the enterprise, including the changes that occur with the passage of time. The purpose of this chapter is to demonstrate how theories based on philosophical realism can contribute to this objective. We discuss how realism-based ontologies (capturing what is generic) combined with referent tracking (capturing what is specific) can play a key role in building the robust and useful corporate memories of the future.

## INTRODUCTION

Corporate memories (CM) are information systems designed to keep track of the history and evolution of an enterprise with the goal of using lessons learned from past experiences to enhance the performance of the business transactions in the

future. Well designed CMs should contain data about both the enterprise and the environment in which it operates. The former, traditionally embodied in what is referred to as an *enterprise model*, consists of data about the organisational structure and operating procedures of the enterprise, its mission and strategic objectives, its staff, their skills and competences, the products and services the company is able to deliver, and, most importantly, data about projects or business transactions brought to a successful (or unsuccessful) end. The latter, the CM's *environment model*, includes data about prospects and clients, competitors and partners, applicable laws and regulations, and techniques and methodologies proposed by outsiders to complement the results of research carried out within the company itself.

For understandable reasons, CM technology is standardly approached from a backward-looking perspective, employing passive knowledge management techniques with the prime goal of making legacy electronic documents more easily accessible. To this end, such documents are manually or semi-automatically annotated with tags that reformulate words or relevant phrases in a document in a more structured and standardised manner (e.g. occurrences of the words *car*, *van*, *bus*, etc. are all tagged with the compound *motor vehicle*), or with meta-tags that add additional context to phrases or paragraphs (e.g. *important*, *motivation*, *marketing*, *outsourced operations* etc.). When these meta-tags are organised in a structure that reflects more or less the way the enterprise itself is structured, they form what is referred to as 'enterprise ontologies'.

CM applications can also, however, be used for the development of more proactive, forward-looking systems, in which data that reflect changes in either the organisation or its environment are able to trigger warnings indicating business opportunities for the enterprise or imminent hazards to its proper functioning. To achieve these goals, however, CM applications must be freed from certain barriers placed around them by traditional knowledge management paradigms. This means, above all, that they must be required to mirror more faithfully those portions of reality which are salient to the workings of the enterprise, including changes that occur with the passage of time. It is especially in the domain of healthcare that work on such proactive technologies is most advanced. The purpose of this chapter is to demonstrate how the proposals to create proactive systems based on electronic healthcare record systems can be generalized in such a way as to achieve analogous objectives in the area of enterprise ontologies and corporate memories.

## BACKGROUND

### Corporate Memories

The word ‘corporate memory’, including its quasi-synonym ‘organisational memory’, is interchangeably used to denote distinct though related entities. Originally, the term referred to a specific type of ‘collective memory’ found in organisations and groups, primarily commercial enterprises, and which, according to social and behavioural scientists descending from Durkheim, is something supra-individual which cannot be reduced to the memories in the minds of single individuals (Wexler, 2002). Collective memory so conceived typically comprehends various kinds of information about (1) external contacts, (2) internal know-how, (3) the types of authority and influence exerted not only by company owners but also employee associations, (4) the behaviour of customers, (5) operational rule sets and routines, and (6) implementation strategies for company operations that determine how the information about all of these things should interact with the company’s primary business (Beckett, 2000). With the advance of computer science, corporate memories became conceived as computer systems which embody a company’s entire stock of knowledge assets, including accumulated know-how (skills), and make the latter available to enhance the efficiency and effectiveness of knowledge-intensive work processes (Kühn & Abecker, 1997).

How to build corporate memory systems is a research topic in its own right, since any such system has to be able to communicate with the majority of computer systems already installed in the company and to re-use the information they contain. Since this involves issues of semantic interoperability, it is no surprise that ontologies have become essential components of corporate memory systems, contributing to a wide variety of tasks. Most prominent, however, are the ontologies that describe organisational aspects of the enterprise, and are therefore called *enterprise ontologies*. This includes ontologies that are designed to deliver background knowledge in applications for electronic business interactions (Haller & Oren, 2006).

### Enterprise Ontologies

Where corporate memories capture primarily what is *specific* for an enterprise, such as information about its employees, projects, business rules, contracts, and so forth, enterprise ontologies capture primarily what is *generic*. The first ontologies of this sort were developed in the course of The Enterprise Project in the United Kingdom (Stader, 1996) and the TOVE project in Toronto, Canada (Fox, 1992).

One of the outcomes of the Enterprise project was *The Enterprise Ontology*, which is described by its authors as a collection of terms and definitions relevant

to business enterprises (Uschold, King, Moralee, & Zorgios, 1998). Approximately 90 terms are defined, grouped in 5 clusters labelled ‘activity’, ‘organisation’, ‘strategy’, ‘marketing’ and ‘time’.

Of more recent date is the *REA Enterprise Ontology* (Geerts & McCarthy, 2002), whose acronym derives from the primary components of the framework’s original domain: (economic) Resources, Events, and Agents, and which is based on the REA accounting model (McCarthy, 1982).

The purpose of the *multi metamodel process ontology* (m3po) is to incorporate and unify the different currently existing workflow metamodels and reference models. They are designed to provide the representational resources for extracting what are called ‘choreographies’ from internal business processes (Haller & Oren, 2006) in such a way as to capture the various sorts of relationships that obtain between participants in business interactions.

### **Why Are Such Systems Not In Use?**

Despite the massive interest in and research activities directed towards both corporate memories and enterprise ontologies, reports on success stories are limited to unverifiable marketing claims or mere speculations. This is for instance witnessed by (Rosenthal, Manola, & Seligman, 2001)’s statement to the effect that ‘*Many initiatives, governmental and commercial, have pursued the grand vision of “transparent access” – making all data available to all consumers (users and applications), in a way the consumer can interpret, anywhere and at any time. Among large-scale enterprises, success stories in achieving such visions seem rare or nonexistent*’ (p. 1), or by papers such as (Hill, 2006), of which the title: ‘*Service Taxonomy and Service Ontologies Deliver Success to Enterprise SOA*’ does indeed imply the existence of actual success even though no evidence is provided in the actual paper.

As pointed out in the literature, there are several reasons for this. In (Partridge & Stefanova, 2001), for instance, it is argued that neither the TOVE nor the Enterprise Ontology meets the criteria of clear characterisation and domain coverage, and that the problem cannot be compensated for by merging them because they do not share a common view of what an organisation is. This provided the motivation to develop a new ontology: the Core Enterprise Ontology (Bertolazzi, Krusich, & Missikoff, 2001), but when this system was analysed by other scholars, then it too was found not to meet certain crucial requirements, which again led to the creation of a new artifact (Osterwalder, Lagha, & Pigneur, 2002). And so on, *ad indefinitum*.

Our research on the (generally low) quality of ontologies has demonstrated that the main reason for failure of ontology projects is the adoption of a methodology rooted in traditional expert-systems-based approaches to knowledge representation and therefore centered around the representation of ‘concepts’ or

'conceptualisations' (Smith, Ceusters, & Temmerman, 2005). The problem with this approach is that, by focusing on the semi-idealized concepts (ideas, meanings, knowledge) in the minds of divergent groups of semi-idealized experts, it does not take into account the concrete reality by which such putative experts are engaged in their day-to-day activities. This is because concepts in the minds of experts are always in one way or another simplifications of the reality to which they are intended to correspond. Representations of concepts in computer systems add a further level of simplification (and thus a further removal from reality) by imposing the restrictions of expressivity needed to guarantee computational tractability of the systems which result. Indeed, when knowledge engineers and information analysts proceed by first defining 'concepts' and 'relationships' and only then connecting these to bodies of data deriving in turn from some area of concrete reality, then they have things precisely the wrong way round. What they should be doing is finding a way to allow the concrete real-world entities to which given systems relate, and about which large amounts of data are typically already on hand, to determine the analysis from the very start and to serve as anchor for this analysis and for the workings of the system in every stage thereafter. Viewing reality always in terms of semi-idealized conceptual surrogates has given rise to several so-called 'ontologies' in which these surrogates themselves, rather than reality, have become the objects of study, so that the quality of one ontology is gauged by the degree to which it conforms to a second ontology (Goossenaerts & Pelletier, 2003). Focusing on reality directly, in contrast, can provide an independent benchmark for the correctness of ontologies, and thus allow systematic measures of quality resting on investigation of the ways in which changes introduced in successive versions of an ontology relate to changes in the reality towards which it is directed (Ceusters & Smith, 2006a). Such measures are indispensable if we are to initiate an evolutionary path towards improvement in ontologies of the sort that we have in other empirical domains.

The predominant focus on conceptualisations which deviate in substantial ways from the structures found in reality applies also to ontologies developed in the context of enterprise engineering. (Huhns & Stephens, 2002), for instance, describes a methodology under which a multiplicity of ontology fragments, encapsulating the semantics employed by several independent parties, can be related together automatically without the use of any single global ontology. Inspection of the examples provided, however, reveals that the resulting unifications contain many erroneous associations. Interestingly, Huhns and Stephens do not consider this to be problematic. Indeed they assert that a *'consensus ontology is perhaps the most useful for information retrieval by humans, because it represents the way most people view the world and its information. For example, if most people wrongly believe that crocodiles are a kind of mammal, then most people would find it easier to locate information about*

*crocodiles if it were located in a mammals grouping, rather than where it factually belonged* (Huhns & Stephens, 2002, p 89).

If ontologies are ever to become useful in mission-critical domains like business or medicine, however, then they must be built on the basis of an approach which maximises the degree to which entities *are* located where they factually belong – and this means an approach that is resolutely grounded in reality. Ontologies which are intended to be used more specifically in the context of enterprise engineering and corporate memory systems must be able to reflect not only how our perceptions and beliefs about reality change in the course of time but also how reality itself changes, and to reflect how the former are related to the latter. If, for the purposes of a given ontology application, it is judged relevant that many people believe that crocodiles are mammals, then this fact should indeed be represented; but it should be represented as a *false belief*, rather than being incorporated into an ontology as a fact on a par with all others.

In the following sections, we describe how to achieve these ends in such a way as to achieve a level of sophistication in ontology development that is able to draw a clear distinction between reality and the conceptualisations thereof on the part of managers, employers, and customers.

## ONTOLOGIES AND FAITHFULNESS TO REALITY

### **Basic Formal Ontology**

The core of our proposal is Basic Formal Ontology (BFO), a framework that is designed to serve as basis for the creation of high-quality shared ontologies especially in the domain of natural science. BFO embraces a methodology which is realist, fallibilist, perspectivalist, and adequatist (Grenon, Smith, & Goldberg, 2004). It holds, in other words, (1) that reality and its constituents exist independently of our (linguistic, conceptual, theoretical, cultural) representations thereof; (2) that our theories and classifications can be subject to revision; (3) that there exists a plurality of alternative, equally legitimate perspectives on reality, and (4) that these alternative views are not reducible to any single basic view.

BFO subdivides reality according to a number of basic dichotomies. **First**, it distinguishes *particulars* from *universals*; the former are entities such as Microsoft Corporation or the specific contract #17896 Microsoft signed with the University of Ohio in 1999; the latter are entities, such as *company* and *contract*, which have the former as their instances. Both universals and instances are restricted to what exists (or existed) in reality, and are thus different from classes and instances as referred to in ontologies adhering to a concept-based view (Smith, 2004). On the concept-based view, “employees of Microsoft Inc.” would be perceived as designating a concept or defined class; according to BFO this phrase, as used at some specific time, designates a particular, namely the specific

collection of persons who are employees of Microsoft Corporation at that time. Whereas under the concept-based view any specific Microsoft employee would be an *instance* of some putative corresponding class or concept, he or she would be a *member* of the collection under BFO.

**Second**, BFO distinguishes, within the realm of particulars, between *continuants* and *occurrents*. Continuants are those entities, such as Microsoft and its current CEO, that endure continuously through a period of time while undergoing changes of various sorts. Occurrents *are* such changes; they are entities which unfold in time through their successive temporal parts or phases – thus they are the entities otherwise called ‘processes,’ ‘actions,’ ‘events’.

The difference between occurrents and continuants is crucial, and any ontology neglecting this distinction is not capable of dealing with changes over time in an appropriate way. While, for instance, a continuant particular may become an instance of distinct universals over time (Bill Gates was once an instance of *child*, later an instance of *adult*; his societal role was once an instance of *student*, later of *CEO*), occurrents cannot undergo such changes because occurrents *are* changes.

**Third**, there is the distinction between *dependent* and *independent* entities, where each dependent entity is defined as being such that it cannot exist without some independent entity which is its bearer. A contract, for example, cannot exist without contracting organisations or persons, and the process of signing a contract cannot exist without some person who signs. Persons themselves, in contrast, are independent: as soon as they exist; they do not depend on the existence of something else in the given sense, although, of course, their coming into existence did depend on other independent entities, for example their parents. The utility of introducing this distinction into an ontology becomes obvious when the ontology is used to annotate data in a repository: when a particular is annotated as being an instance of a dependent entity, then there must be other particulars, perhaps yet unknown to the person who performs this annotation, on which that entity depends. In cases of this sort, the ontology becomes a valuable resource for formalising business rules and database integrity constraints (Hay & Healy, 2000).

**Fourth**, there is the distinction between *fiat* and *bona fide* entities, which is based on the opposition between bona fide (or physical) and fiat boundaries, the latter being exemplified especially by boundaries – such as the boundary of Utah, or of the 20th century – introduced via human demarcation (Smith & Varzi, 1997). Fiat boundaries are overwhelmingly present in the realm of social entities, where they delineate for example markets, market segments, marketing regions, and serve in establishing what is an employee, a minor, a family member for purposes of health insurance coverage, and so forth.

BFO also distinguishes three major families of relations between the entities just sketched: (1) <p, p>-relations, obtaining between particular and particular (for example: Steve Ballmer being the CEO of Microsoft); (2) <p, u>-relations, obtaining between particular and universal (for example: Steve Ballmer being an instance of the universal *person*); and (3) <u, u>-relations, obtaining between universal and universal (for example: *software company* being a subkind of *company*) (Smith, Ceusters, Klagges et al., 2005). The importance of this distinction is exemplified by the fact that relationships such as parthood have distinct properties at the particular and at the universal levels, and that ignoring these distinctions has led to a number of erroneous representations of relations (Donnelly, Bittner, & Rosse, 2006). These distinctions can be handled also in regular concept-based ontologies, but they have thus far characteristically been ignored – not least because concept-based ontologies very often reflect an unsure understanding of the distinction between an instance and a universal.

### **Granular Partition Theory**

The second element of our proposal is Granular Partition Theory, a highly general framework for understanding the ways in which, when cataloguing, classifying, mapping or inventorising a certain *portion of reality* (POR), human beings and other cognitive agents divide up or partition this reality at one or more levels of granularity (Bittner & Smith, 2003). The resultant partitions are composed of *partition units* (analogous to the cells in a grid) and the theory provides a formal account of the different ways in which such units can correspond, or fail to correspond, to the entities in reality towards which they are directed. The theory takes account for example of the degree to which a partition represents the mereological structure of the domain onto which it is projected, and also of the degree of completeness with which a partition represents this domain.

Drawing on this framework, we have proposed a calculus for use in quality assurance of complex representations created for clinical or research purposes in the context of both ontology evolution (Ceusters & Smith, 2006a) and ontology mapping (Ceusters, 2006). The calculus is based on a distinction between three levels (Smith, Kusnierczyk, Schober, & Ceusters, 2006):

1. the level of reality (for example on the side of a specific enterprise, its employees, managers, etc.);
2. the cognitive representations of this reality (for example as embodied in observations and interpretations on the part of sales personnel or business analysts);
3. the publicly accessible concretisations of these representations in artefacts of various sorts, of which ontologies and corporate memories are specific examples.



The representations on levels 2 and 3 are composed in hierarchical fashion out of modular sub-representations built ultimately out of smallest modules called *representational units*, whereby:

1. each module is assumed to be veridical, i.e. to conform to some relevant POR on the basis of our best current understanding (which may, of course, be based on errors);
2. distinct modules may correspond to the same POR by presenting different though still veridical views or perspectives of this reality, for instance one and the same event may be described both as an event of buying and as an event of selling;
3. what is to be represented by the modules in a representation depends on the purposes which that representation is designed to serve.

Relevant portions of reality can include not only physical things (buildings, physical goods) but also mental acts and states (acts of valuation on the part of stockholders, states of willingness of potential customers to buy a certain good) and entities of many other types, including institutions, social roles, social relations of authority or ownership, and so forth.

### **The Referent Tracking Paradigm**

In ontologies and terminologies the representational units are terms from some natural or formal language and are assumed to refer to universals or defined classes (Smith et al., 2006); in corporate memories the representational units must refer also in robust and unambiguous fashion to enterprise-specific entities at the level of instances.

Referent tracking (RT) is a new approach to the handling of data about real world entities introduced in (Ceusters & Smith, 2006b). It allows instances in reality to serve as benchmark for the correctness of the ontologies used to describe them. The RT paradigm has been developed thus far to support the entry and retrieval of data in the Electronic Health Record (EHR), where its purpose is to avoid the problems which arise when statements in an EHR refer to disorders, lesions and other entities on the side of the patient by means of logically complex descriptive phrases such as ‘the fracture in the leg of patient X’ or ‘the tumour in the lung of patient Y’. These problems arise because the phrases in question employ generic terms in ways which may fail to identify the relevant instances unambiguously. (John may have multiple fractures in his leg; or he may have fractured his leg twice at different times in his life.) In (Parsons & Wand, 2000) it is argued that problems in schema integration, schema evolution, and interoperability of databases are precisely the consequence of ambiguities of this sort, which are deeply rooted in the erroneous assumption adhered to in many database design circles according to which entities can be referred to only as instances of pre-specified classes. They make the case that this *assumption of*

*inherent classification* violates philosophical and cognitive guidelines on classification.

Referent tracking avoids such ambiguities by introducing unique identifiers, called IUIs – Instance Unique Identifiers – for each numerically distinct entity that exists in reality and that is referred to in statements in a record. Currently the items uniquely identified for EHR purposes are restricted to entities such as patients, care providers, buildings, machines and so forth. The referent tracking paradigm expands this list beyond the current range to include also fractures, polyps, seizures and a vast variety of other clinically salient real-world instances in all the categories distinguished by the BFO ontology.

In the context of corporate memories, analogously, IUIs would be assigned not merely to the various organizations and persons relevant to the enterprise (companies, employees, customers, and so forth) but also to contracts, applicable laws, meetings, all sorts of business transactions, accidents in manufacturing facilities, deliveries, and so forth. It would include also various types of failures, absences, and other putative negative entities, although these call for special treatment (Ceusters, Elkin, & Smith, 2006). For many entities unique identifiers will exist already in the various information systems of a large corporation. Our proposal is that these identifiers should be consolidated into a single corporate memory store, where they will constitute an evolving dynamic map of the corporation and of all events and processes with which the corporation is involved.

The following requirements have to be addressed if the paradigm of referent tracking is to be given concrete form in a Referent Tracking System (RTS) able to serve the needs of an enterprise: (1) a mechanism for generating IUIs that are guaranteed to be unique strings; (2) a procedure for deciding which particulars should receive IUIs; (3) protocols for determining whether or not a particular has already been assigned a IUI (each particular should receive maximally one IUI); (4) rules governing the processing of IUIs in information systems, including rules concerning the syntax and semantics of statements containing IUIs; (5) methods for determining the truth values of propositions that are expressed through descriptions in which IUIs are employed; (6) methods for correcting errors in the assignment of IUIs, and for investigating the results of assigning alternative IUIs to problematic cases; and (7) methods for taking account of changes in the reality to which IUIs get assigned, for example when particulars change their qualities or when they merge or split.

With respect to (1), IUIs are to be assigned to particulars directly, and thus independently of the universals of which they are instances and of any ontology describing such universals. A strategy consisting of assigning unique IDs to representational units within each ontology, and then adding prefixes to these IDs to denote the particulars which instantiate them, would not work because

particulars can be instances of universals denoted in several ontologies. Moreover, particulars may change over time and so instantiate different universals, or the classification status may change or a given particular may change as errors in a data resource are corrected. The goal of referent tracking is, we recall, to provide a means by which instances in reality can serve as benchmark for the correctness of ontologies. If ontologies themselves are used to generate the referent tracking IDs, then this goal will be defeated from the start.

An RTS can be set up in isolation, for instance within a single department of a large company. Clearly, however, the referent tracking paradigm will serve its purpose most optimally when used in a distributed, collaborative environment such as a large company with several offices dispersed over a wide area. One and the same customer is often served by a variety of departments within a single enterprise, many of them working in different settings, and each of these settings may use its own separate information system. These systems contain different data, but these data often provide information about the same particulars. Under the current state of affairs, it is very hard, if not impossible, to query these data in such a way that, for a given particular, all information available can be retrieved. With the right sort of distributed RTS, such retrieval becomes in very many cases a trivial matter and this even on a meta-company level. It could for instance give considerable added value to services of the kind delivered by a business information service company such as Factiva, which uses a four step automated and manual process to ensure that everything falling under the coverage of its 12,000 information sources is correctly categorized. Customers can receive the data either as an XML feed or a Web service for integration into their corporate intranets, or their CRM or competitive-intelligence systems (Drew, 2006).

### **Services of a Referent Tracking System**

An RTS should offer at least three services: (1) generation of unique identifiers to be used as IUIs, (2) management of the IUIs generated, and (3) provision of access to the IUIs stored.

As to (1), several schemes for generating strings that are guaranteed to be unique are already in use. If RTS services would be offered by a player external to a specific organisation, it might be beneficial that this player not only registers IUIs but also certifies the uniqueness of the strings to be used within a given IUI-repository and guarantees that the assignments claimed to have been made by given authors were indeed made by those authors. Persons assigning IUIs, who will typically play a variety of other roles within the enterprise, will themselves be identified by IUIs, which will enable them to be identified automatically in these several roles and enable also cross-links between the corresponding different groups of entities (including other persons) with which they have to deal.

Service (2) involves what we shall refer to as the *IUI-repository*, whose purpose is to keep track of the identifiers assigned to already existing entities, or reserved for entities that are expected to come into existence in the future. It will do this in such a way that (i) each IUI represents exactly one particular, and (ii) no particular is referred to by more than one IUI. These two requirements are not easy to fulfil, since both depend on the ability and willingness of users to provide accurate information. This, however, introduces no problems different in principle from those already faced by the users of existing systems when called upon to provide information of a non-trivial and occasionally sensitive sort about individuals.

Service (3), here called the referent tracking database (RTDB), should provide access to the information entered into a given corporate memory about the particulars referred to in the IUI-repository. Where the IUI repository is an inventory of concrete entities that have been acknowledged to exist, and, consequently, of what IDs to use if one wants to refer to them, the RTDB is an inventory of descriptions concerning the features of and interrelations between these entities and of the ways in which they change in the course of time. The RTDB, too, does not need to be set up as a single central database but can rely on any paradigm for distributed storage.

The role of the RTDB is to keep track not only of the features and interrelations of given particulars as they change through time but also of the assertions that have been made about such particulars, including those assertions that have been shown to be false (stored, for example, for the purposes of providing an audit trail). The RTDB also helps users to determine whether a particular they encounter for the first time has been registered already in the IUI-repository or whether a new IUI must be created for use in new descriptions. To be sure, this places some additional burden on the person who has to enter the information; but, given that cases such as this are likely to be of high salience, the time perceived as being lost at this stage will likely be recovered when searching for information thereafter.

## APPLYING BFO AND REFERENT TRACKING TO CORPORATE MEMORIES

For the remainder of this paper, we will provide examples of how the theories and paradigms described above can be used to detect and solve a number of problems and inconsistencies that we (and others) encountered in studying the literature on enterprise engineering and corporate memories.

Quite common is the inclusion of representational units in an ontology that do not have a counterpart in reality. This happens at the level both of relationships and of the entities which serve as their relata. Consider the difference between the “Sale” and “Have-Capability” relationships as defined in the Enterprise Ontology

(Uschold et al., 1998). A ‘Sale’ is (acceptably) defined as ‘*a relationship constituting an agreement between two Legal Entities to exchange a Product for a Sale Price*’, in keeping with the Enterprise Ontology’s treatment of relationships as entities in their own right that can thus be instantiated. Two instances of Legal Entity thus enter into a single instance of the Sale relationship. The “Have-Capability” relationship, on the other hand, is defined as ‘*a relationship between a Person and an Activity denoting that the Person is able to perform the Activity*’. The first problem here is the confusion of use and mention: relationships themselves do not ‘denote’; this is the task of the corresponding denoting expressions. But more importantly: being able to engage in an activity does not require that any instance of such an activity exists. Under BFO, properties of this sort would be represented correspondingly as falling within the realm of realizable entities (such as powers, functions, dispositions, orders, plans, algorithms, recipes), in order to do justice to the fact that the existence of a capability does not imply the existence of any realization of this capability (IFOMIS, 2006).

The use-mention confusion – which is common not only among enterprise ontology developers – confuses the level of reality with the level of our representations thereof. Many data dictionaries suffer from this confusion. The ACORD Data Dictionary for Global Insurance Industry, for example, which is used to assist in automating business interactions between insurers and clients (ACORD, 2005), defines a building as ‘*a construction that normally has a roof and walls*’. ‘Air conditioning’, however, it defines as ‘*information necessary to describe a given type of air conditioning in a building.*’ Consistency in providing definitions would dictate that ‘entity’ is used in such a way that it refers always either to *information about* something in reality, or to *that something in reality* itself. ACORD, however, provides a problematic mishmash, in which *buildings*, for example, would contain *information about air conditioning* as parts.

The same confusion is found in (Goossenaerts & Pelletier, 2003): the latter correctly argues that the Enterprise and TOVE ontologies do not emphasize the distinction between things and their changes on the one hand and conceptual entities on the other, drawing their analysis from the work of Bunge (Bunge, 1977) and specifically from its application in the Bunge-Wand-Weber model in the domain of information systems (Wand, Storey, & Weber, 1999). This analysis led them to develop the PSIM Ontology (for: *Participative Simulation environment for Integral Manufacturing renewal*), which was inspired also by earlier work conducted in the European Research Project CIMOSA (AMICE-Consortium, 1989) and from Peircean Semiotics (Hoopes, 1991). The result, however, is not without its own dramatic mysteries and misinterpretations. Thus we read that the PSIM Ontology distinguishes three main categories: Activity, Object and Information (element), whereby an ‘Information (element)’ is defined

as: ‘*a characteristic of either an object or activity or information, which is used to constrain directly or indirectly the involvement of an object in an activity*’ (Goossenaerts & Pelletier, 2003, p. 45). PSIM then classifies as information elements not only ‘*the time needed to perform an activity*’ and ‘*how an activity has to be performed*’, but also ‘*how the enterprise is organised*’, ‘*the way the responsibilities are distributed among the enterprise*’, and even ‘*the weight of a piece of material*’. Weight, for BFO, is a dependent continuant that depends on the material object of which it is the weight, and this independently of whether or not a cognitive being has any sort of information about the matter. Confusions of this sort are a direct result of the concept orientation in ontology.

This concept orientation leads quite often also to a blurring of the distinction between instantiation and subtyping. Where in BFO *instantiation* is a relationship between a particular and a universal, *subtyping* is a quite different relationship holding between one universal and another. Nothing which is an instance can itself have instances, while something that is a subtype, can itself have other subtypes. As is correctly recognised in (Uschold et al., 1998), the distinction between a type of entity, and a particular entity of a certain type, i.e. an instance, is not consistently made when using natural language. This does not, however, mean that it is acceptable that the authors of the Enterprise Ontology ‘*intentionally blurred this distinction*’ in the informal description of their ontology (Uschold et al., 1998, p 35). And when the methodological work underlying the Core Enterprise Ontology allows John Doe to be an instance of “consumer”, and “consumer” to be an instance of “entity” (Bertolazzi et al., 2001), the result is a mistake that is impermissible in any serious ontology work.

Note that it is not just natural language that blurs the mentioned distinction: traditional database design paradigms exhibit the same type of confusion, as do some ontology authoring environments such as Cyc (Foxvog, 2005): a table about cars may contain ‘instances’ such as ‘Volkswagen’ or ‘Audi’. Under a realist paradigm, such a representation can only be the result of a sloppy analysis in which a *car brand* is mistaken for a *car*.

Even more unfortunate are the views adhered to by (Noy & McGuinness, 2001) who claim that ‘*individual instances are the most specific concepts in an ontology*’ (p. 18), or that ‘*deciding whether a particular concept is a class in an ontology or an individual instance depends on what the potential applications of the ontology are*’ (p. 18). As an example, for an expert system intended to give advice on what types of wine pair best with certain types of food, it may not matter whether a specific brand of Elsasser Riesling is represented in the system by means of a class or an instance. But if the latter option is chosen, and this system needs to be used in interactions with restaurants or wine merchants who would like to link their inventory to that expert system, then it will lead to problems if what is a class for the former, is an instance for the latter.

The rigorous identification schemes proposed by the Referent Tracking paradigm are an important first step in doing away with such confusions, and they have been applied in this capacity for example in solving problems related to digital rights management (Ceusters & Smith, 2007). That they can help, too, in the specific case of enterprise engineering is witnessed by a recent case study exploring the possible complementarity of the Demo Engineering Methodology for Organizations (DEMO) and the Object Role Modelling (ORM) paradigm (Dietz & Halpin, 2004). DEMO enables the business processes of organizations to be modelled independently of how these processes are implemented, thereby focusing on the communication acts that take place between human actors in the organization. ORM enables business information to be modelled in terms of *fact types* as well as the business rules that constrain how the fact types may be populated for any given state of the information system and how derived facts may be inferred from other facts. One important feature of ORM is its requirement for the inclusion of at least one identification scheme for each entity type, which functions as an identity criterion for instances of that type. Because of this requirement, data use cases, i.e. samples of information, can be used to seed an initial model. However, if ORM is to be used for the purposes of building an ontology rather than a database schema, then developers should pay attention to the fact that several records in a database may refer to the same entity in reality. This is certainly the case for example when databases maintained in originally distinct organisations are merged because of a company takeover.

## CONCLUSION

For a company to anticipate and manage change for the future, to design appropriate strategies that will create business value for customers, and to improve profitability in current and new markets, its activities must be based on a synoptic view of its present business environment as a complex dynamic whole comprehending the activities, resources, markets, customers, products, services, regulations and costs associated with the enterprise. Such an overview, the key to strategic intelligence, is cultivated for example through the methods used to improve the capabilities of the company's managers and workers to learn about changes in the business or industry environment that are summarized in (Marchand, Davenport, & Dixon, 2000). Corporate memories are crucial to building and sustaining such strategic intelligence, and we believe that ontologies *combined with referent tracking* can play a key role in building the robust and useful corporate memories of the future. Ontology is in essence a philosophical discipline that seeks to capture high-grade terminological knowledge that can provide a sound basis for data schemas and data dictionaries such as are employed by large organizations. The development of ontologies as artifacts for use in computer systems has, unfortunately, been too often conducted in a way that

ignores reality. The referent tracking paradigm, by bringing reality back into business, can solve this problem and thereby save businesses from the ‘conceptual models’ of their IT personnel.

## REFERENCES

- ACORD. (2005). Data Dictionary for Global Insurance Industry (Publication. Retrieved January 10, 2007: <http://www.acord.org/dataDictionary/dataDictionary.htm>)
- AMICE-Consortium. (1989). *Open System Architecture for CIM, Research Reports of ESPRIT Project 688* (Vol. 1). Berlin Springer Verlag.
- Beckett, R. C. (2000). A characterisation of corporate memory as a knowledge system. *Journal of Knowledge Management*, 4(4), 311-319.
- Bertolazzi, P., Krusich, C., & Missikoff, M. (2001). *An Approach to the Definition of a Core Enterprise Ontology: CEO*. Paper presented at the OES-SEO 2001, International Workshop on Open Enterprise Solutions: Systems, Experiences, and Organizations.
- Bittner, T., & Smith, B. (2003). A Theory of Granular Partitions. In M. Duckham, M. F. Goodchild & M. F. Worboy (Eds.), *Foundations of Geographic Information Science* (pp. 117-151). London: Taylor & Francis Books.
- Bunge, M. (1977). *Treatise on Basic Philosophy, Ontology I: The Furniture of the World* (Vol. 3). Boston: Reidel.
- Ceusters, W. (2006). Towards A Realism-Based Metric for Quality Assurance in Ontology Matching. In B. Bennett & C. Fellbaum (Eds.), *Formal Ontology in Information Systems* (pp. 321-332). Amsterdam: IOS Press.
- Ceusters, W., Elkin, P., & Smith, B. (2006). Referent Tracking: The Problem of Negative Findings. In A. Hasman, R. Haux, J. v. d. Lei, E. D. Clercq & F. Roger-France (Eds.), *Studies in Health Technology and Informatics. Ubiquity: Technologies for Better Health in Aging Societies - Proceedings of MIE2006* (Vol. 124, pp. 741-746). Amsterdam: IOS Press.
- Ceusters, W., & Smith, B. (2006a). A Realism-Based Approach to the Evolution of Biomedical Ontologies. In *Proceedings of AMIA 2006* (pp. 121-125).
- Ceusters, W., & Smith, B. (2006b). Strategies for Referent Tracking in Electronic Health Records. *Journal of Biomedical Informatics*, 39(3), 362-378.
- Ceusters, W., & Smith, B. (2007). Referent Tracking for Digital Rights Management. *Forthcoming in International Journal of Metadata, Semantics and Ontologies*.



Dietz, J. L. G., & Halpin, T. A. (2004). Using DEMO and ORM in Concert: A Case Study. *Advanced Topics in Database Research*, 3, 218-236.

Donnelly, M., Bittner, T., & Rosse, C. (2006). A formal theory for spatial representation and reasoning in biomedical ontologies. *Artificial Intelligence in Medicine*, 36(1), 1-27.

Drew, R. (2006, March 20, 2006). In Google's Shadow. *Computerworld*.

Fox, M. S. (1992). *The TOVE Project: Towards A Common-sense Model of the Enterprise* (Technical Report): Enterprise Integration Laboratory.

Foxvog, D. (2005). Instances of Instances Modeled via Higher-Order Classes. In *Proceedings of the Workshop on Foundational Aspects of Ontologies (FOnt 2005)* (pp. 46-54).

Geerts, G., & McCarthy, W. E. (2002). An Ontological Analysis of the Primitives of the Extended-REA Enterprise Information Architecture *The International Journal of Accounting Information Systems*, 3, 1-16.

Goossenaerts, J., & Pelletier, C. (2003). Ontology and Enterprise Modeling [Electronic Version]. Retrieved November 16, 2006 from <http://is.tm.tue.nl/staff/jgoossenaerts/4PublicPdf/PSIM%20book%20ch%205%20Ontol&EM.pdf>.

Grenon, P., Smith, B., & Goldberg, L. (2004). Biodynamic Ontology: Applying BFO in the Biomedical Domain. In D. M. Pisanelli (Ed.), *Ontologies in Medicine* (pp. 20-38). Amsterdam: IOS Press.

Haller, A., & Oren, E. (2006). A process ontology to represent semantics of different process and choreography meta-models [Electronic Version]. Retrieved February 2006 from <http://www.m3pe.org/deliverables/process-ontology.pdf>.

Hay, D., & Healy, K. A. (2000). *Defining Business Rules ~ What Are They Really?* (Final Report No. Revision 1.3): The Business Rule Group.

Hill, M. (2006). Service Taxonomy and Service Ontologies Deliver Success to Enterprise SOA [Electronic Version]. *SOA Webservices journal*, 6. Retrieved November 23, 2006 from <http://webservices.sys-con.com/read/175385.htm>.

Hoopes, J. (1991). *Peirce ON SIGNS. Writings on Semiotic by Charles Sanders Peirce*. Chapel Hill and London: The University of North Carolina Press

Huhns, M. N., & Stephens, L. M. (2002). Semantic Bridging of Independent Enterprise Ontologies. In K. Kosanke (Ed.), *Enterprise Inter- and Intra-Organizational Integration: Building International Consensus* (pp. 83 - 90). Boston, MA: Kluwer Academic Publishers.

- IFOMIS. (2006, December 2006). Basic Formal Ontology. Retrieved January 25, 2007, from <http://www.ifomis.uni-saarland.de/bfo/>
- Kühn, O., & Abecker, A. (1997). Corporate memories for Knowledge Management in Industrial Practice: Prospects and Challenges. *Journal of Universal Computer Science*, 3(8), 929-954.
- Marchand, D. A., Davenport, T. H., & Dixon, T. (2000). *Financial Times—Mastering Information Management, Complete MBA Companion in Information Management*. London: FT Prentice Hall.
- McCarthy, W. E. (1982). The REA Accounting Model: A Generalized Framework for Accounting Systems in a Shared Data Environment. *The Accounting Review*, LVII(3), 554-578.
- Noy, N. F., & McGuinness, D. L. (2001). *Ontology Development 101: A Guide to Creating Your First Ontology* (No. KSL-01-05): Stanford Knowledge Systems Laboratory.
- Osterwalder, A., Lagha, S. B., & Pigneur, Y. (2002). *An ontology for developing e-business models*. Paper presented at the IFIP DSIAGE'2002.
- Parsons, J., & Wand, Y. (2000). Emancipating Instances from the Tyranny of Classes in Information Modeling. *ACM Transactions on Database Systems*, 25(2), 228-268.
- Partridge, C., & Stefanova, M. (2001). A Synthesis of State of the Art Enterprise Ontologies - Work in Progress [Electronic Version]. Retrieved November 22, 2006 from <http://citeseer.ist.psu.edu/632089.html>.
- Rosenthal, A., Manola, F., & Seligman, L. (2001). Getting Data to Applications—Why We Fail—Part 1: Common Fallacies. *The Mitre Information Technology Advisor*, 1(10), 1-2.
- Smith, B. (2004). Beyond concepts: ontology as reality representation. In *Proceedings of the third international conference on formal ontology in information systems (FOIS 2004)* (pp. 73-84). Amsterdam: IOS Press.
- Smith, B., Ceusters, W., Klagges, B., Köhler, J., Kumar, A., Lomax, J., et al. (2005). Relations in biomedical ontologies. *Genome Biology*, 6(5), R46.
- Smith, B., Ceusters, W., & Temmerman, R. (2005). Wüsteria. In R. Engelbrecht, A. Geissbuhler, C. Lovis & G. Mihalas (Eds.), *Connecting Medical Informatics and Bio-Informatics. Medical Informatics Europe 2005* (pp. 647-652). Amsterdam: IOS Press.
- Smith, B., Kusnierczyk, W., Schober, D., & Ceusters, W. (2006). *Towards a Reference Terminology for Ontology Research and Development in the Biomedical Domain*. Paper presented at the KR-MED 2006, Biomedical Ontology in Action., from [http://ontology.buffalo.edu/bfo/Terminology\\_for\\_Ontologies.pdf](http://ontology.buffalo.edu/bfo/Terminology_for_Ontologies.pdf).

Smith, B., & Varzi, A. C. (1997). Fiat and Bona Fide Boundaries: Towards an Ontology of Spatially Extended Objects In *Lecture Notes In Computer Science* (Vol. 1329, pp. 103 - 119). London, UK: Springer Verlag.

Stader, J. (1996). *Results of the Enterprise Project*. Paper presented at the 16th Annual Conference of the British Computer Society Specialist Group on Expert Systems

Uschold, M., King, M., Moralee, S., & Zorgios, Y. (1998). The Enterprise Ontology. *The Knowledge Engineering Review*, 13(1), 31-89.

Wand, Y., Storey, V., & Weber, R. (1999). An Ontological Analysis of the relationship Construct in Conceptual Modeling. *ACM Transactions on Database Systems*, 24(4), 494-528.

Wexler, M. N. (2002). Organizational Memory and Intellectual Capital. *Journal of Intellectual Capital*, 3(4), 393-415.