Ontology and Information Systems

Barry Smith[1]

Philosophical Ontology

Ontology as a branch of philosophy is the science of what is, of the kinds and structures of objects, properties, events, processes and relations in every area of reality. 'Ontology' is often used by philosophers as a synonym for 'metaphysics' (literally: 'what comes after the *Physics*'), a term which was used by early students of Aristotle to refer to what Aristotle himself called 'first philosophy'.[2] The term 'ontology' (or *ontologia*) was itself coined in 1613, independently, by two philosophers, Rudolf Göckel (Goclenius), in his *Lexicon philosophicum* and Jacob Lorhard (Lorhardus), in his *Theatrum philosophicum*. The first occurrence in English recorded by the OED appears in Bailey's dictionary of 1721, which defines ontology as 'an Account of being in the Abstract'.

Methods and Goals of Philosophical Ontology

[1] This paper is based upon work supported by the National Science Foundation under Grant No. BCS-9975557 ("Ontology and Geographic Categories") and by the Alexander von Humboldt Foundation under the auspices of its Wolfgang Paul Program. Thanks go to Thomas Bittner, Olivier Bodenreider, Anita Burgun, Charles Dement, Andrew Frank, Angelika Franzke, Wolfgang Grassl, Pierre Grenon, Nicola Guarino, Patrick Hayes, Kathleen Hornsby, Ingvar Johansson, Fritz Lehmann, Chris Menzel, Kevin Mulligan, Chris Partridge, David W. Smith, William Rapaport, Daniel von Wachter, Chris Welty and Graham White for helpful comments. They are not responsible for the errors which remain.

[2] Sometimes 'ontology' is used in a broader sense, to refer to the study of what might exist, where 'metaphysics' is used for the study of which of the various alternative possibilities is true of reality. See Ingarden (1964).

The methods of philosophical ontology are the methods of philosophy in general. They include the development of theories of wider or narrower scope and the testing and refinement of such theories by measuring them up, either against difficult counterexamples or against the results of science. These methods were familiar already to Aristotle. Some philosophical ontologists conceived ontology as being based on a special *a priori* insight into the essence of being or reality. Here, however, I prefer to look at the entire history of ontology as an endeavor which has some of the features of an empirical science. Seen from this perspective ontology is like physics or chemistry; it is part of a piecemeal, on-going process of exploration, hypothesis-formation, testing and revision. Ontological claims advanced as true today may well be rejected tomorrow in light of further discoveries or new and better arguments.

Philosophical ontology as I shall conceive it here is what is standardly called *descriptive* or *realist ontology*. It seeks not explanation but rather a description of reality in terms of a classification of entities that is exhaustive in the sense that it can serve as an answer to such questions as: What classes of entities are needed for a complete description and explanation of all the goings-on in the universe? Or: What classes of entities are needed to give an account of what makes true all truths? Or: What classes of entities are needed to facilitate the making of predictions about the future? Sometimes a division is made – as for example in the case of Husserl and Ingarden – between formal and material (or regional) ontology. Formal ontology is domain-neutral; it deals with those aspects of reality (for example parthood and identity) which are shared in common by all material regions. Material ontology deals with those features (for example mind or causality) which are specific to given domains.

Philosophical ontology seeks a classification that is exhaustive in the sense that all types of entities are included in its classifications, including also the types of relations by which entities are tied together. In striving for exhaustiveness philosophical ontology seeks a taxonomy of the entities in reality at all levels of aggregation (or, what comes to the same thing, at all levels of granularity), from the microphysical to the cosmological, and including also the middle world (the mesocosmos) of human-scale entities in between. Note that ontology as thus conceived is at odds with the attitude of reductionism, which sees reality in terms of some one privileged level of basic existents. Different schools of reductionism offer different approaches to the selection of the basic existents. One large division is that between what we might call substantialists and fluxists, which is to say between those who conceive reality in terms of substances or things and those who favor an ontology centered on process or function or on continuous fields of variation. Most reductionists are nominalists, which is to say that they deny the existence of universals or multiply-exemplfiied entities and conceive the world as being made up exclusively of individuals.

Reductionists seek to establish the ‘ultimate furniture of the universe’. They seek to decompose reality into its simplest or most basic constituents. They thus favor a criterion of ontological economy, according to which an assay of reality is good to the degree to which it appeals to the smallest possible number of types of entities. The challenge is then to show that all putative reference to non-basic entities can be eliminated in favor of entities on the basic level. The idea is that what is true on the basic level explains those phenomena which appear to obtain on the non-basic levels. The striving for explanatory unification supports reductionism.

Descriptive or realist ontology, in contrast, requires a stand-point of adequacy to all levels of reality, both basic and non-basic.[3] Reductionists seek to 'reduce' the apparent variety of types of entities existing in reality by showing how this variety is generated, for example through permutations and combinations of basic existents. The history of philosophical ontology is indeed marked by a certain trade-off between generativity on the one hand and descriptiveness on the other. By 'generativity' we understand the power of an ontology to yield new categories – and thus to exhaust the domain that is to be covered by ontological investigation – in some recursive fashion. By 'descriptiveness' we understand that feature of an ontology which consists in its reflecting, in more or less empirical ways, the traits or features of reality which exist independently of and prior to the ontology itself. It is generativity which gives an ontology its power; it is descriptiveness which ties an ontology to the world beyond.

All ontologists must find a way to combine as best they can the indispensable virtues of both generativity and descriptiveness. Philosophical ontology can then be enhanced by taking over elements from the methodology of reductionism, for example through the use of the axiomatic method illustrated also in the work of Lesniewski, Woodger, Goodman and others in formal mereology and illustrated also in Part 2 of Carnap's *Introduction to Symbolic Logic* (1958). Indeed in the course of the twentieth century a range of formal tools became available to ontologists for the development and testing of their theories. Ontologists nowadays have a choice of formal frameworks (deriving from formal logic, as well as from algebra, category theory, mereology, set theory, topology) in terms of which their theories can be formulated. These new formal tools allow philosophical ontologists to express intuitive principles and definitions in a clear and rigorous fashion, and they can allow also for the testing of theories for consistency and completeness through the application of the methods of formal semantics.

It is the work of philosophical ontologists such as Aristotle, Ingarden (1964), Chisholm (1996) and Johansson (1989) which will be of primary importance for us here. Their work rests upon the realist presupposition that a single consistent ontological theory can comprehend reality at a multiplicity of different levels of granularity. The taxonomies they propose are in many ways comparable to scientific taxonomies such as those produced by Linnaeus in biology or by Dalton in chemistry, though radically more general than these. All four of the mentioned philosophers are realists about universals, and all four transcend the dichotomy between substantialists and fluxists, since they accept categories of both things and processes, as well as other categories distinct from both of these.

Ontology and Science

Philosophical ontology is a descriptive enterprise. It is distinguished from the special sciences not only in its radical generality but also in its primary goal or focus: it seeks, not predication or explanation, but rather taxonomy. Ontology is (very largely) qualitative. Science is (very largely) quantitative. Science starts, very roughly, with measurement and prediction. There is

---

[3] Though contrast Mäki 2001, pp. 502 f.

an ontology of measure (Bigelow and Pargeter 1990); but ontologists themselves do not measure reality. Philosophical ontology tells us what categories exist within a given domain of reality and thus also what categories are available for the measurement process. Science tells us (for example) how the measurable behavior of entities of a certain class is correlated with the behavior of entities of a second class.

Sciences, by definition, can deal only with the objects which fall within their respective domains. Ontology deals with transcategorial relations – including the relations which hold between entities belonging to distinct domains of science, and also between these entities and the entities recognized by common sense. Already in 1963 Wilfrid Sellars advanced a thesis to the effect that there is a universal common-sense ontology, which he called the 'manifest image' and which he held to be a close approximation to the enduring common core of traditional philosophical ontology (also called the '*philosophia perennis*') initiated by Aristotle.

Strawson (1959) draws in this connection a distinction between two different kinds of ontological investigation. On the one side is what he calls 'descriptive metaphysics', which aims to lay bare the most general features of the conceptual scheme we do in fact employ – which is roughly that of common sense. On the other side is 'revisionary metaphysics', which is prepared to make departures from this scheme, for example in light of developments in science. As Strawson puts it: 'descriptive metaphysics is content to describe the actual structure of our thought about the world, revisionary metaphysics is concerned to produce a better structure.'

Strawson's descriptive metaphysics is certainly related to ontology as here conceived. But it is to be distinguished therefrom in that the very dichotomy between descriptive and revisionary metaphysics masks from view the links between those parts of our ontology that pertain to the reality accessed by common sense, and those parts of our ontology (at different granularities) that pertain to science. Ontology a seeks precisely to do justice in descriptive fashion to all of the various parts and dimensions of reality on all levels of granularity, whether these be accessed by science, by common sense, or by some other means. It should be noted for future reference that Strawson's two type of metaphysics are distinguished from ontology also in this: that they are directed not to reality itself, but rather to the 'conceptual schemes' which we employ when engaging with reality.

Ontology and Taxonomy

An ontology is, in first approximation, a table of categories, in which every type of entity is captured by some node within a hierarchical tree. This ideal lay at the root of Aristotle's thinking on categories, as also of that of his medieval successors, and it has been resurrected in the thinking of contemporary ontologists such as Chisholm, who presents the following table of categories in his (1996):

```
                 Entities
                 /      \
                /        \
               /          \
              /            \
     Contingent            Necessary
      /    \                 /    \
```

```
               /       \              /       \
              /         \            /         \
             /           \          /           \
        States     Individuals    States     Non-states
          /           /    \                   /    \
         /           /      \                 /      \
        /           /        \               /        \
       /           /          \             /          \
  Events    Boundaries    Substances    Attributes    Substance
```

Figure 1: Chisholm’s Tree

In principle all entities in reality would be comprehended along these lines within a single tree, which is then extendible via the drawing of ever finer distinctions. This principle is at work in the taxonomy presented by the seventeenth century English polymath John Wilkins, Bishop of Chester, in his *An Essay toward a Real Character and Philosophical Language* (1668). Here Wilkins proposed a universal taxonomy of forty genera, which he lists as follows:

transcendent relations: General, Mixed, Of Action

unclassified: Discourse, God, World, Element, Stone, Metal

plants: Herb Leaf, Herb Flower, Herb S. Ves., Shrub, Tree

animals: Exsanguinous, Fish, Bird, Beast

parts: Peculiar, General

quantity: Magnitude, Space, Measure

quality: Natural Power, Habit, Manners, Sensible Quality, Sickness

action: Spiritual, Corporeal, Motion, Operation *relation*: Economic, Possessions, Provisions, Civil, Judicial, Military, Naval, Ecclesiastical.

Wilkins' taxonomy is designed to serve as the basis for an ideal language, analogous to the *characteristica universalis* conceived by Leibniz as a language in which it would be possible to express all concepts via systematic composition from a list of simple or basic concepts. Where however Leibniz was, in the terms of our earlier discussion, a generativist, Wilkins' project is carried out against the background of a descriptivist ontology, so that there is for example no attempt to reduce all genera to complexes of atoms or motions or other simples. Wilkins' universal character is distinguished also by the fact that it refers only to existing entities, leaving no room for concepts of fiction or mythology. On the other hand, however, Wilkins' ontology and its associated universal character are unsatisfyingly *ad hoc*. Thus a conspicuously large fraction of his book is devoted to the two categories of Stone and Metal. Wilkins subdivides the former into *common* (silica, gravel, schist), *modic* (marble, amber, coral), *precious* (pearl, opal), *transparent* (amethyst, sapphire) and *insoluble* (chalk, arsenic). The latter he divides into *imperfect* (cinnabar, mercury), *artificial* (bronze, brass), *recremental* (filings, rust) and *natural* (gold, tin, copper).

It was this odd treatment of Stone and Metal which served as the jumping-off point for Borges' essay "The Analytical Language of John Wilkins", which is however devoted not so much to Wilkins' *Real Character*, which Borges had not read, as to a fictional 'Chinese Encyclopedia' (ascribed by Borges to a certain 'Franz Kuhn'), in which it is written that animals are divided into:

1. those that belong to the Emperor
2. embalmed ones
3. those that are trained
4. suckling pigs
5. mermaids
6. fabulous ones
7. stray dogs
8. those included in the present classification
9. those that tremble as if they were mad
10. innumerable ones
11. those drawn with a very fine camelhair brush
12. others
13. those that have just broken a flower vase
14. those that from a long way off look like flies.

There are a number of distinct dimensions along which ontologies can be compared, and as our discussion of the trade-off between generativity and descriptiveness makes clear, there will be no single criterion which we can use to sort the wheat from the chaff. Already on the basis of sheer inspection, however, we can see that there are a number of important respects in which Borges' Chinese classification falls short of the classifications set forth by Chisholm and Wilkins. One such respect, which we shall here take to be of central importance, concerns the degree to which ontology is compatible with the results of the natural sciences. Other criteria for evaluation pertain to the breadth or scope and to the unity of an ontological taxonomy. Ideally, as in the simple table of categories propounded by Chisholm, an ontology should consist of a single all-encompassing taxonomy. As we shall see, however, all the mentioned criteria relate to an ideal case only, an ideal which is in fact unrealizable in the actual practice of ontology in the world in which we live.

Well-Formed Taxonomies

Other criteria which taxonomies must aim to satisfy if they are to serve the needs of ontology have to do with the well-formedness of the taxonomy itself. (Bittner and Smith 2001) We can set forth a preliminary list of principles of well-formedness. These, too, however, are intended to delineate the ideal case only.

1. A taxonomy should take the form of a tree in the mathematical sense.

This means that, as in the case of Chisholm's tree above, it should be a connected graph without cycles. The nodes of the tree then represent categories at greater and lesser levels of generality, and branches connecting nodes represent the relations of inclusion of a lower category in a higher. Corresponding to the inclusion relation between subordinate and superordinate nodes within the tree is the relation of part to whole between the respective totalities of objects out there in the world to which the nodes correspond. The totality of objects belonging to the included category is a sub-totality of the totality of objects belonging to the including category. To insist on the tree structure is to insist, in effect, that from any given node in the tree there is at most one branch issuing upwards. A category is thereby never subordinate to more than one higher category within the tree. (In visual terms a tree has no diamonds.) This means that if two categories represented within a tree are such that their respective families of instances overlap, then one is a subcategory of the other.

The germ of the no diamonds principle is the idea that a classification should involve no double-counting. If, in counting off the cars passing beneath you on the highway, your checklist includes one box labeled *red cars* and another box labeled *Chevrolets*, we will rightly insist that there is something amiss, because you will almost certainly be guilty of counting some cars twice. Another problem is that there is no natural relationship between these two nodes of your classification, which seem as though they ought properly to belong to two distinct classifications made for two distinct purposes.

Inspection reveals that the taxonomies employed by the natural sciences – for example in zoology or botany or chemistry – satisfy, at least ideally, the mentioned condition. Putative counterexamples to the rule are found in the realm of artifacts. For example, a taxonomy of urban structures might employ the two categories: *car parks* and *buildings*, both of which seem to be superordinate categories to *parking ramp*. It is always possible, however, to eliminate such counterexamples by replacing one or the other of the relevant superordinate categories – for example substituting *parking area* for *car park* – in such a way that the overlap is eliminated (Guarino and Welty 2002).

Certainly it is useful for some purposes to employ taxonomies which depart from the tree structure by placing a given category simultaneously on a number of separate branches within a hierarchy in such a way that it inherits information from each branch. Thus a given virus might be a *type of RNA virus* that is also *associated with lymphoma in tortoises*. Such cross-classification confuses two purposes, however. On the one hand is the strictly taxonomical purpose, which attempts to establish at each level within the tree a jointly exhaustive and pairwise disjoint inventory of the entirety of the domain to which the taxonomy applies at a given level of granularity. On the other hand is the task of encoding knowledge about the instances of a category associated with a given node of a tree.

2. A taxonomy should have a basis in minimal nodes, representing lowest categories in which no sub-categories are included.

The term 'basis' here is to be understood in the mathematical sense familiar from the theory of vector spaces. Rule 2. is designed to guarantee that the categories at the lowest level of the tree exhaust the maximal category in the way in which, for example, a chemical classification of the noble gases is exhausted by the nodes Helium, Neon, Argon, Krypton, Xenon and Radon. This rule ensures also that every intermediate node in the tree is identifiable as a combination of minimal nodes.

3. A taxonomy should be *unified* in the sense that it should have a single top-most or maximal node, representing the maximum category.

This maximal category then includes all the categories represented by the nodes lower down the tree. The justification for this principle lies in the fact that a taxonomy with two maximal nodes would be in need of completion by some extra, higher-level node representing the union of these two maxima. Otherwise it would not be one taxonomy at all, but rather two separate and perhaps competing taxonomies, the claims of each of which would need to be considered in their own right.

If the whole of ontology could be represented as a taxonomy in this sense, then we could employ a single term – such as ‘entity’ – as a label for the highest-level category of ontology. Everything which exists would be an entity in the intended sense. (Alternative top-level terms which have been favored by different ontologists include: ‘thing,’ ‘object,’ ‘item,’ ‘element,’ ‘existent.’)

Unfortunately, as Aristotle already recognized, the prospects for ontology as a single taxonomic tree are very poor. There is a variety of cross-cutting ways of establishing ontological categories in reality. All of these ways are compatible (as one can slice a cheese in a variety of different but yet compatible ways). Yet they cannot be combined together to form a single taxonomic slicing.

Ontology as a Family of Trees

How, now, in light of what has been said about taxonomies and trees, are we to conceive ontology? Unfortunately, in spite of the example of Chisholm and Wilkins, it is an unrealizable ideal to suppose that ontology would consist in a single taxonomy comprehending all of reality and satisfying the rules for well-formedness we have mentioned above. The features listed are not simultaneously realizable. Above all, ontology must consist not in one tree but in a family of trees, each reflecting specific views (facets or factors) of the targeted domain – for example (microscopic, mesoscopic, macroscopic) views effected at different granularities.

Different views or facets arise above all because of the different ways in which the categories of entities in reality relate to time: some entities exist *in* time, either in the manner of substances, which endure identically from moment to moment, or in the manner of processes, which unfold themselves in time phase by phase. Other entities (it is commonly held) exist outside time. This holds, for example, of numbers, Platonic forms, and other ideal entities. To put such different kinds of entities together, with Chisholm, into a single ontological tree, would seem to presuppose that there is some (temporal?) order to which they all belong. Another argument against the single tree conception put forward by Aristotle turns on the principles by which the lower levels of a tree are derived from the levels above. For Aristotle this derivation is a form of specification: a human is a *rational* animal, an animal is a *living substance*, and so on. If all of ontology were to take the form of a single tree, then there must be some highest category, say *entity*, of which all lower categories would then be specifications. But what would the highest category be, of which both *animal* and *action* (for example) would alike be specifications?

As Bishop Wilkins' system very clearly reveals, the very complexity of reality already brings with it the necessity to classify the entities in the world according to a variety of different facets or dimensions. If we try to cram the whole of our ontology into a single tree then this results in arbitrary orderings – which differentia should one choose and in which order as one proceeds down the tree? – and either to duplication or omission. Wilkins himself recognized this problem, but justified it on pragmatic grounds.[4] It is Wilkins' tree-structuring that is at fault when stones are categorized into *common*, *modic*, *precious, transparent* and *insoluble*. These are in and of themselves perfectly good categories (thus it is quite reasonable to classify diamonds with rubies rather than with coal). But what Wilkins' classification reveals is that there are different aspects under which stones can be reasonably classified, and the structure of a single tree forces the choice of just one of these aspects, so that one must either ignore all the rest or integrate them in an *ad hoc* manner.

Philosophical ontology is more complex still in virtue of the fact the ontology studies not just taxonomies of reality but also *partonomies*, which is to say assays of the parts of entities of given types. We leave to one side this issue here, noting only that taxonomies and partonomies should not be confused: to say that the category of rabbits is a sub-category of the category of mammals is a quite different sort of statement from the statement that a rabbit's leg is a part of a rabbit.

Ontological Commitment

To create effective representations it is an advantage if one knows something about the things and processes one is trying to represent. (We might call this *the Ontologist's Credo.*) The attempt to satisfy this credo has led philosophers to be maximally opportunistic in the sources they have drawn upon in their ontological explorations of reality. These have ranged all the

[4] Wilkins 1666, Chapter XII, I, Sec. I, p. 289. Wilkins 1666, Chapter XII, I, Sec. I, p. 289.

way from the preparation of commentaries on ancient texts to reflection on our linguistic usages when talking about entities in domains of different types. Increasingly, philosophers have turned to science, embracing the assumption that one generally reliable way to find out something about the things and processes within a given domain is to see what scientists say.

Some philosophers have thought that the way to do ontology is exclusively through the investigation of scientific theories. With the work of Quine (1953) there arose in this connection a new conception of the proper method of ontology, according to which the ontologist's task is to establish what kinds of entities scientists are committed to in their theorizing. The ontologist studies the world by drawing conclusions from the theories of the natural sciences, which Quine takes to be our best source of knowledge as to what the world is like. Such theories are extensions of the theories we develop and use informally in everyday life, but they are developed with closer attention to certain special kinds of evidence that confer a higher degree of probability on the claims made. Quine's aim is to use science for ontological purposes, which means: to find the ontology *in* scientific theories. Ontology is then a network of claims, derived from the natural sciences, about what exists. Each natural science has, Quine holds, its own preferred repertoire of types of objects to the existence of which it is committed. Each such theory embodies only a partial ontology. This is defined by the vocabulary of the corresponding theory.

Note that Quine himself takes ontology seriously. Thus he does not embrace a view according to which ontology *is* the meta-level study of the ontological commitments or presuppositions embodied in the different natural-scientific theories. Ontology is rather these commitments themselves. Quine moves to the meta-level, making a semantic ascent to consider the statements in a theory, only in setting out to establish those expressions which definitively carry its commitments. The latter are marked in his eyes by their special logical form, which is revealed in their canonical representation in first-order predicate logic.

Quine fixes upon the language of first-order logic as the medium of canonical representation, not out of dogmatic devotion to this particular form. He conceives first-order logic simply as a regimentation of corresponding parts of ordinary language from which those features which are logically problematic have been excised. It is then, Quine argues, only the bound variables of the theory as canonically represented that carry its definitive commitment to existence. It is sentences like 'There are horses,' 'There are numbers,' 'There are electrons,' that do this job. His so-called 'criterion of ontological commitment' is captured in the slogan: *To be is to be the value of a bound variable*. This should not be understood as signifying some reductivistic conception of existence itself as a merely logico-linguistic matter. Rather it is to be interpreted in practical terms: to determine what the ontological commitments of a scientific theory are it is necessary to examine the predicates holding of the bound variables used in its canonical formalization.

Quine's approach is thus most properly conceived not as a reduction of ontology to the study of scientific language, but rather as a continuation of ontology in the traditional sense. When viewed in this light, however, it can be seen to be in need of vital supplementation. For the objects of scientific theories are discipline-specific. This means that the *relations* between objects belonging to different disciplinary domains fall out of bounds for Quinean ontology.

Only something like a philosophical theory of how different scientific theories (or their objects) relate to each other can fulfil the task of providing an inventory of all the types of entities and relations in reality. Quine himself would resist this latter conclusion. For him the best we can achieve in ontology lies in the quantified statements of particular theories, theories supported by the best evidence we can muster. We have no way to rise above the particular theories we have; no way to unify their respective claims.

Internal vs. External Metaphysics

Quine is a realist philosopher. He believes in a world beyond language and beliefs, a world which the theories of natural science give us the power to illuminate. There is, however, another tendency in twentieth-century analytic philosophy, a tendency inspired by Kant and associated above all with the names of Carnap and Putnam, according to which ontology is a meta-level discipline which concerns itself not with the world itself but rather only with theories or languages or concepts or systems of beliefs. Philosophical ontology in the traditional sense – ontology as a first-level discipline directed to the world beyond – is impossible. For such an ontology would require what the just-mentioned philosophers call 'external metaphysics', which is to say metaphysics carried out on the basis of what they like to call a God's eye perspective, from which one could view reality as it exists independently of our language and concepts. Since such a perspective is (so the mentioned philosophers argue) for us unavailable, it follows that the best we can achieve is *internal metaphysics*, which means the study of the ontological commitments of specific languages, theories, or systems of beliefs. Strawsonian descriptive metaphysics is one example of such internal metaphysics. Model-theoretic semantics, too, is often implicitly understood in internal-metaphysical terms – the idea being that we can never understand what a given language or theory is really about, but we can build models with more or less nice properties. But we can never compare these models to some reality beyond.

Ontology in the traditional philosophical sense is thus replaced by the study of how a given individual or group or language or science conceptualizes a given domain. It is a theory of the ontological content of certain representations. Traditional ontologists are seeking principles that are true of reality. The practitioners of internal metaphysics, in contrast, are seeking to elicit principles from subjects or theories. The elicited principles may or may not be true, but this, to the practitioner of internal metaphysics, is of no concern, since the significance of these principles lies elsewhere – for instance in yielding a correct account of the taxonomical system used by speakers of a given language or by scientists working in a given discipline.

Ontology Outside Philosophy

In a development that has hardly been noted by philosophers, a conception of the job of the ontologist close to that of the adherents of internal metaphysics has been advanced in recent years also in certain extra-philosophical disciplines, as linguists, psychologists and anthropologists have sought to elicit the ontological commitments (‘ontologies’, in the plural) of different cultures and groups. Exploiting the terminology of Quine, researchers in psychology and anthropology have sought to establish what individual human subjects, or entire human cultures, are committed to, ontologically, in their everyday cognition,[5] in much the same way in which philosophers of science had attempted to elicit the ontological commitments of the natural sciences. Thus they have engaged in inquiries designed to establish how folk ontologies (or folk biologies, folk theories of physics, folk psychologists, and so on) develop through infancy and childhood, or to establish the degree to which given elements of folk ontologies reflect universal features of the human cognitive system.

Note that it was still reasonable for Quine to identify ontology in the traditional sense – the search for answers to the question: what exists? – with the study of the ontological commitments of natural scientists. It is, after all (and leaving to one side the troublesome case of quantum mechanics) a reasonable hypothesis to suppose that all natural sciences are, if not consistent with each other, then at least such that the inconsistencies which arise can be eliminated through the efforts of the scientists themselves. Moreover, the identification of the method of ontology with the isolation of ontological commitments continues to seem reasonable when one takes into account not only the natural sciences but also certain commonly shared commitments of common sense – for example that *tables* and *chairs* and *people* exist. For the common-sense taxonomies of objects can be shown to be in large degree compatible with those of scientific theory, if only we are careful to take into account the different granularities at which each operates (Smith and Brogaard, in press).

Crucially, however, the identification of ontology with the isolation of ontological commitments becomes strikingly less defensible when the ontological commitments of various specialist groups of *non*-scientists are allowed into the mix. For how, ontologically,

[5] See for example Keil 1979, Spelke 1990, Medin and Atran 1999, Xu and Carey 1996.

are we to treat the commitments of Meinongian philosophers, or astrologists, or believers in leprechauns?

Ontology in Information Science

In a related development, also hardly noticed by philosophers, the term ‘ontology’ has gained currency in recent years in the field of computer and information science in a way which has led to a veritable explosion of publications and conferences on the topic of ontology. a term which has become popular especially in domains such as knowledge engineering, natural language processing, cooperative information systems, intelligent information integration, and knowledge management. The philosopher-ontologist, in principle at least, has only one goal: to establish the truth about the domain in question. In the world of information systems, in contrast, an ontology is a software (or formal language) artefact designed with a specific set of uses and computational environments in mind, and often ordered up by a specific client or customer or application program in a specific context.

The work of Quine played an important role, too, in the initial phases of the development of what I shall henceforth refer to as ‘information systems ontology’. It seems that the first use of the term ‘ontology’ in the computer and information science literature occurs already in 1967, in a work on the foundations of data modeling by S. H. Mealy, in a passage which concludes with a footnote referring to Quine’s essay “On What There Is”. Here Mealy distinguishes three distinct realms in the field of data processing:

> the real world itself, ideas about it existing in the minds of men, and symbols on paper or

some other storage medium. The latter realms are, in some sense, held to be models of the former. Thus, we might say that data are fragments of a theory of the real world, and data processing juggles representations of these fragments of theory. No one ever saw or pointed at the integer we call "five" – it is theoretical – but we have all seen various representations of it, such as:

V $(101)_2$ 5_8 5 0.5E01

and we recognize them all as denoting the same thing, with perhaps different flavours. … The issue is ontology, or the question of what exists. (Mealy 1967. p. 525)

This concern with what are effectively ontological questions (what is data? how do data relate to the real world?) arose in reflection of quite specific practical problems which needed to be faced in the late 1960s by those working in the field of database management systems. Just as philosophical ontology has been marked by debates between substantialists and fluxists, so the field of artificial intelligence was marked by debates between the so-called proceduralist and declarativist schools. What is the relative significance of *process* and *content* (or of *procedures* and *data*) in the project of modelling intelligent reasoning and constructing intelligent machines? Proceduralists believed that the way to create intelligent machines was by instilling into a system as much *knowledge how* as possible, via ever more sophisticated programs. The declarativists, on the other side, believed that intelligent machines would best be arrived at by instilling into a system a maximum of content, or of *knowledge that* – knowledge in the form of representations.

In the database management systems field, now, the increasing size and complexity of programs meant in turn increasing difficulties in maintaining such programs and putting them to new uses. Some in the database community saw both the procedural and the declarative elements of computer systems as representations: programs are representations of processes, data structures are representations of objects or things. Recall, however, the Ontologist's Credo, that if one wants to create effective representations it is an advantage if one knows something about the objects and processes one is trying to represent. This means that one must know not only about the specific token objects (customers, payments, debts) recorded in one's database, but also about objects, properties and relations in general, and also about the general types of processes in which objects, properties and relations can be involved. The declarativist response to these problems was to embark upon an effort to represent formally the so-called 'conceptual schemes' involved in given application domains. The idea was to build declarative representations of procedures – for example business processes of ordering or scheduling – in a way designed to enable application systems to re-use program elements more easily, and in a way which would also have the effect of making application systems smaller in terms of code.

All of these tendencies can be seen at work in the idea of the so-called three schema architecture advanced in the database field in the 1970s (Jardine 1977). This distinguishes: 1. *implementation schemas*, describing physical ways of storing the data and object code of the program; 2. *conceptual schemas*, in terms of which declarative representations are formulated; and 3. *presentation schemas*, which are applied at external interfaces for the purposes of communicating to the user. These are three distinct perspectives or views which can be taken of a database. When we take an internal perspective, we describe the physical layout of the data in the computer. When we take a conceptual perspective, we describe the types of information stored, the relationships and operations recognized by the database. When we take an external perspective, we consider the real world to which the database is directed, primarily in terms of the ways in which its outputs will be made available to ultimate users. The three schema architecture thus offers a way for those who are responsible for the maintenance of the physical data, those who are responsible for managing the data, and those who use the data, to refer, each in his own fashion, to the same object.

A database management system offers services for programmers and users designed to ensure that correct data types are employed for given objects or attributes, for example that an age is a number greater than zero and less than 150. All information pertaining to each different object- and attribute-type is controlled by the system in ways designed to facilitate consistency checking and portability from one database to another. In this way all the structural knowledge pertaining to the application domain is captured in one central place.

The step from here to ontology is then relatively easy. The data analyst realizes the need for declarative representations which would have as much generality as possible in order to maximize the possibility of reusability. But at the same time these representations must correspond as closely as possible to the things and processes they are supposed to represent. Thus he starts asking questions like: What is an object/process/attribute/relation? He begins, in other words, to take seriously the Ontologist's Credo. Gradually he begins to see ontology as a theoretical enterprise in its own right – the enterprise of providing a formal representation of the main categories of entities and relations in a given domain that can be shared between different application environments.

The explosion of work in information systems ontology can be seen in this light as reflecting the efforts on behalf of at least some computer and information scientists to look beyond the artefacts of computation and information to that big wide world beyond to which these artefacts relate.

The growth of ontology in some respects parallels the spreading of the paradigm of object-oriented software, where the idea is to organize a program in such a way that its structure mirrors the structure of the objects and relationships in its application domain (Kim 1990). Here, too, one claim that is made on behalf of the programs which result is that they enjoy the benefits of portability.

Ontology in Artificial Intelligence

One early influential use of the term ‘ontology’ in the computer science community was by John McCarthy in his 1980 paper on ‘circumscription.’ McCarthy argues in this paper that the proper treatment of common-sense reasoning requires that common-sense knowledge be expressed in a form which will allow us to express propositions like ‘a boat can be used to cross rivers unless there is something that prevents its use.’ This means, he says, that:

> we must introduce into our *ontology* (the things that exist) a category that includes *something wrong with a boat* or a category that includes *something that may prevent its use*. … Some philosophers and scientists may be reluctant to introduce such things, but since ordinary language allows “*something wrong with the boat*” we shouldn’t be hasty in excluding it. … We challenge anyone who thinks he can avoid such entities to express in his favorite formalism, “*Besides leakiness, there is something else wrong with the boat.*” (p. 31)

McCarthy is here using ‘ontology’ in something very close to the Quinean sense: we know what we are ontologically committed to if we know what kinds of entities fall within the range of the bound variables of a formalized theory.

Another early use of the term is in the writings of Patrick Hayes, a collaborator of McCarthy, for example in his “Ontology for Liquids” (1985a), an early version of which is dated 1978. The term is used also in Hayes’ “Naïve Physics Manifesto”, which advocates a view of AI research as something which should be based not on the procedural modeling of reasoning processes but rather on the construction of systems embodying large amounts of declarative knowledge. Here he takes up the torch of a program outlined by McCarthy already in 1964, rooted in the idea that even a rather simple program – equivalent to an axiom system formulated in first-order predicate logic – could manifest intelligent reasoning. Hayes breaks with McCarthy only in the estimate of the likely size of the knowledge base (or list of predicates and axioms) needed. Thus he proposed abandoning the toy examples that drive McCarthy’s work and building instead a massive theory of physical reality as faced by untutored human beings acting in their everyday intereactions with objects in the world. Thus he concentrates on the formalization of all those manifest physical features which are relevant to the actions and deliberations of human beings engaged in the serious business of living.

Something of the order of 10,000 predicates would, Hayes thought, need to be encoded if the resulting naïve physics was to have the power to simulate the reasoning about physical phenomena of non-experts, and a range of large-scale projects of this type are, he argued, essential for long-term progress in artificial intelligence. Hayes’ “Ontology for Liquids” represents one detailed development of naïve physics in relation to a domain of objects and phenomena which had been almost totally neglected by traditional philosophical ontology.

The methodology here can be traced to the already mentioned Part 2 of Carnap's *Introduction to Symbolic Logic*, and consists in what Carnap calls *applied logic*, which is to say the attempt to formulate axiomatically theories from various domains of science. Where Carnap turns to science, Hayes turns to human common sense. His "Naïve Physics Manifesto" might thus best be conceived as a contribution not to the discipline of ontology in the traditional philosophical sense, but rather to that of knowledge representation. His idea is that the axioms of naïve physics should constitute a computational counterpart of human mental models. Thus they are supposed to be about the real world, not about mental models themselves. The axioms of naïve physics are formulated not by addressing matters psychological, but rather by thinking 'naively' about the world (which is to say: as a normal human actor, rather than as a scientist) and then trying to capture that knowledge of the world formally. Hayes reports that the attempt to formalize his own intuitive understanding of liquids led him to an ontology

> within which one could account for the difficulty that young children have in following conservation arguments; but I didn't *set out* to model this Piagetian phenomenon, and indeed when I told psychologists about what I was doing, I was met with a strange mixture of interest and disapproval, precisely because I was not setting out to test any particular psychological theory about mental structure, which they often found puzzling and unsettling. (Interestingly, the only people who seemed to immediately 'get it' were the Gibsonians, whose own methodology strictly

required a similar kind of focussing on the world being perceived, rather than the perceiver.) (Personal communication)[6]

Thus Hayes takes it as a basic assumption that our mental models in fact are about the physical features of the world itself, and thus that their computational counterparts will also be about the same reality.

Hayes is already in 1979 acutely aware of the fact that any first-order axiomatization of a theory has an infinity of non-intended models (including completely artificial interpretations constructed out of the symbols by means of which the theory itself is formulated).[7] In the first version of his Manifesto he was still confident that it would be possible to do something to overcome this problem of non-intended models and thus to find a way to home in on physical reality itself. In 1985, however, he published his "Second Naïve Physics Manifesto", a revised version of the earlier work, in which he has lost some of this earlier confidence.

Hayes' original "Manifesto" had listed four characteristics which his proposed formalism would have to possess:

1. *thoroughness* (it should cover the whole range of everyday physical phenomena),
2. *fidelity* (it should be reasonably detailed),
3. *density* (the ratio of represented facts to concepts should be fairly high)
4. *uniformity* (it should employ a single formal framework).

In the second version of the Manifesto everything in this list remains the same except that *thoroughness* is renamed *breadth*, and – most important for our purposes – the criterion of

[6] On Gibsonianism see J. J. Gibson 1979.

fidelity is dropped entirely, in a move which marks a further (witting or unwitting) alienation on Hayes' part from the traditional goal of philosophical ontology: that of providing representations adequate to reality. Initially Hayes had been confident that the problem could be alleviated by insisting that the theory be faithful to reality through the addition of lots of extra detail, thus for example by ensuring through extra axioms that every model have an essentially three-dimensional structure.He found, however, that this optimism as concerns the problem of unintended models could not be maintained. In the second version, therefore, he talks not of faithfulness to reality but much more loosely in terms of 'faithfulness to alternative models'. 'If I thought there was any way to pin down reality uniquely, then I would jump at the chance; but I don't think there is (for humans or machines).' (Personal communication)

Even in the later version, however, he still leaves open one route to the solution of the problem of non-intended models, namely by equipping the system with external links to reality. This could be done by having the formal theory be in a creature with a body: 'some of the tokens can be attached to sensory and motor systems so that the truth of some propositions containing them is kept in correspondence to the way the real world actually is.' Alternatively it could be done by having the theory converse with users of natural language like ourselves, whose beliefs were themselves assumed to refer to external entities. We would then 'have no reason to refuse the same honor to the conversing system.' (1985, p. 13)

[7] The results of Gödel and Löwenheim-Skolem demonstrated that any first-order axiomatic theory will have unintended models in the semantic (set-theoretical) sense of 'model'.

If the trick of homing in upon this, our actual world can be carried off in this or in some other way then it would follow in Hayes' eyes that *this actual world would be a model of the theory*. One passage included in both versions is then highly illuminating in this respect. It is a passage in which Hayes advances the thesis that a model can *be* a piece of reality.

> If I have a blocks-world axiomatization which has three blocks, '*A*', '*B*', and '*C*', and if I have a (real, physical) table in front of me, with three (real, physical) wooden blocks on it, then the set of these three blocks can be the set of entities of a model of the axiomatization (provided, that is, that I can go on to interpret the relations and functions of the axiomatization as physical operations on the wooden blocks, or whatever, in such a way that the assertions made about the wooden blocks, when so interpreted, are *in fact* true). There is nothing in the model theory of first-order logic which a priori prevents the real world being a model of an axiom system. (1979, p. 181; 1985, p. 10)

We shall return to this thesis below.

To satisfy the criterion of fidelity a proposed formalism must be 'reasonably detailed'. As Hayes already notes in his first Manifesto, 'since the world itself is infinitely detailed, perfect fidelity is impossible' (1979, p. 172); on the other hand, 'since we want to formalize the common-sense world of physical reality, this means, for us, that a model of the formalization must be recognizable as a facsimile of physical reality' (p. 180).

In his second manifesto, Hayes talk of 'faithfulness to reality' has been replaced by talk of our ability 'to interpret our axioms in a possible world.' To establish whether these axioms are true or not means to develop ‘an idea of a *model* of the formal language in which the theory is written: a systematic notion of what a possible world is and how the tokens of the theory can be mapped into entities … in such worlds’ (1985, p. 10). This implies a new conception of the goal of naïve physics – and thus of the goal of information systems ontology to the extent that the latter is a generalization of the original naïve physics idea. On this new conception, ontology has to do with what entities are included in a model in the semantic sense, or in a possible world. This conception is present also in the writings of John Sowa, who refers to ‘an *ontology* for a possible world – a catalogue of everything that makes up that world, how it’s put together, and how it works’ (1984, p. 294). [8]

The Database Tower of Babel Problem

In the AI community the goal (‘artificial intelligence’) is one of radically extending the boundaries of automation. There we see ontology building – in the work of Hayes, or in the framework of the Cyc project – as a process of extending the frontiers of what can be represented in systematic fashion in a computer, with the analogy to the knowing human subject in the background. In the data modeling community, in contrast, the goal is to integrate the automated systems we already have. Here the problems faced by ontologists are presented by the foibles of the often very tricky and unstable systems used, for example, in

[8] More recently Sowa provides two definitions of the term ‘ontology’. The first is in keeping with the more traditional philosophical sense of the term: “The subject of *ontology* is the study of the *categories* of things that exist or may exist in some domain.” The second seems to involve a confusion between ontology and an epistemologically based study of ontological commitments of a language: “[*an ontology*] is a catalog of the types of things that are assumed to exist in a domain of interest ***D*** from the perspective of a person who uses a language ***L*** for the purpose of talking about ***D***.” See: http://users.bestweb.net/~sowa/ontology/

the different parts of a large enterprise (and these problems are only further compounded by the fact that computer systems can themselves serve as mechanisms for constructing elements of social reality such as deals, contracts, debt records, and so forth).

As time has proved, the most important task for the new information systems 'ontology' relates not to the project of artificial intelligence but still in the realm of database management – and pertains to what we might call the Database Tower of Babel problem. Different groups of data- and knowledge-base system designers have for historical and cultural and linguistic reasons their own idiosyncratic terms and concepts by means of which they build frameworks for information representation. Different databases may use identical labels but with different meanings; alternatively the same meaning may be expressed via different names. As ever more diverse groups are involved in sharing and translating ever more diverse varieties of information, the problems standing in the way of putting such information together within a larger system increases geometrically.

It was therefore recognized early on that systematic methods must be found to resolve the terminological and conceptual incompatibilities between databases of different sorts and provenance. Initially, such incompatibilities were resolved on a case-by-case basis. Gradually, however, the idea took root that the provision of a common reference taxonomy might provide significant advantages over such case-by-case resolution. The term ‘ontology’ then came to be used by information scientists to describe the construction of a reference taxonomy of this sort. An ontology is in this context a dictionary of terms formulated in a canonical syntax and with commonly accepted definitions designed to yield a lexical or taxonomical framework for knowledge-representation which can be shared by different information systems communities. More ambitiously, an ontology is a formal theory within which not only definitions but also a supporting framework of axioms is included (the axioms themselves providing implicit definitions of – or constraints upon the meanings of – the terms involved).

The potential advantages of such ontology for the purposes of knowledge representation and information management are obvious. Each group of data analysts would need to perform the task of making its terms and concepts compatible with those of other such groups only once – by calibrating its results in the terms of a single shared canonical backbone language. If all databases were calibrated in terms of just one common ontology built around a consistent, stable and highly expressive set of category labels), then the prospect would arise of leveraging the thousands of person-years of effort that have been invested in creating separate database resources in such a way as to create, in more or less automatic fashion, a single integrated knowledge base of a scale hitherto unimagined, thus fulfilling an ancient philosophical dream of an encyclopedia comprehending all knowledge within a single system.

The ontological foundation of this Great Encyclopedia would consist of two parts. On the one hand is what is otherwise referred to in the database community as the terminological component (T-box) of the knowledge base. To this would be adjoined the assertional component (or A-box), which is designed to contain the representations of the corresponding facts. Technically the T-Box is that component in a reasoning system that allows, using a logic that is strictly weaker than the first-order predicate calculus, the computation of subsumption relations between terms (relations expressed by sentences like *A rabbit is a mammal*, *A keyboard operator is an employee*, and so on). The A-Box is everything else.

Nicola Guarino, one of the principal figures of this information systems ontology and initiator of the influential FOIS (Formal Ontology and Information Systems) series of meetings, has formulated the matter as follows. An ontology is

> an engineering artefact, constituted by a specific vocabulary used to describe a certain reality, plus a set of explicit assumptions regarding the intended meaning of the vocabulary words. … In the simplest case, an ontology describes a hierarchy of concepts related by subsumption relationships; in more sophisticated cases, suitable axioms are added in order to express other relationships between concepts and to constrain their intended interpretation. (Introduction to Guarino 1998)

The phrase ‘a certain reality’ here signifies in the first place whatever domain one happens to be interested in, whether this be hospital management or car component warehouse inventories. The phrase also however reflects the same sort of tolerant approach to the identity of the target domain of one's ontology as was present earlier in Sowa and in Hayes' second Manifesto. Not only existent objects, but also non-existent objects, would in principle be able to serve as forming 'a certain reality' in the sense Guarino has in mind. 'A certain reality' can include not only pre-existing domains of physics or biology but also domains populated by the products of human actions and conventions, for example in the realms of commerce or law or political administration (including such parts of these realms which involve functioning information systems themselves).

The methods used in the construction of ontologies as conceived by Guarino and others are derived on the one hand from the earlier initiatives in database management systems referred to above. But they also include methods similar to those employed in logical and analytical philosophy, including the axiomatization methods used by Carnap and also the methods used when developing formal semantic theories. They include also the derivation of ontologies from existing taxonomies, databases and dictionaries via the imposition of constraints – for example of terminological consistency and hierarchical well-formedness (Guarino and Welty 2000). Other methods for thd erivation of information systems ontologies have been proposed. Some, for example, havesought to derive such ontologies through the use of statistical clustering techniques applied for example to word corpora for the derivation of lexical hierarchies. (Gangemi *et al.*, 2001) Thus the system WordNet[9] developed at the University of Princeton defines concepts as clusters of terms called synsets. Wordnet itself consists of some 100,000 synsets. A hyponymy relation between synsets is then defined as follows:

[9] http://www.cogsci.princeton.edu/~wn

> A concept represented by the synset {x, x′, …} is said to be a hyponym of the concept represented by the synset {y, y′,…} if native speakers of English accept sentences constructed from such frames as « An x is a kind of y ».

The holy grail is: methods for automatically generating ontologies (REFERENCES). These methods are designed to address the need, given a plurality of standardized vocabularies or data dictionaries relating to given domains, to integrate these automatically – for example by using statistical corpus methods derived from linguistics – in such a way as to make a single database or standardized vocabulary, which is then dubbed an 'ontology'.

Guarino's work is inspired, too, by Aristotle and by other philosophical ontologists in the realist tradition. Like them, but with quite different purposes in mind, heseeks an ontology of reality which would contain theories or specifications of such highly general (domain-independent) categories as: time, space, inherence, instantiation, identity, measure, quantity, functional dependence, process, event, attribute, boundary, etc. The obstacles standing in the way of the extension of such an ontology to the level of categorical details which would be required to solve the real-world problems of database integration are unfortunately prodigious. They are analogous to the task of establishing a common ontology of world history. This would require a neutral and common framework for all descriptions of historical facts, which would require in turn that all events, legal and political systems, rights, beliefs, powers, and so forth, be comprehended within a single, perspicuous list of categories.[10]

[10] The Cyc project is attempting to create a framework of this sort, though we shall see below that there are serious questions as to whether its methodology can realize the goal of providing a framework for genuine integration of data deriving from disparate sources.

Added to this problem of extension are the difficulties which arise at the level of adoption. To be widely accepted an ontology must be neutral as between different data communities, and there is, as experience has shown, a formidable trade-off between this constraint of neutrality and the requirement that an ontology be maximally wide-ranging and expressively powerful – that it should contain canonical definitions for the largest possible number of terms.

One way to address these problems is to emphasize generality at the cost of scope. Another is to abandon the goal of a universal ontology and focus instead on particular domain-specific or regional ontologies, for example, ontologies of geography, or medicine, or ecology.

The relation between the formal and domain-specific ontology is then in some respects analogous to that between pure and applied mathematics. Just as all developed sciences use mathematics, so all domain-specific ontologists should ideally have as their foundation the same robust and widely accepted top-level ontology. The methods used in the development of a top-level ontology are indeed to some degree like those of mathematics in that they involve the study of structures that are shared in common between different application domains. Once theorems have been proved within a given framework of formal ontology, all material ontologies which are specifications of this formal framework will be such that the theorems apply there, too. The relation between formal and material ontology is in this respect analogous to that between pure and applied mathematics.

Examples of Information Systems Ontologies and Their Precursors

KIF

To get some idea of the type of formal theorizing that has developed under the heading of information systems ontology in recent years it will be useful to examine in detail three specific formal theories or frameworks, beginning with KIF (for 'Knowledge Interchange Format'), the work of Mike Genesereth and his colleagues in Stanford. (Genesereth and Fikes 1992). Although not itself conceived for ontological purposes, the language KIF is nonetheless an important milestone in the development of ontology as a solution to the problems of knowledge sharing and knowledge integration. KIF is a language for knowledge representation, a variant of the language of the first-order predicate calculus, motivated by the goal of developing an expressive, flexible, computer- and human-readable medium for exchanging knowledge bases.

The existence of such a language means that each system, provided its syntax is translatable into that of KIF, can internally handle data in its own ways and communicate with its human users in yet other ways, but with the guarantee that the results of the system's operations will be automatically compatible with those of other systems likewise structured in such a way as to be compatible with KIF.

The language has three essential features: 1. a standard set-theoretical semantics (which is, in computer science terms, a descriptive rather than a procedural semantics), 2. logical comprehensiveness – which means that it has all the expressive resources of the first-order predicate calculus, 3. the ability to support the representation of representations, or of knowledge about knowledge.

The semantic side of KIF rests on the technical notion of conceptualism introduced by Genesereth and Nilsson in their (1987). Conceptualizations are defined in terms of sets of objects, properties and relations of certain sorts. Thus for example (and with some crude simplification) the conceptualization involved in a situation where John is kissing Mary might be <{*John*, *Mary*}, {*male_person*, *female_person*, *kissing*}>, where the set {John, Mary}is the universe of discourse of the conceptualization, which is to say a collection of objects hypothesized by the conceptualization to exist in the world, and {*male_person*, *female_person*, *kissing*}is the set of relevant properties and relations of the conceptualization. These properties and relations are then extensionally conceived, which means that they are themselves sets. More precisely, relations and functions are sets of (finite) lists of objects, lists themselves are finite sequences of objects.

A conceptualization is thus an object of a type familiar from standard set-theoretic model theory. Given a conceptualization, the individual terms of KIF denote objects in the associated universe of discourse, the predicate terms of KIF are assigned values from the set of associated properties and relations. Semantics is then relative to conceptualization, and sentences are true or false according to the conceptualization with which one begins.

A universe of discourse is made up of *objects*. Objects themselves are subdivided into *individuals*, on the one hand, and *sets* or *classes* on the other. (KIF includes von Neumann-

Bernays-Gödel set theory as a constituent part.) Each universe of discouse must include at least the following objects:

- complex numbers, which can be seen as couples of real numbers (one real part and one so-called imaginary part). Real numbers are complex numbers whose imaginary part is null. KIF thus includes the real and rational numbers and also the integers. All sorts of arithmetical, trigonometric logarithmic operations can then be defined within the KIF framework;
- an object which is the conventional value of functions for nonsensical combinations of arguments;
- all finite lists of objects and all sets of objects;
- KIF words and expressions (conceived as lists of terms, which may themselves be lists).

It is this last item which allows the representation of representations within the KIF framework by making expressions objects of the universe of discourse and by providing KIF with a truth predicate and tools for manipulating expressions such as operators for quotation and for denoting the denotation of a given term. The analysis in terms of lists means that KIF also has the facility to analyze the internal structure of expressions. Expressions can be referred to via the quotation operator, their properties can be discussed, and expressions may even be quantified over, thus enabling the formulation of axiom schemata. This enables also quantification over parts of expressions. KIF's truth operator can be applied both to quoted and to unquoted expressions under conditions designed to avoid paradoxes. (The central role of lists in KIF also lends it an affinity to programming languages such as LISP.)

KIF's basic universe of objects can be freely extended – for example (as in the case discussed above) by adding the individuals John and Mary – to generate the universe of discourse for a given conceptualization. Given such a universe of discourse, all finite lists of objects in the universe on the one hand are included, together with all sets of objects in the universe on the other.

Examples of KIF Formalism

For notational convenience in what follows I shall avoid the use of lists, with the single exception that the notation '(x, y)' will be used to denote the list of length two of which the first member is x and the second y. In all other cases, whenever a set is defined in the original KIF formalism using lists, I have used a more compact set-theoretic notation. Upper-case letters are used as names for predicates, lower-case letters as names for function.

Logical constants

~: not; $\wedge$: and; $\vee$: or; $\rightarrow$: implies; $\leftrightarrow$: iff; $\exists$: there exists; $\forall$: for all; $\in$: is a member of

Some primitive terms and definitions:

Set(x) := x is a set

List(x) := x is a list

Individual(x) := ~Set(x)

Bounded(x) := x can be a member of a set.

Unbounded(x) := ~Bounded(x)

SimpleSet(x) := $(\text{Set}(x) \wedge \text{Bounded}(x))$

ProperSet(x) := $(\text{Set}(x) \wedge \text{Unbounded}(x))$

Empty(x) := x is the empty set

{x : F(x)} denotes the set of all bounded objects that satisfy F.

for all sets x and y, $x \subseteq y := \forall z\ (z \in x \rightarrow z \in y)$

for all sets x, generalizedUnion(x) := $\{a : \exists t\ (a \in t \wedge t \in x)\}$

for all sets x, y, z, intersection(x, y, … , z) := $\{a : (a \in x \wedge a \in y \wedge \ldots \wedge a \in z)\}$

Some noteworthy axioms and theorems[11]

$\text{Set}(x) \vee \text{Individual}(x)$

$\text{Bounded}(x) \vee \text{Unbounded}(x)$

$x \in y \rightarrow (\text{Bounded}(x) \wedge \text{Set}(y))$

$\text{Relation}(x) \leftrightarrow (\text{Set}(x) \wedge \forall y(y \in x \rightarrow \text{List}(y)))$

Extensionality Property of Sets

$\text{Set}(x) \wedge \text{Set}(y)) \rightarrow ((\forall z\ (z \in x \leftrightarrow z \in y)) \leftrightarrow (x = y))$

Axiom of Regularity

$(Set(x) \wedge \sim Empty(x)) \rightarrow \exists y(y \in x \wedge Empty(intersection(x,y)))$

Axiom-Of-Choice

$\exists s\ (Set(s) \wedge (\forall x\ (x \in s \rightarrow (\exists a\ \exists b\ (x = (a,b)))) \wedge (\forall x\ \forall y\ \forall z\ ((x,y) \in s \wedge ((x,z) \in s \rightarrow (y = z))) \wedge$

$(\forall u\ ((Bounded(u) \wedge \sim(Empty(u))) \rightarrow (\exists v\ (v \in u \wedge (u,v) \in s)))))$

Subset Axiom

$Bounded(x) \rightarrow Bounded(\{y : y \subseteq x\})$

Intersection Axiom

$(Bounded(x) \wedge Set(y)) \rightarrow Bounded(intersection(x, y))$

Union Axiom

$(Bounded(x) \wedge (\forall y\ (y \in x \rightarrow Bounded(x)))) \rightarrow Bounded(generalizedUnion(x))$

Axiom of Infinity

$\exists x\ (Bounded(x) \wedge \sim Empty(x) \wedge (\forall y\ (y \in x \rightarrow (\exists z\ (z \in x \wedge y \subseteq z \wedge \sim(z \subseteq y\ ))))))$

Ontolingua

---

[11] Here and in what follows initial universal quantifiers are taken as understood.

On the basis of KIF, Tom Gruber and his associates at the Stanford Research Institute developed a more serviceable language for ontology representation known as Ontolingua (Gruber 1992, 1995). This is designed to serve as a lingua franca for those involved in building ontologies. Ontolingua is built up on the basis of KIF 3.0, but it has a very distinctive purpose. Where KIF is conceived as an interface between knowledge representation systems, Ontolingua is intended as an interface between ontologies in something like the way in which Esperanto was intended as an interface between the speakers of different languages. It provides an environment and a set of software tools designed to enable heterogeneous ontologies to be brought together on a common platform via translation into a single language.

Ontolingua adds to the original core language of KIF by introducing features which allow the representation of the structural components of an ontology. Above all, it allows an extended treatment of relations (with an almost exclusive focus on binary relations conceived as sets of binary lists), and it introduces the notion of *class*. Classes are formally defined as unary relations, that is as sets of lists of length one, whereby the members of lists are necessarily individuals and are called the instances of the class. In addition Ontolingua includes the notion Individual-Thing (the class of objects that are not sets but can be elements of a set). There is also the class Thing, which is defined as the class encompassing all entities susceptible of being in a class, namely Individual-Things and Simple-Sets. Thing seems to correspond to the class of bounded entities in KIF.

Ontolingua otherwise endorses the general KIF approach, not least in its extensional conception of properties and relations (including functions and classes), all of which are conceived as sets of n-tuples, the latter themselves being conceived as finite lists.

Neither KIF nor Ontolingua embraces the idea of a single shared ontology. There is no suggestion that its authors wanted to incorporate even such notions as time and process, matter and mind. Ontolingua's major goal was rather to collect a large number of distinct, specialized ontologies, and to provide the linguistic resources for moving back and forth between them.

Some sample definitions from Ontologia

For x such that List(x), length(x) denotes the length of the list x

$\text{Holds}(x, y_1, \ldots, y_n) \leftrightarrow (\text{Relation}(x) \wedge (y_1, \ldots, y_n) \in x)$

$\text{Individual-Thing}(x) := {\sim}\text{Set}(x) \wedge \text{Bounded}(x)$

$\text{Class}(x) := \text{Relation}(x) \wedge \forall y\, (y \in x \rightarrow \text{length}(y) = 1))$

$\text{Instance-Of}(x, y) := \text{Class}(y) \wedge \text{Holds}(y, x)$

Note that 'relation' here includes also unary relations (or what would otherwise be referred to as properties in extension).

Description logics

Description logics, also known as terminological logics, reflect the attempt to find a fragment of first-order logic with a maximally high expressive power and yet still a decidable and efficient inference procedure. (See Baader, *et al.*, forthcoming) Description logics are logics designed to realize the maximally effective trade-off between expressive power on the one hand and the ability to serve as engines of efficient reasoning procedures on the other. Description logics are used for example in the querying of (large) knowledge bases; they are the chosen tool underlying the medical ontology GALEN and they underlie much current work under the sponsorship of the US Defense Department within the framework of the DAML-OIL project (to be discussed below). Description logics have received much attention recently in the context of work on the so-called semantic web (Berners Lee, *et al.* 2001), which will consist of formal ontologies represented in some unified description logic.

DAML+OIL arose from EU and US DARPA research programs and is currently undergoing standardization through the W3C WebOnt activity3, to become the Ontology Web Language (OWL). Irrespective of its reasoning capabilities it is becoming a standard language for ontology interchange. As an interchange language it has been designed to encode a wide range of ontologies from taxonomies, frame based ontologies, to ontologies that include logic based concept definitions. This flexibility allows the staged evolution of an ontology within a single representation, greatly simplifying the process.

Within a DAML+OIL ontology each concept is represented as a class. At its simplest, DAML+OIL allows each class to be placed in a taxonomy with the use of the subclass relationship e.g.

class isocitrate dehydrogenase (NAD+) (GO:0004449) 4

subClassOf 'oxidoreductase, acting on the CH-OH group of donors, NAD or NADP as acceptor'

(GO:0016616)

Classes can be further described (or restricted in DAML+OIL terms) by their attributes, specified as property/value pairs, e.g.

class isocitrate dehydrogenase (NAD+) (GO:0004449)

restriction onProperty has_substrate hasClass isocitrate

Both universal and existential quantification can be used to represent such definitions, as 'carbohydrate metabolism is the metabolism of some carbohydrate and only carbohydrate'.

A family of description logics – including `BACK, CLASSIC, CRACK, FLEX, K-REP, KL-ONE, KRIS, LOOM` – has been developed. They are logics which focus on the representation of concepts and on hierarchical classification. Given a sufficiently rich set of concept descriptions a reasoning apparatus employing description logic will generate the corresponding subsumption tree. More generally, the goal of description logics is to derive what are called *is a* relationships (as for example between Dog and Mammal, Person and Entity-with-height, Employee and Entity-With-Social-Security-Number, Bootsie and Dog). Each description logic is formally a subset of first-order logic with certain second-order features. A description logic contains unary predicates, binary predicates, and individual constants. The heart of the description logic is its classes (also 'concepts' in the communities from which description logics evolved). These are defined intensionally in terms of descriptions that specify the properties that the objects of the class must satisfy. Classes should thus be conceived not as collections of elements but rather in a manner dictated by the traditional mode of representation of sets or classes by means of expressions of the form '{x : φx}.'

Cyc and CycL

One of the most influential information systems ontology projects is that of Cyc (Lenat and Guha 1990, http://www.cyc.com), which grew out of an effort initiated by Doug Lenat, one of the pioneers of AI research, to formalize common-sense knowledge in the form of a massive database of axioms covering all things from governments to mothers. ('Cyc' comes from en-cyc-lopedia.)

Cyc started as a research project in the early 80's. In 1995 Lenat created a company, Cycorp, charged with the task of developing further the technology and its applications. Cyc is intended to be able to serve as an encyclopedic repository of all human knowledge. As such it purports to provide a medium for the representation of facts and the inscription of rules about all existing and imaginable things.

Cyc is thus an ontology project in the spirit of Wilkins' *Real Character*, and (as in the case of Wilkins) the resulting ontology has been criticised for its *ad hoc* (which is to say: unprincipled) nature. It takes the form of a gigantic hierarchy, with a topmost node labelled

Thing, beneath which are a series of cross-cutting total partitions including: *Represented Thing* vs. *Internal Machine Thing*, *Individual Object* vs. *Collection*, *Intangible* vs. *Tangible Object* vs. *Composite Tangible and Intangible Object.* Examples of Intangible Objects (*Intangible* means: *has no mass*) are sets and numbers. A *Person* in the Cyc ontology is a Composite Object made up of a Tangible Body and an Intangible Mind.

That Cyc is unprincipled turns on the fact that the partial order from out of which it is built sticks together via logical conjunction too much that is unrelated. Thus the ontology is not divided in systematic fashion into distinct facets or dimensions of reality or into distinct levels of granularity, and nor are its terms stratified into levels of basic terms and defined terms of successively higher-orders of complexity. Cyc has taken some steps towards rectifying this latter defect with the construction of the Upper Cyc Ontology, containing several thousand terms capturing 'the most general concepts of human consensus reality' (http://www.opencyc.org).

Given Cyc's encyclopedic mission, it is of course crucial that CycL should be extendable arbitrarily by the addition of terminology relevant to any domain. Cyc's knowledge-base is compartmentalized into microtheories, but additional microtheories can be introduced at will in order to account for each successive new domain or context. The problem of unprincipledness however means that it is difficult to understand what the constraints might be on the addition of new microtheories or of new terms or axioms.

CycL, the knowledge representation language associated with Cyc, is sometimes presented as a second-order language, sometimes as an (unsorted) first-order language with higher-order capabilities. Thus it also allows quantification over predicates and relations and, more generally, it admits collections, objects of arbitrary order built on the first layer of individuals (collections of individuals, collections of such collections, and so forth). There are relations holding among objects, relations holding among collections in Cyc's technical sense), and also relations holding among CycL sentences themselves. CycL also possesses some of the resources of modal logic and also features derived from natural language such as generalized quantifiers ('every', 'most', 'many').

CycL possesses categories of relations unified by the shared properties of their elements. For instance, '$FunctionalPredicate is the category of relations that are functional in at least one of their argument places. The term 'type' is sometimes used to stand for collection of collections, so that #$RelationshipType is in CycL the collection of collections of relations. There are predicates, such as #$arg2Isa, which are used to stipulate the type of objects that can occupy (in this case) the first and second argument places of a given relation. (These examples are taken from http://www.opencyc.org.)

Cyc itself is a knowledge base written in CycL; thus it is a set of CycL expressions. Cyc is thus referred to as an ontology in the sense that it 'contains' objects, roughly speaking the CycL terms, articulated by axioms, which are CycL sentences (and which themselves can be considered as terms).

The ontology of the Cyc knowledge base is made up of individual objects and set-like objects. The latter are divided further into sets and so-called collections. Individual objects in Cyc include all people, countries, computer programs, and so forth. Cyc also has a rich suite of temporal relations (such as temporallySubsumes, temporallyIntersects, and cotemporality). It has the facility of quantifying over real properties and relations, and this is in addition to quantification over sets of individuals and tuples.

Above the categories of individuals and set-like objects is an all-encompassing category, the collection Thing. Everything is a Thing (in the technical sense of ‘is a’ – written ‘isa’ in Cycl), including Thing itself. Several independent partitions cut through the domain called Things. These allow us to define the Collections of temporal, spatial, intangible Things and moreover of artifacts, agents, organisms and so on. There are also more abstract types of individuals, including objects that are peculiar to Cyc itself, such as Cyc’s own constituent microtheories. The relation isa holds between any Thing T and any Collection of which T is an instance. Apart from isa the most important relation among Things in Cyc is that of generalization (written: ‘genls’), which holds between two Collections C_1 and C_2 when all the instances of C_1 are instances of C_2. (genls is a specialization of the 'subsetOf' relation among sets.)

Individuals are those objects that have parts but no elements. Non-individuals are either sets or collections. The distinction between sets and collections is fundamental corresponds to the two ways in which, in more familiar treatments, sets can be referred to: by extension, that is by enumerating the elements of a set on one hand, and by intension or by providing a criterion for membership in a set on the other. Sets then follow a principle of extensionality that does not hold for collections.

An example illustrating this last point and taken from the documentation runs as follows:

> So in Cyc®, #$USPresidentsNamedRoosevelt, #$USPresidentsWhoWereEach-OthersFifthCousins, and #$TheFirstTwoUSPresidentsWhoseLastNamesBegin-

> WithR would all be different #$Collections, even though they are all comprised of exactly the same elements as the mathematical set {Theodore Roosevelt, Franklin Delano Roosevelt}.

Thus although Collections are sometimes referred to as if they were natural kinds and called 'types', they are in fact associable with arbitary criteria (as the just-mentioned example reveals). Suppose A and B are both presented with the same objects during a certain time interval. There are then two Collections of objects, the Collection of objects that A saw and the Collection of objects that B saw. These collections are distinct, though co-extensional. Put another way, each collection is a generalization (in the sense of 'genls') of the other. Moreover, there is also a third collection, which is the collection of objects that both A and B saw. And if C is some other person who saw none of the mentioned objects, then there is yet another collection consisting of the objects that both A and B saw but which C did not see. Note, too, that although in defining the notion of Collection Cyc appeals to a common feature – i.e., the 'intensional' aspect of a Collection – that instances of a Collection need to share, the constraints which this feature must satisfy are left unexplicated. In the available documentation, it usually takes the form of an appeal to some intuitive understanding of a term found in natural language. The available documentation thus does not allow us to decide whether Collections are to be regarded as natural kinds or as the product of arbitrary composition. This kind of indeterminacy seems however to be a general feature of Cyc, and it seems to go hand with hand with methodological principles according to which the introduction of new terms is governed by informal criteria that appeal to pragmatic considerations and intuition rather than to constraints which are the product of genuine analysis.

Some simple definitions, axioms and theorems of Cyc

Thing(x) := x is a thing

Thing(Thing)

Individual(x) := x is an individual;

Individual(x) $\rightarrow$ Thing(x)

Set(x) := x is a set

Collection(x) := x is a collection

$\sim\exists$x (Set(x) $\wedge$ Collection(x))

SetOrCollection(x) $\leftrightarrow$ (Set(x) $\vee$ Collection(x))

SetOrCollection(x) $\rightarrow$ Thing(x)

Thing(x) $\leftrightarrow$ (SetOrCollection(x) $\vee$ Individual(x))

$\sim\exists$x (SetOrCollection(x) $\wedge$ Individual(x))

$x \in y \rightarrow$ SetOrCollection(y)

$x \subseteq y \leftrightarrow \forall z\ (z \in x \rightarrow z \in y)$

isa(x, y) $\rightarrow x \in y$

isa(x, y) $\rightarrow$ Collection(y)

genls(x, y) $\leftrightarrow$ ($\forall$z (isa(z,x) $\rightarrow$ isa(z,y)))

genls(x, y) $\rightarrow$ (Collection(x) $\wedge$ Collection(y))

$genls(x, y) \rightarrow x \subseteq y$

Axiom of extensionality for sets

$(Set(x) \wedge Set(y)) \rightarrow ((\forall z\ (z \in x \leftrightarrow z \in y)) \leftrightarrow (x = y)))$

Existence axioms

$\exists x\ Thing(x)$

$Thing(x) \rightarrow \exists y\ (SetOrCollection(y) \wedge x \in y)$

$\exists x \exists y\ (Collection(x) \wedge Collection(y) \wedge isa(x,y))$

Problems in information systems ontologies

Some of the problems to which philosophical ontologists might call attention when addressing the literature of information systems ontology are merely terminology. Much of the latter is heavily influenced by a background in work on semantic nets, whose formal structure imposes a division of all entities (or of all that is represented) into two very broad categories:

– concepts, represented by nodes

– relations between concepts, represented by links.

Gradually a third category was introduced, when it became clear that not only concepts would need to be represented but also properties of concepts. In description logic the term ‘role’ was (confusingly or not) selected for this third category. The vocabulary of description logic thus includes in addition to terms for concepts, which are identified as unary predicates and seen as denoting sets of individuals, also terms for roles, which are binary predicates denoting binary relations. Examples of roles are: hasChild, isStudent, isAgent.

There is (from the philosopher's perspective) some cognitive dissonance awakened by the usage in much of the literature of information systems ontology of the terms 'concept' and 'class' as synonyms. This is illustrated in passages such as the following:

> Concepts, also known as classes, are used in a broad sense. They can be abstract or concrete, elementary or composite, real or fict[it]ious. In short, a concept can be anything about which something is said, and, therefore, could also be the description of a task, function, action, strategy, reasoning process, etc. (Corcho and Gomez-Perez 2000, p. 81)

Philosophers and computer and information scientists employ different idiolects, which creates obstacles to mutual understanding. Closer examination of the information systems ontology literature, however, reveals that the terminological running together of 'concept' and 'class' often goes hand in hand with the ontological running together of what are in fact distinct realms (roughly: the theoretical realm and the realm of objects to which our theories are directed). Focusing on concepts means also that each conceptualization, each family of concepts, floats separately from the others. Recall our discussion of the problems facing the construction of a general ontology of world history; this would require a neutral framework for all descriptions of all historical facts, a list of categories which would be accepted by all. We said that Cyc is attempting to create a framework of this sort. As far as one can grasp the methodology of Cyc from published sources, the strategy of Cyc, when faced with disparate standards or systems of laws or names, would be to add corresponding microtheories, more or less at will. Thus (presumably) the description of the historical events surrounding, say, the Louisiana Purchase, would require microtheories of the Continental Napoleonic (codified) legal structures through which the matter was views from the French and Spanish side, and microtheories of the Anglo-Saxon (common) legal structures adopted by the United States. These microtheories, and the corresponding legal vocabularies, would simply be added to the Cyc edified and exist therein side by side. No attempt would be made to build a common framework within which the legal structures embraced by the two systems could be fused or merged or translated into each other. No attempt would be made, in other words, at integration. Ontology from this perspective, simply grows, rather like a spreading vine.

A related problem is the tendency to sacrifice ontological adequacy to software efficiency. This is illustrated by the case of DAML-OIL, with its (current) inability to represent instances of concepts (or members of classes). DAML/OIL is the ontology of the Defense Advanced Research Projects Agency, a combination of the DARPA Agent Markup Language (http://www.daml.org), with the so-called Ontology Inference Layer (OIL: http://www.ontoknowledge.org/oil). This has the goal of exploiting the power and flexibility of XML as a framework for the construction of specialist ontologies. (XML is the universal format for structured documents and data on the world wide web, a matter that is of considerable commercial interest.) The result is dubbed DAML+OIL, and is designed to provide ontology resources distributed on the web.

DAML-OIL, too, is an outgrowth of work on description logic. It is designed to ensure that reasoning can be performed in provably short times relative to competitor systems. It achieves this, however, by accepting severe restrictions on its expressive powers, giving rise to peculiar features such as that for example it has no way of treating individuals. It can only deal with classes/concepts. This is for formal reasons having to do with the desire for computational efficiency, which requires that one keep the algorithms polynomial.

This is how the official DAML/OIL doctrine responds this problem (which, on the face of it, seems very embarrassing):

> Results from research in description and modal logics show that the computational complexity of such logics changes dramatically for the worse when domain-instances are allowed in class definitions … . For this reason, OIL currently does not allow the use of instances in slot-values, or extensional definitions of classes (i.e., class definitions by enumerating the class instances). It is not clear how serious a restriction [this ban on referring to individuals] is for an ontology language, as ontologies should, in general, be independent of specific instantiations – it may be that in many cases, 'individuals' can more correctly be replaced with a primitive class or classes. (Horrocks *et al.*, n. d.)

By 'instance' here is meant: 'contingently existing individual'; by 'specific instantiation' is meant instantiation in a domain of contingently existing individuals. Note the Orwellian aspect of the reasoning: There is this ungainly bunch of things out there that we can't handle with our reasoning system; so we'll just pretend they're not there and rewrite the notion of 'correctly' to suit.

Generally, and in part for reasons of computational efficience rather than ontological adequacy, information systems ontologists have devoted the bulk of their efforts to constructing concept-hierarchies; they have paid much less attention to the question of how the concepts represented within such hierarchies are in fact instantiated in the real world of what happens and is the case. They have neglected, too, the question of the relationship between hierarchies of concepts, on the one hand, and hierarchies of universals or categories (natural kinds, genera and species) on the side of the things themselves.

The Role of Set Theory

One noteworthy feature especially of information systems ontologies built on the basis of KIF is the predominance of set-theory as a tool for purposes of ontology construction. Certainly set theory has great flexibility as a framework for modelling mathematical and other sorts of abstract structures, including many of the structures found in the commercial domains which for a long time served as the primary field of application for information systems ontology. But because sets are themselves abstract (they are entities existing outside the realm of space, time, and causality), a set-theoretical framework must at least be supplemented by other machinery if it is to serve as the basis of a complete ontological theory of the ripe, messy, changing world of concrete objects in which human beings live.

Matters are to some degree improving in this regard, as application domains for ontology become extended to include, for example, the fields of medicine and biology. Thus where early versions of KIF (up to 3.0) assume an extensional treatment of properties and relations, current versions allow distinct properties and relations to have the same extension. Cyc, too, as we have seen , includes radical departures from the extensionalism of set theory. Another example of a non-set-theoretical ontology is provided by the Process Specification Language, PSL, which uses KIF to formulate a simple theory of processes in terms of real-world entities of the following types: activities, activity occurrences, timepoints and objects. (Schlenoff, *et al*. 1999)

Even in such cases, however, set theory is the sole resource that is applied when it comes to the semantics of the languages in question. How, then, are we to understand the ontology expressed in a language like KIF or PSL? Are we to conceive it as a theory of reality, analogous to more traditional philosophical ontologies? Or as a collection of formulas together with a set-theoretical semantics? Some information systems ontologists, such as Patrick Hayes, would answer: Yes to both. This is because they conceive set-theoretical language as used in the semantic context as involving no ontological claim about the nature of the world itself or about the things in it. As Hayes puts it, the application of set-theoretical language

> places no restrictions on the *nature* of the things in the universe; it does not require them to be set-theoretical constructions, any more than an engineer using differential equations to describe a steel surface is somehow committed to saying that the surface is made of fluxions. (Personal communication)

Thus, to use Quine's criterion, the provision of a set-theoretical semantics for an ontology does not in itself commit one to having sets in one's ontology

Hayes himself is interested not in special set-theoretical models for ontological formulas but rather in those models that correspond to reality.[12] Indeed in both versions of his "Manifesto" he goes further and insists, as we saw, that a model for a first-order system can *be* a piece of reality.

But then in either case, if Hayes is right, our models must at least be compared with the corresponding reality in order to ensure the necessary degree of adequacy, Then, however, ontological investigations *of this reality* in something like the traditional philosophical sense become required in any case. And at this point the question then arises whether the detour through semantics is needed, or helpful, at all: why not just move directly to the task of establishing whether the ontological theory is itself adequate to reality.

In any case, the job of the ontologist and the job of the semanticist are independent of each other, and the job of the ontologist must involve the use of tools in addition to those of set theory, since not everything in reality is a set.

Mereology, the formal theory of part and whole (Simons 1987), is a weaker instrument than set theory for purposes of mathematical modelling. It turns out, however, that much of what ontology requires in the form of supplements or alternatives to set theory can most naturally be provided within a mereological framework. A further advantage of mereology turns on the fact that mereology can be applied in the ontological investigation of a given domain of objects even before we have any knowledge of any putative basic level of atoms from out of which this domain might be constructed.

Mereology allows ontologists to begin their investigations with complex wholes as it were on

[12] Only in certain technical contexts might special non-standard models need to be used in order to show up limitations or errors in existing formalizations.

the middle level of reality, and to work upwards and downwards from these, as corresponding coarser and finer grained theories become available. In this way the framework of mereology and also that of mereotopology – the qualitative theory of boundaries, contact and separation on a mereological basis (Smith 1996) – can support the simultaneous development of ontological theories on different levels of granularity, including theories of that mesoscopic reality that is presented to human beings in their everyday perceptions and actions. As Hayes puts it,

> most people associated with database technology, computational ontology work, data modelling and so on are vividly and acutely aware of the need to maintain correspondences with reality; often indeed more acutely than most theoreticians, being sensitive to issues such as the need for time-stamping and the need to reason about plausibility of conflicting information sources; often, millions of dollars or many people's lives may turn on these models being accurate. One does not lightly deny a correspondence with reality when trying to make machines reason about anthrax biopsies (Personal communication)

Certainly it is true that cows, wombats, tropes and subatomic particles, can all be talked of using set-theoretical language. The question is whether the relation between such entities and the other types of entities in the universe can be adequately captured exclusively by means of the resources available in set theory.

Semantic Conceptions of Ontology

As we saw in our discussion of the revised "Naïve Physics Manifesto" above, however, even

Hayes himself occasionally expresses a view of ontology according to which it would relate merely to ‘alternative possible worlds’. This latter tendency is more most clearly manifest in the work of Gruber for whom, as we shall see, semantic investigations take the place of ontology itself conceived as a theory of reality. Thus Gruber, in effect, conceives semantics not as a mere logical tool, but rather as supplying the subject-matter of ontology itself, so that the construction of set-theoretic models becomes an end in itself. An ontology is, in the end, a set-theoretical model, called by Gruber the ‘specification of a conceptualization’, where the latter is itself, as we have seen, a special kind of set-theoretical object.

Divorcing Ontology from Reality – Ontology in Knowledge Representation

Gruber’s work exemplifies a move made by many information systems ontologists away from the principle captured in the Ontologist’s Credo and towards a conception of ontology as a discipline concerned not with reality itself but with well-behaved reality surrogates. The strongest pressure in this direction has been felt in the field of knowledge representation, currently one of the most important areas of ontological research in the information systems field. Many thinkers in the knowledge representation field have come to hold, with Gruber (1995), that: ‘For AI systems what “exists” is that which can be represented’ within whatever formal system one is currently using.

The debate over the correct conception of information systems ontologies would then be an almost exact parallel of the philosophers’ debate over the correct conception of realism and idealism. (Franklin, forthcoming) Briefly, the idealist argues that we can know reality only through our concepts (or language, or ideas, or theories). Hence, he infers, we cannot know reality as it is in itself. This reasoning is on display for example here:

> The difficulty is that whatever we observe, or, more generously, whatever we interact with, is certainly not independent of us. This is the problem of reciprocity. Moreover, whatever information we retrieve from such interaction is … information about interacted-with-things. This is the problem of contamination. How then, faced with reciprocity and contamination, can one get entities both independent and objective? Clearly, the realist has no direct access to his World. (Fine 1986, p. 151)

The flaw in the reasoning is exposed by David Stove when he points out that it has the same logical form as:

> We can eat oysters only insofar as they are brought under the physiological and chemical conditions which are the presuppositions of the possibility of being eaten.
>
> Therefore,
>
> We cannot eat oysters as they are in themselves. (Stove 1991, pp. 151, 161)

Similarly, now, the information systems ontologist asserts:

> We can represent entities in our system only insofar as they are referred to by means of the canonical vocabulary at our disposal.
>
> Therefore,
>
> We cannot represent in our system entities as they are in themselves.
>
> Therefore,

> Ontology must deal not with reality, but rather only with our ‘conceptualizations’.

Information systems ontologies in the sense of Gruber are thus not oriented around the world of objects at all. Rather, they are focused on our knowledge and beliefs – on the concepts or languages we use when we talk about this world. It is in this light that we are to interpret passages such as the following:

> an ontology is a description (like a formal specification of a program) of the concepts and relationships that can exist for an agent or a community of agents. This … is certainly a different sense of the word than its use in philosophy. (Gruber, n.d.)

Good and Bad Conceptualizations

There are a number of problems with the definition of ontology as a specification of a conceptualization. One is this: There are, surely, different specifications – in Hebrew, or in KIF, or in CycL, or in first-order predicate logic – all of which might very well describe what we can intuitively recognize as the same ontology. Thus ontology has nothing to do with the means of specification.

A deeper reason has to do with the confusion of two tasks: the study of reality, on the one hand (which philosophers, at least, would insist is the properly ontological task) and the study of *our concepts of reality* on the other. What would be wrong with a view of ontology – ontology of the sort that is required for information systems purposes – as a study of human concepts or beliefs (or as a matter of ‘knowledge representation’ or ‘conceptual modeling’)? [13] We can get a first

[13] Already by 1986 (see e.g. Alexander *et al.*), the usage of ‘ontology’ as meaning just ‘conceptual model’ had become entrenched in the world of knowledge representation research.

hint of an answer to this question if we recall our treatment of folk biology above. There, it is clear, we find both good and bad conceptualizations – the former reflecting what actually exists in reality, the latter resting on ontological error; the former illustrated by a conceptualization of types of slime mold, the latter by a conceptualization of types of evil spirits.

Given that the notion of conceptualization embraces not only good but also bad (objectless) conceptualizations, it follows that only certain ontologies as specifications of conceptualizations can be *true of* some corresponding domain of reality, while others are such that there is simply no corresponding domain of reality. Information systems ontology becomes also for this reason a pragmatic enterprise. It starts with conceptualizations, and goes from there to the description of corresponding domains of objects (often confusingly referred to as ‘concepts’), but the latter are nothing more than models, surrogate created worlds, devised with specific practical purposes in mind. What is most important, now, is that all of the mentioned surrogate created worlds are treated by the ontological engineer as being on an equal footing. In a typical case the universe of discourse will be specified by the client or customer, and for the purposes of the ontological engineer *the customer is always right* (it is the customer in each case who defines his own specific world of surrogate objects). It is for this reason that the ontological engineer aims not for truth, but rather, merely, for adequacy to whatever is the pertinent application domain *as defined by the client*. The main focus is on reusability of application domain knowledge in such a way as to accelerate the development of appropriate software systems in each new application context. The goal is not truth relative to some independently existing domain of reality – which is after all often hard to achieve – but merely (at best) truth *relative to some conceptualisation*.

Uses of Ontology in Information Science

The project of building one single ontology, even one single top-level ontology, which would be at the same time non-trivial and also readily adopted by a broad population of different information systems communities, is sustained by Cyc, but it has otherwise largely been abandoned. The reasons for this can be summarized as follows. The task of ontology-building proved much more difficult than had initially been anticipated (the difficulties being at least in part identical to those with which philosophical ontologists have been grappling for some 2000 years). The information systems world itself, on the other hand, is very often subject to the short time horizons of the commercial environment. This means that the requirements placed on information systems themselves change at a rapid rate, so that theoretically grounded work on ontologies conceived as modules for translating between information systems has been unable to keep pace.

Work in ontology in the information systems world continues to flourish, however, and the principal reason for this lies in the fact that its focus on classification and on constraints on allowable taxonomies and definitions has proved useful in ways not foreseen by its initial progenitors (Guarino and Welty 2000). Automation requires a higher degree of accuracy in the description of its procedures, and ontology is a mechanism for helping to achieve this. The attempt to develop terminological standards, which means the provision of explicit specifications of the meanings of terms loses nothing of its urgency in application domains such as medicine or air traffic control, even when the original goal of a common ontology embracing all such domains has been set to one side.

Ontology also goes by other names, so that the building of ontologies has much in common with work on what are still called ‘conceptual schemes’ in database design, or on ‘models of application domains’ in software engineering, or on ‘class models’ in object-oriented software design. The designers of large databases are increasingly using ontological methods as part of their effort to impose constraints on data in such a way that bodies of data derived from different sources will be rendered mutually compatible from the start. Ontological methods are used also in the formalization of standards at the level of metadata, where the goal is to provide in systematic fashion information about the data with which one deals, for example as concerns its quality, origin, nature and mode of access.

Ontological methods may have implications also for the writing of better software. If you have gone to the trouble of constructing an ontology for purposes of integrating existing information systems, this ontology can itself be used as a basis for writing software that can replace those old systems, with anticipated gains in efficiency.

Ontological methods have been applied also to the problems of extracting information for example from large libraries of medical or scientific literature, or to the problems of navigation on the Internet, for example, in the already mentioned work on the so-called semantic web, which uses ontology as a tool for taming the immense diversity of sources from which Internet content is derived. Here even a small dose of ontological regimentation may provide significant benefits to both producers and consumers of on-line information.

Ontological methods have been applied also in the domain of natural language translation, where ontologies continue to prove useful, for example as aids to parsing and disambiguation. Sergei Nirenburg and Victor Raskin (2001) have developed a methodology for what they call 'ontological semantics', which seeks to use ontological methods as the basis for a solution to the problem of automated natural language processing, whereby ontology – conceived as a 'constructed world model' – would provide the framework for unifying the needed knowledge modules within a comprehensive system. Thus it uses ontology 'as the central resource for extracting and representing meaning of natural language texts, reasoning about knowledge derived from texts as well as generating natural language texts based on representations of their meaning.' (*op. cit.*)

Efforts continue to be made to use ontology to support business enterprises (Uschold *et al.* 1998, Obrst, *et al.* 2001). Consider a large international banking corporation with subsidiaries in different countries throughout the world. The corporation seeks to integrate the information systems within its separate parts in order to make them intercommunicable. Here again a common ontology is needed in order to provide a shared framework of communication, and even here, within the relatively restricted environment of a single enterprise, the provision of such a common ontology may be no easy task, in virtue of the fact that objects in the realms of finance, credit, securities, collateral and so on are structured and partitioned in different ways in different cultures.

Commercial ontology is not merely a matter of facilitating communication via terminology standardization. It must deal also with the problems which arise in virtue of the existence of conflicting sets of standards in the domains of objects to which terminologies refer. Consider for example the domain of financial statements. These may be prepared either under the US GAAP standard or under the IASC standards which is used in Europe and many other countries. Under the two standards, cost items are often allocated to different revenue and expenditure categories depending on the tax laws and accounting rules of the countries involved. Information systems ontologists have thus far not been able to develop an algorithm for the automatic conversion of income statements and balance sheets prepared on the basis of the two sets of standards. And why not? Because the presuppositions for the construction of such an algorithm simply cannot be found by looking at the two conceptualizations side-by-side-, as it were, and hoping that some way can be found to fuse the two together within a single ontology on the basis of their immanent properties as conceptualizations or on the basis of the associated terminology or reasoning systems. Now will semantic investigations, to the extent that these consist in finding set-theoretical models of the systems in question in the customary manner, i.e. by working from the terminologies outwards towards the models, do the trick. To fuse two systems of the given sort it is necessary to establish how the two relate to some *tertium quid* – the reality itself, of commercial transactions, etc., and to see how the two systems partition this same reality in different ways. This means that one must do ontology in something like the traditional philosophical way – in this case the ontology of assets, debts, net worth and so forth of business firms – before the standard methods of information systems ontology can be applied.

Medical Ontology

A similar problem arises, too, in the field of medical informatics. Here information systems ontologists have worked extensively in providing aids to information retrieval, processing of patient records, hospital management, clinical trial suppose and the like. Significantly, longer time horizons can be assumed to prevail than in strictly commercial environments. Moreover, over-hasty software fixes may here be of greater significance than in other fields. In medicine, too, different nomenclatures (standardized, controlled vocabularies) and classification systems have been developed to assist in the coding and retrieval of knowledge gained through research. Such systems face difficulties in virtue of the fact that the subject-matter of medicine is vastly more complicated than the domain covered by, say, the information system of a large bank. One may thus anticipate that some of the theoretically most important advances in information systems ontology in the future will be made in the area of medical informatics.

Some indication of the problems which need to be confronted in the medical ontology domain can be gained by looking at three alternative medical terminology systems, each of which is often treated as representing some sort of ontology of the medical domain. (See Burgun and Bodenreider 2001)

First is GALEN, for Generalised Architecture for Languages, Encyclopaedias and Nomenclatures in Medicine. GALEN includes, for example, an ontology of medical procedures. Thus the surgical process of open extraction of an adrenal gland neoplastic lesion is represented as follows:

SurgicalDeed which

  isCharacterisedBy (performance whichG

    isEnactmentOf ((Excising which playsClinicalRole SurgicalRole) whichG <

actsSpecificallyOn (NeoplasticLesion whichG

hasSpecificLocation AdrenalGland)

hasSpecificSubprocess (SurgicalApproaching whichG

hasSurgicalOpenClosedness (SurgicalOpenClosedness

whichG hasAbsoluteState surgicallyOpen))>))

Second is the Unified Medical Language System or UMLS, which is maintained by the National Library of Medicine in Washington DC. UMLS comprehends some 800,000 concepts of the biomedical domain arranged in some 134 semantic types, where a concept is defined as a clusters of terms (derived, for example, from different natural languages). UMLS is a fusion of some 5o source vocabularies from which some ten million interconcept relationships have been inherited. The parent-child hierarchy which is the backbone of UMLS is then defined as follows:

> A concept represented by the cluster{x, x′, …}is said to be a child of the concept represented by the cluster {y, y′,…} if any of the source terminologies shows a hierarchical relationship between x and y.

The potentiality for conflict here, given that the UMLS source vocabularies were developed independently of each other (and are of varying quality) is clear.

Finally we can mention SNOMED, or Systematized Nomenclature of Medicine, which is maintained by the College of American Pathologists and is designed as ‘a common reference point for comparison and aggregation of data throughout the entire healthcare process’. SNOMED has been applied especially to the project of developing electronic patient record systems. It comprehends some 121,000 concepts and 340,000 interconcept relationships.

Let us now see how each of these terminology systems localizes *blood* in its concept hierarchy. For the sake of comparison we note that blood in Cyc is categorized as a mixture:

Blood genls Mixture genls TangibleThing

Mixture isa ExistingStuffType

Interestingly, blood in WordNet is categorized as one of the four bodily humors, alongside phlegm, and yellow and black bile.

In GALEN blood is a Soft Tissue (which is subcateogy of Substance Tissue, which is a subcategory of GeneralizedSubstance).

In UMLS – and here we see the effects of constructing UMLS additively, by simply fusing together pre-existing source vocabularies, blood is a Body Fluid *and* a Soft Tissue *and* a Body Substance. Tissue in turn is classified in UMLS as a Fully Formed Anatomical Structure.

In SNOMED, on the other hand, blood is a Body Fluid, which is a Liquid Substance, which is a Substance Characterized by Physical State.

Examination of the hierarchies used especially in UMLS and SNOMED reveals that they are marked by what Guarino (1999) has referred to as *isa* overloading; that is to say, hierarchies are generated in ways which involve at one and the same time readings of *isa* as meaning: identity, categorical inclusion, is part of, or one of several other relations. There is also a confusion of description and characterization. Type-tokens abound as also do mass-count confusions. Hierarchies contain cycles, which must be eliminated by hand, and much of the work devoted to maintaining UMLS and SNOMED thus consists in finding ad hoc solutions to problems which would never have arisen had a robust top-level ontology been established from the start. Then, however, it would have been necessary for those involved in the construction of the given systems to take seriously what we have referred to as the Ontologist's Credo.

A robust terminology system, in medicine or elsewhere, cannot be created simply through the fusion or colligation of existing vocabularies or micro-theories. And the problem of resolving the differences and terminological inconsistencies between distinct systems of financial, or medical, documentation, cannot be solved by examining the separate systems themselves, as purely syntactic instruments. Nor is it enough to build abstract set-theoretical models of the usual semantic sort. Rather, one needs to look at what the terms involved in each system mean in relation to the corresponding concrete objects and processes in reality.

The Closed World Assumption

Clearly, it is for practical reasons not possible to include all the facts pertaining to the objects in a given application domain into a database. Some selection must be made and this, rightly, takes place on pragmatic grounds. Suppose we have a database that includes facts pertaining to object o, and the user asks whether o is F. The programmer has to decide what sort of answer will be shown to the user if the fact that o is F is not recorded in the database. In some systems the answer will be something like 'perhaps'. In some domains, however, it makes sense for the database programmer to rely on what is called the closed world assumption, and then the answer will be 'no'. Here the programmer is taking advantage of a simplifying assumption to the effect that a formula that is not true in the database is thereby false. This closed world assumption 'is based on the idea that the program contains all the positive information about the objects in the domain' (Shepardson 1988, pp. 26-27). Models of systems built on the basis of the closed world assumption are of course much simpler targets from a mathematical and programming point of view than any real-world counterparts. If we attempt to construct an ontology of the ripe, messy exterior reality of ever-changing flesh-and-blood objects, then the closed world assumption can no longer be taken to be valid; yet without this assumption the programmer's job becomes much harder.

The closed world assumption means not only that (to quote Gruber once again) only those entities exist which are represented in the system, but also that such entities can possess only those properties which are represented in the system.[14] It is as if Hamlet, whose hair (we shall suppose) is not mentioned in Shakespeare's play, would be not merely neither bald nor non-bald, but would somehow *have no properties at all* as far as hair is concerned. What this means, however, is that the objects represented in the system (for example people in a database) are not real objects – the objects of flesh and blood we find all around us – at all. Rather, they are denatured surrogates, possessing only a finite number of properties (sex, date of birth, social security number, marital status, employment status, and the like), and being otherwise entirely indeterminate with regard to all those properties and dimensions with which the system is not concerned. Objects of the flesh-and-blood sort can in this way be replaced by tidy tuples. Set-theoretical structures replace reality itself.

These problems are of special significance in the field of medicine. Let us suppose, for

[14] Reiter (1984) formulates the closed world assumption in relation to relational databases as follows: the only possible instances of a relation are those implied by the database. He distinguishes in addition what he calls the domain closure assumption, to the effect that: the individuals occurring in the database are all and only the existing individuals.

example, that there is no mention of diabetes in a patient record within a given database. What should be the answer to the query: 'Does the patient have diabetes?' Here, clearly, the assumption that all relevant information about the domain of discourse is contained in the database cannot be sustained. (Rector and Rogers 2002)

Perhaps we can resolve our puzzle as to the degree to which information systems ontologists are indeed concerned to provide theories which are true of reality – as Patrick Hayes would claim – by drawing on a distinction made by Andrew Frank (1997) between two types of information systems ontology. On the one hand there are ontologies – like Ontek's PACIS and IFOMIS's BFO – which were built to represent some pre-existing domain of reality. Such ontologies must reflect the properties of the objects within its domain in such a way that there obtain substantial and systematic correlations between reality and the ontology itself. On the other hand there are administrative information systems, where (as Frank sees it) there is no reality other than the one created through the system itself. The system is thus by definition, correct.

Consider the world of banking. Here (let us assume) the only operations possible are the ones built into the program and there is no attempt to model connections to an independently existing external reality. In an on-line dealing system a deal is only a deal if it takes place within the system. The world of deals itself exists within the system itself. For many purposes it may be entirely satisfactory to identify a deal with: an event of this and this sort inside the on-line system. But consider: The definition of a client of a bank is 'a person listed in the database of bank clients'. Here an identification of the given sort seems much less tempting. This suggests that Frank's remark has primary relevance only for administrative systems of a very special kind, and not not for those administrative systems which record events or facts obtaining elsewhere. It has application for those operational systems that *do things* of legal or administrative significance within the domain of the system itself. If, now, information systems ontology has (post-Hayes) grown up in an environment where it is precisely the products of such operational systems that have been the primary targets of ontological research, then it is clear why many of those involved might have become accustomed to the idea that ontology is concerned with with some pre-existing reality but rather with entities created by the ontology systems themselves.

There is a higher degree of arbitrariness in the creation of entities (as for example in the case of GAAP or IASC) in the operational systems realm than is to be found in those realms of pre-existing reality customarily dealt with by philosophical ontologists. Moreover, those who have the power to effect fiat demarcations – for example the committees who determine what will count as goods or services in financial statements -- are often themselves muddle-headed or prone to revision or lax in error-checking in ways which render impossible the construction of robust and consistent ontological hierarchies of the resulting domains of administrative objects. This fact, too, lends credence to the anti-theoretical pragmatic approach by which much information systems ontology is marked. The project of theoretical ontology has to a degree been sabotaged by the concentration on the *prescriptions* of others. In producing an ontology in such circumstances one has a choice between accepting the often only opaquely specified word of the imposing authority or attempting via hit or miss to capture a vague specification crisply with the inevitable danger of mischaractarization and obsolescence. Parallel remarks can be made, too, in relation to the construction of ontologies on the basis of linguistic corpora (Guarino 1999). Here, too, there is too often too much that is muddled in the source vocabularies (adding what people are willing to accept is an additive process too). This results in equal amount of muddle in the ontologies generated. No account is taken of the fact that, as we have seen, many of the existing standardized vocabularies themselves embody systematic errors or massive ontological unclarities (some of them generated by earlier sacrifices in due care and attention made for reasons of coding convenience). However carefully such errors and unclarities are merged, they are predestined to yield an end-result that is of dubious merit.

Social and legal systems are, from the perspective of traditional philosophical ontology, something real. Information systems, too, may be part of reality in the same sense. This is so, too, when the information system is doing something of legal or administrative or commercial import (which raises interesting problems, such as how we are to construct models of such systems).

Some approaches presuppose that the building of an information system has two stages. The first stage is to produce a model that describes the world that the system is to record; the second stage is to write a program that would describe this world. This is, to say the least, a very inadequate account of what is involved in system construction when the system and its processes will themselves be part of the world – the world of real reality – to which the system is designed to relate. The tempting idea that an ontology based on some variant of the closed world assumption might work well in the administrative domains while a more traditional ontology is required in domains treated of by natural science, thus needs to be resisted. Administrative and commercial systems are no less firmly integrated into the world of messy flesh-and-blood reality than are the objects described by physics or biology.

Why Information Systems Ontology Failed

Given this background we can point to one further reason why the project of a common ontology which would be accepted by many different information communities in many different domains has failed. Not all conceptualizations are equal. What the customer says is not always true; indeed it is not always sufficiently coherent to be even in the market for being true. Bad conceptualizations abound (rooted in error, myth-making, Irish fairy-tales, astrological prophecy, or in hype, bad linguistics, over-tolerant dictionaries, or antiquated information systems based on dubious foundations). Such conceptualisations deal *only* with created (pseudo-)domains, and not with any transcendent reality beyond.

Consider, against this background, the project of developing a top-level ontology, a common ontological backbone constructed in the purely additive manner by fusing or combining existing conceptualizations or micro-theories constructed elsewhere for any one of a variety of non-ontological purposes. This project now begins to appear rather like the attempt to find some highest common denominator that would be shared in common by a plurality of true and false theories. Seen in this light, the principal reason for the failure of attempts to construct top-level ontologies lies precisely in the fact that these attempts were made on the basis of a methodology which treated all conceptualizations on an equal footing and thus overlooked the degree to which the different conceptualizations have served as inputs to ontology are likely to be not only of wildly differing quality but also mutually inconsistent.

What can Information Scientists learn from Philosophical Ontologists?

Van Benthem has defined artificial intelligence as the continuation of logic by other means. Information systems ontology can be defined, similarly, as the continuation of traditional ontology by other problems. For many of the problems faced by information systems ontologists are analogues of problems dealt with by philosophers in the 2000 year history of traditional ontology or metaphysics. These are not only problems pertaining to the definition of identity, the problem of universals and particular, but also the problem of defining individuals as bundles of concepts

How does Sowa deal with individuals?

Their move echoes the arguments of Fodor (1980) in favor of the adoption by cognitive psychologists of the research program of 'methodological solipsism', according to which only immanentistically conceived mental states and processes can properly figure within the domain of a truly scientific psychology.

It is as if Hamlet, whose hair (we shall suppose) is not mentioned in Shakespeare's play, would be not merely neither bald nor non-bald, but would somehow have no properties at all as far as hair is concerned. Compare our treatment of the closed world assumption with what Ingarden (1973) on the 'loci of indeterminacy' within the stratum of represented objects of a literary work. Ingarden uses the fact that fictional objects are always defined partially – and thus exist with certain loci of indeterminacy – where real objects are determinate down to the lowest possible differences in every dimension of their being, as an argument to the effect that idealistic metaphysical positions which see the world as being constructed in a manner analogous to the construction of fictional worlds must be wrong.

Consider again: The definition of a client of a bank is 'a person listed in the database of bank clients'. There is, incidentally, a counterpart of this sort of approach in doctrines such as conventionalism and operationalism in the philosophy of science, which hold that scientific terms are interpretable only within a specified theoretical context.

Some ontological engineers have recognized that they can improve their models by drawing on the results of the philosophical work in ontology carried out over the last 2000 years.[15] This does not in every case mean that they are ready to abandon their pragmatic perspective. Rather, they see it as useful to employ a wider repertoire of ontological theories and frameworks they are willing to be maximally opportunistic in their selection of resources for purposes of ontology-construction. Guarino and his collaborators use standard philosophical analyses of notions such as identity, part-whole relations, ontological dependence, set-theoretical subsumption and the like in order to expose inconsistencies in ontologies proposed by others, and they go on from there to derive the meta-level constraints which all ontologies must satisfy if they are to avoid inconsistencies of the sorts exposed.

Given what was said above, however, it appears that information ontologists may have sound

[15] In addition to the work of Guarino and his co-authors referred to already above, see: Degen, Heller and Herre 2001, Milton 2000, Milton and Kazmierczak 1998.

pragmatic reasons to take the philosopher ontologist's traditional concern for truth more seriously still. For the very abandonment of the focus on mere conceptualisations and on conceptualisation-generated object-surrogates may itself have positive pragmatic consequences. – not least in terms of greater stability. This applies even in the world of administrative objects – for example in relation to the GAAP/IASC integration problems– where the ontologist is working in a theoretical context where, as we saw, he must move back and forth between distinct conceptualisations, and where he can find the means to link the two together only by looking at their common objects of reference in the real world of actual financial transactions.

On the other hand however the ontological project as thus conceived will face considerable difficulties of its own. Traditional ontology is a difficult business.

To put the point another way: it is precisely because good conceptualizations are transparent to reality that they have a reasonable chance of being integrated together in robust fashion into a single unitary ontological system. The fact that the real world itself plays a significant role in ensuring the unifiability of our separate ontologies thus implies that, if we are to accept a conceptualization-based methodology as one stepping stone towards the construction of adequate ontologies, then we must abandon the attitude of tolerance towards both good and bad conceptualizations. For it is this very tolerance which is fated to undermine the project of ontology itself.

Of course to zero in on good conceptualizations is no easy matter. There is no Geiger-counter-like device which can be used for automatically detecting truth. Rather, we have to rely at any give stage on our best endeavors – which means concentrating above all on the work of natural scientists – and proceed in careful, critical and fallibilistic fashion from there, hoping to move gradually closer to the truth via an incremental process of theory construction, criticism, testing, and amendment, and also through the consideration of theories directed towards the same domain of reality but on different levels of granularity. It will be necessary also, however, to look beyond natural science, for ontology must comprehend also objects (such as societies, institutions and concrete and abstract artefacts) existing at levels of granularity distinct from those which readily lend themselves to natural-scientific inquiry. Our best candidates for good conceptualizations will however remain close to those of the natural sciences – so that we are, in a sense, brought back to Quine, for whom the job of the ontologist is identified precisely with the task of establishing the ontological commitments of scientists, and of scientists alone.

Ontology in information science must in any case find ways to counteract existing tendencies to treat all conceptualizations on an equal footing. Thus it should not, as has been customary, take as its starting point the surrogate worlds which have been constructed inside existing software models or data dictionaries (or inside people's heads, or inside the models of set-theoretic semantics). Rather, as we have seen, it should address reality itself, drawing on the wealth of scientific descriptions of the different dimensions of this reality, with the goal of establishing, not only how these various dimensions of objects, relations, processes and properties are linked together, but also how they are related to the manifest image of common sense.

The holy grain of database fusion and automatic ontology generation assumes additivity of micro-theories and conceptualizations – assumes they are all equal. No automatic solution.

Only way to achieve success is to look at reality, to do ontology (honest toil) in just the way traditional philosophers did ontology.

What Can Philosophers Learn from Information Systems Ontologists?

Developments in modal, temporal and dynamic logics as also in linear, substructural and paraconsistent logics have demonstrated the degree to which advances in computer science can yield benefits in logic – benefits not only of a strictly technical nature, but also sometimes of wider philosophical significance. Something similar can be true, I suggest, in relation to the developments in ontological engineering referred to above. The example of the successes and failures of information systems ontologists can first of all help to encourage existing tendencies in philosophical ontology (nowadays often grouped together under the heading ‘analytic metaphysics’) towards opening up new domains of investigation, for example the domain of social institutions (Mulligan 1987, Searle 1995), of patterns (Johansson 1998), of artefacts (Dipert 1993, Simons and Dement 1996), of dependence and instantiation (Mertz 1996), of holes (Casati and Varzi 1994), and parts (Simons 1987). Secondly, it can shed new light on the many existing contributions to ontology, from Aristotle to Goclenius and beyond (Burkhardt and Smith 1991), whose significance was for a long time neglected by philosophers in the shadow of Kant and other enemies of metaphysics.[16] Thirdly, if philosophical ontology can properly be conceived as a kind of generalized chemistry, then information systems can help to fill one important gap in ontology as it has been practiced thus far, which lies in the absence of any analogue of chemical experimentation. For one can, as C. S. Peirce remarked (1933, 4.530), ‘make exact experiments upon uniform diagrams’. The new tools of ontological engineering might help us to realize Peirce’s vision of a time when operations upon diagrams will ‘take the place of the experiments upon real things that one performs in chemical and physical research.’ The problem of devising ontological theories adequate to the needs of information science provides the analogue of experimental test in a field which has thus far been amenable only to the sort of testing that flows from considerations of the logical and argumentative qualities of a theory.

Finally, the lessons drawn from information systems ontology can support the efforts of those philosophers who have concerned themselves not only with the development of ontological *theories*, but also – in a field sometimes called ‘applied ontology’ (Koepsell 1999) – with the *application* of such theories in domains such as law, or commerce, or medicine. The tools of philosophical ontology have been applied to solve practical problems, for example concerning the nature of intellectual property or concerning the classification of the human foetus at different stages of its development. Collaboration with information systems ontologists can support such ventures in a variety of ways, first of all because the results achieved in specific application-domains can provide stimulation for philosophers,[17] but also – and not least importantly – because information systems ontology is itself an enormous new field of practical application that is crying out to be explored by the methods of rigorous philosophy.

Literature

---

[16] See also Ashenhurst 1996.
[17] See for example Visser and Bench-Capon 1998, Navratil 1998, Grassl 1999.

Alexander, J. H., Freiling, M. J., Shulman, S. J., Staley, J. L., Rehfus, S., and Messick, S. L. 1986 “Knowledge Level Engineering: Ontological Analysis”, *Proceedings of AAAI–86*. *Proceedings of the 5th National Conference on Artificial Intelligence*, Los Altos: Morgan Kaufmann Publishers, 963–968.

Ashenhurst, Robert L. 1996 “Ontological Aspects of Information Modeling”, *Minds and Machines*, 6, 287–394.

Baader, Franz, Calvanese, Diego, McGuinness, Deborah, Nardi, Daniele and Patel Schneider, Peter (forthcoming) *The Description Logic Handbook: Theory, Implementation and Applications*, Cambridge: Cambridge University Press:

Berners-Lee, Tim, Hendler, James and Lassila, Ora 2001 “The Semantic Web”, *Scientific American*, May 2001.

Bigelow, John & Pargetter, Robert 1990 *Science and Necessity*. Cambridge: Cambridge University Press.

Bittner, Thomas 2001 “The Qualitative Structure of Built Environments,” *Fundamenta Informaticae*, 46, 97–126.

Bittner, Thomas and Smith, Barry 2001 “A Taxonomy of Granular Partitions”, forthcoming in Daniel Montello (ed.), *Spatial Information Theory, Proceedings of COSIT '2001*, Santa Barbara, September 2001, Berlin/New York: Springer.

Borges, J. 1981 “The Analytical Language of John Wilkins”, *Borges: A Reader*, New York: Dutton, E. R. Monegal and A. Reid (eds.), 141–43.

Brentano, Franz 1981 *The Theory of Categories*, The Hague/Boston/London: Martinus Nijhoff.

Burgun, Anita and Bodenreider, Olivier 2001 “Aspects of the taxonomic relation in the biomedical domain”, in C. Welty and B. Smith (eds.), *Formal Ontology in Information Systems. Proceedings of the Second International Conference (FOIS '01), October 17-19, Ogunquit, Maine*, New York: ACM Press, 222-233.

Burkhardt, Hans and Smith, Barry (eds.) 1991 *Handbook of Metaphysics and Ontology*, 2 vols., Munich/Philadelphia/Vienna: Philosophia.

Carnap, Rudolf 1954 *Introduction to Symbolic Logic and Its Applications*, New York: Dover.

Casati, Roberto and Varzi, Achille C. 1994 *Holes and Other Superficialities*, Cambridge, Mass.: MIT Press.

Chisholm, Roderick M. 1996 *A Realistic Theory of Categories: An Essay on Ontology*, Cambridge: Cambridge University Press.

Cohn, A. G. 1996 “Calculi for Qualitative Spatial Reasoning”, in *Artificial Intelligence and Symbolic Mathematical Computation* (Lecture Notes in Computer Science 1138), J. Calmet, J. A. Campbell and J. Pfalzgraf (eds.), New York/Berlin: Springer Verlag, 124–143.

Cohn, A. G., Bennett, B., Gooday, J. and Gotts, N. M. 1997 “Qualitative Spatial Representation and Reasoning with the Reason Connection Calculus”, *Geoinformatica*, 1, 1–44.

Corcho, Oscar and Gomez-Perez, Asuncion 2000 “A Roadmap to Ontology Specification Languages”, in Rose Dieng and Olivier Corby (eds.), *Knowledge Engineering and Knowledge Management. Methods, Models and Tools*, Berlin: Springer, 80–96.

Degen, W., Heller, B. and Herre, H. 2001 “Contributions to the Ontological Foundations of Knowledge Modelling”, Institute for Informatics, University of Leipzig, Technical Report 02

(2001).

Degen, W., Heller, B., Herre, H. and Smith, B. 2001 “GOL: Towards an Axiomatized Upper-Level Ontology”, ”, in C. Welty and B. Smith (eds.), *Formal Ontology in Information Systems. Proceedings of the Second International Conference (FOIS '01), October 17-19, Ogunquit, Maine*, New York: ACM Press.

Dement, Charles W. and DeWitt, Stephen 1995 *Corporate Data Integration Tools*, available through the Defense Technical Information Center (http://www.dtic.mil/), WL-TR-95-8011.

Dement, Charles W. and DeWitt, Stephen 1996 *An Adaptable Enterprise Integration Platform*, available through the Defense Technical Information Center (http://www.dtic.mil/), WL-TR-96-8008.

Dement, Charles W. Mairet, Charles E., Dewitt, Stephen E. and Slusser, Robert W. 2001 *Mereos: Structure Definition Management Automation (MEREOS Program Final Technical Report)*, available through the Defense Technical Information Center (http://www.dtic.mil/), Q3 2001.

Dijkstra, Edgar W. 1989 “On the Cruelty of Really Teaching Computing Science”, *Communications of the Association for Computing Machinery*, 32: 12, 1398–1402.

Dipert, Randall R. 1993 *Artefacts, Art Works and Agency*, Philadelphia: Temple University Press.

Fine, Arthur 1986 “Unnatural attitude: realist and instrumentalist attachments to science”, *Mind* 95, 149-79.

Fodor, J. A. 1980 “Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology”, *Behavioural and Brain Sciences*, 3, 63–73.

Forguson, Lynd 1989 *Common Sense*, London/New York: Routledge.

Frank, A.U. 1997 “Spatial Ontology: A Geographical Point of View”, in: O. Stock (ed.), *Spatial and Temporal Reasoning*, Dordrecht: Kluwer Academic Publishers, 135–153.

Franklin, J. (forthcoming) “Stove’s discovery of the worst argument in the world”, *Philosophy*.

Freksa, Christian 1992 “Temporal Reasoning Based on Semi-Intervals”, *Artificial Intelligence*, 54 (1-2), 199–227.

Galton, Antony 2001 *Qualitative Spatial Change*, Oxford: Oxford University Press.

Gangemi, Aldo and Alessandro Oltramari and Nicola Guarino 2001 “Conceptual Analysis of Lexical Taxonomies: The Case of WordNet Top-Level”, in: Chris Welty and Barry Smith (eds.), *Proceedings of the International Conference on Formal Ontology in Information Systems FOIS 2001* (Ogunquit, Maine, October 17-19, 2001), New York: ACM Press, 285-296.

Genesereth, Michael R. and Fikes, Richard E. 1992 *KIF Version 3.0*. Reference Manual, Technical Report Logic-92-1, Stanford: Stanford University.

Genesereth, Michael R. and Nilsson, L. 1987 *Logical Foundation of Artificial Intelligence*, Los Altos, California: Morgan Kaufmann.

Gibson, J. J. 1979 *The Ecological Approach to Visual Perception*, Boston: Houghton-Mifflin, repr. 1986, Hillsdale, NJ: Lawrence Erlbaum.

Grassl, Wolfgang 1999 “The Reality of Brands: Towards an Ontology of Marketing”, in Koepsell (ed.), 313–359. Griffiths, Paul E. 1997 *What Emotions Really Are. The Problem of Psychological Categories*, Chicago: University of Chicago Press.

Gruber, T. R. 1992 *Ontolingua: A Mechanism to Support Portable Ontologies*, StanfordÖ Knowledge Systems Laboratory.

Gruber, T. R. 1993 “A Translation Approach to Portable Ontology Specifications”, *Knowledge Acquisition*, 5, 199–220.

Gruber, T. R. 1995 “Toward Principles for the Design of Ontologies Used for Knowledge Sharing”, *International Journal of Human and Computer Studies*, 43(5/6), 907–928.

Gruber, T. R. n.d. “What is an Ontology?”, http://www-ksl.stanford.edu/kst/what-is-an-ontology.html.

Guarino, Nicola 1995 “Formal Ontology, Conceptual Analysis and Knowledge Representation”, *International Journal of Human-Computer Studies*, 43, 625–640.

Guarino, N. 1997 “Some Organizing Principles for a Unified Top-Level Ontology”, LADSEB-CNR Internal Report 02/97, March 1997.

Guarino Nicola 1999 “The Role of Identity Conditions in Ontology Design”, in C. Freksa and D. M. Mark (eds.), *Spatial Information Theory: Cognitive and Computational Foundations of Geographic Information Science*, Berlin/New York: Springer Verlag, 221–234.

Guarino, Nicola (ed.) 1998 *Formal Ontology in Information Systems*, Amsterdam, Berlin, Oxford: IOS Press. Tokyo, Washington, DC: IOS Press (Frontiers in Artificial Intelligence and Applications).

Guarino, N., Masolo, C., and Vetere, G. 1999 “OntoSeek: Using Large Linguistic Ontologies for Accessing On-Line Yellow Pages and Product Catalogs”, *IEEE Intelligent Systems*, May/June, 70–80.

Guarino, N. and Welty, C. 2000 “A Formal Ontology of Properties”, in R. Dieng and O. Corby (eds.), *Knowledge Engineering and Knowledge Management: Methods, Models and Tools. 12th*

International Conference (EKAW 2000), Berlin/New York: Springer: 97–112.

Guarino, Nicola and Welty, C. 2002 “Evaluating Ontological Decisions with OntoClean”, *Communications of the ACM*, 45(2):61-65. New York: ACM Press.

Hafner, Carole D., and Fridman Noy, Natalya 1997 “The State of the Art in Ontology Design: A Survey and Comparative Review”, *AI Magazine*, Fall 1997, 53–74.

Hayes, Patrick, J. 1979 “The Naïve Physics Manifesto”, in D. Michie (ed.), *Expert Systems in the Micro-Electronic Age*, Edinburgh: Edinburgh University Press, 242–70, cited according to the reprint in Margaret Boden (ed.), *The Philosophy of Artificial Intelligence*, Oxford: Oxford University Press, 1990, 171–205.

Hayes, Patrick J. 1985 “The Second Naive Physics Manifesto”, in Hobbs and Moore (eds.), 1–36.

Hayes, Patrick J. 1985a “Naïve Physics I: Ontology for Liquids”, ”, in Hobbs and Moore (eds.), 71–108.

Hobbs, J. R. and Moore, R. C. (eds.) 1985 *Formal Theories of the Common-Sense World*, Norwood: Ablex.

Horrocks, I. *et al.* (n. d.) *The Ontology Inference Layer OIL*, Technical Report, http://www.ontoknowledge.org/oil/TR/oil.long.html.

Ingarden, Roman 1964 *Time and Modes of Being*, translated by Helen R. Michejda, Springfield, Ill.: Charles Thomas.

Ingarden, Roman 1973 *The Literary Work of Art: An Investigation on the Borderlines of Ontology, Logic, and Theory of Literature*, Evanston: Northwestern University Press, 1973.

Jardine, Donald A. (ed.) 1997 *The ANSI/SPARC DBMS Model: Proceedings of the second SHARE*

Working Conference on Data Base Management Systems, Montreal, Canada, April 26–30, 1976, Amsterdam/New York : North-Holland.

Johansson, Ingvar 1989 *Ontological Investigations. An Inquiry into the Categories of Nature, Man and Society*, New York and London: Routledge.

Johansson, Ingvar 1998 "Pattern as an Ontological Category", in Guarino (ed.), 86–94.

Keil, Frank 1979 *Semantic and Conceptual Development: An Ontological Perspective*, Cambridge, MA: Harvard University Press.

Keil, Frank 1989 *Concepts, Kinds, and Cognitive Development*, Cambridge, MA: MIT Press.

Keil, F.C. 1994 "Explanation Based Constraints on the Acquisition of Word Meaning", in L. Gleitman and B. Landau (eds.), *The Acquisition of the Lexicon*, Cambridge, MA: MIT Press, 1994, 169-196.

Kifer, Michael, Lausen, Georg and Wu, Georg 1995 "Logical Foundations of Object-Oriented and Frame-Based Languages," *Journal of the Association of Computing Machinery*, 42(4): 741–843.

Kim, W. 1990 *Introduction to Object-Oriented Databases*, Cambridge, MA: MIT Press.

Koepsell, David R. 2000 *The Ontology of Cyberspace: Law, Philosophy, and the Future of Intellectual Property*, Chicago: Open Court.

Koepsell, David R. (ed.) 1999 *Proceedings of the Buffalo Symposium on Applied Ontology in the Social Sciences* (*The American Journal of Economics and Sociology*, 58: 2).

Lenat, D. B. and Guha, R. V. 1990. *Building Large Knowledge-Based Systems: Representation and Inference in the Cyc Project*, Reading, MA: Addison-Wesley.

Mäki, Uskali 2001 "Explanatory Unification: Double and Doubtful", *Philosophy of the Social*

Sciences, 31: 4, 488–506.

McCarthy, John 1963 *Situations, Actions and Causal Laws*. Technical Report, Stanford University, reprinted in Marvin Minsky, editor. Semantic information processing. MIT Press, 1968.

McCarthy, J. 1980 "Circumscription – A Form of Non-Monotonic Reasoning", *Artificial Intelligence*, 5: 13, 27–39.

Mealy, G. H. 1967 "Another Look at Data," *Proceedings of the Fall Joint Computer Conference, November 14–16, Anaheim, California* (AFIPS Conference Proceedings, Volume 31), Washington, DC: Thompson Books, London: Academic Press, 525–534.

Medin, Douglas L. and Atran, Scott (eds.) 1999 *Folkbiology*, Cambridge, MA: MIT Press.

Mertz, D. W. 1996 *Moderate Realism and Its Logic*, New Haven, CN: Yale University Press.

Milton, S.K. 2000 *An Ontological Comparison and Evaluation of Data Modelling Frameworks*, Dissertation, The University of Tasmania (School of Information Systems).

Milton, S. K., Kazmierczak, E. 1998 "Comparing Data Modelling Frameworks Using Chisholm's Ontology", *Proceedings of 1998 European Conference on Information Systems*, *Aix-en-Provence*, *France*, *June 4-6, 1998,* Volume I, 260-272.

Mulligan, Kevin 1987 "Promisings and Other Social Acts: Their Constituents and Structure", in Kevin Mulligan (ed.) *Speech Act and Sachverhalt. Reinach and the Foundations of Realist Phenomenology*, Dordrecht/Boston/Lancaster: D. Reidel, 29–90.

Navratil, G. 1998 *An Object-Oriented Approach to a Model of a Cadaster*, M.Sc., Department of Geoinformation, Technical University of Vienna, Vienna.

Nirenburg, Sergei and Raskin, Victor 2001 "Ontological semantics, formal ontology, and ambiguity", in C. Welty and B. Smith (eds.), *Formal Ontology in Information Systems. Proceedings of the Second International Conference (FOIS '01), October 17-19, Ogunquit, Maine*, New York: ACM Press, 151-161.

Obrst, Leo, Wray, Robert E. and Liu, Howard 2001 "Ontological engineering for B2B E-commerce", in C. Welty and B. Smith (eds.), *Formal Ontology in Information Systems. Proceedings of the Second International Conference (FOIS '01), October 17-19, Ogunquit, Maine*, New York: ACM Press, 117-126.

Omnès, Roland 1999 *Understanding Quantum Mechanics*, Princeton: Princeton University Press.

Peirce, C. S. 1933 *Collected Papers*, Cambridge, Mass.: Harvard University Press.

Quine, W. V. O. 1953 "On What There Is", as reprinted in *From a Logical Point of View*, New York: Harper & Row.

Quine, W. V. O. 1969 *Ontological relativity, and Other Essays*, New York: Columbia University Press.

Rector, A. L. and Rogers, J. E. 2002 "Ontological Issues in GALEN", Part I: "Principles", Part II: "High Level Schemata", in *Methods of Information in Medicine*.

Reiter, R. 1984 "Towards a Logical Reconstruction of Relational Database Theory", in M. L. Brodie, J. Mylopoulos, and J. W. Schmidt (eds.), *On Conceptual Modelling*, Berlin/New York: Springer-Verlag, 191–233.

Schlenoff, C., Ciocoiu, M., Libes, D., Gruninger, M. 1999 "Process Specification Language: Results of the First Pilot Implementation", *Proceedings of the International Mechanical Engineering Congress and Exposition*, Nashville, Tennessee, November, 1999.

Schubert, Rainer 2001 “Bones, Holes, and Scales – On the Need for a Spatial Ontology for Anatomy”, Manuscript, University of Hamburg.

Schulze-Kremer, Steffen 1997 “Adding Semantics to Genome Databases: Towards an Ontology for Molecular Biology”, in *Proceedings of the Fifth International Conference on Intelligent Systems for Molecular Biology*, T. Gaasterland, et al. (eds.), Halkidiki, Greece.

Searle, John R. 1995 *The Construction of Social Reality*, New York: Free Press, 1995.

Sellars, W. F. 1963 "Philosophy and the Scientific Image of Man", chapter 1 of *Science, Perception and Reality*, London: Routledge and Kegan Paul.

Shepardson, J.C. 1987 “Negation in Logic Programming”, in J. Minker (ed.), *Foundations of Deductive Databases and Logic Programming*, Los Altos, California: Morgan Kaufmann, 19–88.

Simons, Peter M. 1987 *Parts. An Essay in Ontology*, Oxford: Clarendon Press.

Simons, Peter M. and Dement, Charles W. 1996 “Aspects of the Mereology of Artefacts”, in: Roberto Poli and Peter Simons, ed., *Formal Ontology*. Dordrecht: Kluwer, 1996, 255–276.

Smith, Barry 1996 “Mereotopology: A Theory of Parts and Boundaries”, *Data and Knowledge Engineering*, 20, 287–303.

Smith, Barry 2000 “Objects and Their Environments: From Aristotle to Ecological Psychology “, in Andrew Frank (ed.), *The Life and Motion of Socioeconomic Units*, London: Taylor and Francis, 79–97.

Smith, Barry 2001 “True Grid”, *Conference on Spatial Information Theory* (COSIT ’01), Berlin/New York: Springer Verlag, forthcoming.

Smith, Barry and Brogaard, Berit (in press) “Quantum Mereotopology”, *Annals of Mathematics and Artificial Intelligience*, forthcoming.

Smith, Barry and David M. Mark 2001 “Geographic Categories: An Ontological Investigation”, *International Journal of Geographic Information Science*, 15: 7.

Smith, Barry and Zaibert, Leonardo 2001 “The Metaphysics of Real Estate”, *Topoi*, 20: 2.

Sowa, John F. 1984 *Conceptual Structures. Information Processing in Mind and Machine*, Reading, MA: Addison Wesley.

Spelke, Elizabeth S. 1990 “Principles of Object Perception”, *Cognitive Science*, 14, 29–56.

Stove, David C. 1991 *The Plato Cult and Other Philosophical Follies*, Oxford: Blackwell.

Uschold, M., King, M., Moralee, S., and Zorgios, Y. 1998 “The Enterprise Ontology”, *The Knowledge Engineering Review*, 13.

Visser, P. R. S. and Bench-Capon, T. J. M. 1998 “A Comparison of Four Ontologies for the Design of Legal Knowledge Systems”, *Artificial Intelligence and Law*, 6, 27–57.

Wilkins, John 1668 *An Essay toward a Real Character and Philosophical Language*, London: The Royal Society.

Xu, F., & Carey, S. 1996 “Infants’ metaphysics: The case of numerical identity”, *Cognitive Psychology*, 30(2), 111-153.

View publication stats