# Top-Level Ontology: The Problem with Naturalism

Simon K MILTON[1] and Barry SMITH[2,3]
[1]*Department of Information Systems, The University of Melbourne, Victoria 3010, Australia*
[2]*Institute for Formal Ontology and Medical Information Science, (IFOMIS), Härtelstrasse 16, 04107 Leipzig, Germany*
[3]*Department of Philosophy, University at Buffalo, Buffalo, NY 14260, USA*

**Abstract**. Much work in ontology has been based on ideas derived explicitly from philosophy. In particular there are two philosophy-based high-level ontologies which have been used to analyse tools and techniques for designing information systems, both deriving from the tradition of realism – one from naturalism (Bunge), the other from Aristotelian common-sense realism (Chisholm). Does it matter which underlying philosophy is used in such cases? We use data modelling as a vehicle to explore this question but what we have to say relates quite generally to information systems analysis and design. We find that naturalism cannot account for the multiple perspectives common in real-world situations and we accordingly recommend that common-sense realism be adopted as underlying philosophy.

## Introduction

Ontology in information systems and computer science comes in many flavours, but, for present purposes we can distinguish two categories of ontology: First, we have highly general ontologies – sometimes called 'top-level' ontologies, based on ideas taken over from philosophy and used to analyse information systems and their design tools and methods [1-6]. Second, we have ontologies restricted to specific domains such as medicine, accounting, or geography. Top-level ontologies are used to provide the theoretical underpinnings for representation and modelling in information systems in ways designed to bring benefits in the form of more reliable applications, better quality data-creation, and also help in error-detection. Domain ontologies are aimed at facilitating automated data-sharing between complex information systems in specific fields and also at supporting the automatic construction and population of ontologies developed in these fields.

The two types of research are clearly related. For example, a top-level ontology can be used to help establish what categories of entities there are in a specific domain. Similarly, each specific domain-ontology must use some high-level categories of the sort captured in a top-level ontology. The more standard the high-level categories employed, the more readily the domain-specific ontology can be aligned with other domain-specific ontologies that share similar commitments. Good top-level ontology can thus support data integration and information systems interoperability [7]. At the same time the efforts directed towards constructing domain-specific ontologies can lead to improvements in top-level ontologies and in the philosophical theories which underlie them.

Unfortunately, no consensus has as yet emerged which might serve to unify the various activities directed towards ontology, although there are promising signs. Two top-level ontological theories have been applied in analytical studies of modelling tools and techniques in information systems. Both of these theories come from the philosophical tradition of realism, one from naturalism (Bunge), and the other from Aristotelian common-sense realism (Chisholm). Both have been found to be supported by most modelling tools [1, 4]. In this paper we examine the philosophies of these two ontologies and show why it matters which underlying philosophy is selected in work on information systems. We use data modelling as a vehicle to explore the issues, but the same points could be made in relation to other information systems topics, including ontology engineering, systems requirements engineering, and process modelling.

## 1 Ontology

An ontology describes what is fundamental in the totality of what exists. It defines the most general categories to which we need to refer in constructing a description of reality, and it tells us how these categories are related. It thus describes reality without specifying the particulars of any category. It must further be able to be used to describe reality at any point in time (either now, or in the future, or in the past). This account is in keeping with definitions of ontology found in the standard philosophical literature, for example:

- The study of being in so far as this is shared in common by all entities, both material and immaterial. [Ontology] deals with the most general properties of beings in all their different varieties [8]
- Ontology, understood as a branch of metaphysics, is the science of being in general, embracing such issues as the nature of existence and the categorial structure of reality. … Different systems of ontology propose alternative categorial schemes. A categorial scheme typically exhibits a hierarchical structure, with ‘being’ or ‘entity’ as the topmost category, embracing everything that exists. [9]

Philosophers also construct ontologies for domains such as medicine, geography or accountancy, with categories that are sufficient to support the representation of the entities existing in the corresponding domains, for example, in medicine, by means of categories such as: disease, gene, and cell. Where such endeavours are principally driven by philosophical theory we can categories the domain-ontologies which result as ‘theory focused’

Philosophers are interested in ontological *theories*, but the disciplines of artificial intelligence (AI) and computer science [10, 11] use ontology in a highly pragmatic way:

- in its most prevalent use in AI, an ontology refers to an engineering artifact, 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. [12]

Such ontologies are generally built on the bases of bottom-up methodology by considering specific application domains such as banking, taxation, or the French revolution. These often use a mixture of ontological tools and principles selected for pragmatic purposes and at times they rest on foundations which are, from the philosophical point of view, highly problematic. Where groups building domain-specific ontologies collaborate in building top-level ontologies to unify their separate approaches, the results are often confined to mere syntactic regimentation, illustrated by ventures such as XML or OWL [13]. The Semantic Web represents an effort to unify a number of ontologies in different domains written by different groups within a single scheme.

## 2 The Significance of Top-level Ontology

Both top-level and domain-specific ontologies can be categorised as being motivated either by theory or by practice. (See Table 1) In the former case appeal is made to one or other philosophical school in establishing high-level ontological commitments. In the latter case there occurs a pragmatic merging of a number of ontologies built by different groups working in specific domains.

Table 1: Ontologies in Computer Science and Information Systems

| | **Top-Level** | **Domain-Specific** |
|---|---|---|
| **Theory Focus** | DOLCE, Basic Formal Ontology (BFO), Bunge-Wand-Weber, Chisholm | BFO based Medical and Geospatial Ontologies; DOLCE- and ONIONs-based Medicine, Fishery, and Legal ontologies |
| **Pragmatic Focus** | Standard Upper Merged Ontology (SUMO) | Unified Medical Language System (UMLS) |

All top-level ontologies assume a philosophical position in the form of commitments to certain categories of entities and relations. Such positions restrict the contexts in which the ontology may apply. Thus a top-level ontology which does not have room for the category of processes/occurrents will be inapplicable in a domain such as physics or biology that is marked by continuous change. Moreover, if the theoretical position implicit in an applications ontology falls short of being fully coherent then this may lead to absurdities such as 'New York City is a concept' [14] or 'an order can only be a shared property of a customer and stock' [15]. On the face of it, the problems posed by such statements may seem trivial; however, the virtues of consistency and coherence of high-level categories associated with a good general ontology can be exploited to support high-powered reasoning and error-checking across the information expressed; they can thereby help to guarantee the long-term usefulness of associated domain-specific ontologies.

In information systems the issue of a coherent top-level ontology is of particular interest because computer systems run by organizations may involve several domains simultaneously. An information system ontology designed to support all activities taking place in a hospital, for example, must have elements that belong to the domain of medicine together with terms that relate to other domains such as billing, financial planning, building management, stock ordering, and so forth. This requires that those responsible have some idea of how the various domain-specific ontologies relate to each other, and this means that they need to appeal to something like a top-level ontology to ensure coherence. Further problems are created by the fact that the relationship between a top-level ontology and given domain-specific ontologies is not always clear-cut.

In business information systems one often needs to construct a data model to reflect a variety of stakeholder views; each such view refers to reality from the perspective of a different group and in ways that are grounded in practical commercial (legal, administrative) concerns. The differing views of the different groups need somehow to be integrated into a unified framework. For example, the accounts department is interested in one range of entities such as stock or bill-of-materials, whereas the engineering department is interested in components or sheer-stress-value. The two groups slice reality in different ways, and the treatment of each corresponding sub-ontology must somehow be made consistent.

## 3 Conflicting Perspectives: Bunge vs. Chisholm

In light of the above, we argue that effective assessment and development of tools for information systems analysis and design must use an appropriate top-level ontology. The Bunge-Wand-Weber ontology, based on theories advanced by the philosopher Mario

Bunge [16, 17], and Roderick Chisholm's ontology [18], are the two examples of top-level ontologies we shall address in this paper.

Both of these top-level ontologies employ very general terms and both present a highly general view of the whole of reality. The Bunge ontology employs terms like: thing, property, state, state transition, system. The Chisholm ontology has: individual, attribute, class, relation, substance, boundary, state, set. Analytical studies of data modelling tools and techniques have applied these top-level ontologies.

## 4 The Story of 'H'

Data modelling in the information systems world is the process by which different perspectives on reality are reconciled. Its goal is the construction of effective socio-technical systems to support human activity, especially in large organizations. Data modellers construct models comprehending those kinds of data that an organization believes it needs to know about in order to ensure effective functioning. And what organizations need to know about most of all are people.

The way in which different perspectives are reconciled has traditionally been viewed as relatively straightforward. However, there are cases where perspectives cannot be reconciled so easily at all. We will describe one such case, which is designed to compare the relative strengths of data modelling rooted in the science-based naturalistic ontology of Bunge, on the one hand, and in common-sense realism on the other.

An Australian man named Horace residing in Sydney has for some time been living under the name 'Helen' while preparing for a sex change operation. Horace is employed by a company in Australia that holds data about Horace's sex. Horace identifies as female but has not yet had the operation to reassign his bodily organs. The database entry shows Horace's sex as <u>female</u> because, according to the Australian standard [19], a person with transgender issues who self-identifies as female must be classified as female. The other categories in this standard are:

- <u>male</u>, which includes both those who self-identify as male and those who have had female-to-male reassignment procedures;
- <u>intersex or indeterminate</u>, a category that is reserved for those who have specific genetic conditions involving the possession of reproductive organs or sex chromosomes that are not exclusively male or female and for those whose sex has not yet been determined for whatever reason;
- <u>not stated/inadequately described</u>, which is used for data transfer only;

Helen after the operation is clearly the same person as Horace in his earlier life; but she is also clearly changed in specific ways. Even before the operation she was recognised as female by the company's database. From here on we will refer to him/her as 'H'.

The company has a subsidiary in the USA, and shortly before her sex change operation, H moves from Sydney to San Francisco, where her data is added to the company's Californian database. In the USA there are different categories for sex [20] compared with those found in Australia. In addition to male and female there is: <u>not known and not specified</u> (coded in the same way as the Australian <u>not stated/inadequately described</u>). The transgender classification of <u>intersex or indeterminate</u> is not included in the US standard.

Suppose that in the US, H is classified according to her birth certificate and genitalia and is therefore registered in the database as male. Is the person identified in the Australian database as <Helen, female> the same person as the person known to the US database as <Horace, male>? And how do both of these relate to H after her sex reassignment surgery? The company is constructing an enterprise-wide information system and needs to reconcile its two databases. There is work to be undertaken at two levels. First, how can the two data standards be reconciled in general? Second, how is the resulting system to deal with the problems raised by people like H? It may be that the US entry is not linked with the Australian entry, in which case the implication will be drawn from the new combined database that there are two distinct persons working for the company, where only one exists in reality.

The situation is already difficult enough; examining the reality of H a little closer, however, we see still more complexity. Genetically, H has one X and one Y chromosome and is therefore male; but she identifies as female, being classified as male in the US and as female in Australia. But, suppose that part of the company is engaged in clinical studies of its employees and is interested in H's sex from a genetic perspective. It will discover that it is impossible to determine this reliably from the data found in either of the company's databases, thus foiling H's attempt in enrolling in company-sponsored medical trial investigating the genetic basis of certain sex-relevant diseases. Later, H falls in love with another employee, Mike, whom he wishes to marry (after the operation) by taking advantage of his classification as female in order to satisfy the marriage requirements in Australia. How will H and Mike fare under the company's information systems?

Problems like these are not isolated. As many as one in 500 women suffer from *androgen insensitivity syndrome* where genetic sexual status is unclear so that classification must be effected according to the outward physical appearance of the sufferer. Genetic tests would reveal in such cases a male configuration of chromosomes but female genitalia in addition to other transgender complications. Many of those affected by this syndrome are unaware of their ambiguous ontological status.

Which philosophical framework, now, can help us to maintain an ontology that can serve as basis for resolving problems such as those described above? In order to answer this question, we must explore more fully the two philosophical traditions at issue.

## 5 Naturalism and Common-sense Realism: The Two Ontologies

Mario Bunge developed his ontology in the late nineteenseventies motivated by a dislike of traditional philosophical ontology due what he saw as its less than scientific approach. He held that only through natural science can we know what there is in reality, and that the appropriate tool to describe reality is that of set theory. His ontology constitutes a significant contribution to the philosophy of science. However, it assumes that only through natural science can we examine and document what exists in the world, and that all complex entities are to be analysed in set-theoretic terms. This has two implications. First, when paradigm shifts in science occur, then a Bunge-style ontology must be rewritten entirely. While radical shifts of this sort are rare, local changes in understanding – as evidenced by the various theories of the atom popular in the late eighteenth century and developments in genomics and protenomics in our own day – are much more frequent.

Yair Wand and Ron Weber have used Bunge's work as the basis for an ontology to support information systems. They have noted that Bunge's ontology is "oriented towards the physical world and therefore does not provide for human perceptions and social context" [21]. But they nevertheless praise Bunge's ontology because it is consistent with the naturalistic thesis according to which "the best methods of inquiry in the social sciences or philosophy are… those of the natural sciences." Naturalism insists not merely that natural science be used "in recognising what is real" [8] but that "our ontology is constrained by the result that all physical bodies are composed entirely of particles" [8].

Unfortunately, a position along these lines is methodologically difficult to defend in the realm of information systems. This is first of all because here human beings and their intentional states play an important role and the latter are in some circumstances "attributable to individuals only relative to an observer" something which is "inconsistent with the objectivity of the methods of natural science" [8, 22]. Second, it is difficult to do justice in the Bunge framework to those aspects of reality which are a matter of social construction, since the latter – for example entities such as profits, debts, taxes – despite being real are not amenable to natural scientific inquiry [23].

Roderick Chisholm wrote the definitive version of his complete ontology in his *A Realistic Theory of Categories*, published in 1996, where he seeks to show how realism can go hand in hand with an ontology of endurant persons within a single coherent theory. His ontology is based principally on Aristotelian realism and on the principle of the primacy of the intentional, which means that it is intentional states, not language, which determine the ontological commitments of human beings. Chisholm takes account in particular of perception in determining an individual's ontological commitments. Importantly, his ontology is compatible with the thesis that there is one world to which all individuals relate but at the same time many perspectives towards that world which people may legitimately adopt.

Chisholm belongs to a common-sense realist stream of philosophy that is more amenable to the reality of intentionality and of social phenomena. This philosophy holds that we know most, if not all, of those things which ordinary people know. Its theory of the common-sense world falls short of the striving for quantitative exactitude that is characteristic of the natural sciences. Rather it is focused on the mesoscopic reality of those entities relevant to human action and perception. It is delineated by our beliefs about what happens in mesoscopic reality not in every case but rather only *in most cases and most of the time* [24]. In this it is in keeping, too, with the methodological views of Aristotle.

The common-sense-based approach does not dismiss the view of the world that is rooted in physics. Rather it holds that it is possible to develop an ontology that is able to accommodate both the world of physics (and other parts and dimensions of the natural world) and the world as it is apprehended by common sense – effectively by developing a suitably sophisticated theory of granular perspectives [25]. In contrast to naturalistic ontologies such as Bunge's, common-sense ontology thus has the further advantage that it does not need to be rewritten as a whole in the wake of paradigm shifts in one or other branch of natural science.

Common-sense realism holds that there is only one world towards which our cognition relates on different levels of granularity, and that the objects in this world, on almost all levels of granularity, exist independently of our cognitive relations to them. It concedes that our natural cognitive experiences are in many cases unable to be verified; but it points out that common sense is itself aware that errors are often made. "The thesis that there is only one world towards which natural cognition relates must thus be understood as being compatible with the thesis that there are many different ways in which the world can appear to human subjects in different sorts of circumstances" [24].

## 6 Which Realism for Social Reality?

Comparing the two ontological positions (naturalism and common-sense realism) which is better able to help us in relation to the conundrum of H?.

### *6.1 Bunge on H*

Bunge's naturalism would deal with the case in hand by asserting that there is an intrinsic property possessed by H, the quality of sex as determined by natural scientific inquiry. Further, he would assert that we cannot have more than one distinct perspective on H's sex in the sense that, of the three perspectives reflected by the Sydney and San Francisco databases and H's genetic sex, at least two must be wrong. But what is this intrinsic property? Is it H's genetic sexual identity? Is it the sex dictated by H's genitalia? Is it H's (psychological) self-identification?

H's genetic sex is male, and he is classified as male as far as the company and the law in California are concerned; but his preference in life is to self-identify as female and he is classified as such in Australia. How, in the naturalistic ontology, can these different perspectives be reconciled? How could such an ontology permit perspectives on H's legal sexual identity to vary according to the jurisdiction in which H is being handled? H is, after all, the same object, wherever he is legally registered. Clearly a naturalist perspective is not adequate to serve as a basis for the construction of information systems flexible enough to keep track of cases like H.

### *6.2 Common-Sense Realism on H*

Common-sense realism, in contrast, holds that we need to recognise different perspectives (legal, social, biological) that run in tandem with each other and capture different aspects of reality, and more specifically of the endurant substances in reality that change their qualities over time. The fine-grained genetic/scientific view of Horace and Helen will maintain at all stages the value for H as <u>male</u> in the sense of having an X and a Y chromosome. The database administrators in Sydney will have H changing from male to female and accordingly register her as a transgendered individual. Database administrators in San Francisco will classify H as male throughout. None of these perspectives can alter the fact that there is a person H, an enduring substance, who has a certain genetic makeup, who lived for a certain time as a male, and who underwent gender reassignment.

Sex as registered in a database reflecting legal classification standards is what is commonly known as a *phase sortal* [26], analogous to 'student' or 'adult'. It is the nature of phase sortals that several conflicting phase sortals can apply to a single person in different phases of life, and that the value attributed to an individual for a sortal of this type may vary over time. It is very difficult, perhaps impossible, to change genetic makeup. Genetic sex is consequently an example of an essential sortal. Its value cannot be easily changed over time. Naturalism, which deals with universals and kinds in reality in set-theoretic terms, cannot handle distinctions of these sorts, and nor can it handle situations involving multiple perspectives operating in parallel. Common-sense realism, however, with its acceptance of phase as well as essential sortals, can handle such situations and the associated multiple perspectives. In consequence we must conclude that information systems should use common-sense realism in preference to naturalism when analyzing or understanding modelling tools and methods or when seeking a general ontology to unify domain ontologies. We have argued elsewhere [27] that common-sense ontology also brings other practical benefits in enabling data to be modelled more accurately and also in enabling data to be transferred from one information context to another in such a way as to remain true to the reality beyond the database.

## 7 Conclusions

We have considered the relationship between top-level and domain-specific ontologies, arguing that a coherent top-level ontology applying rigorous theory from philosophy is for many reasons an important objective. Of those top-level ontologies that are heavily influenced by philosophical theory and adhere to a reasonably consistent ontological position, two of these have been used in information systems to analyse systems analysis and design tools and methodologies. We believe that we have demonstrated that of these only common-sense realism is able to handle the complexity of the human reality with which information systems have to deal.

## Acknowledgements

This paper was written under the auspices of the Wolfgang Paul Programme of the Alexander von Humboldt Foundation, of the EU FP6 Network of Excellence in Semantic Datamining, and the Project Forms of Life of the Volkswagen Foundation.

## References

[1] Wand, Y. and R. Weber, *Research Commentary: Information Systems and Conceptual Modeling - A Research Agenda.* Information Systems Research, 2002. **13**(4): p. 363-376.

[2] Wand, Y. and R. Weber, *An Ontological Model of an Information System.* IEEE Transactions on Software Engineering, 1990. **16**(11): p. 1282–1292.

[3] Wand, Y. and R. Weber, *An Ontological Evaluation of Systems Analysis and Design Methods*, in *Information Systems Concepts: An In-depth Analysis*, E.D. Falkenberg and P. Lindgreen, Editors. 1989, Elsevier Science Publishers B.V.: Amsterdam. p. 79–107.

[4] Milton, S.K. and E. Kazmierczak, *An Ontology for Data Modelling Languages: A Study Using a Common-Sense Realistic Ontology.* Journal of Database Management, 2004. **15**(2): p. 19-38.

[5] Smith, B., *Basic Concepts of Formal Ontology*, in *Formal Ontology in Information Systems*, N. Guarino, Editor. 1998, IOS Press. p. 19-28.

[6] Niles, I. and A. Pease. *Towards a Standard Upper Ontology*. in *2nd International Conference on Formal Ontology in Information Systems*. 2001. Ogunquit, Maine.

[7] Ceusters, W., B. Smith, and J.M. Fielding, *LinkSuiteTM: formally robust ontology-based data and information integration*, in *Database Integration in the Life Sciences*. 2004, Springer Verlag: Berlin.

[8] Kim, J. and E. Sosa, eds. *A Companion to Metaphysics*. Blackwell Companions to Philosophy. 1995, Blackwell Publishers: Oxford. 540.

[9] Honderich, T., ed. *The Oxford Companion to Philosophy*. 1995, Oxford University Press: Oxford. 1009.

[10] Vet, P.E.v.d. and N.J.I. Mars, *Bottom-Up Construction of Ontologies.* IEEE Transactions on Kowledge and Data Engineering, 1998. **10**(4): p. 513–526.

[11] Vickery, B.C., *Ontologies.* Journal of Information Science, 1997. **23**(4): p. 277–286.

[12] Guarino, N. *Formal Ontology and Information Systems*. in *Formal Ontology in Information Systems*. 1998. Trento, Italy: IOS Press, Amsterdam.

[13] OWL, OWL Web Ontology Language, http://www.w3.org/TR/owl-guide/:http://www.w3.org/TR/owl-guide/, Date Accessed: May 2004

[14] US National Library of Medicine: UMLS SEMANTIC NETWORK, http://www.nlm.nih.gov/research/umls/META3.HTML:http://www.nlm.nih.gov/research/umls/META3.HTML, Date Accessed: May 2004

[15] Weber, R., *Ontological Foundations of Information Systems*. Coopers and Lybrand Accounting Research Methodology Series. Vol. Monograph #4. 1997, Blackburn, Victoria: Buscombe Vicprint.

[16] Bunge, M., *Treatise on Basic Philosophy: Vol. 3: Ontology I: The Furniture of the World*. 1977, Boston: Reidel.

[17] Bunge, M., *Treatise on Basic Philosophy: Vol. 4: Ontology II: A World of Systems*. 1979, Boston: Reidel.

[18] Chisholm, R., *A Realistic Theory of Categories—An Essay on Ontology*. 1 ed. 1996: Cambridge University Press. 146.

[19] Australian Standard 5017-2002 (Health Care Client Identification) as used in Institute of Health and Welfare Meta-Data Standard, http://www.aihw.gov.au/pls/nhik/nhik_data_elements.data_element_details?pCommand=PREVIOUS&pResultSetID=145&pLinkType=DD&pLink_ID=ALL&pde_seq_id=1227:http://www.aihw.gov.au/pls/nhik/nhik_data_elements.data_element_details?pCommand=PREVIOUS&pResultSetID=145&pLinkType=DD&pLink_ID=ALL&pde_seq_id=1227, Date Accessed: May 2004

[20] *ISO Standard 5218:1977 Information interchange; Representation of human sexes.* 2004.

[21] Wand, Y., *Ontology as a Foundation for Meta-modelling and Method Engineering.* Information and Technology Software, 1996. **38**: p. 182–287.

[22] Winch, P., *The Idea of a Social Science and its Relation to Philosophy*. 1958, London: Routledge and Kegan Paul.

[23] Smith, B. and J. Searle, *The Construction of Social Reality: An Exchange.* American Journal of Economics and Sociology, 2003. **62**(1): p. 285-305.

[24] Smith, B., *Formal Ontology, Commonsense and Cognitive Science.* International Journal of Human-Computer Studies, 1995. **43**(12): p. pp. 641–667.

[25] Bittner, T. and B. Smith, *A Theory of Granular Partitions*, in *Foundations of Geographic Information Science*, M. Duckham, M.F. Goodchild, and M.F. Worboys, Editors. 2003, Taylor & Francis Books: London. p. 117-151.

[26] Wiggins, D., *Sameness and Substance Revisited*. 2001, Cambridge: Cambridge University Press.

[27] Verschelde, J.-L., M.C. Dos Santos, T. Deray, B. Smith, and W. Ceusters, *Ontology-assisted database integration to support natural language processing and biomedical data-mining.* Journal of Integrative Bioinformatics, 2004(1).