

## **DATA MODELLING LANGUAGES: AN ONTOLOGICAL STUDY**

**Simon Milton**

Department of Information Systems, The University of Melbourne, Vic. 3010 Australia  
Tel.: +61 3 8344 0666, Fax: +61 3 9349 4596  
s.Milton@dis.unimelb.edu.au

**Ed. Kazmierczak**

Department of Computer Science and Software Engineering, The University of Melbourne, Vic. 3010 Australia  
Tel.: +61 3 8344 9106  
ed@cs.mu.oz.au

**Chris Keen**

School of Information Systems, The University of Tasmania, GPO Box 252-87 Hoabart 7001 Australia  
Tel.: 61 3 6226 2900, Fax: +61 3 6226 2913 (fax)  
chris.keen@infosys.utas.edu.au

### **ABSTRACT**

*There are many data modelling languages used in today's information systems engineering environment. Some of the data modelling languages used have a degree of hype surrounding their quality and applicability. We would like to understand exactly what makes some data modelling languages successful and in some way suggest how useful data modelling languages will be in the context of an organisation and why. We are also interested in a theory capable of unifying the disparate range of languages. To do these things we select a theory based on ontology using which data modelling languages can be investigated. In this context theory should allow us to understand, compare, evaluate, and strengthen data modelling languages. The theory may also be used to suggest how useful various data modelling languages may be in an organisational setting. In this paper we present Chisholm's ontology which we use to investigate data modelling languages. We show how Chisholm's ontology can be used as a unifying theory of data models, develop methods for comparing data modelling languages based on this theory and summarise our findings. In conclusion, we evaluate the methods and the theory and examine avenues for future research. In this paper we present a deeper understanding of method together with analysis of new data modelling languages.*

### **1. INTRODUCTION**

One class of modelling tool used in the analysis and design of information systems is the class of data modelling languages. Data modelling languages are the class of tools and techniques responsible for

designing data and static structures (Eg. ER) (Yourdon, 1989). Many languages have been proposed during the last three decades (Hull and King, 1987, Peckham and Maryanski, 1988, Banerjee et al., 1987, Cattell, 1991, Dobbie, 1990, Lécluse et al., 1988, Rumbaugh et al., 1991, Boman et al., 1997). There is significant variety in the data modelling languages, but, they all aim at model the nature and structure of reality for the purpose of creating information models. This raises at least two questions. Firstly, how do we rationally choose between these modelling languages? Secondly, how can we learn from these languages in order to design better languages for the future? One way to answer these two questions is to find a reference separate from information systems in which the underlying meta-models of the languages can be evaluated. Such an external reference point can act as a unifying framework for comparing the meta-models. In this paper we suggest one such reference theory and show how such a comparison can be performed. Therefore, the chosen theory needs to have a basis in common with data modelling languages such that it can be used to evaluate them meaningfully. Note, there are other reference theories that could yield equally meaningful comparisons.

It has been suggested that the mature philosophical study of ontology is a suitable source of theories (ie. specific ontologies) to help understand the representational parts of the analysis and design of information systems including data modelling. There has been some research utilising an ontology based upon one by Mario Bunge (Bunge, 1977, Bunge, 1979, Weber, 1997) to examine systems analysis and design methodologies (Green, 1996, Wand and Weber, 1989, Wand and Weber, 1990, Wand and Weber, 1993, Weber, 1997, Wand, 1996, Rohde, 1995). Recently, a commonsense realistic ontology by Roderick Chisholm (Chisholm, 1996, Chisholm, 1992) has been published. We selected this specific ontology for this study based on what appeared to be a commonality with a representative group of classical data modelling languages. In this study we are interested in using Chisholm's ontology in a conceptual comparison with a number of representative data modelling languages to examine the commonality more deeply. We have selected Chisholm's ontology because *realistic* ontologies tend to have elements common to all of the modelling languages, such as defining and describing objects/individuals/entities and their relationships. The only other ontology used in the literature (Bunge's ontology) is also a realistic ontology and it has been our working hypothesis that realistic ontologies provide good support for the analysis of meta-models. Also, by deliberately choosing a different ontology from that which has already been used, we hope to verify the findings of previous studies.

Our contribution is three-fold. Firstly, we develop a qualitative method for conceptually comparing a range of data modelling languages with an ontology. Secondly, we apply the method using Chisholm's ontology, a newly written ontology, to a representative range of data modelling languages. Thirdly, we have reason to believe that Chisholm's ontology is suitable to be developed as a unifying theory for data modelling languages. This work is a refinement of earlier work (Milton, 1998) in that in this paper we examine more data modelling languages, we refine the method we use in the conceptual study of data modelling languages, and we take our analysis further.

## 2. METHOD FOR USING ONTOLOGY IN CONCEPTUAL COMPARISONS OF DATA MODELLING LANGUAGES

In this section we introduce the method of using ontology in conceptual comparisons with data modelling languages. We begin by discussing ontology generally. We then outline the basis upon which we build the method before describing the method itself and the analysis associated with it.

An ontology is a world view and describes what is fundamental in 'what there is'. An ontology defines the terms used to construct a description of reality and how the terms are related. For our purposes, we find the following definition most helpful:

### □ Definition 1—Ontology

“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.” (Honderich, 1995)

As a philosopher understands it, the study of ontologies deals with the ‘categorical structure of reality’ (Honderich, 1995). An ontology also provides a description of fundamental terms, which one uses to describe reality, and the ways in which these terms relate to the categories. In these ontologies basic questions are asked concerning the constitution of reality and what fundamental categories of things exist in reality, together with the terms that one needs to construct a description of a ‘state of affairs’.

Other groups of researchers view ontologies as being restricted to a particular domain such as those used in artificial intelligence research (Vickery, 1997). The world view of a civil engineer is an example of a domain-specific ontology. In these types of ontology, categories are constructed that represent the vocabulary of the universe of discourse in which the group is likely to be interested.

We are interested in ontologies that are of the more general type and not confined to one problem domain, or that describe categories of a restricted sort. This is because data modelling languages are not restricted to one class of modelling problem and are needed to build representations of reality which are widely applicable in a similar way to ontologies constructed by philosophers.

An ontology tells us what terms are needed to describe ‘what there is’ (physical and non-physical) in reality and, more deeply, the ontology also provides us with a concept that defines each term. Similarly, a data modelling language consists of terms that are used to model ‘what there is’. The language also conceptually defines the meaning of terms. For example, the ontology may use the term ‘individual’ which allows description of ontologically separate entities from the world (such as Prime Minister Tony Blair). In a similar way the entity-relationship modelling language (ER) provides for the term ‘entity’ that allows the description of separate entities from reality (such as President Clinton). These two terms, individual and entity, have concepts that are described in the literature. We can use these concepts as a basis for comparison. That is, we design our method around the need for a deep understanding of the differences not just the presence or absence of a term (the technique used in ontological studies to date).

We are aiming to explore data modelling languages using an ontology utilising this basis of concepts. This is a significant advance on previous ontological studies that have been based on structural comparisons and not on comparison of concepts.

## 2.1. Conceptual Basis of the Method

Data modelling languages and most ontologies are semi-formal or descriptive providing a rich description of terms making up the ontology or the data modelling language. It is for this reason that our method is based on the concepts associated with terms from an ontology and from data modelling languages.

### □ Definition 2—Concept

“[A] concept is a way of thinking about something—a particular object, or property, or relation, or some other entity.” (Dancy and Sosa, 1992)

There is a concept associated with each term that gives meaning to the term. It is through concepts that a rational analysis of data modelling languages using an ontology can be made. For example the term ‘entity’ in ER relates to terms such as ‘individual’ from an ontology, or ‘object’ from an object model. It is only through examining concepts associated with these terms that a sensible analysis can be made of the similarities and differences between the data models.

A concept is expressed using what is called a ‘possession condition’ (Dancy and Sosa, 1992) so that the meaning of a concept can be conveyed to another person.

### □ Definition 3—Possession condition

A possession condition is “[a] statement which individuates a concept by saying what is required for a thinker to possess it”. (Dancy and Sosa, 1992)

Umberto Eco's semiotic theory of codes (Eco, 1976) can be used to further understand the relationship between terms and their associated meaning. In a semiotic sense, terms signify (Eco, 1976) concepts, and we adopt the reading that the terms and concepts that we have described fit Eco's theory of codes (Eco, 1976), in that they are examples of sign-functions, where the term takes the role of expression, and the concept its content.

□ Definition 4—Sign function

“A sign-function is realized when two functives (expression and content) enter into a mutual correlation” (Eco, 1976).

Some terms in our ontology and data modelling languages are described using compound concepts that can be subdivided into several parts. Each part defines further aspects to the meaning of the whole term. It is the concept that is compound. A concept can have a core without which the entire concept is meaningless.

□ Definition 5—Core concept

A core concept is one such that its absence will render the definition of the entire concept meaningless.

Analysis of a term can be undertaken at the part or the whole level of a term. In Eco's semiotic theory, each term spans a semantic field (Eco, 1976). Each term from an ontology or data modelling language spans part of a semantic field (Eco, 1976) through its associated concept. Therefore, through the idea of a semantic field we can express the similarities and differences between concepts in the ontology and those in the data modelling languages and thereby discuss and explain our findings.

Since the terms and concepts (their meaning) that we are interested in are stable, we need not explore Eco's semiotic theory beyond his theory of codes. The semiotic theory plays the role of analysing the results of the method and is discussed in section 2.3. Firstly we need to explain the method.

## **2.2. Method**

The method is a pairwise comparison of concepts from the ontology with concepts from each of the data modelling languages and a discussion of the similarities and differences. In conducting the pairwise comparisons we are testing each language against the selected and independent view of reality. In order to conduct such a comparison we must first extract definitions for concepts from the ontology and from the data modelling languages.

Consequently, the method has two parts.

The first part is the extraction of concepts and their associated definitions from the ontology and from the candidate data modelling languages. There is no algorithmic way of conducting such an extraction. All key sources in the literature, for the ontology and the data modelling languages, are utilised and care is taken to faithfully represent concepts from each.

The second part is the pairwise comparison of concepts from the ontology with concepts from each of the data modelling language. Each pairwise comparison is a bidirectional evaluation, firstly using concepts from an ontology to evaluate a data modelling language followed by the reverse (Milton, 2000). Each pairwise comparison of concepts from an ontology with concepts from a data modelling language results in a series of graded indicators of agreement of the data modelling language with the ontology (the benchmark). We use Eco's theory of codes and the idea of coverage of semantic field to more fully explain the nature of the graded indicator. In the following sub-section we discuss the analysis of results from the pairwise comparison with particular focus on the graded indicator. There is no algorithmic method of conducting the pairwise comparison and the method is consequently inherently subjective, that is, it relies on the person performing the method to interpret nuances of the pairwise comparison. We have been careful to faithfully describe the similarities and differences we have found.

### 2.3. Analysis

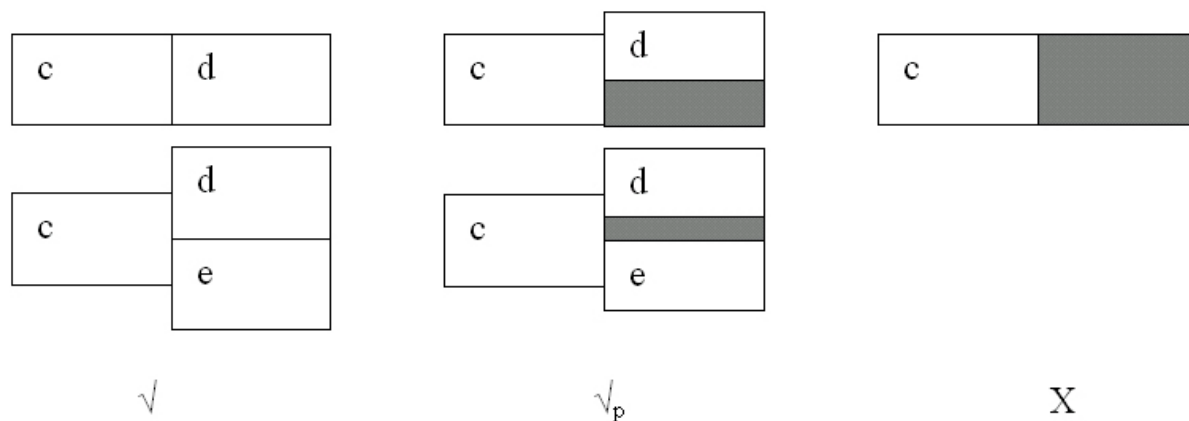
When comparing a concept *c* (from the ontology) with a specific data modelling language, there are three broad categories of results.

Firstly, the data modelling language may have total overlap with respect to *c*. Total overlap may be provided by one concept (for example, *d*) or perhaps by several concepts (for example, *d* and *e*). That is, there may be one concept or several concepts that together provide total overlap, in terms of semantic field, with the concept from the ontology.

The second possibility is where the overlap is partial. It may be that the concept from the modelling language can be extended to support the full generality of the concept from the ontology. It may also be the case that there is little prospect for extension due to certain concepts in the data modelling languages being contradictory (Eco, 1976) with respect to the ontology in ways that make extension difficult.

Finally, it may be that there is no overlap at all between the data modelling language and *c* from the ontology.

Figure 1 shows the three categories pictorially. While the coverage of a specific concept is depicted in this figure as a sharp rectangle, the nature of semantic fields dictates that the edges are quite blurry.



**Figure 1:** Degree of overlap in coverage of semantic field

It is useful to attach a symbol to each of these categories of results so that an idea of the results of the comparison can be conveyed easily. The three symbols for full coverage, partial coverage and no coverage are ( $\sqrt{\quad}$ ) ( $\sqrt{p}$ ) and (X) respectively. It is important, however, to keep in mind that the comparison is qualitative and therefore is subject to some interpretation. Further, the description of differences is important additional to the indicative results.

In this paper we conduct a conceptual comparison of five data modelling languages with Chisholm’s ontology. We describe Chisholm’s ontology in the next section before presenting the results in section 4.

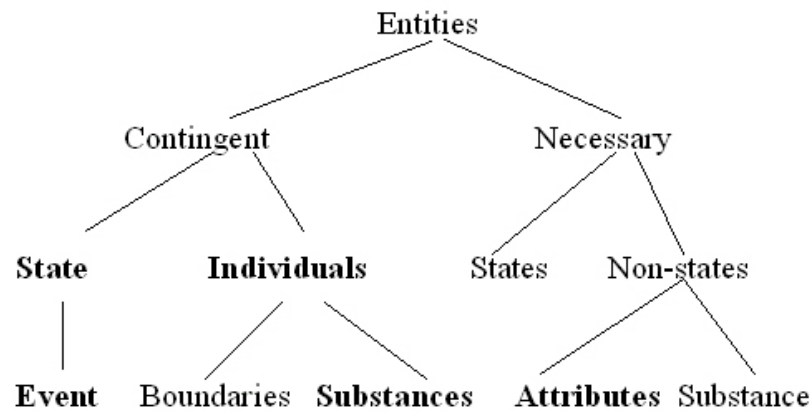
### 3. CHISHOLM’S ONTOLOGY

We must necessarily be brief in introducing Chisholm’s ontology. Interested readers are encouraged to read the essay and monograph by Chisholm (Chisholm, 1992, Chisholm, 1996) for a much more comprehensive treatment of the ontology.

Chisholm has written extensively in the areas of ontology, metaphysics, and epistemology (Chisholm, 1989a, Chisholm, 1957, Chisholm, 1976, Chisholm, 1979, Chisholm, 1982, Chisholm, 1989b) and these works provide a backdrop to the 1992 and 1996 treatments.

Chisholm’s categories are organised into a taxonomy which is shown in Figure 2. Chisholm adheres to the theory that establishes the dichotomy dividing the world into entities that are ‘contingent’ and don’t have to

exist, and ‘necessary’ entities that must exist in order for his theory to be consistent (Honderich, 1995). The categories that are similar to those used in data modelling languages, and consequently play important roles in our methods, are shown in bold type-face in this figure.



**Figure 2:** The categories for chisholm’s ontology (Chisholm, 1996)

Chisholm’s ontology centres on individuals and the attributes they exhibit. Other terms that we need are defined only using attributes in relation to individuals. That is, ‘individual’ and ‘attribute’ have descriptions that show not only their individual nature, but also their roles in the terms class and relation. These are important terms that are used by Chisholm to make sense of ‘what there is’ and are therefore important for our goals.

#### *Individual*

An individual is a discernable and transient object. It need not be material (or physical) in nature. Examples of individuals are an accountant named Freda, the annual financial statements for Ericsson, and Orly International Airport.

Individuals are identified using attributes that only they exemplify, and may have constituents thereby giving them structure. Constituents may be other individuals (called parts) or may be boundaries (the other constituents). For example, consider Orly Airport. It has several rent-a-car franchises, bars, restaurants, departure gates, each of these are parts of Orly Airport and are also individuals. In this example, most of these can be further sub-divided.

#### *Attribute*

An individual may exemplify attributes. Orly Airport is very busy; Nokia’s balance sheet is good; Freda, our accountant, is of age 43. Some attributes may never be exemplified and others cannot be exemplified. For example, Orly Airport may never be green. We can be sure that Orly Airport can never be a liquid.

Two attributes may be equivalent in the sense that if one is exemplified by an individual the other is also. This relationship is called conceptual entailment. This can be illustrated by considering Orly Airport. The attribute very busy may involve a conceptual entailment with the attribute of having over a certain number of aircraft movements an hour.

Chisholm allows for compound attributes. Compound attributes may consist of other compound attributes or simple attributes. He suggests that an attribute *may* be the conjunction or disjunction of several attributes. For example, the attribute of ‘being good’ with respect to Nokia’s financial statements may be the conjunction of being in surplus (profit) and being of good credit rating. Chisholm also indicates that there may be alternative mechanisms for providing compound attributes.

Philosophically and logically, it makes little sense to talk about when an attribute came into being. Chisholm asserts that attributes are enduring, thus avoiding the problem of declaring when an attribute comes into being. For example, when did the attribute ‘being green’ first come into being? Since we cannot know and

since raising its genesis brings about certain problems it is better to adopt the position that attributes are non contingent, they exist perpetually.

*Classification*

In Chisholm’s ontology, attributes are used to restrict membership of sets and classes. Chisholm reduces discussion of classes to discussion of attributes by adopting Russell’s reduction of classes to attributes (Russell, 1908). This has the effect of building classes and sets from individuals through the exemplification of attributes and not by constructing elaborate class structures. In this way flexibility is maximised. For example, suppose we are maintaining a taxonomy of plants. Periodically, the taxonomy may change quite drastically without a change in the majority of attributes exhibited by the plants involved. Using Chisholm’s ontology classes can change radically through a change in membership criteria based on attribute exemplification.

Classes and sets can be selected based upon attributes that are conjunctions and disjunctions of other attributes, and in this sense complex class relationships can be realised, that are essentially class structures. Classes are reflections of attributes exemplified by individuals due to the fact that they exemplify the attributes that are used to select the class.

*Relation*

Individuals may be related. Chisholm says: “To know what relations are, we must understand the concept of the direction of relations” (Chisholm, 1996). Chisholm means that relations may not be reciprocated. For example, I may be interested in a job with Nokkia. Nokkia may not be interested in employing me. It is for this reason that relations are unidirectional. Further, in the ontology, relations are attributes in the form of ordered pairs. This results from the fact that ordered pairs are related to sets of a specific form; Therefore ordered pairs can be reduced to being represented by attributes through mechanisms already introduced.

For an ordered pair to represent unidirectional relations, attributes that uniquely describe and thereby identify each individual need to be found. For example, suppose that Freda (our accountant) is recruited to audit Nokia’s books then an attribute being an ordered pair of identifying attributes for Freda and Nokia would have to be exhibited by Freda. A reciprocal attribute representing the reverse relation would need to be exhibited by Nokia, if the reciprocal applies. In the simplest case an individual may be related to another (binary). More complex relations between three individuals (ternary) or more (n-ary) are allowed. Mathematically it is proven that these all can be reduced to a series of binary relations (Quine, 1960).

Concept	Description
Individual Core	Chisholm allows for discernible and transient objects. These are called individuals. Individuals come into being (are created) and pass away (destroyed). In this sense they are transient.
Identity	Each individual possesses an attribute (or several attributes) that uniquely identifies it.
Structure	Individuals may have constituents. These are either other individuals (known as parts) or boundaries (the other constituents.) Individuals that make up parts of others are still thought of as being individuals.
Attribute Core	Attributes are exhibited by individuals. They are central to Chisholm’s ontology, after individuals. Further, attributes are enduring, in the sense that they don’t come into being and don’t pass away. Further, attributes must be loosely coupled with individuals.
Equivalence	Attributes can be equivalent in the sense that if something exhibits one attribute then it exhibits the other.
Complexity	Attributes may be simple or complex. Complex attributes are combinations of either simple or other complex attributes. The mechanism suggested by Chisholm is one involving conjunction and disjunction of attributes. He feels there may be other ways of providing for this complexity.
Classification Core	Classes and sets are provided using attributes, in the ontology. Specifically, it is through the attributes that membership of classes is determined.
Relation Core	Individuals may be related. Specifically, relations are attributes (an ordered pair). The ontology requires that attributes that identify the participating individuals are required. The relations are unidirectional (not bidirectional).

**Table 1:** summarises the relevant terms and concepts from chisholm’s ontology.

## 4. RESULTS

We selected five representative data modelling languages for the comparison. Specifically, the classic Entity-Relationship (ER) model (Chen, 1976), the Functional Data Model (FDM) (Kerschberg and Pacheco, 1976, Shipman, 1981), the Semantic Data Model (SDM) (Hammer and McLeod, 1990, Hammer and McLeod, 1981), NIAM (Nijssen and Halpin, 1989), and Object Modelling Technique (OMT) (Blaha and Premerlani, 1998) compatible with UML. Due to space constraints we don't describe them here but instead refer readers to previous work (Milton and Kazmierczak, 2000, Milton, 2000). Table 2 shows the indicative results for the comparison of Chisholm's ontology with the data modelling languages.

Ontological Concept	ER	FDM	SDM	NIAM	OMT
<b>Individual</b>	√ <sub>p</sub>	√ <sub>p</sub>	√	√ <sub>p</sub>	√
Core	√	√	√	√	√
Identity	√	√	√	√	√
Structure	X	√ <sub>p</sub>	√	X	√
<b>Attribute</b>	√ <sub>p</sub>	√ <sub>p</sub>	√ <sub>p</sub>	√ <sub>p</sub>	√ <sub>p</sub>
Core	√ <sub>p</sub>	√	√ <sub>p</sub>	√	√ <sub>p</sub>
Equivalence	X	X	X	X	X
Complexity	√ <sub>p</sub>	√	√	√ <sub>p</sub>	√
<b>Classification</b>	√ <sub>p</sub>	√	√ <sub>p</sub>	√ <sub>p</sub>	√ <sub>p</sub>
<b>Relation</b>	√ <sub>p</sub>	√	√	√ <sub>p</sub>	√ <sub>p</sub>

**Table 2:** Indicative results of the comparison of selected data modelling languages using chisholm's ontology

Chisholm's ontology views the world as a collection of individuals and relations between them, and the ontology uses attributes to describe both individuals and relations. Attributes are universals and endure, and, consequently they are loosely coupled with individuals. Attributes are also used to determine class and set membership. Our comparison suggests that this is to a large extent a similar world-view as those imparted by the languages and there is a significant level of agreement with the ontology and the modelling languages that we've studied, but the data modelling languages lack the full generality of Chisholm's ontology. The major differences between the ontology and some data modelling languages are in the structure of individuals, the nature relations and of classification, and the relationship between individuals and attributes. We discuss each of these below.

The structure of individuals, that is, part-whole relationships defined on individuals in the ontology, is supported by some but not all data modelling languages. Defining structure in this way, rather than at the level of classes of individuals, is useful in situations where there is significant diversity in the structure of individuals to the extent that it cannot be generalised to the level of classes. Those modelling languages that do not support the ontology in this way insist that individuals exist only in classes. Each such class contains only those individuals which exemplify the attributes that govern that class. Consequently, the class structure requires that attributes and individuals are tightly coupled. For example, in E-R, entities can only exist in classes that exhibit a fixed range of attributes. In Chisholm's ontology, however, individuals and attributes always exist outside classes.

The second point of departure is that some data modelling languages do not support the loose-coupling between individuals and attributes. The practical significance of loose-coupling of individuals and attributes is that it easily allows for individuals to change exemplifications of attributes. We expect this to be useful in cases where change in exemplification is expected to be frequent or where there is high diversity in



exemplification. Nevertheless most have a reasonable coverage of the core and complexity parts of attribute. Attribute equivalence is completely absent from all data modelling languages. Attribute equivalence is something which has less obvious impact in practical situations and appears to be more of philosophical importance. More investigation is required to assess the pragmatic utility of this feature.

Classification and relations are concepts recognised by all data modelling languages. Classification in the ontology is evident through the attributes exemplified by members of classes. In the ontology, classes are related to each other by the intersections and unions of the attributes used to select them and thereby can simulate class hierarchies. This approach to classification structure is entirely different from the most common classification approaches used by most data modelling languages where instead, rich and rigid class hierarchies are prevalent. It is the tight coupling of classes of individuals and attributes that leads to the need for rigid class hierarchies. The opposite, supported by the ontology, allows for a sea of individuals from which classes are built. This means greater flexibility when (and if) required but also impacts on the processing efficiency for implementation models (should the idea be extended to that context).

The ontology suggests that relations are unidirectional. The ontology, in requiring unidirectional relations, states a clear truth - that not all relations are reciprocated. Relations are not bidirectional in three of the data modelling languages but, however, the concept is fully supported by FDM and SDM.

The structuring of individuals, the loose coupling between individuals and attributes, the formation of classes are all critical for situations in which the information model must evolve over time or where frequent changes to the model are required. We hypothesise that modelling languages without the flexibilities gained by the structuring of individuals, loose coupling and the formation of classes, will be harder to change as the situation they model changes. In contrast, Chisholm's ontology already takes into account the changing nature of reality by allowing for each individual to change structure, and to change the range of attributes they exemplify (a consequence of loose coupling of individuals and attributes), and to be a member of disparate classes (the way classes are formed).

We can see from the results that the modelling languages share the world view of the ontology to a reasonably large degree. The areas of departure tend to be of the nature of a difference in flexibility rather than complete absence of support. Also, all concepts had a high degree of emphasis with respect to their core. On the basis of this, we can say that the ontology is a good candidate for a unifying theory for data modelling languages. Work is proceeding to further deepen our understanding of this theory and to formalise it in a way appropriate for data modelling languages.

There are some issues that need investigation. The importance of conceptual entailment and the implications of the universality of attributes need to be investigated in practical situations. Apart from the universality of attributes, the area of most concern is that of classification. Clearly, the rigidity of classes and the presence of rigid class hierarchies is not supported in the ontology. As modelling features there appears to be little support in traditional realistic philosophy for such an approach.

## **5. CONCLUSIONS AND FUTURE WORK**

In this paper we have examined Chisholm's ontology as a potential unifying framework with data modelling languages. We have also shown that it can be used to construct methods that are effective in conceptual comparisons with data modelling languages. We have found that the ontology shares to a significant degree the world view of the data modelling languages we have studied. It is for this reason that we judge Chisholm's ontology to be a suitable candidate for further investigation as a unifying theory and this is subject to ongoing research.

It is important that we reflect upon the role of that ontology as theory in the way we discussed in the introduction, and also reflect on the quality of the methods based on Chisholm's ontology.

We selected an ontology by Roderick Chisholm based on what appeared to be a commonality with a representative group of classical data modelling languages. We have found that because of the quality of terms and concepts found in the ontology it was very useful in analysing the data modelling languages.

Further, there is a good degree of overlap between the five data modelling languages that we have studied and Chisholm's ontology. The overlap is not total but does cover the core parts of terms and concepts from the ontology. We found that there were parts of the theory that were not appropriate. Specifically, we found that those aspects concerning cosmology, theology, and spatial boundaries to be inappropriate for our comparison. An ontology should consider those areas, but clearly data modelling languages should not. Chisholm's ontology also considers dynamics. This is an area for ongoing research into dynamic and process models in information systems.

We have learned by applying the methods that data modelling languages should be able to model a reasonably diverse range of states of affairs. Chisholm's ontology, in its role as theory, has also predicted areas in which the modelling languages can be extended or strengthened for them to apply more broadly than they have thus far. These predictions need further investigation.

The method developed as part of this study represent a significant advance over previous approaches in that they recognise the important role played by terms and concepts in qualitatively comparing data modelling languages with ontologies. We have argued that studies of data modelling languages can be undertaken using ontological terms and concepts. In the method we analyse the relationship between terms in a data modelling language and an ontology using a restricted part of a broad semiotic theory. By using semiotic theory in this way we highlight the importance of meaning of terms in the ontology and data modelling languages. This in itself is an important advance over past approaches.

The comparison of the ontology with each of the modelling languages requires a deep understanding of the ontology and of each data modelling language. Further, it is on the basis of the terms described in the ontology that the comparison with each candidate modelling language is undertaken. It takes time to understand an ontology as dense as Chisholm's and to relate each term and its associated concept to the modelling languages. We cannot conceive of an easier way in which to undertake such a work. Further, there is clearly a subjective element to the methods and consequently the results of the comparison may vary to a degree between researchers. This, however, is the very nature of the type of research and is not of itself grounds enough to doubt its applicability. There is no algorithmic comparison of concepts. However, given that we are interested in comparing, contrasting, evaluating, and ultimately improving our data modelling languages, our approach is quite appropriate. Further, the approach has the potential to greatly enhance our understanding of the nature of data modelling languages.

The subjective nature of this work raises important questions about the validity of the analysis we have presented in this paper. We can say several things. Firstly, in all subjective methods such as we have presented here, there are always questions about how valid the findings are. Consequently, it is important to explore the findings drawn in a practical arena. This work is currently in progress. Also the methods can be refined by considering commentary from other investigators, for example, by reviewing the selection of concepts used in the comparison, but ultimately, we are still faced with a qualitative analysis. Beyond theoretical work completed, we are interested in examining the role of the ontology in suggesting how clients approach various parts of the modelling process. Consequently, we are undertaking research that examines the attitudes to classification structures by practitioners and clients using Chisholm's ontology as a theoretical reference.

Concluding, we have reason to believe that both Chisholm's ontology and the methods developed for applications of ontologies of this style have much to offer research into data modelling in information systems. We also invite the investigation of other ontologies with data modelling languages.

## **REFERENCES**

- Banerjee, J., Chou, H.-T., Garza, J. F. and Kim, W. (1987) Data Model Issues for Object-Oriented Applications. *ACM Transactions on Office Information Systems*, 5 (1), pp. 3–26.
- Blaha, M. and Premerlani, W. (1998) *Object-Oriented Modeling and Design for Database Applications*, Prentice Hall, Upper Saddle River.

- Boman, M., Bubenko, J. A., Johannesson, P. and Wangler, B. (1997) *Conceptual Modelling*, Prentice Hall, London.
- Bunge, M. (1977) *Treatise on Basic Philosophy: Vol. 3: Ontology I: The Furniture of the World*, Reidel, Boston.
- Bunge, M. (1979) *Treatise on Basic Philosophy: Vol. 4: Ontology II: A World of Systems*, Reidel, Boston.
- Cattell, R. G. G. (1991) *Object Data Management*, Addison-Wesley, Menlo Park.
- Chen, P. (1976) The Entity-Relationship Model—Toward a Unified View of Data. *ACM Transactions on Database Systems*, 1 (1), pp. 9–36.
- Chisholm, R. (1957) *Perceiving: A Philosophical Study*, Cornell University Press, Ithaca.
- Chisholm, R. (1976) *Person and Object: A Metaphysical Study*, Open Court, La Salle.
- Chisholm, R. (1979) Objects and persons: revisions and replies. *Grazer Philosophische Studien*, 7/8 317–388.
- Chisholm, R. (1982) *The Foundations of Knowing*, University of Minnesota Press, Minneapolis.
- Chisholm, R. (1989a) *On Metaphysics*, University of Minnesota Press, Minneapolis.
- Chisholm, R. (1989b) *Theory of Knowledge*, Prentice-Hall, Englewood Cliffs.
- Chisholm, R. M. (1992) In *Language, Truth, and Ontology*(Ed, Mulligan, K.) Kluwer Academic Publishers, Dordrecht, pp. 211.
- Chisholm, R. M. (1996) *A Realistic Theory of Categories—An Essay on Ontology*, Cambridge University Press.
- Dancy, J. and Sosa, E. (Eds.) (1992) *A Companion to Epistemology*, Blackwell Publishers, Oxford.
- Dobbie, G. (1990) *Object Oriented Database Systems: A Survey*, Department of Computer Science, The University of Melbourne, Melbourne.
- Eco, U. (1976) *A Theory of Semiotics*, Midland, Bloomington.
- Flew, A. (1989) *An Introduction to Western Philosophy: Ideas and Arguments from Plato to Popper*, Thames and Hudson, London.
- Green, P. (1996) *An Ontological Analysis of Information Systems Analysis and Design (ISAD) Grammars in Upper Case Tools*, Unpublished PhD Thesis, Department of Commerce The University of Queensland.
- Hammer, M. and Mcleod, D. (1981) Database Description with SDM: A Semantic Database Model. *ACM Transactions on Database Systems*, 6 (3), pp. 351–386.
- Hammer, M. and Mcleod, D. (1990) In *Research Foundations in Object-oriented and Semantic Database Systems*(Eds, Cárdenas, A. F. and McLeod, D.) Prentice Hall, Englewood Cliffs, pp. 341.
- Honderich, T. (Ed.) (1995) *The Oxford Companion to Philosophy*, Oxford University Press, Oxford.
- Hull, R. and King, R. (1987) *Semantic Database Modelling: Survey, Applications, and Research Issues*. *ACM Computing Surveys*, 19 (3), pp. 201–260.
- Kerschberg, L. and Pacheco, J. E. S. (1976) *A functional database model*, Pontificia Univ. Catholica do Rio de Janeiro, Rio de Janeiro, Brazil.
- Kim, J. and Sosa, E. (Eds.) (1995) *A Companion to Metaphysics*, Blackwell Publishers, Oxford.
- Lécluse, C., Richard, P. and Velez, F. (1988) O2, and Object-Oriented Data Model, In *Proceedings of Advances in Database Technology 88* , pp. 556–562.

- Milton, S. (2000) An Ontological Comparison and Evaluation of Data Modelling Frameworks, Unpublished PhD Thesis, School of Information Systems, The University of Tasmania, Hobart. (<http://www.dis.unimelb.edu.au/staff/simonm/thesis/>)
- Milton, S. and Kazmierczak, E. (2000) Enriching the Ontological Foundations of Modelling in Information Systems, In Proceedings of Workshop on Information Systems Foundations: Ontology, Semiotics, and Practice (Ed, Dampney, C.) Lighthouse Press, Macquarie University, Sydney, pp. 55-65.
- Milton, S., Kazmierczak, E. and Keen, C. (1998) Comparing Data Modelling Frameworks using Chisholm's Ontology, In Proceedings of 6th European Conference on Information Systems , Vol. I Aix-en-Provence, pp. 260–272.
- Nijssen, G. M. and Halpin, T. A. (1989) Conceptual Schema and Relational Database Design: A Fact Oriented Approach, Prentice-Hall, New York.
- Peckham, J. and Maryanski, F. (1988) Semantic Data Models. *ACM Computing Surveys*, 20 (3), pp. 153–189.
- Quine, W. V. O. (1960) *Word and Object*, MIT Press, Cambridge Massachusetts.
- Rohde, F. (1995) An Ontological Evaluation of Jackson's System Development Model. *Australian Journal of Information Systems*, 2 (2), pp. 77–87.
- Rumbaugh, J., Blaha, M., Premerlani, W., Eddy, F. and Lorenzen, W. (1991) *Object-Oriented Modeling and Design*, Prentice-Hall, Englewood Cliffs, New Jersey.
- Russell, B. (1908) Mathematical Logic as Based on the Theory of Types. *American Journal of Mathematics*, XXX 222-263.
- Shipman, D. W. (1981) The Functional Data Model and the Data Language DAPLEX. *ACM Transactions on Database Systems*, 6 (1), pp. 140–173.
- Vickery, B. C. (1997) Ontologies. *Journal of Information Science*, 23 (4), pp. 277–286.
- Wand, Y. (1996) Ontology as a Foundation for Meta-modelling and Method Engineering. *Information and Technology Software*, 38 182–287.
- Wand, Y., Monarchi, D. E., Parsons, J. and Woo, C. (1995) Theoretical Foundations for Conceptual Modelling in Information Systems Development. *Decision Support Systems*, 15 (1995) 285–304.
- Wand, Y. and Weber, R. (1989) In *Information Systems Concepts: An In-depth Analysis*(Eds, Falkenberg, E. D. and Lindgreen, P.) Elsevier Science Publishers B.V., Amsterdam, pp. 79–107.
- Wand, Y. and Weber, R. (1990) An Ontological Model of an Information System. *IEEE Transactions on Software Engineering*, 16 (11), pp. 1282–1292.
- Wand, Y. and Weber, R. (1993) On the Ontological Expressiveness of Information Systems Analysis and Design Grammars. *Journal of Information Systems*, 1993 (3), pp. 217–237.
- Wand, Y. and Weber, R. (1995) On the Deep Structure of Information Systems. *Information Systems Journal*, 1995 (5), pp. 203–223.
- Weber, R. (1997) *Ontological Foundations of Information Systems*, Buscombe Vicprint, Blackburn, Victoria.
- Yourdon, E. (1989) *Modern Structured Analysis*, Prentice-Hall, New Jersey.