An Ontology of Data Modelling Languages: A Study Using a Common-Sense Realistic Ontology

Simon K Milton[1] and Ed. Kazmierczak[2]

[1]Department of Information Systems   [2] Department of Computer Science and Software Engineering

The University of Melbourne
Victoria, Australia, 3010
smilton@unimelb.edu.au
ed@cs.mu.oz.au

Abstract

Data modelling languages are used in today's information systems engineering environments. Many have a degree of hype surrounding their quality and applicability with narrow and specific justification often given in support of one over another. We want to more deeply understand the fundamental nature of data modelling languages. We thus propose a theory, based on ontology, that should allow us to understand, compare, evaluate, and strengthen data modelling languages. In this paper we present a method (conceptual evaluation) and its extension (conceptual comparison), as part of our theory. Our methods are largely independent of a specific ontology. We introduce Chisholm's ontology and apply our methods to analyse some data modelling languages using it. We find a good degree of overlap between all of the data modelling languages analysed and the core concepts of Chisholm's ontology, and conclude that the data modelling languages investigated reflect an ontology of commonsense-realism.

Introduction

Data models have been used in information engineering environments for many decades for the precise purpose of building representations of reality. To date, there have been many different data modelling languages proposed with the most popular being the Entity-Relationship Model (Chen, 1976) but also including the Functional Data Model (Kerschberg & Pacheco, 1976; Shipman, 1981), the Semantic Data Model (Hammer & McLeod, 1981), NIAM (Nijssen & Halpin, 1989), and Object Modelling Technique (Blaha & Premerlani, 1998). Each new modelling language has often been accompanied with claims of its superiority and at times hype when compared with the others. There has been little beyond opinion to substantiate such claims and yet all notations purport to do similar things. We have two research questions:

Q1: How well do data models represent reality? and

Q2: What are the similarities and differences between data modelling languages?

We need a theory to help us answer these questions.

The mature philosophical study of ontology has been used as a source of theory to investigate tools and techniques used in the analysis and design of information systems. A key development in the use of ontology for the study of information systems has been the work of Wand and Weber (Weber, 1997), based on Bunge's ontology (Bunge, 1977, 1979). Part of the focus of this research has been to investigate the representational power of data modelling languages (Green, 1996; Rohde, 1995; Wand, 1996; Wand & Weber, 1989, 1990, 1993; Weber, 1997). Our work is motivated by the search for semantic methods to answer the research questions above. The work of Wand and Weber, while ground breaking, is based on structural comparison of elements of grammar and concludes only presence or absence of a 'construct'.

The conclusions drawn are very much based upon whether or not the data modelling language supports the ontological construct.

Our work seeks to develop semantic methods that not only detect the presence or absence of a construct but also allow us to judge the level of agreement or disagreement between a data modelling language

The contribution of this paper is three-fold. Firstly, we develop qualitative methods (1) 'the method of conceptual comparison', for conceptually evaluating individual data modelling languages through ontologies and (2) 'the method for conceptual comparison', for comparing a range of data modelling languages with an ontology based on a number of individual evaluations. These methods help answer the two research questions and are detailed in the method section. Our methods are, to some extent, independent of the individual ontology chosen as the basis of comparison. As a by-product we are starting to investigate the dominant ontology within data modelling languages. Secondly, we apply the methods using Chisholm's ontology (Chisholm, 1996) to a representative range of data modelling languages.

We follow this introduction with a deeper discussion relating ontologies and data modelling languages. We then examine the realism assumed in Chisholm's ontology and relate it to that contained within Bunge's ontology (the ontology upon which BWW is based.) Following this we describe Chisholm's ontology, which is the ontology used in this study, before describing the methods applied. Finally we present the results and conclude.

Relating Ontology and data Modelling Languages

We begin by defining ontology and discussing our ontological view of data modelling languages. Our treatment of ontology stems from the philosophical tradition and a good definition of ontology is found in Honderich (Honderich, 1995).

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 categorical structure of reality. ... Different systems of ontology propose alternative categorical schemes. A categorical scheme typically exhibits a hierarchical structure, with 'being' or 'entity' as the topmost category, embracing everything that exists". (p. 634)

Any ontology uses technical "expressions" to define what is real and it is these same "expressions" that are used to describe a specific 'state of affairs'. Each expression has meaning and may be defined formally or informally. It is important therefore to distinguish between the *labels* or words given to each technical expression and the *meaning* behind that expression. For example *individual* is a label given to a technical expression in an ontology and has a specific meaning in that ontology. For the purpose of distinguishing expressions from their intended interpretation we will use *Term* to mean the label used in the ontology and *Concept* to mean the interpretation that the ontology gives to that term. A concept is defined as follows.

Definition 2 – Concept

"[A] concept is a way of thinking about something – a particular object, or property, or relation, or some other entity." (Dancy & Sosa, 1992)

Now, to describe an ontology we require the following elements:

1. A categorisation of what constitutes reality, and what fundamental categories of things exist. As an example Chisholm presents his categorisation in a taxonomy (shown later in Figure 1).
2. A set of terms, each associated with a concept that fully defines the term, with which to construct descriptions of 'what there is'.
3. A set of fundamental terms labelling the fundamental categories (in 1) and a means of relating all terms back to those labelling the fundamental categories.

For example, Chisholm's ontology uses the terms *individual*, *attribute*, *event*, *relation* and *set/class* but only only *individual*, *attribute* and *event*. Are fundamental and label categories in Chisholm's taxonomy. *Set/class* and *relation* derive their definitions from individual and attribute and are thus related back to the fundamental categories.

Data modelling languages also use terms, to signify modelling elements. For example, the Entity Relationship model (Chen, 1976) uses the terms *entity*, *relationship* and *attribute* to signify specific modelling concepts used for modelling information systems. The terms, and their underlying concepts, used to describe a given data modelling language constitute an ontology of the type described above. We claim that the terms and concepts used to define a data modelling language constitute an ontological meta-model for the modelling language.

We can now compare the ontological meta-model for a data modelling language with a philosophical ontology (acting as an external, or reference, ontology) thus yielding a qualitative semantic analysis of the similarity and difference between the world view that can be captured by the data model and that embodied in the ontology.

Concepts are the basis for comparing an ontology with an ontological meta-model for a data modelling language. A concept is expressed using a possession condition.

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 & Sosa, 1992)

Some concepts are compound. A compound concept has a core without which the entire concept is meaningless. In this sense the core of a concept is *necessary* for the concept's meaning to be intact (Honderich, 1995).

Definition 4 - Core Concept

A core concept is one such that its absence would render the definition of an entire concept meaningless (Milton, 2000; Milton, Kazmierczak, & Keen, 2001).

For example, the concept of a knife would be meaningless without a blade and some handle that fits within the palm of one's hand. If absent, other parts of the concept do not have the effect of rendering the concept meaningless.

By way of example concept consider the term *entity* as used in the Entity-Relationship (ER) model. The concept for this term in ER is "... something which involves information. It is usually identifiable. Each entity has certain characteristics, known as attributes. A grouping of related entities becomes an entity set." (Thalheim, 2000)

Other similar definitions can be found in many texts, including the classical paper by Chen (Chen, 1976). For data modelling languages, the concept associated with a term can be synthesised from seminal sources.

In the case of entity from the classic ER modelling language, we can show the compound concept in Table 1.

Concept	Description

Entity	Core	ER allows for significant entities, or objects (either physical or conceptual) to be modelled. These must be grouped into entity classes. Each entity cannot depend upon other entities to be classed as an entity.
	Identity	Each member of an entity class must have an identity (a key).

Table 1 – The concept "entity" from the classic ER modelling language

We can observe if the models built using our data modelling languages exhibit the world view embodied by an ontology by observing how well concepts found in the data modelling language's ontological meta-model matches concepts from the ontology. This in turn is achieved by comparing the concepts in the ontological meta-model with those of the ontology. An examination of the results would reveal the overlap between the ontology and the data modelling language. The method for conducting such conceptual evaluations and comparisons is described in the method section. Before this, we discuss the philosophical heritage and content of Chisholm's ontology, and its categories. We do this respectively in the following two sections.

Commonsense Realism: Chisholm's Philosophy

In this section we discuss the objectives for a philosopher when writing an ontology. Chisholm's ontology is then discussed and found to be one of Commonsense Realism. We conclude by placing Bunge's ontology (used as a basis for BWW used in information systems) in context with Chisholm's ontology.

Ontology defines the 'sum total of reality' (Honderich, 1995). It examines the nature of existence (what exits), and the categories into which these things fit. Existence means what exists outside an individual's mind and asks 'what is real?' Things that exist may be concrete (physical) or abstract. Each ontologist must answer these questions when defining their categories. They do so by stating their philosophical approach and by relating their ontology to the stream of philosophical arguments.

Chisholm (Chisholm, 1996) asserts that his approach is one of critical commonsensism:

"Our approach to philosophy is what Charles Sanders Peirce has called 'critical commonsensism.' This approach is based on faith in one's own rationality. Reason, as Peirce put it, not only corrects its premises, 'it also corrects its own conclusions' " (p. 4)

"Commonsensism is the view that we know, most, if not all, of those things which ordinary people think they know and that any satisfactory epistemological theory must be adequate to the fact that we do know such things." (Dancy & Sosa, 1992)

Critical commonsensism differs from commonsensism in that it demands a more rigorous standard of support for knowledge to be acquired, requiring the term 'critical'. Chisholm's ontology is also categorised as being one of "extreme realism" (Chisholm, 1996) in that in addition to individuals, abstract things such as attributes are also real.

"Realism in any area of thought is the doctrine that certain entities allegedly associated with that area are indeed real. Common sense realism — sometimes called 'realism', without qualification — says that ordinary things like chairs and trees and people are real. Scientific realism says that theoretical points like electrons and fields of force and quarks are equally real. And psychological realism says mental states like pains and beliefs are real." (sic) (Dancy & Sosa, 1992)

However, Chisholm's critical commonsensism combined with his form of extreme realism means that his ontology adheres to a philosophy that may be a little different from commonsense realism. Barry Smith (Smith, 1995), also a prominent realist and who refers to

Chisholm in his discussion, defines the school more clearly than Dancy & Sosa (Dancy & Sosa, 1992), and in his article he outlines his support for the thesis that commonsense realism in its various guises seems to be useful in cognitive science.

> "The thesis that there is only one world towards which natural cognition relates is a central plank of what philosophers in the course of history have identified as the doctrine of common-sense realism. It is a doctrine according to which:

> a. we enjoy in our everyday cognitive activities a direct and wide-ranging relational contact with a certain stable region of reality called the commonsense world

> b. our everyday cognitive activities rest upon a certain core of interconnected beliefs called "common sense" which is in large part true to the common-sense world as it actually is, not least in virtue of the fact that such beliefs and our associated cognitive capacities have arisen through interaction with this world;

> c. this common-sense world exists autonomously, which is to say independently of our cognitive relations to it. Indeed from the perspective of common-sense realism the common-sense world exists entirely independently of human beings. Partial evidence for this thesis is provided by the fact that palæontology and related disciplines describe this world as it was before human beings existed. Of course this world would lack theoretical interest in a universe populated exclusively by creatures with cognitive capacities radically different from those of human beings. But what these disciplines describe is, nonetheless, such as to exist independently." (Smith, 1995)

A careful reading of the extract reveals that it leaves room to subsume the scientific reality or outlook while still allowing for a commonly held view or socially agreed reality and it reassures us that we don't require human cognition for this world to exist. Additionally, this school of philosophy allows for a difference between the reality and the appearance of reality. Often this is called the error that is involved in making sense of reality.

> "Thus common-sense is not, in spite of its reputation, naïve; it draws a systematic distinction between reality and appearance, or in other words between the way the world is and the way the world seems or appears via one or other of the sensory modalities and from the perspective of one or other perceiving subject in one or other context. 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." (Smith, 1995)

It is important, however, not to ignore the success in describing reality through science, and so we need to describe the relationship between the world as seen in common sense and that described by physics, the most closely related science. Later in this paper, Smith relates commonsense realism with physics:

> "The common-sense realist must confront the question of the relation between the common-sense world and the world that is described in the textbooks of standard physics. Here again a number of different philosophical alternatives have been mapped out in the course of philosophical time, including that view that it is the common-sense world that is truly autonomous while the world of physics is to be awarded the status of a cultural artifact. Here in contrast, we assume a thesis to the effect that the commonsense world overlaps substantially with physical reality in the more standard sense." (Smith, 1995)

This is important because physics, and indeed science generally, cannot be discounted in the quest for determining what there is. Science has been very successful in determining what there is in the physical world and is in part responsible for constructing models of physical

reality. As paradigm shifts in science come and go, so will the models we use to describe physical reality. Paradigm shifts clarify our understanding of the physical world often importantly at the margins. We are experiencing the ramifications of just such a paradigm shift that began a little under a century ago. Issues such as string theory, quarks, and quantum mechanics generally have arisen from this paradigm shift, just as notions of mass, force, momentum, gravity, and planetary motion accompanied the Newtonian paradigm shift of several centuries ago. One therefore needs to be careful not to overstate the importance of science when examining ontology. Nevertheless, commonsense realism has a place for such development and clearly refuses to discard scientific realities in order to allow for social structures and understanding.

To date, the only ontology that has been adapted and used in information systems is a scientific ontology by Mario Bunge (Bunge, 1977, 1979). It is scientific in that it uses the results of the natural sciences and of systems theory in its design. Consequently, it requires reworking when paradigm shifts occur in science or when our understanding of how the natural world works changes significantly. It also takes a mechanistic or deterministic view towards the world and its physical and social structures.

Being a scientific ontology, Bunge's ontology is set at a level much more finely grained than an ontology based on common-sense realism. The two are clearly related as we explained earlier in this section. Both are likely to have a role in information science research with commonsense realism likely to be influential when considering socio-technical or social design issues and scientific realism is likely to be influential when considering purely technical or computerised parts of information systems.

Nevertheless the two ontologies have different but complementary realisms and it is therefore valuable to consider Chisholm as an alternative ontology as invited by earlier researchers.

Chisholm's Ontology

Roderick Chisholm has written extensively in the areas of ontology, metaphysics, and epistemology (Chisholm, 1957, 1976, 1979, 1982, 1989a, 1989b, 1996) and these works provide a backdrop to his 1996 monograph. We are obviously restricted in covering Chisholm's ontology in this paper. Interested readers are encouraged to refer to the monograph by Chisholm (1996), or his earlier paper (Chisholm, 1992), for a comprehensive treatment of the ontology.

Chisholm's categories are organised into the taxonomy shown in Figure 1. Chisholm adheres to the theory that establishes the dichotomy dividing the world into entities that are 'contingent' and don't have to exist, and those that are 'necessary' entities and must exist in order for the theory to be consistent (Honderich, 1995). This latter group is also often referred to as abstract entities that nevertheless exist. This is part of his realism and is reflected in the first branch in Figure 1 where the entire universe consists of entities divided in those that are contingent and those that are necessary. The fundamental categories that are relevant to all types of modelling in information systems are shown in bold typeface. Additional terms are real and are defined in the ontology (relation, set/class) and are related back to the terms that label the fundamental categories.

In studying data modelling (a sub-set of information systems modelling) states and events are not relevant. These are relevant for studies of the process models of the Object Modelling Technique (OMT) and other modelling languages that model states and changes in states.

Necessary substance and its state (God and His state (Chisholm, 1996)) are also clearly not in the realm of this study. Similarly, we do not need to consider issues of boundaries between spatial individuals in data modelling (where do I end and the chair upon which I sit begin?) and are therefore outside the scope of this work.
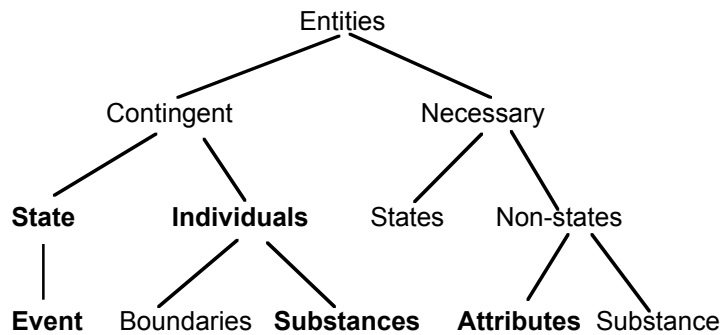

Figure 1—The Categories for Chisholm's Ontology (Chisholm, 1996)

Consistent with our earlier discussion, the nodes in the taxonomy are labelled by terms forming a subset of those found in the ontology, for example, the terms *individual* and *attribute*. In fact individuals and attributes are central to Chisholm's ontology. Further, other terms that Chisholm's ontology requires to make sense of 'what there is' are defined with reference to these fundamental terms. For example, 'individual' and 'attribute' have descriptions that show not only their own nature, the terms *class* and *relation* and related (and defined) in terms of attributes. In this section we introduce the terms from Chisholm's ontology that we use in ontological studies of data modelling languages. The paragraphs describing each term convey the concept associated with the term. We conclude with a tabular summary of the concepts.

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. This is called mereology (Honderich, 1995). 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, and departure gates. Each of these is a part of Orly Airport and each is also an individual. In this example, most of these parts can be further sub-divided.

On the other hand spatial substances have boundaries. A boundary is a surface, line, or point. For example, Orly Airport may have as its constituent surfaces that help to identify it as a spatial object. That surface is a boundary and is in turn made up of a number of surfaces, lines, and points. Boundaries of spatial substances are not of interest in data modelling.

Attribute

An individual may exemplify *attributes*. Each attribute may be exemplified by many individuals. 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 cannot be a liquid.

If two attributes are considered to be equivalent then it is the case that where one attribute is exemplified by an individual then so is the other. This is called *conceptual entailment* in the ontology. 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*, which in turn 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, other than conjunction and disjunction.

Philosophically and logically it makes little sense to talk about when an attribute came into being. In Chisholm's ontology, 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

Classes and sets may be part of a state-of-affairs. In Chisholm's ontology, attributes are used to restrict membership of sets and classes. Indeed, Chisholm's ontology reduces the discussion of classes to the 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 an individual's attributes and not by constructing elaborate class structures. 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. The central point remains that individuals come together to form classes and are fundamental to the ontology. 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

Chisholm's ontology allows for *relations* between individuals. 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, or alternatively must be carefully considered from the viewpoints of all individuals concerned. For example, I may be interested in a job with Nokia. Nokia may not be interested in employing me. It is for this reason that relations are unidirectional. Further, in the ontology, relations are be represented by ordered pairs of identifying individuals. This comes from the fact that ordered pairs are related to sets of a specific form and therefore can be reduced to a discussion of attributes in the manner noted

above. Therefore relations, despite being needed to describe a state-of-affairs, do not constitute a separate fundamental category but instead are related to attribute.

For an ordered pair to represent unidirectional relations, attributes need to be found that uniquely describe and thereby identify each individual. 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 that in turn represents the relation. A corresponding attribute representing the reverse relation would need to be exhibited by Nokia, if the relation were to be reciprocal. 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 discernable 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 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.
Conceptual Entailment	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 bi-directional).

Table 2—Relevant Concepts from Chisholm's Ontology

Method

The role of our methods is to compare and to contrast the ontological meta-model embodied in the data modelling languages with a reference ontology (in this case Chisholm's ontology). In the discussion below we assume that we have not yet selected the ontology to be used and refer to a fixed but arbitrary ontology as our reference ontology. The reader can however, consider Chisholm's ontology as our intended target.

We present two methods for evaluating data modelling languages against a specific ontology: (1) the method of conceptual evaluation which can be applied to each specific data modelling language, and in turn forms the basis of the second method which is, (2) the method of

conceptual comparison. The relationship between the two is explained in this section. In this paper we present the results of an application of the latter method.

The Method of Conceptual Evaluation

The aim of the method of conceptual evaluation is to compare the ontology embodied in a data modelling language with the reference ontology selected from the range of ontologies available. In conducting a conceptual evaluation we are seeking to provide qualitative answers, for specific data modelling languages, to questions such as:
- How well does the data modelling language capture reality relative to an ontology?
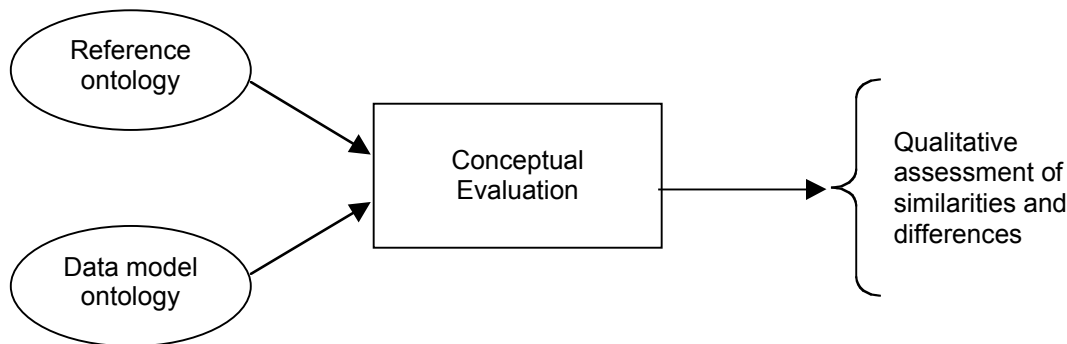- How similar are a range of data modelling languages?

Figure 2 – the method of conceptual evaluation

As indicated in Figure 2 the inputs to the method of conceptual evaluation are the reference ontology and the ontology derived from the meta-language of the data modelling language. The output of the method is a list of similarities and differences between the two sets of concepts and a qualitative analysis of those similarities and differences. It is not necessary for either the ontology or the data modelling language to be described using a mathematical formalism. It is possible that both are at best semi formal with natural language descriptions of concepts found in each.

The method of conceptual evaluation has four basic steps. Step 1: Determine the set of concepts from the reference ontology to be used in a forward evaluation. This set of concepts we call the *reference concepts*. Step 2: Determine the set of concepts from the ontology embodied in the data modelling language to be used in a backward evaluation. This set of concepts we call the *data modelling concepts*. Step 3: Perform a forward and backward evaluation of the two sets of concepts and tabulate the results. Step 4: Perform the analysis step in which the results are analysed. We explain the steps below.

The first step is to determine the basic set of concepts on which the forward evaluation will be based. The method does not prescribe which set of concepts from the reference ontology should be chosen. For example, one may wish to study the concept of state in OMT (Rumbaugh, Blaha, Premerlani, Eddy, & Lorensen, 1991) by reference to Chisholm's ontology, and consequently the fundamental concepts of states and events from Chisholm's ontology may be the only relevant concepts that need to be considered for such a limited study. Nevertheless, the chosen concepts must be appropriate for the modelling language under study. In this study, for example, only the static or structural concepts are required because that is the common nature of the data modelling languages under examination.

The second step resembles the first and involves determining the set of concepts from a particular data modelling language. Each data modelling language will have a different group of concepts using which sense is made of reality. For example, the ER model uses different terms from that used by OMT. It is likely that there will be a degree of similarity in the concepts associated with terms from those languages.

The third step involves the comparison of concepts from each of the reference ontology and the ontology embodied in concepts from a data modelling language. It is performed utilising concepts from the data modelling language as well as the reference ontology according to our philosophy – the reference ontology is not the only ontology that could be chosen and is not the only theory in that other reference ontologies could be used. Nor is the ontology *necessarily* better than that embodied in any data modelling language. Further, the comparison is at the level of concepts thus moving beyond the specific names or terms used to signify the concepts. Additionally, this step is highly subjective – there is no other way to undertake a conceptual evaluation of this nature.

The presentation of the results of the evaluation utilises semiotic theory for two reasons. Firstly, terms and concepts are clearly semiotically related. Secondly, comparison of concepts is semantic with semiotic theory providing an ideal basis for explaining semantic differences in terms.

The relationship between terms in an ontology and their concepts are explained through semiotics: Each term through its associated concept in a reference ontology or the ontology of a data modelling language, spans part of a *semantic field* (Eco, 1976), or *conceptual plane* (Cruse, 2000; Culler, 1976). Alternatively, each term from an ontology possesses an *essential depth* (Liska, 1996) which similarly evokes the conceptual span of a term. In this paper we adopt the term 'semantic field' to label these ideas and use it to express the similarities and differences between concepts in the reference ontology and those embodying the ontology from the data modelling language. Specifically, we use a graded indicator to express the similarities and difference. 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, two concepts 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. Finally, it may be that there is no overlap at all between the data modelling language and c from the ontology.

Figure 3 shows the three categories of results 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 boundaries between semantic fields are quite imprecise. This emphasises the fact that the comparison is conceptual and that concepts may be partially covered and that a simple presence absence is not ideal for ontological evaluations of this nature.
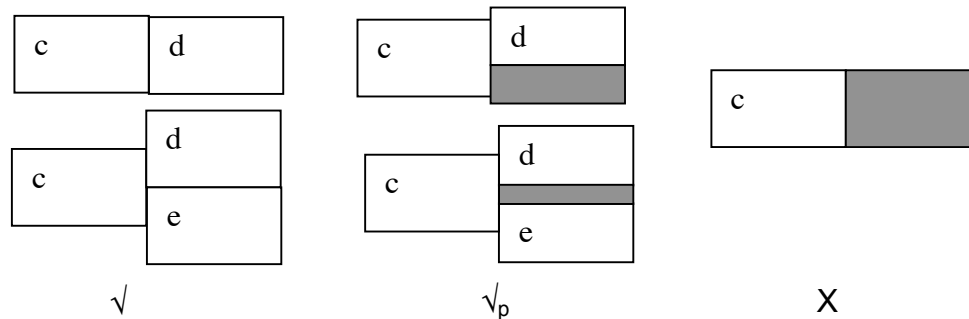
Figure 3—Degree of Overlap in Coverage of Semantic Field

Each of these categories of results can be indicated using symbols so that an idea of the results of the comparison can be conveyed easily in tabular form. This is called the indicative results. The three symbols for full coverage, partial coverage and no coverage are ($\sqrt{}$) ($\sqrt{}_p$) and (X) respectively.

For compound concepts a summary for the complete concepts can be calculated. This is shown in Table 3 below. The summary (concept level) *result* is shown as the 'additive of results' for the parts of the complex concept made up of parts *a* and *b*.

Concept					
Result	X	$\sqrt{}_p$	$\sqrt{}_p$	$\sqrt{}_p$	$\sqrt{}$
Part A	X	$\sqrt{}$	$\sqrt{}$	$\sqrt{}_p$	$\sqrt{}$
Part B	X	$\sqrt{}_p$	X	$\sqrt{}_p$	$\sqrt{}$

Key: X – no coverage   √p – partial coverage   √ – full coverage

Table 3—results for a compound concept, with parts *a* and *b*

The second dimension of final step in the method is the qualitative result of evaluating a data modelling language using an ontology revealing the story behind the indicative results from step 3. The analysis of the qualitative results presents issues beyond the direct comparison of concepts and discusses issues such as the nature of the gaps in coverage that are evident from the results as presented in Step 3 and the implications of these on the data modelling language under study.

Our method is related to work by Wand and Weber (Wand & Weber, 1989) where they undertook an informal comparison between tools and an ontology. This informal comparison later progressed to be a more formalised (Wand & Weber, 1993) understanding of the representational clarity with which a model of reality is created from human perceptions using a specific tool. In this, they examine the grammars that information systems analysis and design methods provide to describe aspects of the real world. A grammar in this context "generates a language, which is a set of strings over some alphabet. … In these grammars, sentences provide a graphical representation of some real-world phenomena" (Wand & Weber, 1993).

In their latter more formal treatment, Wand and Weber use ideas of ontological clarity and ontological completeness to establish a measure of the ontological expressiveness of an analysis and design tool as represented by its *grammar* as it compares with the *grammar* representing the ontology. These measures are defined using construct (or term) mapping to and from the ontology for clarity and completeness respectively. Their measures are based on the presence or absence of terms in a grammar representing the modelling tool when compared with terms in a grammar from the ontology. Both grammars and terms are required to be expressed mathematically.

These concepts bear a degree of similarity to those used in our method. However, they do not explore the more fundamental question of the *qualitative* differences or similarities in world-view between the ontology and the various tools under examination that are uncovered by examining the subtle differences in *meaning* of the various terms found in the ontology and in the tools. Instead their comparison is based on the mapping (or failure in mapping) of mathematical constructs to and from the ontology. Their approach also requires that the ontology selected be precise and defined mathematically. Not all ontologies are capable of being defined mathematically due to mathematics failing to adequately represent reality. The modelling tool under investigation must similarly be precisely defined. Formalising the modelling tool in a grammar compromises the meaning attached to terms. Terms may have their meaning restricted by formalising them. Data modelling languages often lack formality. It is for these reasons that we have included semiotic theory to express the results of each ontological evaluation of data model. We further note that as a result, and of necessity, the method is highly qualitative and has a degree of subjectivity.

The Method of Conceptual Comparison

The method of conceptual comparison seeks to compare a number of data modelling languages by analysing the results of conducting a series of conceptual evaluations against the selected reference ontology. The method consists of repeated applications of the method of conceptual evaluation against a number of data modelling languages. The results indicate the degree to which the reference ontology is reflected in the ontology of a range of data modelling languages and utilises the reference ontology as a benchmark against which the data modelling languages can be assessed. In conducting the series of conceptual evaluations we are testing each language against the selected and independent view of reality as represented by a reference ontology.

As a direct consequence, the method of conceptual comparison can be used to determine how wide spread and to what degree a reference ontology is reflected in the ontologies implicit in a range of data modelling languages. The analysis of results sheds light on the ontological overlap or dominance of the reference ontology with a range of data modelling languages.

In the following section we present the results of such a conceptual comparison with a range of data modelling languages the range of languages spanning twenty years of scholarship. In accordance with the method, it is constructed from a number of conceptual evaluations using Chisholm's ontology and the ontology implicit in the respective data modelling language.

Results

We selected five representative data modelling languages for a conceptual comparison with Chisholm's ontology. These data modelling languages span the period from the beginning of the semantic data modelling to its extension into the world of object data modelling. Apart from the classical Entity-Relationship (ER) model (Chen, 1976), we also selected the Functional Data Model (FDM) (Kerschberg & Pacheco, 1976; Shipman, 1981), the Semantic Data Model (SDM) (Hammer & McLeod, 1981), NIAM (Nijssen & Halpin, 1989), and Object Modelling Technique (OMT) (Blaha & Premerlani, 1998).

Table 4 shows the indicative results for the comparison of Chisholm's ontology with the data modelling languages.

Ontological Concept		ER	FDM	SDM	NIAM	OMT
Individual		√p	√p	√	√p	√
	Core	√	√	√	√	√
	Identity	√	√	√	√	√
	Structure	X	√p	√	X	√
Attribute		√p	√p	√p	√p	√p
	Core	√p	√	√p	√	√p
	Conc. Entailment	X	X	X	X	X
	Complexity	√p	√	√	√p	√
Classification		√p	√	√p	√p	√p
Relation		√p	√	√	√p	√p

Key: X – no coverage   √p – partial coverage   √ – full coverage

Table 4—Indicative Results of the Comparison of Selected Data Modelling Languages Using Chisholm's Ontology

Summarising from an earlier section, Chisholm's ontology views the world as a collection of individuals and relations between them. Individuals have structure and represent ontologically distinct entities. Attributes in the ontology are used to describe and identify individuals and using identity of individuals, describe relations. Further, attributes are *universals* in a philosophical sense and endure, and, by inference they are loosely coupled with individuals. Attributes are also used to determine class and set membership. Relations are also seen as being uni-directional.

Our conceptual comparison, summarised by Table 4, suggests that the world-view described by the reference ontology is to a large extent a similar world-view to 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 departures are in the structural aspects of individuals and more subtly, attitudes to the nature of attributes and relations and the implications of a lacking of loose coupling between individuals and attributes (particularly implications concerning classification). We examine each in turn.

Individual structure represented by part-whole relationships defined at the individual level is supported by some but not all data modelling languages. Those modelling languages not showing support provide a crude mereology at the class level. This approach is useful in cases where structure cannot be generalised to the class level due to a high degree of diversity in the structure of individuals grouped. Apart from this difference the remainder of the concept 'individual' is supported.

The core of the concept 'attribute' is partly supported by most data modelling languages. This is due to the lack of support by most data modelling languages of loose coupling between attributes and individuals by three data modelling languages. Loose coupling allows for a high degree of diversity in the types of attributes that similar individuals exemplify. Consequently, individuals where such diversity is evident are no longer constrained to classes of like individuals all exemplifying a fixed range of types of attributes. Some data modelling languages (FDM and NIAM) support the loose coupling evident in Chisholm's approach; the remainder do not. Nevertheless most have a high degree of coverage of the core and complexity parts of attribute. Attribute equivalence is completely absent from all data modelling languages.

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. Essentially, in the ontology there exists a sea of individuals from which classes are built. This provides flexibility in cases where the classes of individuals that are required change drastically without very much changing in the nature of the individuals themselves.

The concept of relations between individuals is supported by all data modelling languages. However, the reference ontology also requires relations to be unidirectional, thus allowing for non-reciprocation of relations. We have found that relations are not bi-directional in several of the data modelling languages but the concept is fully supported by FDM and SDM.

The consequence of the departures from the ontology by the data modelling languages that we have observed is that it is likely one can model a narrower range of situations using the studied data modelling languages than Chisholm's ontology. Further, Chisholm's ontology has the potential to change our view of data modelling by its increased flexibility achieved through bi-directional relations and through its loose-coupling of attributes with respect to individuals. In turn, this has positive implications for the flexibility of models that are subject to radical change. It is the formation of classes through attributes as a direct consequence of loose coupling that is most beneficial for flexibility.

We can see from the results that the modelling languages share the world-view of the ontology to a large degree. The areas of departure are of the nature of a difference in emphasis rather than complete absence of support. In the case of complex concepts all modelling languages support the core to a high degree of coverage. On the basis of this, we can say that the world-view held through the ontology is substantially similar to that held by this representative range of data modelling languages.

Conclusions, Future Work and Reflection

In this work we have compared Chisholm's ontology with the ontologies implicit in a representative range of data modelling languages using qualitative methods. We have found that there is a significant degree of overlap between Chisholm's ontology and the data modelling languages selected. Indeed, we have found that the ontology's core elements are reflected in the range of data modelling languages selected. Recall that Chisholm's ontology consists of individuals and their structure, the attributes that they exemplify and the relationships that exist between individuals.

At the beginning of the paper we posed two research questions. They were:
- how well do data models represent reality, and
- what are the similarities and differences between data modelling languages.

We discuss these in the following sub-sections.

Representing reality

We can conclude based on the results presented in this paper that the data models generally overlap with the core concepts of Chisholm's ontology to a large degree and that the world-view encapsulated in Chisholm's ontology is broadly consistent with the world-view implicit in the data modelling languages at least as far as the terms used for comparison are concerned.

Chisholm's ontology is one of commonsense realism and is categorised by Chisholm as being a realistic ontology. We have found, in our reading (Audi, 1995; Flew, 1989; Honderich, 1995), that the terms used by Chisholm are widely supported, and the style of realism upon

which his ontology is based has a high degree of consensus between philosophers. We are confident that this core of consensus, which is considerable, forms a good starting point from which to progress towards a detailed view of reality that is shared between the data modelling languages. We have found the following forms such a consensus and can be summarized as a realistic core:

       1) Individuals that are ontologically independent
       2) Attributes that are exemplified by individuals
       3) Part-whole relationships exist between individuals (mereology)
       4) Relations exist between individuals
       5) Classes of individuals are selected on the attributes exemplified

       All of the data modelling languages showed good support for all of these concepts. Consequently, we conclude that the data modelling languages studied, by having a large degree of overlap with a typical commonsense realist ontology, represent reality well.

<u>Similarities and differences between data modelling languages</u>

       The similarities between the data modelling languages is summarized by the realistic core (listed 1-5 above.) However, we can see three areas where there are differences between the modelling languages and the view of Chisholm's ontology.

       Firstly, in the ontology, attributes are considered to be quite separate from the individuals that exemplify them, and they endure. One can describe this as loose coupling between individuals and attributes. Some, but not all, of the data modelling languages contrastingly group individuals into homogeneous classes and by so doing restrict the range of attributes that each individual can exhibit thus exhibiting a tight coupling. This tight coupling is also counter to the realistic core.

       Secondly, in data modelling languages the data modelling equivalent of individuals are seldom allowed to be members of several classes simultaneously. The only exception to this is in cases where a class hierarchy around a common family of classes (with OMT) is established or where individuals are distributed or dispersed across several classes (with SDM). Both of these are different from the ontology in which individuals can be member of more than one class simultaneously because classes are established based upon attributes that member individuals exhibit. The classes are not repositories for individuals, and instead they filter individuals.

       Thirdly, in contrast to class hierarchies there is support for part-whole (or aggregation) structures in the ontology, known in philosophy as mereology. This also supports related research that uses a different ontology. Many of the modelling frameworks supported this to at least a crude degree, such as with the ER modelling framework where the part-whole relationships are expressed at the class level. Others had more sophisticated methods of providing this, such as with most object modelling frameworks or with SDM where individual-based mereology is present.

<u>Future Work and Critical Analysis</u>

       There are some issues that need further investigation. The importance of conceptual entailment and the implications of the enduring and abstract nature of attributes need to be investigated in practical situations because of their potential to influence efficiency and effectiveness of implemented databases. There is also a question about the nature and usefulness of rigid class hierarchies found in many modelling languages.

The ontological evaluation of 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 condensed as Chisholm's and to relate each term and its associated concept to modelling languages. We cannot conceive of an easier way in which to undertake such a work. Further, there is clearly a critically important interpretive dimension 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 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 by permitting analysis using qualitative ontologies such as Chisholm's.

References

Audi, R. (Ed.). (1995). *The Cambridge Dictionary of Philosophy*. Cambridge: Cambridge University Press.

Blaha, M., & Premerlani, W. (1998). *Object-Oriented Modeling and Design for Database Applications*. Upper Saddle River: Prentice Hall.

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

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

Chen, P. (1976). The Entity-Relationship Model—Toward a Unified View of Data. *ACM Transactions on Database Systems, 1*(1), 9–36.

Chisholm, R. (1957). *Perceiving: A Philosophical Study*. Ithaca: Cornell University Press.

Chisholm, R. (1976). *Person and Object: A Metaphysical Study*. La Salle: Open Court.

Chisholm, R. (1979). Objects and persons: revisions and replies. *Grazer Philosophische Studien, 7/8*, 317–388.

Chisholm, R. (1982). *The Foundations of Knowing*. Minneapolis: University of Minnesota Press.

Chisholm, R. (1989a). *On Metaphysics*. Minneapolis: University of Minnesota Press.

Chisholm, R. (1989b). *Theory of Knowledge* (3rd ed.). Englewood Cliffs: Prentice-Hall.

Chisholm, R. (1992). The Basic Ontological Categories. In K. Mulligan (Ed.), *Language, Truth, and Ontology* (pp. 211). Dordrecht: Kluwer Academic Publishers.

Chisholm, R. (1996). *A Realistic Theory of Categories—An Essay on Ontology* (1 ed.): Cambridge University Press.

Cruse, D. A. (2000). *Meaning in Language: An Introduction to Semantics and Pragmatics*. Oxford: Oxford University Press.

Culler, J. (1976). *Saussure*: Fontana.

Dancy, J., & Sosa, E. (Eds.). (1992). *A Companion to Epistemology*. Oxford: Blackwell Publishers.

Eco, U. (1976). *A Theory of Semiotics*. Bloomington: Midland.

Flew, A. (1989). *An Introduction to Western Philosophy: Ideas and Arguments from Plato to Popper* (Fully revised edition of the original 1971 volume ed.). London: Thames and Hudson.

Green, P. (1996). *An Ontological Analysis of Information Systems Analysis and Design (ISAD) Grammars in Upper Case Tools*. Unpublished Unpublished PhD Thesis, The University of Queensland.

Hammer, M., & McLeod, D. (1981). Database Description with SDM: A Semantic Database Model. *ACM Transactions on Database Systems, 6*(3), 351–386.

Honderich, T. (Ed.). (1995). *The Oxford Companion to Philosophy*. Oxford: Oxford University Press.

Kerschberg, L., & Pacheco, J. E. S. (1976). *A functional database model*. Rio de Janeiro, Brazil: Pontificia Univ. Catholica do Rio de Janeiro.

Liska, J. J. (1996). *A General Introduction to the Semeiotic of Charles Sanders Peirce*. Bloomington, USA: Indiana University Press.

Milton, S. K. (2000). *Ontological Studies of Data Modelling Languages*. Unpublished PhD Dissertation, The University of Tasmania.

Milton, S. K., Kazmierczak, E., & Keen, C. (2001). *Data Modelling Languages: An Ontological Study*. Paper presented at the 9th European Conference on Information Systems, Bled, Slovenia.

Nijssen, G. M., & Halpin, T. A. (1989). *Conceptual Schema and Relational Database Design: A Fact Oriented Approach*. New York: Prentice-Hall.

Quine, W. V. O. (1960). *Word and Object*. Cambridge Massachusetts: MIT Press.

Rohde, F. (1995). An Ontological Evaluation of Jackson's System Development Model. *Australian Journal of Information Systems, 2*(2), 77–87.

Rumbaugh, J., Blaha, M., Premerlani, W., Eddy, F., & Lorensen, W. (1991). *Object-Oriented Modeling and Design*. Englewood Cliffs, New Jersey: Prentice-Hall.

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), 140–173.

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

Thalheim, B. (2000). *Entity-Relationship Modeling: Foundations of Database Technology*. Berlin: Springer-Verlag.

Wand, Y. (1996). Ontology as a Foundation for Meta-modelling and Method Engineering. *Information and Technology Software, 38*, 182–287.

Wand, Y., & Weber, R. (1989). An Ontological Evaluation of Systems Analysis and Design Methods. In E. D. Falkenberg & P. Lindgreen (Eds.), *Information Systems Concepts: An In-depth Analysis* (pp. 79–107). Amsterdam: Elsevier Science Publishers B.V.

Wand, Y., & Weber, R. (1990). An Ontological Model of an Information System. *IEEE Transactions on Software Engineering, 16*(11), 1282–1292.

Wand, Y., & Weber, R. (1993). On the Ontological Expressiveness of Information Systems Analysis and Design Grammars. *Journal of Information Systems, 1993*(3), 217–237.

Weber, R. (1997). *Ontological Foundations of Information Systems* (Vol. Monograph #4). Blackburn, Victoria: Buscombe Vicprint.