COMPARING DATA MODELLING FRAMEWORKS USING CHISHOLM’S ONTOLOGY

Simon Milton
Management Information Systems, Deakin University
Simon.Milton@deakin.edu.au

Ed Kazmierczak
Computer Science and Software Engineering
The University of Melbourne
eka@cs.mu.oz.au

Chris Keen
Information Systems, The University of Tasmania
Chris.Keen@utas.edu.au

ABSTRACT

Data modelling frameworks consist of a series of constructs which are used to create an abstraction of the world. Ontologies consist of concepts through which reality can be described. The aim of this paper is to use an ontology as a basis for comparing and evaluating data modelling frameworks. We begin by establishing a method for comparing modelling frameworks using ontology, and then examine the role Chisholm’s ontology can play in such a comparison. We then perform a comparison of three semantic data models: the Entity-Relationship Model, the NIAM model, and the Functional Data Model.

In undertaking the comparison we introduce Chisholm’s ontology, which is a new ontology in the Information Systems literature. Chisholm’s ontology has given us new insights into the role of attributes and into the changing nature of relationships, although in this paper we do not discuss the latter. All the data modelling methods supported the ontology to varying extents, with the Functional Data Model offering the best support. However we found that all the models lacked the full generality of Chisholm’s ontology. The disagreement between current data models and Chisholm’s ontology provides insight into the requirements for a truly flexible data modelling technique.

1. INTRODUCTION

Data modelling involves establishing an abstraction of data and information in the world which is acceptable and relevant for an organisation and its purposes.
Many modelling frameworks have been proposed during the last two decades, beginning with semantic models (see (Hull and King 1987; Peckham and Maryanski 1988)) and progressing through to object models (see (Banerjee et al. 1987; Cattell 1991; Dobbie 1990; Lécluse et al. 1988; Rumbaugh et al. 1991)).

In this paper we consider ontology as the basis for evaluating data models. Ontology is useful for this purpose because of its role in Philosophy—that of considering the “furniture of the world.” Philosophers who construct ontologies seek ideal concepts which are needed to construct an understanding of the real world. It is in this way that ontologies perform an analogous role to that played by data modelling frameworks. Our motivation for studying ontologies is that we are interested in determining the effectiveness of specific data models in modelling the real world.

The aim of this paper is to compare a number of data models against a selected ontology, and thereby evaluate their effectiveness. Specifically we identify a candidate ontology. We then adapt Mill’s methods of agreement and difference (Neuman 1997) to compare the three data models against the ontology. The chosen ontology is used to determine the important features on which the comparison is based.

Previous comparisons of data models have not used ontologies as a basis for comparison, for example, the surveys in (Hull and King 1987; Peckham and Maryanski 1988). Previous work involving ontology in information systems concerned systems analysis and design (Wand 1996, Wand et al. 1995; Wand and Weber 1990; Wand and Weber 1993; Wand and Weber 1995). Our needs of an ontology differ from Wand and Weber’s in that we are considering only data modelling frameworks for database design and consequently we are not concerned with overall system issues. In considering data models we are chiefly interested in discovering important elements which need to be present in a data model in order that it can successfully represent the world as perceived by an organisation.

This paper begins by considering our method for evaluating data models. We then describe Chisholm’s ontology and three data modelling frameworks. Finally we compare the selected models with Chisholm’s ontology using our adapted version of the method of agreement and method of difference.

2. A METHOD FOR EVALUATING DATA MODELS

2.1 Adapting Mill’s methods of agreement and difference

We will use an adaptation of Mill’s methods of agreement and difference to compare data models (Neuman 1997). In these methods, social cases are inspected for similarities and differences respectively to establish causal links in qualitative data. Further, the inferred causations may be extended to form limited generalisations. The methods both involve determining a set of qualitative features and a set of characteristics. Features are elements of a situation which the researcher believes may be important in distinguishing
between cases. Characteristics are the set of properties that the features can exhibit in a given situation. In the method of agreement the researcher seeks to uncover common characteristics for a given outcome in order to determine which features are likely causes for that outcome. In the method of difference the researcher seeks to verify that for different outcomes, different characteristics are observed across the same features.

Data modelling frameworks concern data modelling concepts. Each of these concepts are by their nature indistinct and subject to interpretation, that is, in a given situation two modellers may construct quite different but equally valid models. Our interest is in a qualitative comparison of data modelling concepts offered by the frameworks. That is, we are interested in the extent to which a data modelling concept can capture aspects of a real world situation.

Modification of the method is required because we seek to compare data modelling frameworks (through the concepts that they provide) with an ontology and are not seeking to find causal relationships. We modify Mill’s methods of agreement and difference by seeking a graded measure of agreement with the chosen ontology, where the features of this comparison are selected from the ontology. We use a graded measure because Mill’s methods simply provides us with yes/no, or binary information; Either the model contains the concept or it does not. We are seeking qualitative measure of agreement/difference. This is because a framework may contain a feature but not in the full generality provided by the ontology. Consequently, the characteristics are that the data modelling framework supports the feature, does not support the feature, or supports the feature through modification or to some extent.

2.2 Choosing an ontology

An ontology describes what concepts or entities make up reality and to this end, no ontology is better than another. Many ontologies have been proposed but we are only interested in those which can be used to compare data modelling frameworks.

We are guided in the choice of an ontology by the use to which it is to be put. Firstly, our selected ontology should be concerned with identifying elements of a similar nature to data models. Secondly, it should not be set at a level which loses all conceptual reference by being too microscopic. Thirdly, it must be involved with establishing an understanding of how significant elements are linked or related. These requirements are difficult to formalise, but may be used by an investigator to help select an ontology.

We have found that the ontology recently published by Chisholm (Chisholm 1992; Chisholm 1996) serves our purposes best. Two others stand out as being significant (Bunge 1977; Bunge 1979; Kattsoff 1956). Bunge (Bunge 1977; Bunge 1979) has a clear and unambiguous scientific objective, that of describing quantum mechanical and other natural phenomena. We examined it for our purposes because of its past use in other research into information systems. Wand and Weber (Wand 1989; Wand 1996; Wand et al. 1995; Wand and
3. CHISHOLM'S ONTOLOGY

In this section we introduce Chisholm’s ontology (Chisholm 1992; Chisholm 1996). It is a daunting task to summarise an ontology in a paper and so we restrict ourselves to the elements which are fundamental to the ontology. Moreover emphasis is given to those elements which have a clear role in the comparison. Before considering the elements we introduce the categories established by Chisholm in his ontology. We do this because it gives us structure into which the elements can be seen to fit. The discussion of these categories is not critical for an understanding of the ontology. We conclude with a discussion of the pragmatics of using Chisholm’s ontology with data models. Before we commence it is worth noting that Chisholm uses *entity* differently from its usage in data modelling frameworks. In Chisholm’s ontology, entities describe all things which make up the furniture of the world, with *individuals* representing significant substances within that world.

3.1 Categories

Chisholm divides the world of entities into contingent entities which don’t have to exist, and necessary entities which always exist. The taxonomy is shown in Figure 1.

Necessary entities encompass attributes, necessary substances and their states. Attributes and (necessary) substances are examples of enduring things and do not pass away or come into being. Observe that the existence of necessary substances has clear theological ramifications.

![Figure 1. Chisholm’s categories](image-url)
The sub-categories of contingent entities consists of individuals, and their states. Individuals can either be substances, or boundaries. Boundaries are spatial elements which define the bounds of individual substances. Substances can encompass physical things as well as legendary (fictitious) or conceptual things. Our interest will be occupied by individual substances, which may be said to be in certain states. In turn events may be described as being a sub-category of states. These are not central to our comparison to existing data models, as our comparison does not concern dynamic issues. Dynamic issues will be examined in future work.

### 3.2 Attributes

Attributes are the key ontological feature. Individuals have attributes associated with them by the intent and belief of observers. If an observer believes that an individual has a particular attribute, and that belief is shared, then that individual is said to **exemplify** that attribute. This is called the *primacy of intention* because intent, or belief, is the fundamental notion in the ontology. It is not necessary that an individual has an attribute, but rather that someone believes that an individual has an attribute—a subjectivism. Some attributes are exemplified, others are not and some cannot be exemplified.

One can say that an individual exemplifies Being-\(F\). In this sense \(F\) is a predicate which may be replaced by an English predicate. Additionally, propositions in English can be reduced to attributes. If a proposition is true, then there is an attribute which establishes the truth of the proposition. A fact is a true proposition. We can represent both facts and propositions using attributes.

Attributes may be related through conceptual entailment. \(P\) conceptually entails \(Q\) can be interpreted to mean that if \(P\) is an attribute which has been determined to be exemplified in an individual then \(Q\) is an attribute which is also exemplified in that individual.

Attributes may be simple or compound. Compound attributes are formed through disjunctions and conjunctions of compound or simple attributes. A conjunction of attributes conceptually entails, and exemplifies, each of the attributes in the conjunction simultaneously. A disjunction of attributes exemplifies one attribute or the other attribute but not both simultaneously.

### 3.3 Relations and classes

Relations exist between individuals. Relations need not be reciprocated, although many are. This means that relations are directed from one individual to another.

Relations are represented by an attribute which is an ordered pair of attributes (which can be compound). Indeed the attribute representing the relation is said to order things. An attribute relating one thing to another, consists of an
attribute which uniquely describes the first individual and an attribute which uniquely describes the second individual. Moreover, one can have dyadic (two term), triadic (three term), tetradic (four term), and in general n-ary (n term) relations, but all of these relations can be represented as a series of dyadic relations.

Classes and sets are also represented in Chisholm’s ontology using attributes. Chisholm shows that discussions involving classes and sets can be reduced to discussions involving attributes. Indeed, an attribute (which may be compound), can collectively establish a class. This was suggested first by Russell (Russell 1908), modified by Carnap (Carnap 1963) and presented in the Ontology:

"The class of x's such that x is F is so and so" may be replaced by "For every attribute P, if P and being-F are exemplified by the same things, then P is so and so."

There is a set of axioms which establish the fundamental properties of sets and classes in Chisholm’s ontology, but for the sake of brevity we omit them here. Interested readers are referred to the ontology (Chisholm 1996).

Chisholm’s ontology also considers theology, perceptions and appearances, stories, times and the temporal, and events. Issues concerning times and the temporal as well as events will be covered later in this paper, when we discuss extensions to data modelling capabilities.

### 3.4 Pragmatics of data modelling

When modelling data, several issues are important. Firstly, a unified view of data for a group must be established. This view reflects the outlook and ambitions for that group. Secondly, different sub-groups within the group have different views to place upon the model. Thirdly, the model must adhere to business rules (or more generally rules to which data must conform as established by the organisation).

Models are important for prediction and analysis, and as such must be able to provide answers to a number of questions of interest to, or required by, the group. Finally a model must be extensible. This calls for a degree of flexibility in any modelling framework.

Chisholm’s ontology provides for individuals, attributes/properties, relations (as attributes), classes/sets (reduced to attributes), and events/states and data modelling frameworks possess many, or all, of these. Further, different sub-groups may believe that the same individual exemplifies different attributes and consequently different sub-groups may view the individual differently. Multiple views are sustained in Chisholm’s ontology by using attributes in different ways. There are enough constructs within Chisholm’s ontology regarding attributes for this to be possible.
Business or social rules by which the group operates can be catered for either in the structure of a data model, or in the dynamics of a data model. Most ontologies deal with either dynamic, or static issues but seldom both. Chisholm’s ontology deals with both static and dynamic concerns.

Answers to questions of interest to a group can be found in Chisholm’s ontology because Chisholm’s ontology allows for transitive relations between individuals in the same way that data models do.

Finally, Chisholm’s ontology allows for change, for example, individuals can change the attributes they exemplify and classes can change their composition. This is in part made possible through treating attributes as fundamental elements in the ontology and by describing state and change.

### 4. THE DATA MODELS

We have selected three semantic data modelling frameworks from the literature. The Entity Relationship Model (ER) (Chen 1976), the Functional Data Model (FDM) (Shipman 1981), and NIAM (Nijssen and Halpin 1989). These cover a selection of modelling schemes from entity-attribute-relationship approaches through to fact-based approaches.

In each case we describe the data modelling framework as tersely as possible. Should readers desire more detail, then they are directed to the references given.

#### 4.1 Entity-Relationship Model

Entity-Relationship models (sometimes called Entity-Attribute-Relationship models) were first proposed in the late 1960s, and the most popular of which is the model proposed by Chen (Chen 1976). Chen’s model has since been augmented with extra constructs (Teorey et al. 1986) to form the Extended Entity-Relationship Model.

E-R models are centred around three fundamentals concepts: entity-type, relationship-type, and attribute-type. E-R is popular because of its easy to understand concepts but suffers from the drawback of rigidity. Members of entity-type are identifiable objects (either physical or conceptual) from the world. Some E-R approaches allow disjoint entity sets—entities may be members of more than one entity type. Attributes are properties held by all members of a given entity-type or relationship-type. Relationship-types are defined over one or more entity-types. In other words these relate entities. This is a concept which is fundamental to all E-R approaches.

Relationship-types are of several dimensions, characterised by the number of entity-types allowed to participate in the relationship-type. Additionally functionality (1-1, 1-m, m-n nature of the relationship-type) together with partial or total participation in a relationship is described as cardinality (min, max levels of participation in relationships for members of entity-types).
4.2 Functional Data Model

The functional data model consists of entities and functions. Entities are identifiable objects (as with Entity Relationship models) but also include properties. Functions map entities to entities and are used to express a relationship between entities or to associate attributes with entities. These relationships are uni-directional. Bi-directional relationships are modelled using inverse functions.

Observe that in the Functional Data Model attributes are separate from entities and the association of attributes to entities is achieved via a function.

Generalisation/specialisation is achieved through sub-type super-type definitions thus establishing a hierarchy. Operators (union, intersection, difference) enable the construction of lattices of entity types.

One key strength of FDM is its simplicity. All attributes and relationships are described using functions.

4.3 NIAM

NIAM (Nijssen and Halpin 1989) is a fact oriented design method. A subset of the method concerns data modelling.

A NIAM data model consists of facts, constraints, and derivation rules. Facts concern elements of a data model, including entities, roles (relationships), and labels (properties or attributes). Scenarios are used to ascertain the players and roles in a typical situation in order to establish facts about entities, roles, and labels. They are encapsulated in typical descriptions or statements. These facts may be in the form of relationships between entities, or may be simple attributions (properties). Each of these are examples of roles. Simple properties and binary relationships don’t involve complications such as connectives or quantifiers.

5. RESULTS AND DISCUSSION

In this section we use Chisholm’s ontology to compare the chosen data models introduced in the last section. The comparison proceeds according to the method in Section 2.1, and is based on several features of Chisholm’s ontology. Chisholm’s ontology is also subject to the comparison, and it is expected that it fits all features exactly. The data models in turn are then examined to determine agreement/difference with Chisholm’s features.

Our first task is to determine features for comparison. Chisholm’s ontology establishes several features with a number of ancillary (or resultant) features. These are summarised in Table 1. The results of the comparison are summarised in Table 2.

The first feature is Entity Support (ES). All of the modelling frameworks studied support entities but these are qualitatively different from entities (individuals) in Chisholm. In the data models entities combine to form classes
of like individuals. Sometimes these classes are referred to as entities. In Chisholm’s ontology the individual is the fundamental element and classes of individuals are formed through attributes. Further work is needed to determine which method is better in practice.

Table 1. Features for the comparison

<table>
<thead>
<tr>
<th>Feature</th>
<th>Qualitative summary of emphasis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entity Support (ES)†</td>
<td>Model supports significant perceivable or observable entities or things which may be described in part through the perception of attributes or properties (below).</td>
</tr>
<tr>
<td>Attribute Support (AS)†</td>
<td>Attributes must be modelled with one overriding requirement—attributes are displayed by entities (and perceived by observers), but must not be tightly coupled with that entity. In other words, attributes form a fundamental and independent part of the modelling technique.</td>
</tr>
<tr>
<td>Attribute Construction (AC)†</td>
<td>Attributes must be able to be formed through the combination of other attributes. These combinations may be conjunctive or disjunctive.</td>
</tr>
<tr>
<td>Class/Set Support (CS)</td>
<td>The membership of classes or sets is based on attributes. Importantly, classes can be described through attributes. This means that entities can be members of several classes simultaneously and can be moved freely between classes.</td>
</tr>
<tr>
<td>Relationship Support (RS)</td>
<td>Relations must be attributes (or provided in the same way as attributes). Additionally, they must be directional.</td>
</tr>
</tbody>
</table>

† These form fundamental (or core) features. All others are resultant (or consequential)

Table 2. Results of the comparison

<table>
<thead>
<tr>
<th>Variable</th>
<th>E-R</th>
<th>FDM</th>
<th>NIAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>AS</td>
<td>*</td>
<td>√</td>
<td>^</td>
</tr>
<tr>
<td>AC</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>CS</td>
<td>X</td>
<td>*</td>
<td>X</td>
</tr>
<tr>
<td>RS</td>
<td>X</td>
<td>√</td>
<td>X</td>
</tr>
</tbody>
</table>

√—Support for the feature
X—No support for the feature
*—Support given, but attributes are tightly coupled to entities
•—Support not given, but could be provided relatively easily
^—Loose coupling is provided, but generally ignored in the method
The second feature is Attribute Support (AS). Chisholm’s ontology contains an important emphasis with respect to attributes. It is critical that attributes be *loosely* coupled with individuals. The only data modelling framework which supports loose coupling of attributes and individuals is the Functional Data Model. Both NIAM and the Entity-Relationship Model use attributes. The Entity-Relationship Model couples attributes tightly with entities (individuals) and relationships. NIAM on the other hand discusses decoupling attributes from entities but does not pursue this further in the method.

Furthermore, Chisholm argues that attributes are enduring. That is, it makes no sense to discuss the beginning or ending of attributes. It seems reasonable that the attributes themselves are enduring even if no observer perceives them for a long time. It is an entirely different issue than that of modellers being able to construct a data model with or without such a requirement of endurance. Furthermore, if we do take the view that we need to think of attributes as conceptually enduring, we need not *implement* a data model that way. If we take this interpretation of enduring attributes then the Functional Data Model supports enduring attributes, but NIAM and ER do not directly support enduring attributes. We are examining this issue in ongoing research.

Attributes in Chisholm’s ontology may have internal structure. We call this feature Attribute Construction (AC). Recall from Section 3.2 that this means that attributes may be simple or compound. All three data modelling frameworks could be extended to provide support for this feature as described in the ontology. The Entity-Relationship Model and NIAM both support a limited form of compound attributes. The Functional Data Model similarly offers limited support, but could in principal be extended to provide much better support for compound attributes because in the Functional Data Model attributes are decoupled from entities. Conceptual entailment is part of Chisholm’s ontology. None of the models studied support this feature. Although decoupling of attributes from individuals is important to support conceptual entailment.

None of the modelling frameworks support classes as envisaged by Chisholm (the feature Class Support - CS) although the Functional Data Model could be extended to provide this capability. In Chisholm’s ontology attributes determine membership of classes. A class structure could be imposed on a Functional Data Model by choosing the set of attributes which determine membership of the class and using the functions to determine the entities which are members. In contrast NIAM and the Entity-Relationship Model subordinate attributes with respect to entities and consequently entities determine attributes. Thus it is not easy to see how to add support for flexible classes based on attributes to these two data modelling frameworks.

The remaining feature from Chisholm’s ontology is Relationship Support (RS). Chisholm’s ontology views relations as attributes which are unidirectional. Only the Functional Data Model satisfies this requirement. It is difficult to see how relationships of this nature can be provided in NIAM or in the Entity-Relationship Model. Chisholm’s approach to relations raises questions regarding the criticality of modelling relations as attributes. It is
likely that a significant part of this issue is philosophical and to a lesser extent concerns modelling. Research into this is continuing.

Notwithstanding these issues, Chisholm’s ontology views the world as a collection of entities (individuals) and relations between them where relations change over time. The ontology uses attributes to describe both entities and relations. Most important, attributes are loosely coupled with entities. Our comparison suggests that this is to a large extent a similar world-view as those imparted by the data models, but the data models lack the full generality of Chisholm’s ontology.

6. CONCLUSIONS AND FUTURE WORK

We have found that there is a good level of agreement between the data modelling frameworks that we have studied and with Chisholm’s ontology. Some key differences have emerged especially with respect to the nature of entities and attributes.

The consequence of these differences is that one can model a wider range of situations using Chisholm’s ontology. Further, Chisholm’s ontology has the potential to change our view of data modelling by its flexibility. This flexibility is achieved through its relations and loose-coupling of attributes with respect to entities (individuals). In turn, this has positive implications for the flexibility of models which are subject to radical or ongoing change. We have found that E-R and NIAM do not support such flexibility. This is principally because of tight coupling between entities and attributes found in E-R and by practice in NIAM. We found, however, that the Functional Data Model captures the fundamental nature of Chisholm’s ontology more closely than the other models and has more potential to be able to support other elements presently not supported. Further research is needed to more deeply compare the Functional Data Model with Chisholm’s ontology.

REFERENCES


