Ontological Assumptions in Information Modeling

John M. Artz
The George Washington University, USA

INTRODUCTION

Information modeling is a technique by which a database designer develops a conceptual model of a database depicting the entity classes that will be represented in the database. There are three competing ontological assumptions that guide the modeling process. The broadest characterization of these assumptions is realism vs. conceptualism, with social realism occupying a middle ground. The realist believes that object classes exist in the real world, waiting to be discovered. The conceptualist believes that object classes are constructed in the mind of the modeler, based on observations about the application domain and the objectives of the information model. The social realist believes that classes exist as shared meanings among stakeholders in an application domain. This article explores these assumptions and then reviews selected literature in information modeling to determine which assumptions are held by key authors. It concludes that most authors hold inconsistent views, and this inconsistency provides some important insights into information modeling while presenting serious problems for practitioners and students of information modeling.

BACKGROUND

Perhaps one of the most perplexing problems in information modeling is the ontological status of entity classes. This raises the question, Do entity classes exist in the world, or are they constructed in the mind of the modeler? This seemingly esoteric question is important because the way in which one answers it has a significant impact on how one approaches the process of information modeling.

If entity classes exist in the world, independent of the mind of the observer, then the job of the information modeler is to discover those classes and record them in an information model. Hence, information modeling is a discovery process rather than a constructive process. If two modelers examine a domain and come up with different models, then one is wrong. One or the other (or possibly both) must bring their models into conformance with the real world. An information model can be validated by ensuring its conformance with the real world. This is the realist position. To the realist, the challenges in information modeling are how to discover the existing entity classes and accurately represent those classes in an information model. Validation is not a problem for the realist because a model is valid if it correctly represents the real world. Realism can be detected when writers use the phrases “real world” or “as they exist in the real world” or, more subtly, when they refer to “natural classes” or “natural data relationships.”

If classes do not exist in the world, then it is the job of the information modeler to construct them. Thus, information modeling becomes a process of construction rather than discovery. The conceptualist position holds that classes exist only in the mind of the observer and are constructed according to objectives (probably not explicit) that guide the process of abstraction. The modeler selects certain facts from the application domain and constructs classes based on individual objects with similar attributes. Conceptualists believe that different models of an application domain cannot be determined to be correct or not correct. They can only be more or less useful for meeting the model objectives. This leads to problems in both construction and validation of the information model. Construction is difficult because most of the literature on information modeling focuses on description of entities rather than construction of them. The literature is strangely silent on how to construct a set of entity classes to meet a set of modeling objectives. Validation is also a problem because the model cannot be compared with entities existing in the real world, it can be evaluated only with respect to the objectives of the model, again an area in which the literature is strangely silent. The conceptualist position creates serious problems for information modelers because it requires that modeling objectives be defined before model construction, and it requires some method of evaluating a model with respect to a set of objectives. Conceptualism can be detected in the literature when authors talk about “abstraction,” “the problem to be solved,” “objectives,” or the possibility of “multiple representations” or “multiple models.”

An intermediate position is social realism, which assumes that entity classes exist as shared meanings within a social context. This is a realist position in that the classes exist independent of the mind of the modeler.
Presumably if several modelers were to examine the same application domain they would eventually discover the same shared meanings and hence would produce the same information model. Validation is less problematic under the social realist position in that the model can be compared with the social reality, that is, the model agrees with what people in the application domain believe to be the entity classes or it does not. The social realist discovers the entity classes by talking with users and recording their usage of key words. Consensus is an important factor in the social realist approach because social reality is a shared understanding. If people do not agree on meanings, then the realism assumption breaks down because different modelers may very well come away with different understandings, depending on who they spoke to. Domain experts are important to the social realist position because the domain expert is the gatekeeper to the social reality. Social realism can be detected in the literature when authors talk about “language,” “shared meanings,” “domain or subject matter experts,” or modeling as a process of “consensus.”

Realism is a shaky assumption from a philosophical perspective, but desirable from a pragmatic perspective. If entity classes exist in the world, where do they reside? Although there are ample instances of an entity class, nobody has ever seen the class itself, nor will they. Entity classes exist only in the mind of the observer and have no real existence in the application domain. The modeler examines the application domain and, through a cognitive process of abstraction, derives a set of entity classes. Yet this process of abstraction is poorly understood and difficult to explain, so modelers act as though the classes actually exist in the world and are being discovered. From a pragmatic perspective, realism is a desirable assumption because it reduces the class construction process to one of simple discovery and provides an easy means of validation by requiring that the model simply conform to the real world.

Conceptualism is a much more justifiable position from a philosophical perspective, yet a nightmare in practice. Conceptualism recognizes the role of modeler and his or her cognition in the class construction process. Yet in practice it presents some severe problems. Since classes are constructed, how does the process of construction work? What criteria are used in class construction? Once classes are constructed, how do we know that the right classes have been constructed? One answer is that classes are constructed by grouping objects with similar attributes. But that position raises the question of whether or not attributes exist in the world and opens up, once again, the three positions just described with regard to attributes. Another answer might be to say that the classes are right if they meet the objectives of the model, but that answers the question by raising two more: How do we define modeling objectives and how do we determine if a set of classes meets those modeling objectives?

Most authors nod towards conceptualism, using terms such as problem solving or multiple models but back off when it comes to the actual process of modeling, where they will often fall back to a realist position by talking about modeling “the real world.” Recognizing the faultiness of the realist position, several authors have adopted an intermediate position of social realism. The more rigorous ones adopt social realism with respect to attributes, but the best that can be said is the literature is confusing and few authors have taken and articulated a consistent philosophical position.

**ANALYSIS OF THE LITERATURE**

The ontological assumptions made by practitioners are rarely articulated. They are more often manifest in their behavior. Practitioners may even claim to hold one belief while acting as though they held a conflicting view. Hence, in order to gain a sense of the variety of assumptions that are held in the field of information modeling, it is necessary to look at the recorded literature—widely read texts and papers—to see what assumptions are being put forth.

Peter Chen’s (1976) original article on the entity-relationship model begins by establishing a clearly realist perspective: “The entity-relationship model adopts the more natural view that the real world consists of entities and relationships. It incorporates some of the important semantic information about the real world” (p. 9-10). This perspective is picked up by later authors. Andleigh and Gretzinger (1992) claimed “the Entity Model describes the real-world relations for the information system” (p. 383), while Teorey (1990) referred to the model as including “the natural data relationships,” again a strong indication of realism. Yet only a couple of paragraphs later in Chen’s paper, he refers to a conceptual data model as “information concerning entities and relationships which exist in our minds” (p. 10) and “conceptual objects in our minds” (p. 14), showing a clearly conceptualist perspective. In discussing whether a given object should be an entity or a relationship, Chen defers to the enterprise administrator, who should decide so that “the distinction is suitable for his environment” (p. 10), suggesting an objectives-driven conceptualist view or a social-realist view, depending on the meaning of the word suitable. He goes on a bit later to say, “If we know an entity is in the entity EMPLOYEE, then we know that it has the properties common to other entities in the entity set” (p. 11), suggesting extreme class realism. Since nothing is said, in the
Related Content

Kyoung-Il Bae and Soon-Young Huh (2004). *Advanced Topics in Database Research, Volume 3* (pp. 272-293).
[www.igi-global.com/chapter/federated-process-framework-transparent-process/4364?camid=4v1a](www.igi-global.com/chapter/federated-process-framework-transparent-process/4364?camid=4v1a)

Evaluation of Component-Based Development Methods
[www.igi-global.com/chapter/evaluation-component-based-development-methods/23021?camid=4v1a](www.igi-global.com/chapter/evaluation-component-based-development-methods/23021?camid=4v1a)

Regression Testing of Database Applications
[www.igi-global.com/article/regression-testing-database-applications/3278?camid=4v1a](www.igi-global.com/article/regression-testing-database-applications/3278?camid=4v1a)

Data Mining: New Arsenal For Strategic Decision-making
[www.igi-global.com/article/data-mining-new-arsenal-strategic/51212?camid=4v1a](www.igi-global.com/article/data-mining-new-arsenal-strategic/51212?camid=4v1a)