IDEA GROUP PUBLISHING
701 E. Chocolate Avenue, Suite 200, Hershey PA 17033-1240, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.idea-group.com
ITB12822

This paper appears in the book, *Emerging Trends and Challenges in Information Technology Management, Volume 1 and Volume 2* edited by Mehdi Khosrow-Pour © 2006, Idea Group Inc.

Philosophical Foundations of Information Modeling

John M. Artz, Department of Information Systems and Technology Management, The George Washington University, Washington, DC 20052, jartz@gwu.edu

INTRODUCTION

For the past few decades there have been articles appearing, in the Information Systems research literature, criticizing the field for a lack of theory, no core set of concepts, and no accepted paradigm. [Culnan], [Weber, 1987]. Weber [2003, pg. viii] points out, "After a long period of discernment, we found we could identify only one class of phenomena, for which theories sourced from other disciplines seemed deficient—namely, phenomena associated with building conceptual models and designing databases." So perhaps the core of Information Systems has something to do with information models or information modeling. This is plausible since data modeling and information modeling are, perhaps, the only intellectual developments that are unique to information systems. Yet the theories in these areas are sketchy at best. But theory doesn't just appear spontaneously. It must be developed. One of the ways in which this can be done it to take general theories from other areas, apply them to phenomena in Information Systems, and then advance them by making them more specific to Information Systems. This mini-paper will introduce four concepts from metaphysics that are highly relevant to information modeling and as such may provide some foundation from a philosophical perspective. These four concepts are (1) the Concept of Identity, (2) the Problem of Universals, (3) Teleology, and 4) Correspondence versus Coherence views of truth.

INFORMATION MODELING

Information modeling is the first step in database design, sometimes referred to as conceptual database design. In this step, in the design process, the information modeler examines the domain of interest and determines the classes of entities that will be represented in the database and the relationship between those classes. When the modeler is finished, the conceptual database design is represented in an information model which contains entity class descriptions, attributes of the entity class, relationships between entity classes. In the construction of the information model, a variety of philosophical assumptions are made that address which classes should be represented in the model, where those classes come from, how those classes are identified or constructed, and the truth status of the information in the database. Following are four philosophical concepts that lie at the foundation of those assumptions.

CONCEPT OF IDENTITY

The concept of identity is both deceptively simple and profoundly important. It is one of the most fundamental concepts in all of metaphysics and goes right to the heart of how we organize our knowledge. In its simplest form the Concept of Identity addresses the problems that arise when we use the word the 'same' [Stroll]. In the case of information modeling, we group entities into entity classes because those particular entities are deemed to be of the 'same' kind. This notion is reinforced by the relational principle of entity integrity which asserts that a relational table should not contain more than one kind of entity and the entity relationship modeling rule that requires internal consistency. But where do kinds come from and how do we know when an instance of an entity is of a particular kind? This leads us to the next concept - The Problem of Universals.

PROBLEM OF UNIVERSALS

When we use the word 'same' to refer to 'same kind' we are organizing the world into categories. Categories are useful because they help us organize our knowledge efficiently. When I point to a tree and call it a tree, I am assigning it to a category. By doing this, I can apply my general knowledge of trees to the specific tree at which I am pointing. The thing at which I am pointing is actually an instance of a tree, but we do not make that distinction in normal speech. Yet, philosophically, we do make that distinction. The instance at which I am pointing is called a particular and the category to which I assign it is called a universal.

The Problem of Universals attempts to address the question – Where do universals come from? There really isn't an easy answer to this question and philosophers have provided a variety of answers over the centuries. From an information modeling perspective we would ask - are the categories or entity classes real and hence discovered, or are categories constructed and if so according to what criteria? The Problem of Universals is fundamental to information modeling because the process of constructing entity classes is no more or less than the Problem of Universals. And understanding what has been said about the Problem of Universals provides great insight into the process of information modeling.

TELEOLOGY

Teleology is the study of or understanding of things in terms of their ultimate purposes. Aristotle, who first introduced teleological thinking, believed that an adequate understanding of a phenomenon required an understanding of four causes: formal, material, efficient, and final. The final cause is the ultimate purpose of the thing and Aristotle felt that you could not understand a thing if you did not understand its *telos* or ultimate purpose. This position was somewhat extreme in the case of physical objects and led to some problematic interpretations such as – fire rises because it wishes to return to the sun, or objects fall because they wish to return to the center of the earth. This attribution of purpose to physical objects does not sit well with one's modern sense. And indeed, it should not. Galileo dismissed teleological explanations from his view of astronomy leading eventually to the modern view that physical objects simply follow the laws of nature and do not have any ultimate purpose.

One can easily dismiss teleological thinking from physics, but human systems must always be understood in terms of purposes or they are not understood at all. The most extreme example of teleological thinking is in the construction of artifacts such as an information system. The point here is that we should always understand the purpose for which we are building an information system or the purpose for which we are designing a database. If we do not understand the ultimate purpose of a

thing we are designing it is unlikely that we will design the right thing. Hence, all database design and information systems development should be teleological.

CORRESPONDENCE VS. COHERENCE VIEWS OF TRUTH

Next we must ask the question—what does it mean for something to be true - and we must consider three positions on this question: naïve adaptation, correspondence and coherence. The naïve adaptation view of truth assumes that something is true if one has no reason to question it. Many things that we know fall into this category. However, if we decide to question something we may continue to believe it is true if we find that our belief actually does correspond to what we find in the world. This is the correspondence view of truth. The correspondence view of truth [Prior] asserts that in order for a thing to be true it must correspond to a things or events in the real world. So instead of just believing something, we check it out and find that our belief actually does correspond to reality. But sometimes reality is messy. So we organize our experiences into a coherent framework in order to make sense out them. This is the coherence view of truth. The coherence view of truth says that the world is a messy and chaotic place. In order for something to be true it has to make sense. So I make the world coherent by creating, organizing, and sense-making categories and relationships. So the coherence view of truth [White] suggests that truth is not something derived from the world, rather it is an order superimposed on the world. All three of these views of truth can be seen as options in the database design process.

How does the concept of truth apply to a database? The answer is quite simple. When one executes a query in a database and gets an answer, that answer should be the same as the answer they would get had they interrogated the real world. The point of a database, after all, is to model some aspect of the real world so that database users can interrogate their model of the world rather than having to interrogate the world each time. Now we can consider each level of truth with respect to databases.

The naïve adaptation level of truth is a good model for many databases. The records in the database often reflect the way information is processed rather than the information itself. So the extent to which the records actually model the domain is, to some extent, coincidental. However, as long as the information provided by the database is reasonably useful and satisfies the basic processing needs of the organization nobody is likely to question it further.

In order to meet the correspondence criterion tables should correspond to classes of entities. Rows should correspond to specific occurrences of entities. Attribute values should be facts about specific entity occurrences. Foreign keys should correspond to regular relationships between entities in one class and entities in another class. The strength of correspondence is that when a user asks a question of the database, the answer they get is not just a calculation in the database but it is also true in the real world.

Finally, the aspect of the real world that is being modeled may not be sufficiently refined to gain the full benefits of modeling it in a database. This occurs when the semantics of the domain are ambiguous. For example, in the university environment the word *Course* is often used both for a *Course* and a *Course Offering*. This leads to problems in modeling the domain. Let's say that Professor Smith is teaching two sections of one course, and one section of another course. If you ask the question – how many courses is Professor Smith teaching this semester – you cannot get a clear answer. The answer may be two if you actually mean *Course*. But it could also be three if you mean *Course Offerings*. Often times the semantics of a domain are too muddled to model directly and in these cases we have to go beyond the correspondence criteria to the coherence criteria.

BRINGING IT ALL TOGETHER

The Concept of Identity provides insight into the Problem of Universals which is the central metaphysical problem encountered in information modeling. Thus, the Concept of Identity and the Problem of Universals provide two important philosophical foundations for information modeling. When classes (or universals) are formed, there are a wide variety of options for the grouping that may be formed. The only way we can know if one set of classes is superior to another set of classes is to know the purpose of the database which requires a Teleological approach to information modeling. One of the purposes of a database may be to model the world as it is (despite existing flaws in semantics) which adopts the Correspondence Theory of Truth. The other possibility is to use the database to superimpose order on the world in order to gain greater benefits from the database which adopts the Coherence Theory of Truth.

REFERENCES

Culnan, M. (1986) The Intellectual Development of Management Information Systems, 1972-1982: A Co-citation Analysis. Management Science. 32(2). Pp. 156-172.

Prior, A. N. (1967) Correspondence Theory of Truth. In Encyclopedia of Philosophy. Paul Edwards (ed). Macmillan Publishing, New York. Vol 1. Pp. 223-232.

Stroll, A. (1967) Identity. In Encyclopedia of Philosophy. Paul Edwards (ed). Macmillan Publishing, New York. Vol. 4. pp. 121-124.

Weber, R. (1987) Toward a Theory of Artifacts: A Paradigmatic Base For Information Systems Research. Journal of Information Systems. Spring, pp 3-19.

Weber, R. (2003) Still Desparately Seeking for the IT Artifact (editor's comments). MIS Quarterly, (27) 2. pp. iii-xi.

White, A. R. (1967) Coherence Theory of Truth. Encyclopedia of Philosophy. Paul Edwards (ed). Macmillan Publishing, New York. Vol 1. Pp. 130-133.