# A Meta-Analysis Comparing Relational and Semantic Models

*Keng Siau, University of Nebraska – Lincoln, USA*

*Fiona Fui-Hoon Nah, University of Nebraska – Lincoln, USA*

*Qing Cao, Texas Tech University, USA*

## ABSTRACT

*Data modeling is the sine quo non of systems development and one of the most widely researched topics in the database literature. In the past three decades, semantic data modeling has emerged as an alternative to traditional relational modeling. The majority of the research in data modeling suggests that the use of semantic data models leads to better performance; however, the findings are not conclusive and are sometimes inconsistent. The discrepancies that exist in the data modeling literature and the relatively low statistical power in the studies make meta-analysis a viable choice in analyzing and integrating the findings of these studies.*

*Keywords:* *Data Model Comprehension, Data Modeling, Relational Data Model, Semantic Data Models, User Performance*

## 1. INTRODUCTION

Databases are used in every sector of the industry including business, health care, education, government, and libraries (Liao & Palvia, 2000; Firat *et al*., 2009; Aiken *et al*., 2011; Siau & Rossi, 2011). As databases continue to become ever more critical components of both large and small systems, the success of development projects will become increasingly dependent on the accuracy of the data models (Ramakrishnan & Gehrke, 2002; Siau *et al*., 2010). We have witnessed enormous growth in the number and importance of database applications in the last three decades. Aiken *et al.* (2011) stressed that the amount of data available these days and the value from information that organizations must derive from these data have made data modeling and database management a critical and vital issue for the competitiveness and survival of organizations. Data modeling techniques have continued to evolve since the first database management system was introduced in the 1960s (Watson, 2005). Subsequently, hierarchical and network database models were introduced in the 1970s. The relational data model and semantic data models, such as entity-relationship model (ER) and extended entity-relationship model (EER), have been widely used in practice since the early 1980s (Hoffer *et al.*, 2010) and the trend continues even today.

Object oriented modeling became popular in the late 1990s and early 2000s. Studies on object oriented approaches (e.g., Siau & Cao,

DOI: 10.4018/jdm.2011100103

2001; Siau & Loo, 2006; Taha & Elmasri, 2009; Gemino & Parker, 2009; VanderMeer & Dutta, 2009; Evermann & Wand, 2009; Siau & Tian, 2009; Shank *et al*., 2010; Siau, 2010; Tan *et al*., 2011) started to appear in academic publications in the 2000s. Despite the emergence of object oriented modeling techniques, relational and ER/EER models and their query languages continue to be important and widely studied (e.g., Sengupta & Ramesh, 2009; Niu *et al*., 2009; Sakr, 2009; Ayanso *et al*., 2009; Huh *et al*., 2010; An *et al*., 2010; Bera *et al*., 2010; Shlezinger *et al*., 2010). Further, relational database management systems are widely used and relational data model is still the dominant model in industry.

In the relational data model, a relation is a subset of the Cartesian product of an ordered list of domains (Codd, 1970). The power of the relational data model lies in its rigorous mathematical foundations and simple user-level paradigm. Semantic models use the concepts of entities or objects and relationships among them. Semantic models provide flexible structuring capabilities and allow data constraints to be specified explicitly. Object-oriented data models, dynamic data models, and active data models are the new trends in data modeling (Embley, 1998). Despite much research interest in semantic and object oriented data modeling, the relational database is still the predominant database used in the industry (Coronel *et al*., 2009), while ER modeling remains highly popular as a tool for conceptual data modeling (Post, 2004).

Many empirical studies have examined the effect of relational versus semantic data modeling on user performance (Batra *et al*., 1990; Carlis & March, 1983; Chan *et al*., 1993; Chan & Lim, 1998; Hardgrave & Dalal, 1995; Jarvenpaa & Machesky, 1989; Juhn & Naumann, 1985; Kim & March, 1995; Leitheiser *et al*., 1996; Shoval & Even-Chaime, 1987; Siau *et al*., 1995, 2004). Much of the research in data modeling suggests that the use of semantic data models leads to better performance; however, the findings are not conclusive and are sometimes inconsistent. The discrepancies in the data modeling literature and the relatively low statistical power of the studies make meta-analysis a viable choice in analyzing and integrating the findings of these studies (Chan & Lim, 1998).

The objective of this research is twofold. First, a narrative review was carried out to compile empirical studies that compared relational data modeling and semantic data modeling. This provided a basis for an assessment of inconsistencies in the data modeling literature. Second, a meta-analysis was carried out as it offers powerful tests for detecting significant relationships by increasing the statistical power.

The rest of the paper is organized as follows: Section 2 provides a comprehensive literature review on relational and semantic data modeling, and summarizes prior studies comparing relational and semantic data modeling. Section 3 presents the research questions and model. Section 4 discusses the theories used in this research and lists the hypotheses for this study. Section 5 describes the research methodology and details of the research procedure. Section 6 presents the results and discussion. The last section, Section 7, lists the limitations of the research and concludes the paper.

## 2. LITERATURE REVIEW

This section of the paper reviews the basic concepts of relational and semantic data modeling and provides a narrative review of the data modeling literature.

### 2.1. Relational and Semantic Models

Numerous data modeling approaches have been introduced in the last three decades. Two of the most popular approaches are the relational and semantic models. The relational data model is the most popular technique for managing large collections of data (Watson, 2005). A data relational model consists of a set of relations, where a relation is a two-dimensional table arranged in rows and columns. A distinguishing characteristic of the relational data model is that there is no explicit linkage between tables. The