**Data Models for End-user Information System Design and Querying**

Laurel King
*University of Hawaii*
*laurelk@hawaii.edu*

Martha E. Crosby
*University of Hawaii*
*crosby@hawaii.edu*

**Abstract**

*The following study was designed to better understand how end users model their data in database design. It seeks to find ways to facilitate end-user conceptual modeling in order to create better data models for the databases they design and use. End users and novices often use a bottom-up modeling approach from instances of the data for which they want to design a database. Conceptual modeling is at a higher level of abstraction than logical or physical structural database design and encourages a non-data-driven top-down approach to database design. Techniques to support end-user conceptual modeling are explored as potential aids to better database design.*

## 1. Introduction

In many organizations, end users are developing and modifying their own information systems using readily available software. These systems often do not include conceptual design tools, so the users have the option to adapt template databases to his or her needs, or import data and organize it into a logical design as best they can. Traditionally, system analysts and database developers have used conceptual models to design systems according to the stakeholder requirements they have gathered from end users, as well as to communicate and validate the specifications of a system with its intended users. The system users usually have more domain knowledge than the analysts and developers, and can help validate the design if the representation is effective in communicating the conceptual design. When users develop their own databases, they are often more skilled in the domain of the database than in designing databases. By studying how end users model their data and processes, data modeling techniques can be developed to help them organize their knowledge of the target domain better during the design process.

Analysts and designers create conceptual models of proposed information systems to design systems that better meet the requirements of stakeholders. Topi and Ramesh define conceptual models as “a set of constructs that can be used to create an abstraction of reality” that is data and not process oriented, unambiguous and independent of the database structure [18]. They list four different uses of conceptual models compiled from several studies: (1) as a communication tool between users and analysts, (2) as a formal and accepted conceptual foundation of an organizational system, (3) as a foundation for applications developed by end users, and (4) as an essential part of the system documentation for system maintenance [2, 5, 10].

Accurately defined conceptual models allow necessary alterations to be detected and accomplished before the logical and physical system is developed when changes are much more complicated and expensive to make. The importance of the conceptual design cannot be overstated, however, there are conflicting objectives for the data model because users, analysts, and developers use them for different tasks. The tasks of specifying the information requirements, validating the resulting conceptual design, and creating the logical and physical design of the chosen database system require different levels of detail and abstraction of the conceptual model. Topi and Ramesh suggest a framework that separates the user-analyst and analyst-developer communication via the conceptual model. The topic of this study involves yet another category of potential usage where the end users become the user, analyst and developer of their own database applications using common database software packages. The scale is usually much smaller, there is less complexity and the software allows less ability to customize the resulting database. Nevertheless, it is an important area in that the ability to successfully capture data and convert it into knowledge is dependent on the quality of the database created.

## 2. Previous Studies

Most previous studies have looked at the analyst-developer usage of conceptual models as opposed to its role as a user-analyst communication tool. Furthermore, there is often a blurring of the distinction between the conceptual model and the logical model, which is based on the particular database management system (DBMS) selected to implement the conceptual model. This study uses the logically independent definition of a conceptual model, but some of the previous studies have compared logical data models, which were used as conceptual models.

The study by Juhn and Naumann tested MBA students' model comprehension using database design validation tasks and found semantic (graphic) models to be better understood, except for tasks to identify relationships, and most interestingly, that such models also seemed to lead to a more systematic approach to designing a database [10]. A better understanding of how and when external representations are effective for end-user database design and usage will add to our knowledge of representations used in the user-analyst validation process of systems designed by experts as well. Although we still lack a clear understanding of how and if external graphical representations interact with one's internal representations to facilitate learning and problem solving, external representations are thought to support problem solving by computational off-loading [14].

There are two basic types of data modeling formalisms, the entity-attribute-relationship (EAR) model and the object-role (OR) model [4, 11]. Previous research has shown semantic (graphic) conceptual models to be generally superior in database modeling to text-based models when the user is trained in the target model, but they are not easily converted to a logical design for the database [2, 6, 7, 16]. At the same time, there is evidence that when the user is not trained in the specific model, highly graphical models can be harder to understand than less graphical models [13]. CASE tools which help in the conceptual design and conversion to logical design have had mixed success in improving design quality due to undetected modeling errors [19]. When end users develop databases, is it more important to have a realistic conceptual model or easier conversion to the logical design? The role of the data model is different when end-users develop their own databases to organize their data. For non-experts, the grammar of the representation must be easy to understand and also facilitate the systematic design of the database. Can representations be customized to help the end-user create better designs and improve querying ability by starting with conceptual designs as opposed to logical designs?

Recent work on conceptual modeling has led to better ontology and methods to evaluate data models by comparing the resulting scripts of data models for quality and equivalence, but much more research is needed to understand and evaluate conceptual modeling *grammars*, *methods*, *scripts*, and *contexts* of use [20]. The current study is focused on the conceptual modeling and querying within the *context* of end users.

Batra and Wishart compared the performance of novice designers trained with rule-based methods or commonly occurring design patterns to databases and found the rule-based method better for the simpler and the more difficult designs, but indicate that pattern-based training may be better for problems of medium complexity [3]. Batra identified eleven basic patterns found most frequently in database design textbooks [1]. These basic patterns may be used as templates to help aid end users in designing a database with rule-based and pattern-based methods.

By analyzing the end user's conceptualization of the data requirements and use of representations, it may be possible to design an interface that will enforce external rules to reduce the cognitive effort of trying to remember formal conceptual modeling formalisms. Distributed cognitive tasks such as designing a database or query should benefit from having external rules or physical constraints to automatically enforce the required rules, thus making conceptual modeling easier for end users. Without constraints, internal rules must be memorized and use working memory that could otherwise be used for problem solving [21]. The interface could be implemented in an expert system that guides the user through the decision process and creates a flexible database representation to meet the individual needs of the user.

Research on the use of external representations has shown that different types of representations can improve learning and problem solving, but individual differences and whether the representation is self-constructed or not interact with their effectiveness [8, 9, 12, 21]. Recently, Siau and Tan have presented a framework to improve the quality of data models by using cognitive mapping techniques for user-analyst communication [15]. The cognitive maps they refer to are external representations created by techniques to identify and externalize subjective beliefs, and not the internal schema or mental model used for problem solving as defined by Tolman [17]. Siau and Tan list three types of cognitive maps that may prove to be useful in end-user conceptual modeling as well as user-analyst communication as they propose. They are causal maps drawn in networks, semantic maps for organizing sub-ideas around a main idea, and concept maps that link related concepts in a network configuration. These are used to solicit ideas and information from users, but they may be useful for end user database design as well.

## 3. Methodology

In this study, university students who have taken one or two courses in database or information retrieval will perform database modeling and search tasks to see how they use data representations in an attempt to elicit their internal representations of the database design and search tasks. These students are thought to sufficiently represent the end-users that would try to design a database to improve data access and manipulation in their unit or organization after they enter the workforce. Each participant will be given a questionnaire to determine their verbal-visual preference in learning, their database training, and their confidence in database design and usage.

Participants will be given case studies that describe database tasks and requirements in situations involving typical database domains. A specific type of external representation will accompany the tasks or the participant will be asked to construct a representation to see which leads to optimal design and query performance. They will be asked to draw/diagram or write the entities, relationships and cardinalities in the way they picture the database according to the case study description.

Their performance will be measured on the accuracy of the design and their answers to questions. Self-constructed models will be analyzed for accuracy and similarities to established conceptual and logical models. They will be given one point for each correct structural aspect of the model and lose one point for each error. Spelling errors will be ignored.

The participants will also be asked to perform database search tasks on the correct relational model of the database they attempted to design. The questions will ask them to select the information that would be needed to search for a particular combination of data. For example they would specify the table, combination of tables, filtering, etc., necessary to select the requested data, but will not have to know the syntax for the exact query needed to extract the data.

Participants will be given an opportunity to discuss the tasks and how they created or used the data model after the experiment. This feedback will be analyzed to better understand their approach to the data modeling and search tasks. The interaction of performance and type of representation will be analyzed in terms of other individual characteristics provided, such as verbal-visual preference and database training.

## 4. Results

Results are expected to improve our understanding of end-user conceptual modeling, and what types of data representations improve database design. We hope to further our understanding of end user representations of database tasks, and how database performance and learning are supported by the external representation of the data. The results could inform the design of a database interface that uses artificial intelligence to guide the user through the design process and offer a flexible database representation that meets the needs of the individual user.

## 5. References


[1] D. Batra, "Conceptual Data Modeling Patterns: Representation and Validation," *Journal of Database Management*, vol. 16, pp. 84-106, 2005.

[2] D. Batra, J. A. Hoffer, and R. P. Bostrom, "Comparing Representations with Relational and EER Models," *Communications of the ACM*, vol. 33, pp. 126-138, 1990.

[3] D. Batra and N. A. Wishart, "Comparing a rule-based approach with a pattern-based approach at different levels of complexity of conceptual data modelling tasks," *International Journal of Human-Computer Studies*, vol. 61, pp. 397-419, 2004.

[4] H. Biller and E. Neuhold, "Concepts of the conceptual schema," in *Architecture and Models in Data Base Management Systems*, G. M. Nijssen, Ed. Amsterdam: North Holland, 1977, pp. 1-30.

[5] D. Campbell, "Entity-relationahip modeling: one style suits all," *Database*, vol. 23, pp. 12-18, 1992.

[6] H. C. Chan, K. Siau, and K.-K. Wei, "The Effect of Data Model, System and Task Characteristics on User Query Performance - An Empirical Study," *The DATA BASE for Advances in Information Systems*, vol. 29, pp. 31-49, 1998.

[7] H. C. Chan, K.-K. Wei, and K. L. Siau, "User-Database Interface: The Effect of Abstraction Levels on Query Performance," *MIS Quarterly*, vol. 17, pp. 441-464, 1993.

[8] R. Cox, "Representation construction, externalized cognition and individual differences," *Learning and Instruction*, vol. 9, pp. 343-363, 1999.

[9] M. Hegarty and M. Kozhevnikov, "Types of Visual-Spatial Representations and Mathematical Problem Solving," *Journal of Educational Psychology*, vol. 91, pp. 684-689, 1999.

[10] S. H. Juhn and J. D. Naumann, "The Effectiveness of Data Representation Characteristics on User Validation," presented at Proc. of the Sixth International Conference on Information Systems, Indianapolis, IN, 1985.

[11] Y.-G. Kim and S. T. March, "Comparing data modeling formalisms," *Communications of the ACM*, vol. 38, pp. 103-115, 1995.

[12] R. E. Mayer, "Comprehension as affected by structure of problem representation," *Memory & Cognition*, vol. 4, pp. 249-255, 1976.

[13] J. C. Nordbotten and M. E. Crosby, "The effect of graphic style on data model interpretation," *Information Systems Journal*, vol. 9, pp. 139-155, 1999.

[14] M. Scaife and Y. Rogers, "External cognition: how do graphical representations work?," *International Journal of Human-Computer Studies*, vol. 45, pp. 185-213, 1996.

[15] K. Siau and X. Tan, "Improving the quality of conceptual modeling using cognitive mapping techniques," *Data & Knowledge Engineering*, vol. 55, pp. 343-365, 2005.

[16] K. L. Siau, H. C. Chan, and K. K. Wei, "Effects of Query Complexity and Learning on Novice User Query Performance With Conceptual and Logical Database Interfaces," *IEEE Transactions on Systems Man and Cybernetics Part A:Systems and Humans*, vol. 34, pp. 276-281, 2004.

[17] E. Tolman, "Cognitive maps oin rats and men," *Psychological Review*, vol. 55, pp. 189-208, 1948.

[18] H. Topi and V. Ramesh, "Human Factors Research on Data Modeling: A Review of Prior Research, An Extended Framework and Future Research Directions," *Journal of Database Management*, vol. 13, pp. 3-19, 2002.

[19] I. Vessey, S. L. Jarvenpaa, and N. Tractinsky, "Evaluation of vendor products: CASE tools as methodology companions," *Communications of the ACM*, vol. 35, pp. 90-105, 1992.

[20] Y. Wand and R. Weber, "Research commentary: Information systems and conceptual modeling - A research agenda," *Information Systems Research*, vol. 13, pp. 363-376, 2002.

[21] J. Zhang and D. A. Norman, "Representations in Distributed Cognitive Tasks," *Cognitive Science*, vol. 18, pp. 87-122, 1994.