

# VALIDATION OF A METHOD FOR REPRESENTING LARGE ENTITY RELATIONSHIP MODELS: AN ACTION RESEARCH STUDY

Daniel L. Moody

Department of Computer and Information Science  
Norwegian University of Science and Technology  
(NTNU)  
Trondheim, Norway N-7491  
email: dmoody@idi.ntnu.no

School of Business Systems,  
Monash University  
Melbourne, Australia 3800  
email: dmoody@infotech.monash.edu.au

## ABSTRACT

*One of the most serious limitations of the Entity Relationship (ER) Model in practice is its inability to cope with complexity. Once data models exceed a certain threshold size, they become difficult to understand (end user's viewpoint) and also to document and maintain (analyst's viewpoint). A number of approaches have been proposed in the literature to address this problem, but so far there has been no systematic empirical research into the effectiveness of these methods. This paper describes an action research study in which a method for representing large ER models was tested in a large application development project in one of Australia's largest commercial organisations. The research was successful in achieving both practical and research outcomes—it resulted in change of data modelling practices in the organisation, and the method was refined significantly as a result of experiences in practice. However a major problem experienced in this study was that the size of the project imposed constraints on the evolution of the method. Because of the number of people involved, it was difficult to make changes to the method “on the fly” and to experiment with variations of the method, as is customary in action research.*

## 1. INTRODUCTION

### The Problem of Complexity in Data Models

One of the most serious limitations of the Entity Relationship (ER) Model in practice is its inability to cope with complexity (Feldman and Miller, 1986; Gilberg, 1986; Simsion, 1989; Teory et al, 1989; Wand and Weber, 1993; Gandhi et al, 1994; Akoka and Comyn-Wattiau, 1996; Allworth, 1996; 1999). The two major practical problems with large data models are:

- Understanding (end user's perspective): when data models exceed a certain size, they become difficult for end users to understand.
- Documentation and maintenance (analyst's perspective): when data models exceed a certain size, they become difficult to document and maintain.

Neither the standard ER Model or the Extended Entity Relationship (EER) model provide explicit abstraction mechanisms for managing the complexity of data models (Weber, 1997). A number of methods have been proposed in the literature to address this issue (e.g. Martin and McClure, 1985; Feldman and Miller, 1986; Gilberg, 1986; Simsion, 1989; Teory et al, 1989; Gandhi et al, 1994; Allworth, 1996; 1999), but so far, none of these have been widely accepted in practice. Consequently, it remains an open research issue (Thalheim, 1999).

## Previous Research

A previous paper (Moody, 1991) defined a method for representing large data models based on the organisation of a street directory (using the approach of *analogical reasoning*). This represents the research idea being tested by this study. The method is briefly summarised here to provide context for discussion of the empirical results. The proposed method organises large data models into a number of different levels of abstraction (Figure 1):

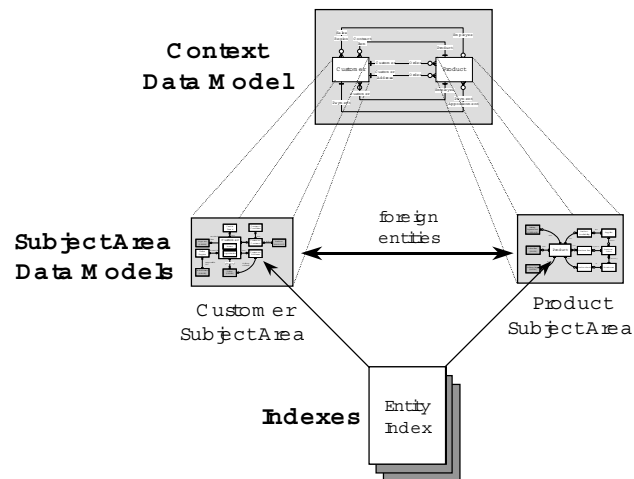


Figure 1. Levelled Data Model Architecture

- A high level diagram, called the *Context Data Model*, provides an overview of the model and how it is divided into subject areas. Each subject area is shown as an entity, with *boundary relationships* (relationships between entities from different subject areas) shown as relationships between them. This corresponds to the key map in a street directory.
- A set of named *Subject Area Data Models* show subsets of the data model in full detail. These correspond to detail maps in a street directory. *Foreign entities* are used to show cross-references between subject areas, and correspond to inter-map references in a street directory. These are shown as shaded rectangles, with their primary subject area in brackets.
- A range of *indexes* are used to help locate individual objects (entities, relationships and attributes) within each subject area.

The resulting model is called a Levelled Data Model. The model may be organised into any number of levels, depending on the size of the underlying data model, resulting in a hierarchy of models at increasing levels of detail. At each level, the diagrams are shown in ER form.

## Research Objectives

The broad research questions addressed by this study are:

- Research Question 1: Does the method improve end user understanding and simplify documentation and maintenance of large data models? That is, is the method effective in achieving its objectives?
- Research Question 2: How can the method be improved?

Because of its open-ended nature, the second research question cannot be easily answered using traditional hypothesis-testing research approaches.

## **2. RESEARCH DESIGN**

### **The Need for Empirical Validation of Methods**

It is essential for IS design methods to be tested in practice—ultimately, the scientific merit of any method is an empirical rather than a theoretical question (Rescher, 1973; Ivori, 1986). However real world validation of methods in IS design research is very poorly done. Wynekoop and Russo (1997) conducted a review of IS design research published in the leading IS journals over the past three decades. The results of the analysis showed a heavy reliance on normative research, largely focusing on the development of new methods or modifications to existing methods. They concluded that there was a “lack of serious empirical research into the efficacy of methods in practice” and a “need for validation of methods in organisational contexts using real practitioners”.

This is certainly the case in the literature on large data model representation methods. So far, there has been no systematic empirical research into the effectiveness of these methods in practice. The authors of the methods argue that their approaches are effective but in most cases, no empirical evidence is provided. Most evidence of successful use of these methods is anecdotal and in many cases reports the direct experience of the author (Shanks, 1996).

### **Research Method Selection**

A major barrier to the empirical validation of IS design methods is that it is very difficult to get new approaches, especially those developed in academic environments, accepted and used in practice. Practitioners who have developed familiarity and expertise with existing techniques are reluctant to adopt academic approaches that are theoretically sound but unproven in practice (Bubenko, 1986; Wynekoop and Russo, 1997; Avison et al, 1999). Action research provides a method for testing and refining research ideas by applying them in practice (McCutcheon and Jurg, 1990; Jönsson, 1991; Hatten et al, 1997). It provides the opportunity to test out research ideas “in an organisational context using real practitioners”, as recommended by Wynekoop and Russo (1997). One of its major advantages is that it can help to overcome the problem of persuading practitioners to adopt new techniques, and overcome the cultural divide that exists between information systems academics and practitioners (Checkland, 1991; Moody and Shanks, 1998a; Avison et al, 1999; Moody, 2000).

Action research is particularly useful in exploratory research to develop solutions to practical problems (Dick, 1997). It thus provides an appropriate way to evaluate a method in its early development phases (in this case, when it has first been proposed). At this stage, traditional hypothesis-testing approaches are of limited value, as they typically provide only accept/reject responses to questions (Chow, 1988). This provides minimal feedback for the purpose of improving the method (Research Question 2). Action research allows the method to evolve as a result of experience in practice, as part of an ongoing learning and reflection process (Checkland and Scholes, 1990; Baskerville and Wood-Harper, 1996).

### **The Action Research Process**

Action research is usually carried out in a number of discrete cycles, which function as “mini-experiments” carried out in practice. Through reflection on previous action, a theory of the form “if I do X, then Y will occur” is proposed, which is applied in practice, and then evaluated in a cyclic manner (Oosthuizen, 2000). One of the most widely used approaches to action research is that developed by Kemmis and McTaggart (1988), which emerged from the field of education. Each action research cycle consists of the following steps:

- Plan: Develop a plan of action to improve current practice. The plan must be flexible to allow adaptation for unforeseen effects or constraints.
- Act: The participants act together to implement the plan.

- Observe: The action is observed to collect evidence which allows thorough evaluation of outcomes. A variety of data collection methods may be used to evaluate the results of the intervention (Holter and Schwartz-Barcott, 1993; Stringer, 1996).
- Reflect: Participants reflect on what went wrong, what went right and how to improve the idea in the next cycle. Each cycle may lead to improvement of the original idea ( $M_1$ ), resulting in a sequence of successively refined and improved ideas  $M_2, M_3...$

The first research question (evaluation of method effectiveness) is addressed in the *observe* phase of each action research cycle, while the second research question (method improvement) is addressed in the *reflect* phase.

### Theoretical Framework

Checkland (Checkland, 1991; Checkland and Holwell, 1998) argues that a critical component of any action research study is an explicit theoretical framework, declared in advance, in terms of which learning will be defined. This framework is likely to evolve over the course of the study, and is the equivalent of a theoretical model used in positivist research. Figure 2 defines the theoretical framework used in this action research study. While not specifying formal hypotheses or measures, it defines the nature of the intervention and the expected outcomes, and therefore defines a *causal model* of the action research situation.

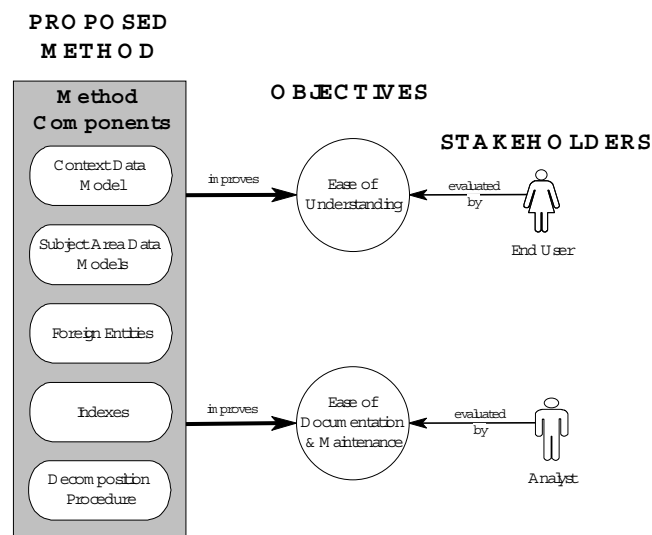


Figure 2. Theoretical Framework

At a high level, this model is similar in structure to the theoretical models used in positivist research. The proposed method corresponds to the independent variable (the intervention that is applied) and the objectives correspond to the dependent variables (the desired outcomes of the intervention). The theoretical framework provides the basis for data collection, evaluation of outcomes and reflection. In each action research cycle, learning is reflected in changes to the theoretical framework.

## 3. BACKGROUND TO THE SITUATION

### Organisational Context

The organisation involved in this study was a large Australian bank, and one of the largest commercial organisations in Australia. The bank had initiated a large application development project to redevelop its core banking systems. This was the largest development project ever undertaken in the bank, and was estimated at over 250,000 man days and over \$250 million in development effort. It was also

one of the largest application development projects ever undertaken in Australia. The project had just completed the business case analysis phase and was awaiting approval from senior management to begin the requirements definition phase. A high level scoping data model had been developed during the business case phase, consisting of over 300 entities. This was represented on a single diagram.

### **Justification for Project Selection**

This project provided an ideal test of the proposed method for two reasons:

- **Project size:** given that this was one of the largest application development projects ever undertaken in Australia, this would provide a thorough test of the proposed method's ability to handle complexity.
- **Number and diversity of project participants:** According to Rescher (1977), a method is not a successful method unless it is usable by other people. There were over 100 people involved in this project with a wide range of experience levels, which would provide a test of the method's applicability by both experts and novices.

## **4. THE INTERVENTION**

### **Introduction of the Method**

The proposed method was presented to the Project Director and the five project team managers at their weekly management meeting. As a result of this meeting, it was agreed to adopt the proposed method as a project standard. The project development methodology, training materials and repository structure were modified to incorporate the concepts of the method. The existing data model was divided into subject areas, following the rules defined in the method (Moody, 1991). The models were documented in the repository, and each subject area assigned to a particular project team. In all, twenty-three subject areas were identified and distributed approximately equally among the five project teams. In addition, requirements for five additional subject areas were identified after consultation with the project management team. These represented gaps in the existing data model where analysis work had yet to begin.

### **Ongoing Use of the Method**

The method was used by members of the central data architecture group and business analysts attached to project teams. Many of the business analysts were business representatives with no prior analysis experience. Throughout the requirements definition phase, subject area data models were developed and maintained as relatively independent units. Each subject area was assigned to a single project team, who were responsible for all changes to it, and for achieving user signoff of the subject area prior to the end of the requirements phase.

### **Final Data Architecture**

When all subject areas had been signed off by business users, they were packaged together into a single data architecture. The data architecture was handed over to the database design team as the basis for developing the physical database design. The final data architecture consisted of over 500 entities, and was organised into three levels of abstraction (Figure 3). An alphabetical entity index was also produced, listing each entity and its subject area reference. The diagrammatical conventions used in the final data architecture were significantly different to those proposed in the original method.

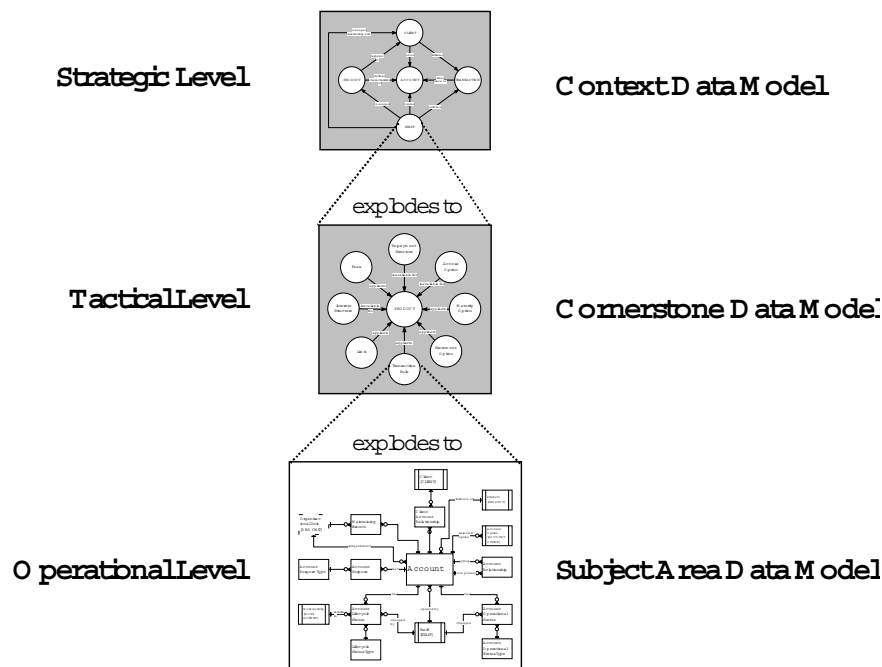


Figure 3. Final Data Architecture

## 5. EVALUATION OF THE METHOD (OUTCOMES)

This section evaluates the effectiveness of the method in achieving its objectives. This addresses Research Question 1.

### Stakeholder Analysis

An important part of achieving rigour in action research is to use multiple sources of information. This can be done by increasing the number and diversity of informants (Dick, 1999). Decisions about who to involve in the process can be made using *stakeholder analysis* (Pouloudi, 1999). A stakeholder is defined as anyone who is affected by, or can influence the action or outcome. The effectiveness of the method in achieving its objectives was evaluated from the point of view of all relevant stakeholders:

- Analysts: these represent *direct users* of the proposed method. This included members of the central data architecture group and business analysts attached to project teams. The data architects were all experienced in data modelling, and acted as participants in the research. The business analysts included both novice and experienced analysts, and applied the method using the standards defined and provided feedback on its usefulness—they acted as informants only.
- End users: these represent *indirect users* of the method, in that they use the output of the method. End users play an important role in the data modelling process, because they have to verify that the data model accurately represents their requirements. Their ability to do this effectively has a major impact on the quality of the final system and its ability to meet user requirements (Kim and March, 1995). As the “customer”, they also have a major stake in the outcome of the process (the final system).
- Database designers: these are also *indirect users* of the method. They are important stakeholders in the data modelling process, as they are responsible for implementing the model—for translating it into a physical database. They also have a major impact on the outcome of the process, as if the data model is not implemented correctly, the system is unlikely to meet user requirements. (Moody and Shanks, 1998b).

- **Project managers:** these are not users of the method at all, but have a major stake in the outcome—they are responsible for the overall success or failure of the project. These were the primary stakeholders in the decision to adopt the proposed method on the project, so they were involved in the intervention from the beginning.

While only analysts and end users were considered as stakeholders in the original theoretical framework (Figure 2), project managers and database designers emerged as important stakeholders during the course of the study. This illustrates how in action research, unlike traditional positivist research, the theoretical framework can evolve during the course of the study.

### **Documentation and Maintenance of Data Models (Analyst’s Viewpoint)**

Overall, the method was well received by analysts on the project. Most felt that it was easier to develop and verify data models with end users on a subject area by subject area basis. Also, data models could be easily drawn on normal sized paper, and reduced time spent battling with the drawing tool (which was quite primitive) to make the diagrams readable. The major practical problem experienced in documenting models was the need to manually keep the diagrams consistent with the repository definitions and with other subject areas (foreign entity links). Clearly, this would be improved by a drawing tool that was integrated with the corporate repository and provided explicit support for the concept of levelling.

The method was also positively received by analysts on other development projects. Because this was such a large project, there were frequent meetings with other development teams, and the method began to be used (unofficially) in application development projects across the organisation. Analysts on the project involved in this study had little choice about whether to use the method, as it had been mandated as a project standard. The fact that analysts outside the project began using the method of their own volition provides strong evidence that it was perceived to be useful.

### **End User Validation (End User’s Viewpoint)**

Subject areas proved to be useful in validating the data model with end users and formed “logical units” for signoff of the data model. Separate user reviews were conducted for each subject area, with workshop participants chosen for their specialist knowledge in that area. As a result, users were able to verify a manageable sized “chunk” of the model without having to deal with the model in its entirety. This was a much more efficient process than the existing practice in the organisation, in which a single data model review was typically conducted at the end of each project. It was not uncommon for review sessions to consist of 30 or more people, and to take several days. This was much more cost-effective in terms of users’ time and enabled data models to be reviewed more thoroughly.

### **Translation to Database Design (Database Designer’s Viewpoint)**

The proposed method received a mixed reaction from the database design team. This was not surprising, as database designers had not been considered in the design of the method and were not even included as stakeholders in the theoretical framework (Figure 2). The division of the data model into subject areas was found to be useful for managing the database design work, and the same subject area boundaries were retained during the Logical Design phase. However they had problems with single directional foreign entity links, and asked for foreign entities to be shown in both directions (see discussion in the next section). They also did not find the higher level diagrams useful (Context Data Model and Cornerstone Data Models) and asked for a consolidated data model showing all entities and relationships on a single diagram. However observation of their work practices showed that they worked mainly at the level of individual subject areas, and used the consolidated diagram simply as a way of visually plotting their progress in completing the design work (by colouring in entities).

### Management of Analysis Work (Project Manager’s Viewpoint)

An unexpected benefit of the method was discovered in managing the data modelling process. This was not one of the original objectives of the method, and project managers were not considered as stakeholders in the theoretical framework. However the method proved to be extremely useful from a project management point of view:

- Division of labour: Subject areas were used as “units of work” for managing the analysis work. The “divide and conquer” approach helped to reduce quite a daunting analysis task to manageable pieces.
- Interdependencies between project teams: An important issue in large application development projects is the need to clearly define interdependencies between different parts of the system (Brooks, 1979). Using the proposed method, a single project team was responsible for all changes to each subject area, with foreign entities defining the interfaces between them. The fact that each entity was assigned to one and only one subject area helped to clearly delineate project team responsibilities.
- Monitoring progress: The Subject Area Responsibility Matrix (Figure 4) showed the list of subject areas together with who was responsible and their completion status. This was used as one of the major tools for monitoring the progress of analysis in weekly status meetings.
- Planning and estimation: Subject areas were also used as the unit of analysis for estimation and planning. When a new area of analysis was identified, a new subject area was created and an estimate made of the effort required to complete it.

## 6. LEARNING ABOUT THE METHOD (REFLECTION)

During this study, the method evolved considerably as a result of feedback from those using it (both direct and indirect users). This section summarises the changes that were made to the method. This addresses Research Question 2.

### Subject Area Responsibility Matrix

When the method was initially introduced, a Context Data Model was developed, but this was highly complex and difficult to make sense of. Almost every subject area was related to every other subject area, leading to a spider’s web of crossed lines. This diagram was quickly abandoned as it was not found to be useful. Instead, a simple matrix was produced, listing each of the subject areas, project team responsible data architect responsible and completion status (Figure 4).

Project Team	Subject Area	Data Architect Responsible	Completion Status (%)
Lending	Repayments	A. Poodle	75%
	Loan Application	A. Poodle	45%
	Loan Security	A. Poodle	0%
	...	...	...

Figure 4. Subject Area Responsibility Matrix

This was an *ad hoc* extension to the method to meet the needs of the situation, but proved to be a useful tool for project management purposes. Unlike the Context Data Model, this is a tabular rather than a graphical representation, but provided a one page summary of the model and also progress in completing the requirements analysis task.

## Foreign Entities

A number of changes were made to the conventions for showing foreign entities during the course of the project.

### *Foreign Entity Symbol*

Firstly, the symbol for showing foreign entities was changed because of limitations of the drawing tool. In the original method, foreign entities were shown as shaded rectangles. Because the drawing tool did not support shading, foreign entities were represented instead by the “Predefined Process” flowcharting symbol. This suggests that it really does not matter what symbol is used for foreign entities, so long as they are clearly distinguishable from other entities.

### *Single vs Bi-Directional References*

In the original method, boundary relationships were duplicated on adjoining subject areas to preserve links between subject areas and enable navigation between them. However two problems were identified with this:

- Many of the data models became quite complex because of the number of foreign entities that needed to be shown.
- Duplication of boundary relationships led to additional maintenance effort and the possibility that relationships could become inconsistent.

To address these issues, the conventions for showing foreign entities were changed so that each boundary relationship was shown on a single subject area. The rule used was that an entity was only included as a foreign entity if it appeared as a foreign key within one of the entities on that the subject area. This reduced the average size of subject area diagrams (including foreign entities) by about 20%, and meant that each relationship in the model was shown on only one subject area. This represents the *minimum level of redundancy* possible without losing information. This naturally appealed to the analysts, who were highly focused on minimising redundancy in modelling requirements.

However at the end of the requirements phase, a number of problems were found with to this change:

- Verification of models: Use of this convention led to confusion in user review sessions as a result of relationships that people knew existed, but were not shown on the subject area: for example, “why isn’t client related to account?”
- Navigation: This convention also caused problems in navigating between subject areas—it was possible to follow a foreign entity link and not be able to return to where you began. This is equivalent to showing inter-map references in only one direction in a street directory.
- Translation to database design: Problems were also encountered in the translation of the data model to database design. The database designers needed to be able to see all relationships to a given entity (not just foreign key relationships), in order to understand access paths and referential integrity constraints.

As a result of these problems (and an explicit request from the database designers), bi-directional foreign entity links were re-introduced at the end of the requirements definition phase. This is an example of a “mini-experiment” that didn’t work. However the failure of this mini-experiment provides strong evidence that the level of redundancy in the original method was indeed the “optimal” level of redundancy. Reducing redundancy beyond this point simplifies documentation and maintenance (direct users’ viewpoint) but reduces understandability (indirect users’ viewpoint). This proposition could be tested experimentally. This shows how action research can lead to the formulation of research questions that can be evaluated using traditional hypothesis testing methods. It also shows that a failed action research cycle can provide just as useful information as a successful one. Finally, it emphasises the importance of considering all stakeholders in evaluating the effectiveness of a method. A change to a method may be seen as an improvement by some stakeholders but not by others.

## Context Data Model

In the original formulation of the method, the Context Data Model was shown in the form of an ER diagram. This was seen as a strength of the method because the same diagrammatical conventions were used at all levels of the model—this is also claimed as an advantage by a number of methods previously proposed in the literature (Feldman and Miller, 1986; Simsion, 1989; Teory et al, 1989). However while this seems to provide a degree of theoretical “elegance”, it was found to be a disadvantage in practice. Use of the same symbol to represent both entities and subject areas led to frequent confusion in interpretation, as they have quite different semantics.

Another problem was that in the original method, all boundary relationships were shown on the Context Data Model. When this was applied to the initial scoping data model, it resulted in a diagram that was highly complex and confusing, with relationships between almost all subject areas.

To address these issues, the representation of the Context Data Model was changed in the following ways.

- Different symbols (circles) were used to represent subject areas to clearly distinguish them from entities. Circles were chosen as these were the most easily distinguishable from the rectangles used to represent entities.
- Relationship cardinalities were removed from the Context Data Model. Instead, arrows were used to indicate which direction the relationship should be read.
- Entity labels on relationships (used to indicate the entities involved in boundary relationships at the next level) were also removed.
- Only the most important relationships between subject areas were shown to simplify the diagram. Just as the subject areas “summarise” the entities at the next level, the subject area relationships “summarise” the relationships between subject areas.

The use of different symbols to represent subject areas represents a major insight from this action research study, which contradicts previous research in this area. It may be a general principle that symbols used to represent aggregates should be clearly distinguishable from those used to represent elementary concepts. Experimental studies show that use of the same or similar symbols to represent different concepts leads to confusion in interpretation (Nordbotten and Crosby, 1999). Again, this is a proposition that could be tested using traditional hypothesis-testing approaches.

## Intermediate Level Diagrams

Prior to this action research study, the conventions for intermediate level diagrams had not been explicitly defined. While in theory, the method was designed to handle models of any size, it had not been applied to any examples that required more than two levels. In this study, the Cornerstone Data Models represent intermediate level models, and use the same diagrammatical conventions as the Context Data Model. The justification for this is that cornerstones and subject areas are both instances of the same underlying semantic construct (subject area). While it is important to distinguish between primitive elements and aggregates, it is less important to distinguish between aggregates at different levels of abstraction.

Cross-references between diagrams at this level were not shown—there is no equivalent concept to a foreign entity on intermediate level diagrams. The justification for this was that all boundary relationships were shown on the Subject Area Data Models and did not need to be duplicated at this level. “Foreign subject areas” were briefly experimented with, but were found to make the Cornerstone Data Models much more complex, and did not seem to add to understanding or navigation. As a result, the higher levels of the data model form a “pure” hierarchy (no overlap of elements), while the bottom level models (Subject Area Data Models) form a lattice structure (duplication of foreign entities).

## Decomposition Procedure

Compared to many approaches previously proposed in the literature, the proposed method was far less prescriptive in how to decompose a data model into subject areas. However in reviewing the model at the end of the requirements phase, a great deal of variation was found in the size and composition of subject areas, which reduced the conceptual integrity and understandability of the model. This indicated the need for more precise guidelines for grouping entities into subject areas. While it is important to provide scope for judgement, it is also important for different analysts using the same method to produce consistent results. In large projects especially, explicit guidelines are needed to promote consistency and conceptual integrity. As a result, the decomposition guidelines were refined considerably.

## External Entities

In such a large project, there were many interfaces to external systems. An extension was made to the method to show references to entities in other projects or existing systems—these were called *external entities*. External entities were shown using the foreign entity symbol, but with dotted lines.

## Indexes

In the original version of the method, indexes were proposed for all data model components: entities, relationships and attributes. In this study, reports were developed to generate these automatically from the repository. However the only index that was found to be useful was the entity index. This listed entity names in alphabetical order together with its primary subject area and any secondary subject areas (subject areas on which it appeared as a foreign entity). Whenever someone wanted to find a relationship or attribute (which was relatively rare), this was more easily done via an electronic lookup in the repository.

## Exiting the Problem Situation

Following completion of this project, the method was used on another large application development project. Only minor changes were made to the method as a result of this second study, which suggests that the method had become relatively stable. The method was then adopted as a corporate data modelling standard and incorporated into the bank's development methodology.

## 7. CONCLUSION

This action research study represents a genuine “real world” test of the proposed method. It was applied on a project of realistic size and complexity, and was used intensively by a large number of people, both experts and novices, over a one year period.

### Practical (Action) Outcomes

This was a highly successful action research project in terms of achieving practical outcomes. It led to improved end user validation of data models, simplified documentation and maintenance and also helped in the management of the analysis and subsequent database design work. The method was received positively by most stakeholders and was subsequently adopted by the organisation as a corporate standard. In this sense, this was a highly successful intervention, in that it resulted in change in practices not just in one project but in the organisation as a whole. The client's willingness to act on conclusions is an indication of the validity of the results of the project (Argyris et al, 1985).

### Research (Knowledge) Outcomes

The action research study was also successful from a research viewpoint, in that a great deal was learned about the method, and it was improved significantly as a result. As might be expected in the first real world test of the method, a large number of changes were made to the method as originally

proposed. Figure 5 summarises the changes to the method and the theoretical framework as a result of the action research study. Elements changed as a result of the study are shaded, and new elements added are shown using dotted lines.

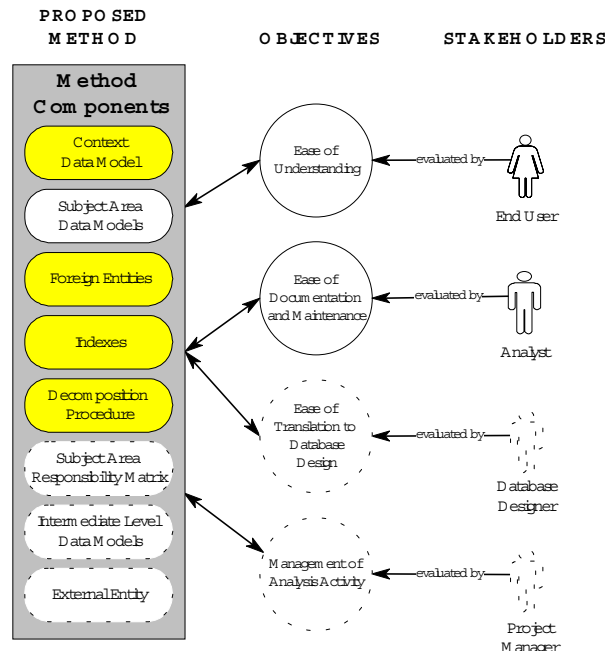


Figure 5. Revised Theoretical Framework

#### *Changes to the Theoretical Framework*

Two new stakeholders and objectives were identified: project managers (management of analysis activity) and database designers (ease of translation to database design).

#### *Changes to the Proposed Method*

As shown in the diagram, almost all of the original components of the method changed to at least some extent and two new components were added.

- Context Data Model (modified): the representation of the Context Data Model changed radically as a result of this study.
- Foreign entities (modified): A new symbol was introduced for foreign entities due to limitations of the drawing tool. Also, one way links were introduced to reduce complexity of diagrams and maintenance overheads. However this led to problems in end user understanding and translation to database design, so bi-directional links were reintroduced.
- Indexes (modified): only the entity index was included in the final data architecture.
- Decomposition Procedure (modified): The rules for decomposing data models into subject areas were refined in order to increase consistency between different analysts using the method.
- Subject Area Responsibility Matrix (new): this was initially introduced as an *ad hoc* replacement for the Context Data Model, but was found to be useful on an ongoing basis for monitoring progress in completing the analysis task.
- Intermediate level diagrams (new): the issue of how to represent data models at more than two levels of abstraction was clarified as part of this study.
- External entities (new): external entities were introduced to show references to entities in external projects or systems.

## **Internal Validity**

The following strategies were used to improve the internal validity of the results:

### *Seeking of Disconfirming Evidence*

The fact that the method changed so much as a result of this study is evidence of the fact that disconfirming evidence was sought and acted upon.

### *Use of Multiple Informants*

A wide range of stakeholders were involved in using the method and evaluating its effectiveness:

- Data architects: direct users
- Business analysts: direct users
- End users: indirect users
- Database designers: indirect users
- Project managers: non-users

### *Participation*

The members of the central data architecture team were involved as participants in the research and contributed to the development and refinement of the method.

### *Use of Multiple Cycles*

A number of different versions of the method were used during the project, some of which turned out to be improvements, others which didn't.

### *Change of Practice*

The proposed method became a pervasive part of project and organisational practices. The fact that the organisation was prepared to adopt the method as a corporate standard provides strong objective evidence that the method was effective in achieving its objectives.

## **External Validity**

The study has high external validity because of the size of the project and the number and diversity of participants involved. The data model developed in this project consisted of over 500 entities, which is more than five times the size of the average application data model (Maier, 1996). Clearly, if the method works successfully on a project of this size, it is likely to work on a project of any size.

## **Other Issues**

In many ways this project was an ideal test of the method because of its size and complexity. Use of such a large project provides highly convincing evidence that the method works in practice (external validity). However one of the problems experienced was that the size of the project imposed constraints on the evolution of the method. Because of the number of people involved, it was difficult to make changes to the method “on the fly” and to experiment with variations to the method. Every time any change was made to the method, there was a need for communication of the change to everyone affected, changes to training materials, changes to standards and changes to existing models. In hindsight, it may have been better to have applied the method for the first time on a smaller project. However given the nature of the research question, a small project would not have provided a realistic test of the method, so this problem was a “Catch-22” situation.

**REFERENCES**

1. AKOKA, J. and COMYN-WATTIAU, I. (1996): "Entity Relationship And Object Oriented Model Automatic Clustering", *Data And Knowledge Engineering*, **20**, 1,
2. ALLWORTH, S. (1996): "Using Classification Structures To Develop And Structure Generic Industry Models", *Proceedings Of The First Australian Data Management Conference*, D.L. Moody, (Ed.), Melbourne, Australia, Australian Data Management Association (DAMA).
3. ALLWORTH, S. (1999): "Classification Structures Encourage the Growth of Generic Industry Models", *Proceedings of the Eighteenth International Conference on Conceptual Modelling (Industrial Track)*, D.L. Moody, (Ed.), Paris, France, Springer.
4. ARGYRIS, C., PUTNAM, R. and SMITH, D.M. (1985): *Action Science: Concepts, Methods and Skills for Research and Intervention*, Jossey-Bass, San Francisco.
5. AVISON, D., LAU, F., MYERS, M. and NIELSEN, P.A. (1999): "Action Research", *Communications of the ACM*, **42**, 1, January.
6. BASKERVILLE, R.L. and WOOD-HARPER, T. (1996): "A Critical Perspective on Action Research as a Method for Information Systems Research", *Journal of Information Technology*, **3**, 11, pp. 235-246,
7. BROOKS, F.P.J. (1979): *The Mythical Man-Month: Essays On Software Engineering*, Addison-Wesley.
8. BUBENKO, J.A. (1986): "Information Systems Methodologies - A Research View", *Information Systems Design Methodologies: Improving The Practice*, T.W. Olle, H.G. Sol and A.A. Verrijn-Stuart, (Eds.), North-Holland.
9. CHECKLAND, P.B. (1991): "From Framework Through Experience To Learning: The Essential Nature of Action Research", *Information Systems Research: Contemporary Approaches And Emergent Traditions*, H.E. Nissen, H.K. Klein and R. Hirschheim, (Eds.), North-Holland.
10. CHECKLAND, P.B. and HOLWELL, S. (1998): *Information, Systems and Information Systems: Making Sense of the Field*, John Wiley & Sons, New York.
11. CHECKLAND, P.B. and SCHOLLES, J. (1990): *Soft Systems Methodology In Action*, Wiley, Chichester.
12. CHOW, S.L. (1988): "Significance Test or Effect Size", *Psychological Bulletin*, **103**, 1,
13. DICK, B. (1997): "Approaching an Action Research Thesis: An Overview". *Action Research and Evaluation On Line (AREOL)*, B. Dick, R. Passfield and P. Wildman, (Eds.), Available on-line at <http://www.scu.edu.au/schools/gcm/ar/arp/guide.html>.
14. DICK, B. (1999): "Sources of Rigour in Action Research: Addressing the Issues of Trustworthiness and Credibility", *Association for Qualitative Research Conference "Issues of Rigour in Qualitative Research"*, Duxton Hotel, Melbourne, Victoria.
15. FELDMAN, P. and MILLER, D. (1986): "Entity Model Clustering: Structuring A Data Model By Abstraction", *The Computer Journal*, **29**, 4,
16. GANDHI, M., ROBERTSON, E.L. and VAN GUCHT, D. (1994): "Levelled Entity Relationship Models", *Proceedings Of The Thirteenth International Conference On The Entity Relationship Approach*, P. Loucopoulos, (Ed.), Manchester.
17. GILBERG, R.F. (1986): "A Schema Methodology For Large Entity-Relationship Diagrams", *Proceedings Of Fourth International Conference On The Entity Relationship Approach*, P.P. Chen, (Ed.).
18. HATTEN, R., KNAPP, D. and SALONGA, R. (1997): "Action Research: Comparison with the concepts of The Reflective Practitioner and Quality Assurance", *Action Research Electronic Reader (on-line)*, I. Hughes, (Ed.), <http://www.behs.cchs.usyd.edu.au/arow/Reader/rmasters.htm>, The University of Sydney.
19. HOLTER, I.M. and SCHWARTZ-BARCOTT, D. (1993): "Action Research: What is it, how has it been used and how can it be used in nursing?", *Journal of Advanced Nursing*, **18**,
20. IVARI, J. (1986): "Dimensions Of Information Systems Design: A Framework For A Long Range Research Program", *Information Systems Journal*, June,
21. JÖNSSON, S. (1991): "Action Research", *Information Systems Research: Contemporary Approaches And Emergent Traditions*, H.E. Nissen, H.K. Klein and R. Hirschheim, (Eds.), North-Holland.
22. KEMMIS, S. and MCTAGGART, R.E. (1988): *The Action Research Planner*, Deakin University, Melbourne, Australia.

23. KIM, Y.-G. and MARCH, S.T. (1995): "Comparing Data Modelling Formalisms", *Communications of the ACM*, **38**, 6,
24. MAIER, R. (1996): "Benefits And Quality Of Data Modelling-Results Of An Empirical Analysis", *Proceedings Of The Fifteenth International Conference On The Entity Relationship Approach*, B. Thalheim, (Ed.), Cottbus, Germany, Elsevier.
25. MARTIN, J. and MCCLURE, C.L. (1985): *Diagramming Techniques for Analysts and Programmers*, Prentice-Hall, Englewood Cliffs, N.J.
26. MCCUTCHEON, G. and JURG, B. (1990): "Alternative Perspectives on Action Research", *Theory into Practice*, **24**, 3, Summer.
27. MOODY, D.L. (1991): "A Practical Methodology For The Representation Of Large Data Models", *Proceedings Of The Australian Database And Information Systems Conference*, University Of N.S.W., Sydney, Australia.
28. MOODY, D.L. (2000): "Building Links Between IS Research and Professional Practice: Improving the Relevance and Impact of IS Research", *International Conference on Information Systems (ICIS'00)*, R.A. Weber and B. Glasson, (Eds.), Brisbane, Australia.
29. MOODY, D.L. and SHANKS, G.G. (1998a): "Evaluating and Improving the Quality of Entity Relationship Models: An Action Research Programme", *Australian Computer Journal*, November.
30. MOODY, D.L. and SHANKS, G.G. (1998b): "What Makes A Good Data Model? A Framework For Evaluating And Improving The Quality Of Entity Relationship Models", *Australian Computer Journal*, August.
31. NORDBOTTEN, J.C. and CROSBY, M.E. (1999): "The Effect of Graphic Style on Data Model Interpretation", *Information Systems Journal*, **9**,
32. OOSTHUIZEN, M. (2000): "Action Research". *Research Methods for Students and Professionals: Information Management and Systems*, K. Williamson, (Ed.), Charles Sturt University, Centre for Information Studies.
33. POULOU DI, A. (1999): "Information Technology for Collaborative Advantage in Healthcare Revisited", *Information & Management*, June.
34. RESCHER, N. (1973): *The Primacy of Practice*, Basil Blackwell, Oxford.
35. RESCHER, N. (1977): *Methodological Pragmatism: Systems-Theoretic Approach to the Theory of Knowledge*, Basil Blackwell, Oxford.
36. SHANKS, G.G. (1996): *Building And Using Corporate Data Models (PhD Thesis)*, Department Of Information Systems, Monash University, Melbourne, Australia.
37. SIMSION, G.C. (1989): "A Structured Approach To Data Modelling", *The Australian Computer Journal*, August.
38. STRINGER, E.T. (1996): *Action Research - A Handbook for Practitioners*, SAGE publications, London.
39. THEORY, T.J., WEI, G., BOLTON, D.L. and KOENIG, J.A. (1989): "ER Model Clustering As An Aid For User Communication And Documentation In Database Design", *Communications Of The ACM*, August.
40. THALHEIM, B. (1999): "The Strength of ER Modelling". *Current Issues in Conceptual Modelling*, P.P. Chen, J. Akoka, H. Kangassalo and B. Thalheim, (Eds.), Springer-Verlag (Lecture Notes in Computer Science).
41. WAND, Y. and WEBER, R.A. (1993): "On the Ontological Expressiveness of Information Systems Analysis and Design Grammars", *Journal of Information Systems*, October.
42. WEBER, R.A. (1997): *Ontological Foundations Of Information Systems*, Coopers And Lybrand Accounting Research Methodology Monograph No. 4, Coopers And Lybrand, Melbourne, Australia.
43. WYNEKOOP, J.L. and RUSSO, N.L. (1997): "Studying Systems Development Methodologies: An Examination Of Research Methods", *Information Systems Journal*, **7**, 1, January.