2009

Current Trends and Future Directions In The Practice Of High-Level Data Modeling: An Empirical Study

Barbara Anglim
Monash University

Simon Milton
The University of Melbourne

Jayantha Rajapakse
Monash University

Ron Weber
Monash University

CURRENT TRENDS AND FUTURE DIRECTIONS IN THE PRACTICE OF HIGH-LEVEL DATA MODELING: AN EMPIRICAL STUDY

Barbara Anglim*, Simon K. Milton^, Jayantha Rajapakse*, Ron Weber*

*Faculty of Information Technology, Monash University, Australia.

^Department of Information Systems, The University of Melbourne, Australia.

Abstract

Many organizations now purchase and customize software rather than build information systems. In this light, some argue that high-level data modeling no longer has a role. In this paper, we examine the contemporary relevance of high-level data modeling. We addressed this issue by asking 21 experienced data-modeling practitioners to reflect on their work and to give their opinions on trends and future directions in high-level data modeling. We analyzed transcripts of our interviews with them using Klein and Myers's (1999) framework for qualitative research. We found considerable variation in the practice of high-level data modeling. We also found that high-level data modeling is still considered important, even though organizations ultimately may purchase off-the-shelf software. The reason is that high-level data modeling assists organizations to obtain clarity about IT project scope and requirements, thereby reducing the risk that costly implementation mistakes will be made.

Keywords: high-level data-modeling practice**,** conceptual data-modeling practice, logical data-modeling practice, enterprise systems, package selection.

1 INTRODUCTION

Historically, organizations have built information systems in-house to solve business problems, exploit business opportunities, support business processes, and enable new products or services to be delivered. In this context, part of the practice of building an information system involved analyzing the data needs of an organization so that databases could be designed and implemented to support the system. This requirement gave rise to the traditional database design life cycle of steps reported in contemporary textbooks (e.g., Teorey et al. 2006; Hoffer et al. 2005) and attempts to develop relevant theory to support database design (e.g., Batra 2007; Codd 1970; Wand and Weber 2002).

The steps commonly articulated in database design (e.g., Teorey et al. 2006; Hoffer et al. 2005) and more broadly addressed in systems analysis (e.g., Kendall and Kendall 2004) include two that are often called *conceptual* data modeling and *logical* data modeling. Together, these steps are sometimes called *high-level* data modeling. They are undertaken before *physical* data modeling, which involves taking into account a target database management system and a hardware/software platform.

In this research, our focus is on conceptual and logical data modeling (i.e., high-level data modeling). Both steps involve constructing and using data models to represent, clarify, define, and relate important business phenomena. Often the distinction between the two steps revolves around the need in *logical* data modeling to consider key challenges with and to avoid common flaws in database implementation. In contrast, conceptual data modeling focuses much more on the meaning, use, and definition of and relationships among key business entities.

The motivation for our research is that the need for, purposes of, and activities involved in high-level data modeling appear to be changing significantly. For instance:

- Many information systems are now considered to be a commodity. As a result, organizations often no longer *build* information systems. Instead, they *buy* ready-made packages to support their information systems needs (Davenport 1998; Shang and Seddon 2002). Often these packages require extensive configuration and customization (Markus and Tanis 2000).
- To support managerial decision-making, many organizations now build data warehouses and use various business-intelligence tools that obtain data taken directly from business transactions (Kimball et al. 1998; Wixom and Watson 2001).
- The strategic importance of data is increasing, especially at the corporate level (e.g., Shanks 1997). For instance, acquisition and merger activities are unlikely to be successful if information systems cannot be integrated.
- Patterns are playing an increasingly important role in contemporary high-level modeling (e.g., Batra, 2005; Hay, 2006; Silverston 2001a, 2001b). Furthermore, a number of industry-standard data models are now appearing (e.g., in health services, http://www.hl7.org/, telecommunications, http://www.tmforum.org/, and supply-chain management, http://www.supply-chain.org/).

These changes affect the practice of high-level data modeling. Historically, some research has been done on data modeling *practices* (e.g., Batra and Marakas 1995; Floyd 1986; Necco et al. 1987). A paucity of research exists, however, in relation to contemporary high-level data modeling *practice*. Some work has been done on design and development methodologies (e.g., Kautz et al. 2004) and process modeling (e.g., Chang et al. 2001), but little work has been done on data modeling broadly or high-level data modeling specifically.

In response to a paucity of research on conceptual modelling *practice*, Davies et al. (2006) conducted a survey of conceptual modeling practitioners in Australia. They had two objectives (p. 359): *(a)* "to obtain empirical data that conceptual modeling is indeed being performed now and into the foreseeable future in IS practice in Australia"; and *(b)* "to find out what are the principal tools, techniques, and purposes for which conceptual modeling is performed currently in Australia." Their focus was conceptual modeling *broadly,* including process and workflow modeling additional to data modeling. They found that (p. 376) "[d]atabase design and management remains the highest average purpose for use of modeling techniques," and that the reasons for continued use of all conceptual-modeling approaches included "communication…to/from stakeholders, internal knowledge…of techniques, user expectations management, understanding models integration into the business, and tool/software deficiencies." The limited qualitative data that came from the survey constrained the depth with which current practice and future trends in data modeling practice could be studied.

In this research, our goal was to explore current practices and trends in high-level data modeling in some depth. We conducted field research in which we undertook semi-structured interviews with 21 highly experienced practitioners to obtain their views about high-level data modeling. We sought a deep understanding of the reasons behind the answers they provided to our questions. Specifically, we used the following broad questions as a "scaffold" for the interviews we conducted with them:

- Why is high-level data modeling undertaken?
- What high-level data-modeling activities are undertaken?
- How are high-level data-modeling activities undertaken?
- Who is involved in undertaking high-level data-modeling activities?
- What quality-assurance measures are used in high-level data modeling?
- What is the future of high-level data modeling?

We canvassed these questions with our interviewees in a fairly informal way. Our concern was to use the questions to elicit and trigger their opinions and reflections rather than to impose a rigid structure on our interactions with them. We wanted to understand the nature and purposes of the activities they undertook and the meaning they ascribed to high-level data modelling (Neumann 2007, pp. 276-278).

The remainder of the paper proceeds as follows. First, we discuss our research method. Next, we present our results in relation to each of our research questions. We then summarize our findings. Finally, we present some implications of and some strengths and limitations of our research.

2 RESEARCH METHOD

Because our research is exploratory, we adopted an interpretative research method. In this light, we followed Klein and Myers' (1999) principles for the conduct of interpretative research (Table 1). During 2007 and 2008, we carried out interviews with 21 experienced data-modeling professionals engaged in high-level data modelling. Each interview lasted about 90 minutes. All interviews were transcribed. Using techniques from Miles and Huberman (1994), we then sought to make sense of interviewees' experiences with and views about high-level data modeling. Specifically, we employed a descriptive matrix to relate comments made by our intereviewees to our broad research questions.

Klein and Myers Principle	Approach to Addressing the Principle
Contextualization	Our experience allowed us to understand interviewees' responses and reflections and to engage actively with interviewees.
Interaction between researchers and interviewees	We have substantial academic and practical data-modelling experience. This experience enabled us to establish rapport with interviewees, which facilitated disclosure and a deeper understanding of issues that the questions surfaced.
Hermeneutic circle	Analysis of interview transcripts followed multiple cycles. We constantly reflected independently and as a research team on the transcripts until we felt we understood underlying themes that interviewees were seeking to convey.
Abstraction and generalisation	By applying principles in Miles and Huberman (1994), we linked emerging themes back to our research questions.
Dialogical reasoning	A research assistant with data-modeling training but no commitment to a particular theory or method of data modeling collated an initial list of key themes from the transcripts before we explored them deeply.
Multiple interpretations	We recognized that interviewees sometimes held different views about data-modeling phenomena and sought to understand these conflicting views from multiple perspectives.
Suspicion	We used our experience to see when interviewees brushed over (or avoided) questions and rephrased our questions to try to elicit more-helpful responses.

Table 1. How we followed Klein and Myers' (1999) principles for interpretive research

Table 2 provides an overview of our interviewees. All had constructed at least five major data models in at least two different industries or areas of government. Some were highly regarded among their peers (both nationally and internationally) because of their substantial data-modeling expertise.

Interviewee	Number of High-Level Models Completed	Industries Where Work Undertaken
I#1	10+	Banking, health, telecommunications.
I#2	Unknown, but many	Many; over 30 years' experience.
I#3	30-40	Half-dozen different industries.
I#4	100+	Many, including manufacturing, telecommunications, banks, insurance, social security, community services, land title offices, defense
I#5	6+	Banking, insurance, telecommunications, higher education.
I#6	200+	Many, including government, construction, manufacturing, retailing.
I#7	10-12	Health insurance, life insurance, property, telecommunications, finance, retail.
I#8	100+	Many; more than 20.
I#9	50+	Many; over 30 years' experience, including fire management, game

		and wildlife licensing, minerals and petroleum, primary production, chemical standards, banking, law enforcement, community services, telecommunications.
I#10	40+	Many, including government, financial services, utilities.
I#11	Unknown, but many	Many; over 12 years' experience; health, utilities, finance, telecommunications, insurance.
I#12	Unknown, but many	Many; over 14 years' experience.
I#13	5-6	Primarily telecommunications.
I#14	20-25	Over 15 years experience; 8+ industries.
I#15	30+	Over 10 years experience; 5+ industries.
I#16	12, reviewed many	Over 15 years experience; Government mostly social security.
I#17	50	Over 15 years experience; Many government departments in different areas, limited business.
I#18	10-20	Four different areas of government.
I#19	20	Several areas of government mainly in health and welfare.
I#20	50+	Several areas of government mainly in social security.
I#21	200+	Six different areas of government.

Table 2. Overview of research interviewees

We sought interviewees by making presentations to local chapters of the Data Management Association (DAMA) and Data Warehousing Association (DWA) in three cities. We explained the nature of our research, its purposes, the assistance we required, and the qualifications of individuals from whom we sought assistance. We obtained contact details from individuals who indicated they were willing to assist us. In some cases, to comply with ethics restrictions, we used 'snowballing' whereby individuals asked other data modelers to contact us to indicate their willingness to participate in our research. Subsequently, we scheduled interviews with all who had agreed to assist us.

3 RESULTS

The following subsections present our findings in relation to each of our major research questions. We provide illustrative quotations to support our findings. Quotations are labelled as, for example, I#4, which means a quote from Interviewee 4's transcript.

3.1 Reasons for Undertaking High-Level Modeling

Previous research has considered conceptual modeling as a precursor to logical modeling, which in turn leads to physical data modeling. This sequence assumes that an application supported by a database will be built. Interviewees had alternative views on reasons why conceptual modeling was *now* undertaken by organizations.

Scoping for proposed projects was the most-common reason given for preparing a conceptual model. I#4, who is a highly experienced modeler, stated that rigor is unimportant in conceptual modeling. He commented that the primary purpose of the conceptual models he prepared was to scope the project. Furthermore, he argued that conceptual models are not prepared with the objective of "*getting towards implementation or working out business rules.*" I#6 also considered that articulating rigorous business requirements was unnecessary at the conceptual-model level. Instead, conceptual modeling was done just to "*scope the next level of activity.*" I#2 describes a model as a "*means to the end*" with a project's scope report as a likely associated key deliverable. I#1 described "*scope as possibly the key part of the conceptual model.*"

I#5 said she created conceptual models for two purposes. First, she used them to build an enterprise data model, which she described as not necessarily leading to the immediate purchase of a package or development of a bespoke system. Instead, a conceptual model was a resource that had "*many* *purposes.*" Second, sometimes she used an application conceptual data model for a "*particular application*" development. Others reinforced this view. For example, I#3 described how a quality conceptual model and organizational overview helped direct an organization's selection and purchase of package software and development of applications. In particular, the model and overview helped avoid the pitfalls of systems that evolve independently. He argued that it was better to have the applications "*logically conforming to some overarching model than to have them evolve independently and then try and put them together, which is where we're at today*." Similarly, I#1 spoke of using high-level data models to find "*pain points*" with proposed systems and thereby to manage risk.

The need for integrated information systems, particularly data warehouses, was the driver for some to undertake entity definition work at the conceptual level. In these contexts, interviewees stressed the need for entity definition resolution. Interviewees found potential in conceptual models for transcending individual projects. They were long-term resources that could be used to help integrate and plan an organization's informational needs, projects, applications, and data-warehouse activities.

Without a conceptual model to guide configuration, I#3 argued that configuring a package becomes more expensive and error prone. He commented that lack of a conceptual model to provide guidance during package configuration led to arguments about the meaning of data among different departments in an organization. He gave an example where an organization found it could migrate only 30 percent of data from six billing systems to a new system because of the lack of agreed data definitions. The result was seven systems that lacked interoperability. He was not alone in this observation.

I#1, who has supervised many teams on major projects for decades, argued that high-level data modeling is needed to help avoid major problems for organizations by guiding the evolution of data needs. "*I see the design documents as a key resource for the company.*" While business evolves over time, I#7 argued that conceptual modeling done correctly often needed only fine tuning over time.

3.2 Types of High-Level Data Modeling Activities Undertaken

Interviewees had different views about the nature of conceptual versus logical models. Each particular view appears to be linked to the depth of the project they were undertaking, the domain in which they were modeling, and the complexity or size of the organization they were modeling.

Although interviewees varied in the activities they undertook at the two levels (and sometimes varied in the modeling activities they undertook at these two levels between projects), most considered preparation of conceptual models to be a phase that was distinct from preparation of logical models. Specifically, they indicated the logical level involved adding more details to the model (e.g., keys and attributes) and resolving many-to-many relationships. During the preparation of a logical model, I#4 said that he added more detail to a conceptual model and "*nailed down the rules.*" I#6 commented that rigor in a conceptual model was unimportant. He indicated that his logical or "subject" models had keys and relationships, but his conceptual or "enterprise" models had no keys or relationships.

Those interviewees who constructed enterprise conceptual models did not fully articulate their models. For instance, a fully articulated ER (entity-relationship) diagram would have keys, attributes, and fully resolved many-to-many relationships. On those occasions where he used ER diagrams, I#1 was clear that relationship resolution did not occur at the conceptual-modeling level. Instead, his focus was on identifying "*scope*" and "*pain points*" and achieving some agreement about entity definition.

Most interviewees indicated that the largest amount of work at the conceptual-model level involved getting agreement on definitions. I#4 said this task as "*difficult.*" I#3 commented that obtaining agreement on name definitions was not always possible, and he described it as an ongoing job. Nonetheless, he argued that obtaining agreement on name definitions was important because it meant that business requirements between departments could be understood better and more easily integrated. I#15 thought it important to get the right stakeholders with expertise because it was "*a question of confidence in the answer*" and he could "*manage expectations.*" I#7 endeavored to have one agreed definition for each entity and attribute to ensure "*multiple departments are speaking the* *same language."* She said it is a mistake to allow different departments or applications to have different definitions. I#10 spoke about the importance of consensus about data definitions: "*Build up a glossary and say if we've got two different terms for them and maybe let's think of a term that covers the general, both of them but just trying to agree on the same language*."

I#5 indicated she focuses on definitions needing resolution. She also looks *for key performance indicators*, as these often needed definitional agreement. I#5 indicated she had recently completed a project to help the client get a "*structured business vocabulary*" as a first step. The client then followed through with an articulation of business rules using the agreed vocabulary.

Whether entity definition occurred at the conceptual or logical level varied among interviewees and the types of projects they undertook. I#8 indicated he did not undertake entity definitions during conceptual modeling but rather left this task for "*techies.*" Instead, he gave a "*description of what these areas of information are.*" Most interviewees considered that attaining some agreement on entity definitions was a key part of conceptual model work, however. For example, at the conceptual-model level, I#1 argued the main deliverables were the model and entity definitions. At the logical level, he would include some key attributes. Many interviewees thought a large part of conceptual modeling was establishing business vocabulary or metadata in the form of data definitions relating to business rules. The client organization could then use these definitions as a business "language" for communication across divisions and applications.

I#9 had a different view, however, about the link between conceptual models, logical models, and physical models. He commented, "*they (a client) took that as a logical design which they put into physical code and they got it straight from the conceptual model. Now is that blurring the boundaries?*" Even during his conceptual modeling work, he included a high level of detail in the models he prepared, primarily because he relied substantially on previously developed data modeling patterns. I#12 also included high levels of detail in the models he developed. I#15 had changed his mind about the role for logical data models by saying that "*the physical needs to be supported.*"

Model completion usually was not expected. I#5 said she only represented a few types of phenomena in a conceptual model. Sometimes she might prepare just a simple conceptual model with a few key entities and definitions as a basis for communication among stakeholders in some domain. At other times in small projects, she prepared a conceptual model that included attributes and relationships (which she described as eventually constituting a logical model). She described her most-recent project as being supervised by an experienced modeler who expected rigor. On this project, she had to include detail like relationships at the conceptual level. Moreover, her conceptual model had to link coherently to a logical model that included keys and other "technical information."

Some interviewees said they undertook sub-typing of entities in their models. For instance, I#3 described sub-typing at the logical level as time consuming and expensive. Nonetheless, he felt the clarification of business rules saved much time in implementation. I#14 avoided sub-typing, however, because when he explained "*sub-class, sub-typing, and bringing in sort of universal data modeling type concepts*" to business personnel, "*they get lost*." By way of contrast, he emphasized the need for abstraction: "*So you've got to move up a level there and think of that level and that's important*."

3.3 Ways in Which High-Level Data Modeling is Undertaken

All except one interviewee used workshops to garner the business information required for the data models they needed to construct. Attendees at the workshops were almost always representatives of the business. Some interviewees emphasized that workshops also helped to obtain agreement among client representatives about priorities and definitions of business objects or metadata. For instance, I#11 commented: "*Often it's good for us to get into the workshop and try and thrash out some of the nitty grittys with a lot of business units because often they have conflicting priorities*."

Helping participants feel comfortable during workshops was a reason behind some interviewees' choice of material or methodology. For instance, rather than using software tools to build a model, I#7 indicated she uses whiteboards and hand-drawings to help client representatives feel relaxed and to build the model in conjunction with them. I#10 follows a similar approach: "*Really, to be truthful I tend to start off with a whiteboard and some people sitting around and saying what are the concepts*."

Some interviewees use standard data model patterns that they customize to suit the context in which they are working. For instance, I#1 argued, "*You do not need to reinvent the wheel.*" He had specialized in banks and telecommunication companies where transactional relational databases are common. He had used the *Financial Services Data Model*, which is a proprietary data model developed for the banking and finance sector.

I#9 was the strongest advocate for the use of patterns. He commented that he used patterns "*I guess, in ever-increasing amounts of detail. I can knock together, and this is going to frighten you, I can knock together an enterprise-wide conceptual model in probably 30 to 60 minutes. And you say you're joking, but I can interact with senior management and start to talk and their eyes light up when they realize, hey you understand and all I'm doing is patterns*." I#14 is also a strong believer in patterns: "*I'm a great believer in utilizing what other people have done*." I#12 did not use patterns, but he reused models he had developed during previous engagements with clients.

I#8 said that he started with a straw-man pattern and looked for disagreement with it, because he frequently had only a few days to complete a model. He commented that he often had already decided what package was likely to suit his client's business. As a result, he would look to make only slight adjustments to the patterns he used. He saw patterns as a fast means of getting input from end users. I#10 also sometimes used a straw-man approach: "*The other way is probably a bit more common…come up with a straw-man yourself and then go through and sort of say well is this right*."

One of the more-experienced interviewees, I#4, replied that he used patterns only occasionally and used them only in the sense of capturing previous experience. He indicated that he discarded them quickly, however, if they did not fit the context. I#7 found dangers when presenting models prior to eliciting business requirements. She warned that end users might simply accept a pattern without sufficient reflection on its accuracy and completeness. As a result, she emphasized the importance of starting with a clean whiteboard and working with end users to create their own model.

I#3 was the most outspoken against patterns. He argued that it was "best practice" *not* to assume similarities among models and to commence modeling with an open mind. He referred to recent experiences with patterns in which he found that "*a US telco is very different from an Australian telco.*" Despite the fact that telecommunication companies often offer similar products, he found that they applied different business rules that required different models. He had found that patterns did not save time and led to important omissions in data definitions. I#5 commented that some proprietary data models were too expensive for some organizations to purchase.

I#19, who has worked primarily in government, was keen to identify "reference sets" in his modeling workshops. He indicated that he "*would source (reference data) from somewhere else*." Examples he gave included "countries of the world," "area codes," and "telephone prefixes."

3.4 Participants in High-Level Data Modeling

Interviewees were split as to whether IT professionals should be included in workshops. If IT professionals were present, however, interviewees held a common view–namely, that IT professionals should play little or no role in the discussions that ensued in a workshop. Instead, IT professionals should play the role primarily of observers.

I#5 voiced one perspective saying she *"does not deal with IT folks… I may have some present but only as observers."* I#18 preferred *"to concentrate on the business problem."* Others had regrets like I#6: *"I wish I did work with people at that (the technical) level, but consultants come in."*

3.5 Quality-Assurance Practices in High-Level Data Modeling

Throughout our interviews, it became clear that quality-assurance practices like audits, walkthroughs, reviews, metrics, and documentation standards received little attention during high-level data modeling. Our findings stand in contrast to the modeling standards now in force in organizations such as the U.S. Defense Department (where UML has been adopted as the conceptual-modeling standard). Interviewees stressed that model correctness and completeness were important for high-level data modeling, but that standards were not defined clearly. For instance, interviewees selected a modeling methodology depending on the project, personal leanings, or their client's preferences. Moreover, conformity with standards was often impossible because of time constraints imposed on them.

We asked interviewees how their models were validated and updated. Many indicated they were hired as consultants and were not given time to validate and correct their models. For the most part, any validation that occurred was an outcome of interactions among attendees at modeling workshops.

I#7 aims to get consensus about correct and complete business requirements and data definitions despite the fact that "*a lot of companies are not willing to invest that time.*" She indicated that some organizations were willing to devote only an hour or two to model validation. I#8 acknowledged that getting agreed data definitions was pivotal to quality and validation concerns. Nonetheless, he argued that trying to get agreement on definitions quickly was "*counterproductive.*" His view is that during modeling he must "*understand business and not techno stuff.*" He prefers to get a "*description of what these areas of information are*" rather than an agreed definition because "*we can't let things drag on.*"

Interviewees involved with data-warehousing projects seemed to have more opportunities to validate and update their models. For instance, I#5 said she was responsible for updating her conceptual models if errors and omissions were found during logical-level modeling. Similarly, I#7 remains engaged until the "*physical-build*" stages of a warehousing project. She oversees quality checks that she suggests for her logical model. Her model is updated when errors and omissions are found.

With governments, I#4 indicated he has been given the chance to undertake quality reviews. He went through many cycles of feedback with stakeholders until differences relating to the models he had prepared were resolved. Nonetheless, for most of his projects, achieving correct and complete high-level data models was not considered important. He commented, "*you do not really worry about getting it perfect,*" because once the model is used it becomes evident how it should be "*tweaked.*" I#8 went further and stated that aiming to get most data definitions right at the conceptual level is "*counterproductive.*" He felt definitions and relationships should be left for "*techies*."

Interviewees were asked about the tools they used to enhance the quality of their models, such as tools to check syntax and reverse engineer models based on existing systems. They responded that they used whatever tools their client organization wanted or provided. Few mentioned use of expensive tools. Those who did often had large government organizations as their clients. For instance, I#6 described liking the "*Mega*" tools, and commented that "*System Architect*" was "*brilliant.*" Nonetheless, he indicated that these tools often require full-time personnel to support their use because they are highly complex.

I#6 said that he preferred to use the types of tools that allow his models to be linked with other models (including physical models) and ultimately with the data that populates his models. He described the tools as not so much CASE (computer-aided software/system engineering) tools but as tools to manage business areas on many levels. He stated that use of these tools enhanced the quality of models because they facilitate validation and update of the models–they allow a system's requirements and architecture to be integrated and considered from several perspectives. In relation to the projects he undertook, I#6 pointed out a trend toward increasing demands of cohesiveness between (a) different organizations models, and (b) business and technology. He observed that some data-modeling tools were evolving to support this trend.

I#5 sees herself as exceptional among modelers in her attempts to provide metadata about each object in her models. She spoke of a procedure whereby a *"series of assertions"* derived from a model was presented to stakeholders for validation as a thorough, best-practice technique. Her experience with using this approach, however, was that business people were unable to understand it or refused to devote time to it. Instead, like I#16, she recommended using a walkthrough with stakeholders that focused primarily on entities. I#18 used quasi-assertions in which he would "*explain each one of those relationships both backward and forward and find out… is this correct or is it not correct?*"

Like I#5, I#3 seeks to cross-reference metadata definitions with business rules to test the validity of his models. He was the only interviewee who expressed his ability to undertake model validation in a fairly formal way. He used the most-extensive procedures to validate his models. For him, quality was also about exploring business requirements rather than using presupposed patterns. He uses three validation techniques. (1) He presents his model to stakeholders for feedback and questions and resolves errors, omissions, and ambiguities. (2) He describes the business rules in a spreadsheet and asks stakeholders to check their validity. (3) He asks stakeholders to check his model through scenarios. He particularly focuses on the validity of subtypes in his models.

In short, our interviewees indicated that they had only limited influence over quality-assurance procedures. Unlike other areas of IT practice where quality checks, metrics, validations, updates, and reviews are now standard, they often are not features of high-level data-modeling environments. Moreover, quality problems appear to be compounded by a general lack of stakeholder input and interaction and the distribution of modeling activities across different IT professionals.

3.6 Future of High-Level Data Modeling

Interviewees were emphatic and unanimous about the ongoing need for high-level data modeling. Several argued that "*erroneous ideas*" about the declining importance of high-level data modeling were "*driven from the package industry.*" I#7 argued that good high-level data models save millions of dollars by avoiding the purchase of inappropriate, sometimes expensive, software systems. I#1 indicated that a package might not fit all business requirements. He believes that knowledge of the overall business requirements is necessary to make an informed decision about package purchases, especially expensive packages.

I#5 believes a resurgence in high-level data modelling is occurring, driven partly by greater use of Service Oriented Architectures (SOAs), which are a relatively low-cost approach to aid application integration across organizations. I#5 is working increasingly with "*information architects*," who have been appointed by their organizations to achieve the goal of better data management. I#5 argued that pressures to develop high-level models quickly were counterproductive to the growing need of "*providing an enterprise asset for the organization.*"

I#3 commented that systems integration and interoperability were now a high priority for many organizations. He argued that high-quality, high-level data models help with the development of applications systems and the purchase of packaged software. They enable organizations to avoid the pitfalls associated with bespoke systems that are developed independently. Moreover, they help ensure that packages can be integrated with other systems in the organization. In the absence of high-quality high-level data models, I#3 argued "*some really expensive failures*" can result.

I#7 argued that high-level data models were increasingly important to effective strategic planning. Some types of organizational strategy required systems integration. In this regard, I#6 said that the need for integrated information systems required organization-level data management, which in turn was linked to a trend toward increasingly big and complex organizations and applications.

I#6 argued that the assets of an organization are "*in the requirements and design,*" because nowadays one can "*auto generate code*" or "*select packages*." Therefore, he believes that increasing emphasis will be placed on developing high-quality design documents and data models. He also stressed the importance of governance to support high-level data modeling. He has observed many times that models "*fell apart*" when good governance structures and procedures were not in place and working. He explained that a clear governance structure effectively allows discrepancies among high-level data models to be resolved and an organization to own changes that it makes to its data models. He now prefers to work only for organizations that have effective governance in place. I#6 argued that effective governance occurs, however, only when a leader recognizes the importance of clearly stating business process requirements and is willing to devote resources to support quality business analysis.

I#21 was circumspect about the role that high-level data models could play in package selection by helping to drive a "*side-by-side functionality capability check*" between the package and requirements. I#19 cited a specific example of how data modeling helped build a requirements document where the package vendor "*knew exactly what our requirements were at a data level, at the process level, at the use-cases.*" Another project took nine months longer than expected where the vendor "*saw us coming a mile away… because the requirements were not clear*" after I#19 had warned the internal client "*you need to have models for this*."

Where companies continue to build bespoke systems, agile methodologies are now in the ascendancy. Those data modelers who used agile methodologies identified advantages and challenges. I#20 sees a key role for high-level data modeling and observed that sometimes "*what falls off the bottom of agile prototyping is the underlying model,*" which means it is difficult to extract data into warehouses. I#21, who is also experienced with agile methodologies, said that "*we use a tool to gather the stories*" and "*exercise the model… on how well it supports the business functionality… (and if it) can actually support the business processes*."

4 SUMMARY OF FINDINGS

Table 3 is a descriptive matrix that summarizes major *themes* (Miles and Huberman 1994, p. 246) we identified in the responses provided by interviewees to our research questions.

Research Question	**Findings/Themes**
Why undertake high-level data modeling?	To scope projects, build enterprise data models, guide the evolution of an organization's data needs, select and configure packages, manage project risk, and develop applications.
What high-level data-modeling activities are undertaken?	Building conceptual data models in which entities and sometimes relationships are defined. Building logical data models in which attributes and keys are defined. Gaining agreement on entity definitions.
How are high-level data-modeling activities undertaken?	Via workshops and sometimes by using patterns, past models that the modeler has developed, industry-standard data models, and CASE tools (often those tools mandated by the client).
Who is involved in high-level data-modeling activities?	Modeler and most often key business stakeholders.
What quality-assurance measures are used in high-level data modeling?	Quality-assurance activities are limited due to lack of client resources and enthusiasm.
What is the future of high-level data modeling?	To articulate business requirements before package selection, help manage data that needs to be integrated in service-oriented architectures, and improve the effectiveness of strategies that lead to system integration in large, complex organizations by clearly defining data to be integrated.

Table 3. Summary of results

In our view, these *themes* enable us to characterize contemporary high-level data-modeling practice in the following way:

- Relative to the past, most high-level data modeling is now undertaken with much-less emphasis on database design. Instead, the goals are to *(a)* identify critical business concepts and their relationships, and *(b)* provide guidance for projects, especially in terms of a project's scope.

- Group techniques dominate contemporary high-level data modeling. Experienced modelers play the important role of facilitating stakeholder discourse. They assist stakeholders to share ideas and information, negotiate outcomes, and establish shared vocabularies. IT professionals are seldom included as active stakeholders in group activities.
- High-level data modeling increasingly is used to underpin package selection, data-warehouse design, and enterprise planning.
- Quality assurance in relation to high-level data models often is not undertaken as a distinct activity. Instead, it is an active process that occurs, somewhat subliminally, during workshops. To some extent, it is facilitated by using industry-standard data models and data-model patterns.
- Experienced data modelers see an ongoing and more-important role for high-level data modeling.

5 IMPLICATIONS, STRENGTHS, AND LIMITATIONS

Our findings have implications for research, teaching, and practice. For research, it is clear that the objectives of and practices associated with high-level data modeling are changing. A deeper understanding is needed of the motivation for and nature of these changes and the likely ways they will unfold. Our findings show that variations exist in the views of and activities undertaken by practitioners. A deeper understanding of why these variations exist is required. Ultimately, theories are needed to account for the variations.

For teaching, we believe high-level data modeling still needs to be an important part of the curriculum taught to intending IT professionals. Moreover, it appears that conceptual modeling, logical modeling, and physical modeling increasingly are becoming distinct areas of expertise and practice. While students need to understand how these three types of data modeling are related, it may be advantageous for them to develop higher levels of competence in one particular type of data modeling.

For practice, we believe the management of organizations should revisit the role that high-level data modeling plays in their planning activities and the operations of their organizations. In some organizations, this role seems to have become diminished, mainly because of some of the rhetoric around package software (e.g., that packages obviate the need for detailed requirements specifications and thus data models). Our findings show that high-level data modeling is transcending its traditional roles in system development and playing an increasingly important role in strategy and planning.

A strength of our research is that it is one of few studies investigating how highly skilled practitioners undertake high-level data modeling. It is also novel because it explores the under-researched role of high-level data modeling in the post-bespoke-systems era. Because we have used an interpretive approach to the analysis of data collected in our research, we believe we have obtained rich insights into the current practice of high-level data modeling. Thus, our research should inform both researchers and practitioners.

A limitation of our research is that we have obtained only the opinions and reflections of our interviewees. We did not observe them directly, and we recognize that espoused theories may not reflect theories-in-use (Argyris and Schön 1974). Furthermore, because we interviewed only *active* experienced data modelers, our results potentially are biased toward emphasizing the ongoing importance of high-level data modeling.

References

Argyris, C., and Schön, D. *Theory in Practice: Increasing Professional Effectiveness*, Jossey-Bass Publishers, San Francisco, 1974.

Batra, D. "Conceptual Data Modeling Patterns: Representation and Validation," *Journal of Database Management* (16:2), 2005, pp. 84-106.

Batra, D. "Cognitive Complexity in Data Modeling: Causes and Recommendations," *Requirements Engineering* (12), 2007, pp. 231-234.

Batra, D., and Marakas, G.M. “Conceptual Data Modeling in Theory And Practice,” *European Journal of Information Systems* (4:3), 1995, pp. 185–193.
Chang, S., Kesari, M., and Seddon, P. “A Content-Analytic Study of the Advantages and Disadvantages of Process Modelling,” in: J. Burn, C. Standing, P. Love, (Eds.), *14th Australasian Conference on Information Systems*, Perth, 2003.
Checkland, P., and Holwell, S. *Information, Systems and Information Systems: Making Sense of the Field*, John Wiley & Sons, Chichester, 1998.
Codd, E.F. “A Relational Model of Data for Large Shared Data Banks,” *Communications of the ACM* (13:6), June 1970, pp. 377-387.
Darke, P. & Shanks, G. “Understanding Corporate Data Models,” *Information and Management* (35), 1999, pp. 19-30.
Davenport, T. “Putting the Enterprise into the Enterprise System,” *Harvard Business Review*, (76:4), July-August, 1998, pp. 121-131.
Davies, I., Green, P., Rosemann, M, Indulska, M., and Gallo, S. “How Do Practitioners Use Conceptual Modeling in Practice,” *Data & Knowledge Engineering* (58), 2006, pp. 358-380.
Floyd, C. “A Comparative Evaluation of System Development Methods,” in *Information Systems Design Methodologies: Improving the Practice*, North-Holland, Amsterdam, 1986.
Hay, D.C. *Data Model Patterns: A Metadata Map*, Morgan Kaufmann Publishers, 2006.
Hoffer, J.A., Prescott, M.B., and McFadden, F.R. *Modern Database Management*, Pearson Prentice Hall, NJ, 2005.
Kautz, K., Hansen, B., and Jacobsen, D. “The Utilization Of Information Systems Development Methodologies in Practice,” *Journal of Information Technology, Cases and Applications* (6:4), 2004, pp. 1-20.
Kendall, K.E., and Kendall, J.E. *Systems Analysis and Design*, Pearson Prentice Hall, NJ, 2004.
Kimball, R., Reeves, L., Thornthwaite, W., and Ross M. *The Data Warehouse Lifecycle Toolkit: Expert Methods for Designing, Developing and Deploying Data Warehouses*, John Wiley & Sons, New York, 1998.
Klein, H.K., and Myers, M.D. “A Set of Principles for Conducting and Evaluating Interpretive Field Studies in Information Systems” *MIS Quarterly* (23:1), March 1999, pp. 67-94.
Markus, M.L. and Tanis, C. “The Enterprise Systems Experience - From Adoption to Success,” in *Framing the Domains of IT Research: Glimpsing the Future Through the Past*, R.W. Zmud (ed.), Pinnaflex Educational Resources, Cincinnati, Ohio, 2000, pp. 173-207.
Miles, M.B. and Huberman, A.M. *Qualitative Data Analysis*, 2nd Edition, Sage Publications, 1994.
Necco, C.R., Gordon, C.L., and Tsai, N.W. “Systems Analysis And Design: Current Practices,” *MIS Quarterly* (11:4), December 1987, pp. 461–475.
Neuman, W.L. *Basics of Social Research: Qualitative and Quantitative Approaches*, 2nd Ed., Pearson Education, Boston, 2007.
Shang, T, and Seddon, P.B. “Assessing and Managing the Benefits of Enterprise Systems: The Business Manager's Perspective,” *Information Systems Journal* (12 :4), October 2002, pp. 271-299.
Silverston, L. *The Data Model Resource Book: Revised Edition: Volume 1: A Library of Universal Data Models for All Enterprises*, John Wiley & Sons, New York, 2001a.
Silverston, L. *The Data Model Resource Book: Revised Edition: Volume 2: A Library of Universal Data Models by Industry Types*, John Wiley & Sons, New York, 2001b.
Shanks, G. “The Challenges of Strategic Data Planning in Practice: An Interpretive Case Study,” *Journal of Strategic Information Systems* (6*)*, 1997, pp. 69-90.
Simsion G. C. and Witt G. C. *Data Modelling Essentials: Analysis, Design and Innovation*, 2nd Ed., Coriolis, Arizona, 2005.
Teorey, T.J., Lightstone, S., and Nadeau, T. *Database Modeling & Design: Logical Design*, Morgan Kaufmann Publishers, San Francisco, 2006.
Wand, Y., and Weber, R. “Research Commentary: Information Systems and Conceptual Modelling - A Research Agenda,” *Information Systems Research* (13:4), December 2002, pp. 363-376.
Wixom, B.H., and Watson, H.J. “An Empirical Investigation of the Factors Affecting Data Warehousing Success,” *MIS Quarterly* (25:1), March 2001, pp. 17-41, 2001.