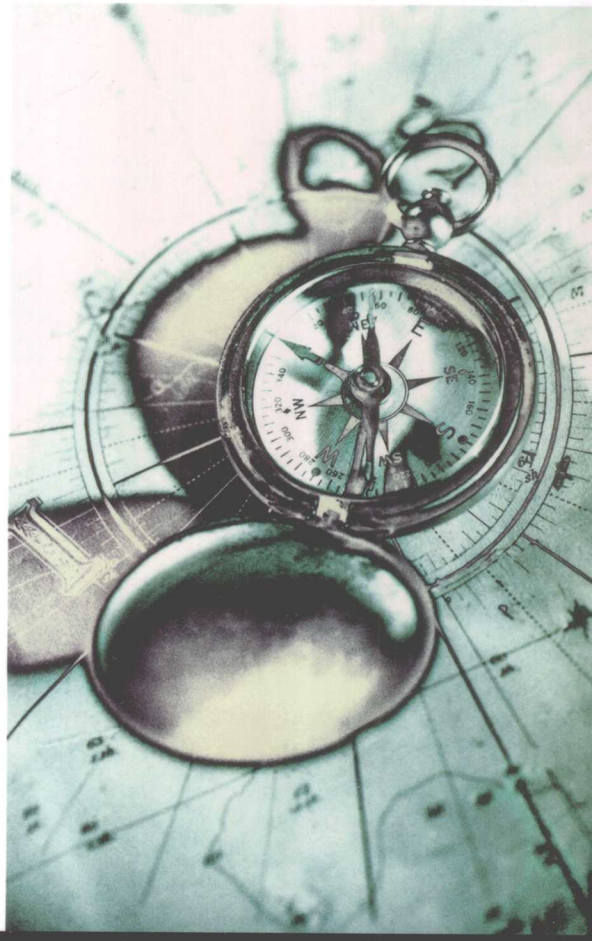
"An excellent introduction from someone who knows his subject and knows how to teach it"
Graeme Simsion, University of Melbourne

Data Modeling Made Simple

A Practical Guide for Business & Information Technology Professionals



Steve Hoberman
Author of the *Data Modeler's Workbench*

Acknowledgements

I'd like to thank Jeani, Dave, and Wayne for their help with this text.

Jeani Wells is a top-notch data modeler. Her experience and ability to keep things simple had a positive influence on this text. David Wells' belief in the need for a strong relationship between business and IT helped shape this book. After conferring with him, I strove even harder to make this book relevant not only to IT experts but also to business professionals. Wayne Eckerson can make words come alive. I thank him for his insights on the text, and especially the title.

My current manager at Mars, Larry Priester, is a model manager. He knows how to motivate a team without losing sight of the individual. I thank him for his strong belief in the development of his employees.

Thanks to Craig Brelsford from Precision Editorial Services for copy-editing Data Modeling Made Simple. Thanks to Eric Dome for lending his brain on the cover. Thanks to Roger Hedden for the Cobra photo shoot.

As was the case with my first book, my wife, Jenn, worked extra hard to free up my "free" time so I could focus on writing. While I was busy writing, Jenn was busy changing diapers, mowing the grass, shopping, cleaning, and so on. I'll have to start writing again quick before that next dirty diaper comes along.

I thank my daughters, Sadie and Jamie, for reminding me to keep things simple.

I thank my dad for instilling in me a strong work ethic and a desire to understand how things tick, and my mom for setting the example of a teacher who loves to share knowledge.

Foreword

A data model is the brain of an application, a conceptual framework that represents the business as accurately as possible. It defines the players, actions, and rules that govern the ways in which business processes work, representing them in a standard syntax that both humans and applications can interpret. In essence, a data model turns business concepts into computer code so that applications and systems can process information on our behalf. Without data models, we wouldn't be able to automate many of the processes that drive our organizations.

Given the pivotal role played by data models, it should be no surprise that they often determine whether an application is effective. A poorly designed data model can wreak havoc on even the most elegantly designed application. Poor performance, inaccurate query results, inflexible rules, and inconsistent metadata are just some of the results of a poor data model. A poor data model can hamstring an application.

On the flip side, a good data model serves as a lingua franca between business and information-technology professionals. It provides a shared understanding of the business that aligns business and technical professionals at the outset of a project. Conceptual and logical data models capture the ways in which technical professionals think a business process works. Business professionals can examine those assumptions and offer corrections and refinements before code is created.

I can't think of anyone better suited to explaining how data models work in plain, simple English than Steve Hoberman. Many skilled data modelers revel in the arcane art they practice and may as well be orbiting Pluto as working with business professionals. Not so Steve, who has demonstrated a mastery of making data modeling fun and easy in the courses he teaches for The Data Warehousing Institute. Although an accomplished data modeler himself—as is evident in his other book, *The Data Modeler's Workbench*—Steve is even more skillful in connecting with his audience. His enthusiasm and energy to communicate data modeling techniques is beyond compare. Steve is one of our most beloved and effective faculty members.

Meeting an important need. I'm extremely glad that Steve has decided to write *Data Modeling Made Simple*, because the need for this kind of book is huge. Given the importance of data models to the success of applications, it is surprising that so many business people (and more than a few technical people) lack understanding about them. This book will go a long way toward raising awareness of the importance and role of data models to organizations among business and technical professionals.

Specifically, business professionals who are sponsoring an application or have been assigned to the project team will find this book a useful primer on the topic. Technical professionals who are new to application design and development will find the book a quick

and easy way to learn the fundamentals of data modeling. And professors who want to help their students grasp data modeling concepts, terminology, and success criteria will want to add this book to their required reading list.

Wayne W. Eckerson
Director of Research and Services
The Data Warehousing Institute

Introduction

The writer does the most who gives his reader the most knowledge and takes from him the least time.

Charles Caleb Colton

The first book I wrote on data modeling, *The Data Modeler's Workbench,* was more than 500 pages long. I questioned how many readers read the entire book to extract all the bits of knowledge they needed to efficiently produce high-quality models. I then started thinking about those books that are most valuable to us—those books that make us want to read every chapter. Instruction manuals fit into this category. Whether for a car, digital recorder, or computer, reading the manual is almost essential to understanding how something works.

Data Modeling Made Simple was written with the same thought in mind. It is an instruction manual. In it, I provide information of immediate practical value to the reader, and I present that information as clearly and succinctly as I can. Theory and advanced topics I leave for the bigger books. Just as the instruction manual for an automobile avoids spending hundreds of pages on how an engine works, because such information is of limited use to the general reader, this book avoids spending pages on the theory, history, and mathematics behind modeling and focuses only on what is most useful.

Here are the main questions this text will answer:

- What is a data model?
- What is so special about data models?
- What are entities? What is the difference between an independent and a dependent entity?
- What are data elements? What are primary, foreign, and alternate keys?
- What are relationships and cardinality?
- What makes a definition great?
- What are the three types of subject area model?
- What is the logical data model? What are normalization and abstraction?
- What is the physical data model? What are denormalization, surrogate keys, indexing, partitioning, views, and dimensionality?
- What is the best approach to building the models? What are the top-down, bottom-up, and hybrid approaches to completing the modeling deliverables?
- How do I validate a data model? What is a scorecard?
- How do I keep my modeling skills sharp?
- What is the best data modeling tool?
- What is the future role of the data modeler?

There are more than 30 exercises throughout the text. When I teach data modeling, I rely on exercises to help reinforce concepts and keep the students alert. You will find exercises in this text for this same purpose. The more exercises you complete, the more you will get out of this book.

You will find the answers to the exercises on my Web site at www.stevehoberman.com.

Contents

ACKNOWLEDGEMENTS I

FOREWORD III

INTRODUCTION V

CHAPTER 1: WHAT IS A DATA MODEL? 7

CHAPTER 2: WHAT IS SO SPECIAL ABOUT DATA MODELS? 11

COMMUNICATION 11
FORMALIZATION 13
SCOPE 14
FOCUS 16

CHAPTER 3: WHAT ARE ENTITIES? 19

INDEPENDENT 20
DEPENDENT 21

CHAPTER 4: WHAT ARE DATA ELEMENTS? 27

DOMAIN 28
KEYS 29

CHAPTER 5: WHAT ARE RELATIONSHIPS? 33

CARDINALITY 35
RECURSION 37
IDENTIFYING AND NON-IDENTIFYING RELATIONSHIPS 40
RELATIONSHIP LABELS 41

CHAPTER 6: WHAT MAKES A DEFINITION GREAT? 43

CLARITY 44
COMPLETENESS 44
ACCURACY 45

CHAPTER 7: WHAT IS THE SUBJECT AREA MODEL? 47

BUSINESS SUBJECT AREA MODEL (BSAM) 48
APPLICATION SUBJECT AREA MODEL (ASAM) 51
COMPARISON SUBJECT AREA MODEL (CSAM) 53

CHAPTER 8: WHAT IS THE LOGICAL DATA MODEL? 55

NORMALIZATION 55
STARTING WITH CHAOS 57
FIRST NORMAL FORM (1NF) 60
SECOND NORMAL FORM (2NF) 68
THIRD NORMAL FORM (3NF) 71
ABSTRACTION 74

CHAPTER 9: WHAT IS THE PHYSICAL DATA MODEL? 81

DENORMALIZATION 81
SURROGATE KEYS 93
INDEXING 95
PARTITIONING 96
VIEWS 97
DIMENSIONALITY 99

CHAPTER 10: WHAT IS THE BEST APPROACH TO BUILD THE MODELS? 105

PURPOSE + AUDIENCE = DELIVERABLES 105
DELIVERABLES + RESOURCES + TIME = APPROACH 107

CHAPTER 11: HOW DO I VALIDATE A DATA MODEL? 111

1. HOW WELL DOES THE MODEL CAPTURE THE REQUIREMENTS? 113
2. HOW COMPLETE IS THE MODEL? 116
3. HOW STRUCTURALLY SOUND IS THE MODEL? 117
4. HOW WELL DOES THE MODEL LEVERAGE GENERIC STRUCTURES? 119
5. HOW WELL DOES THE MODEL FOLLOW NAMING STANDARDS? 120
6. HOW WELL HAS THE MODEL BEEN ARRANGED FOR READABILITY? 121
7. HOW GOOD ARE THE DEFINITIONS? 122
8. HOW WELL HAS THE REAL WORLD CONTEXT BEEN INCORPORATED? 123
9. HOW CONSISTENT IS THE MODEL WITH THE ENTERPRISE? 124
10. HOW WELL DOES THE METADATA MATCH THE DATA? 125

CHAPTER 12: TOP 3 MOST FREQUENTLY ASKED QUESTIONS 127

HOW DO I KEEP MY MODELING SKILLS SHARP? 127
WHAT IS THE BEST DATA MODELING TOOL? 128
WHAT IS THE FUTURE ROLE OF THE DATA MODELER? 128

SUGGESTED READING 131

BOOKS 131
WEB SITES 131

Praise for *Data Modeling Made Simple*

"Data Modeling Made Simple is a must read for all professionals new to data modeling as well as those that want to 'speak the language' and understand the concepts. Steve writes as though he is right there with you, walking you through the terminology, explaining the symbols, and telling you what you should consider doing before, during and after you have modeled your data."

Robert S. Seiner
President, KIK Consulting & Educational Services, LLC
Publisher, The Data Administration Newsletter, tdan.com

"Data Modeling Made Simple is an excellent training guide for anyone entering the data modeling field. Steve Hoberman takes the fundamental concepts of data modeling and presents them in an easy to understand and entertaining manner that I'm sure you will find valuable."

David Marco
President, *EWSolutions*

"How does one who is not a formally trained 'data modeler' understand the basics of data modeling? Steve Hoberman has created an informative, fun, easy to follow, and practical book sharing data modeling concepts which are essential for any professional involved in information technology. Mr. Hoberman clearly answers key questions behind the what, why and how of data modeling and reinforces the explanations with appropriate examples, analogies and exercises."

Len Silverston, Best-Selling Author of
The Data Model Resource Book, Volumes 1 and 2

Data Modeling Made Simple

A Practical Guide for Business & IT Professionals

Steve Hoberman

Technics Publications, LLC
New Jersey
www.technicspub.com

Published by:
Technics Publications, LLC
Post Office Box 161
Bradley Beach, NJ 07720 U.S.A.
Orders@technicspub.com
www.technicspub.com

Edited by Precision Editorial Services (www.precisionedit.com)

ISBN, printed. 0-9771400-0-8
First Printing 2005
Second Printing 2007
Third Printing 2008
Printed in the United States of America

Library of Congress Control Number: 2005906929

To Dad, who taught me the value of hard work—and the even higher value of family.

CHAPTER 1: What is a data model?

I gave the steering wheel a heavy tap with my hands as I realized that, once again, I was completely lost. It was about an hour before dawn, I was driving in France, and an important business meeting awaited me. I spotted a gas station up ahead that appeared to be open. I parked, went inside, and showed the attendant the address of my destination.

I don't speak French and the attendant didn't speak English. The attendant did, however, recognize the name of the company I needed to visit. Wanting to help and unable to communicate verbally, the attendant took out a pen and paper. He drew lines for streets, boxes for traffic lights, circles for roundabouts, and rectangles for his gas station and my destination.

With this custom-made map, which contained only the information that was relevant to me, I arrived at my address without making a single wrong turn. The map was a model of the actual roads I needed to travel.

A model is a representation of something in our environment. It makes use of standard symbols that allow one to grasp the content quickly. In the map he drew for me, the attendant used lines to symbolize streets and circles to symbolize roundabouts. His skillful use of those symbols helped me visualize the streets and roundabouts.

Models are all around us. An organizational chart is a model of a reporting structure in a company. A blueprint is a model for a building. A table of contents is a model of the contents in a book. A data model, as the name makes clear, is a model of data—data that can be as complex as or more complex than those roundabouts in France.

Data, as defined by the U.K. Ministry of Defense are "A representation of facts, concepts, or instructions in a formalized manner suitable for communication, interpretation, or processing by humans or by automatic means." We not only hear the word *data* hundreds of times a day, but we encounter an almost infinite amount of data through the activities and processes in which we participate. Sometimes the sheer quantity of data is overwhelming, making even a cursory level of understanding unreachable.

I have seen many definitions for the term *data model*. Some are extremely technical, using terms such as *predicate logic* and *set theory*. Other definitions essentially say that "a data model is a model of data." Still other definitions explain not what a data model is, but what it is used for (such as the role of a data model in software development). The error is understandable. A data model is, after all, an important deliverable for any application being built on a database.

My definition of a data model is the following:

A data model is a diagram that uses text and symbols to represent groupings of data so that the reader can understand the actual data better.

A spreadsheet groups data in columns. There is a column for last name, another for first name, and so on. A data model takes this idea a step further, showing not only the column headings but also the way in which the headings relate to each other. For example, a data model will show not only "first name" and "last name," but also how first name and last name relate to each other.

A claims data model for an insurance company, for example, most likely will display claim and policy information as well as the ways in which each type of information relates to the other. Claim number, policy effective date, claim amount, deductible, and hundreds of other possible groupings of data will be diagrammed, along with the ways in which they relate to one other.

Business cards contain a wealth of data about people and the companies for which they work. In this book, we will illustrate many concepts by using a business card as a model. By building a business card data model, we will see firsthand how much knowledge we gain of the contact-management area.

I once opened the drawer in my nightstand (a scary proposition, as it had not been cleaned since the mid-1980s) and grabbed a handful of business cards. I laid them out and picked four that I thought would be the most fun to model. I chose my current business card; a business card from an internet business that my wife and I tried to start years ago when dot-com was hot; a business card from a magician who performed at one of our parties; and a business card from one of our favorite dining establishments. I changed the names and contact information to protect the innocent and reproduced these in fig. 1.1.

Fig. 1.1 Four business cards from my nightstand