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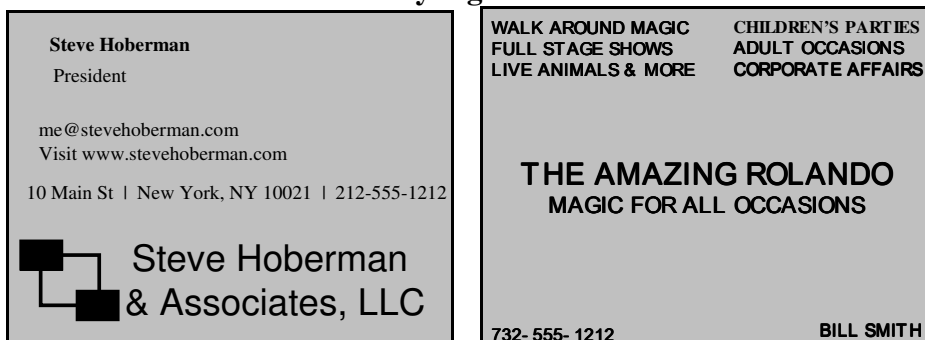
A spreadsheet groups data in columns. There is a column for last name, another for first name, and so on. A data model takes this idea a step further, showing not only the column headings but also the way in which the headings relate to each other. For example, a data model will show not only "first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

WEB MARKETING DIGITAL PHOTOGRAPHY **findsonline.com** **Internet Auction Experts** We'll get you the Steve and Jenn most for your item! (973) 555-1212	**Jon Smith** **John Doe** *Raritan River Club* *FINE FRESH SEAFOOD* **58 Church Avenue** Fax (908) 555-1212 **New Brunswick, NJ 08901** Banquet 554-1212 **(908) 333-1212**

Assuming our goal with this exercise is to understand the information on the business cards better, let's begin by listing some of the data:

- `Steve Hoberman & Associates, LLC`
- `Bill Smith`
- `Steve`
- `Jenn`
- `FINE FRESH SEAFOOD`

For the sake of brevity, let's stop here. Even though we are dealing only with four business cards, listing all the data would do little to aid our understanding. Now imagine that, instead of limiting ourselves to just these four cards, we looked through all the cards in my nightstand—or worse yet, the entire contact-management area. We would quickly become overloaded with data.

A data model groups data together to make them easier to understand. For example, we would examine the following set of data and realize that they fit within a group called "company name":

- `Steve Hoberman & Associates, LLC`
- `The Amazing Rolando`
- `findsonline.com`
- `Raritan River Club`

Taking this exercise a step further, we can organize all the data on the cards into the following groups:

- `person's name`
- `person's title`
- `company name`
- `e-mail address`
- `Web address`
- `mailing address`
- `phone number`

```
-  logo (the image on the card)
```
So, are we finished? Is this listing of groups a data model? Not yet. We are still missing a key ingredient: the ways in which these groups relate to one other. The model will also be able to show these relationships, including showing the relationships from this conversation about the e-mail address:

Each e-mail address can be assigned to at most one person or at most one company. The e-mail address me@stevehoberman.com, for example, can be assigned only to me and no one else. On the flip side, each person or company can have zero, one, or many e-mail addresses. I actually have several e-mail addresses, and so do many people I know.

A data model for our business card example is shown in fig. 1.2. Mindful that you may not know what all the symbols in the model mean yet, I have given the figure the caption "a preview of things to come." By the time you are about halfway through this book, you should be able to understand the symbols fully.

Fig. 1.2 A preview of things to come