EXAMPLE 1-1. THE PLANT DATABASE

Figure 1-1 shows a small portion of a database table recording information about plants. Along with the botanical and common names of each plant, the developer decides it would be convenient to keep information on the uses for each plant. This is to help prospective buyers decide whether a plant is appropriate for their requirements.

plantID	genus	species	common_name	use1	use2	use3
1	Dodonaea	viscosa	Akeake	shelter	hedging	soil stability
2	Cedrus	atlantica	Atlas cedar	shelter		
3	Alnus	glutinosa	Black alder	soil stability	shelter	firewood
4	Eucalyptus	nichollii	Black peppermint gum	shelter	coppicing	bird food
5	Juglans	nigra	Black walnut	timber		
6	Acacia	mearnsii	Black wattle	firewood	shelter	soil stability

***ure 1-1.** The plant database*

plantID	genus	species	common_name
1	Dodonaea	viscosa	Akeake
2	Cedrus	atlantica	Atlas cedar
3	Alnus	glutinosa	Black alder
4	Eucalyptus	nichollii	Black peppermint gum
5	Juglans	nigra	Black walnut
6	Acacia	mearnsii	Black wattle

Table Plants

plant	use
1	soil stability
1	hedging
1	shelter
2	shelter
3	firewood
3	soil stability
3	shelter

Table Uses

1-2. *An improved database design to represent Plants and Uses*

EXAMPLE 1-2. RESEARCH INTERESTS

An employee of a university's liaison team often receives calls asking to speak to a specialist in a particular topic. The liaison team decides to set up a small spreadsheet to maintain data about each staff member's main research interests. Originally, the intention is to record just one main area for each staff member, but academics, being what they are, cannot be so constrained. The problem of an indeterminate number of interests is solved by adding a few extra columns in order to accommodate all the interests each staff member supplies. Part of the spreadsheet is shown in Figure 1-3.

personID	...	...	...	interest 1	interest 2
152				Computing education	
275				Computer visualisation	Simulation
282				Scientific visualization	Statistics
292				Visualisation of data	Computing education
890				Databases	Scientific visualisation

***ure 1-3.** Research interests in a spreadsheet*

We are able to see at a glance the research interests of a particular person, but as was the case in Example 1-1, it is awkward to do the reverse and find who is interested in a particular topic. However, we have an additional problem here. Many of the research interests look similar but they are described differently. How easy will it be to find a researcher who is able to "visualize data"?

Figure 1-4 shows a portion of the data as it was recorded in a spreadsheet.

	A	B	C	D	E	F
1	farm	field	date	sample	springtail	fungus_beetle
268	1	ADhc	Aug-11	1	2	0
269	2	ADhc	Aug-11	2	2	0
270	1	ADhc	Aug-11	3	7	0
271	1	ADhc	Aug-11	4	3	2
272	1	ADhc	Aug-11	5	3	0
273	1	ADhc	Aug-11	6	3	9
274	1	ADhc	Aug-11	7	2	1
275	1	ADhc	Aug-11	8	6	1
276	1	ADhc	Aug-11	9	2	1
277	1	ADhc	Aug-11	10	5	3
278	1	ADhc	Aug-11	11	0	0
279	1	ADhe	Aug-11	1	0	6
280	1	ADhe	Aug-11	2	1	1
281	1	ADhe	Aug-11	3	5	2

ure 1-4. *Insect data in a spreadsheet*

The information about each farm was recorded (quite correctly) elsewhere, thus avoiding that data being repeated. However, there are still problems. The fact that field ADhc is on farm 1 is recorded every visit, and it does not take long to find the first data entry error in row 269. (The coding used for the fields raises other issues that we will not address just now.)

about fields and visits into separate tables not only reduces problems with repeated information, but allows more data (soil types for fields, weather conditions for visits) to be easily added. The Counts table still suffers the same problems as the tables in Examples 1-1 and 1-2, but that can be addressed. We will return to this example in Chapter 4.

field	farm	soil
Adhc	1	
Adhe	1	
Mvhe	2	
MVhc	2	

Table Fields

visitID	field	date	conditions
113	Adhc	Aug-06	Fine
114	Adhe	Aug-06	Fine
115	Adhc	Sep-06	Rain
116	Adhe	Sep-06	Overcast

Table Visits

visitID	sample	springtail	fungus_beetle
113	1	2	0
113	2	2	0
113	3	7	0
113	4	3	0
113	5	0	2
113	6	3	1

Table Counts

Figure 1-5. *An improved database design for the insect problem*

ID	Name	S001	S002	S103	S104	S202	S310	S331	GPA
982208	Jo Brown	A+	A	A	A+	A	B+	B+	8.6
986667	Helen Green	A	A	A+	A	A	B+	B+	8.5
987645	Peter Smith	A	B+	A-	A-	B+	A-	B	7.5

ure 1-6. *Report required for students' results*

A database table was designed to exactly match the report in Figure 1-6, with a field for each column. The first year the database worked a treat. The next year the problems started. Can you anticipate them?

Some students were permitted to replace one of the papers with one of their own choosing. The table was amended to include columns for option name and option mark. Then some subjects were replaced, but the old ones had to be retained for those students who had taken them in the past. The table became messier, but it could still cope with the data.

What the design couldn't handle was students who failed and then reenrolled in a subject. The complete academic record for a student needed to be recorded, and the design of the table made it impossible to record more than one mark if a student completed a subject several times. That problem wasn't noticed until the second year in operation (when the first students started failing). By then, a fair amount of effort had gone into development and data entry. The somewhat curious solution was to create a new table for each year, and then to apply some tortuous logic to extract a student's marks from the appropriate tables. When the original developer left for a new job, several years' worth of data were left in a state that no one else could comprehend. And that's how I got my first database job (and the database coped with changing requirements over several years).

TESTING YOUR UNDERSTANDING

Exercise 1-1

A school is planning some outdoor activities for its students. The staff wants to create a database of how parents can help. The secretary sets up the database table in Figure 1-7 to keep the information.

last_name	first_name	phone	contribution	contribution2
Smith	Jane	4623598	Food preparation	Driving
Green	Rob	8965431	Transport	
Henry	James	9576342	Camping Gear	Cooking
Wang	Li	9612345	Cooking	

***ure 1-7.** Initial database table for recording parent contributions*

What problems can you foresee in making good use of this information?

Suggest some better ways that this information could be stored.

Exercise 1-2

A small library keeps a roster of who will be at the desk each day. They have a database table as shown in Figure 1-8.

week_start	Mon	Tue	Wed	Thur	Fri
17/10/2011	Jane	Sue	George	Sue	Jane
24/10/2011	Jane	Sue	Linda	Sue	Lee
31/10/2011	Sue	Sue	Lee	George	George

ure 1-8. *An initial database table to record roster duties*

What problems can you foresee in making good use of this information?

Suggest some better ways that this information could be stored.