


Chapter 4

**Developing,
managing and using
customer-related
databases**

Chapter objectives

By the end of this chapter, you will understand:

1. the central role of customer-related databases to the successful delivery of CRM outcomes
2. the importance of high quality data to CRM performance
3. the issues that need to be considered in developing a customer-related database
4. what data integration contributes to CRM performance
5. the purpose of a data warehouse and data mart
6. how data access can be obtained by CRM users
7. the data protection and privacy issues that concern public policy makers.

Introduction

In this chapter we discuss the importance of developing an intimate knowledge and understanding of customers. This is essential to achieving CRM success. Strategic CRM, which focuses on winning and keeping profitable customers, relies on customer-related data to identify which customers to target, win and keep. Operational CRM, which focuses on the automation of customer-facing processes such as selling, marketing and customer service, needs customer-related data to be able to deliver excellent service, run successful marketing campaigns and track sales opportunities. Analytical CRM mines customer-related data for strategic or tactical purposes. Collaborative CRM involves the sharing of customer-related data with organizational partners, with a view to enhancing company, partner and customer value. Customer-related databases are the foundation for the execution of CRM strategy. Proficiency at acquiring, enhancing, storing, distributing and using customer-related data is critical to CRM performance.

What is a customer-related database?

You may have already noted that this chapter is not about customer databases. Rather, it is about customer-related databases. Why? Companies typically do not have a single customer database; instead, they have a number of customer-related databases. Large organizations, such as financial services companies, can have 20 or more customer systems, each with a separate database. These databases capture customer-related data from a number of different perspectives. Customer-related databases

might be maintained in a number of functional areas – sales, marketing, service, logistics and accounts – each serving different operational purposes. Respectively, these databases might record quite different customer-related data – opportunities, campaigns, enquiries, deliveries and billing. Customer-related data might also be maintained by different channel managers – company-owned retail stores, third-party retail outlets and online retail, for example. Similarly, different product managers might maintain their own customer-related data. Customer-related data can have a current, past or future perspective, focusing upon current opportunities, historic sales or potential opportunities. Customer-related data might be about individual customers, customer cohorts, customer segments, market segments or entire markets. They might also contain product information, competitor information, regulatory data or anything else pertinent to the development and maintenance of customer relationships.

Developing a customer-related database

Most databases share a common structure of files, records and fields (also called tables, rows and columns). Files (tables) hold information on a single topic such as customers, products, transactions or service requests. Each file (table) contains a number of records (rows). Each record (row) contains a number of elements of data. These elements are arranged in common sets of fields (columns) across the table. The modern customer-related database therefore resembles a spreadsheet. There are six major steps in building a customer-related database, as shown in Figure 4.1.¹

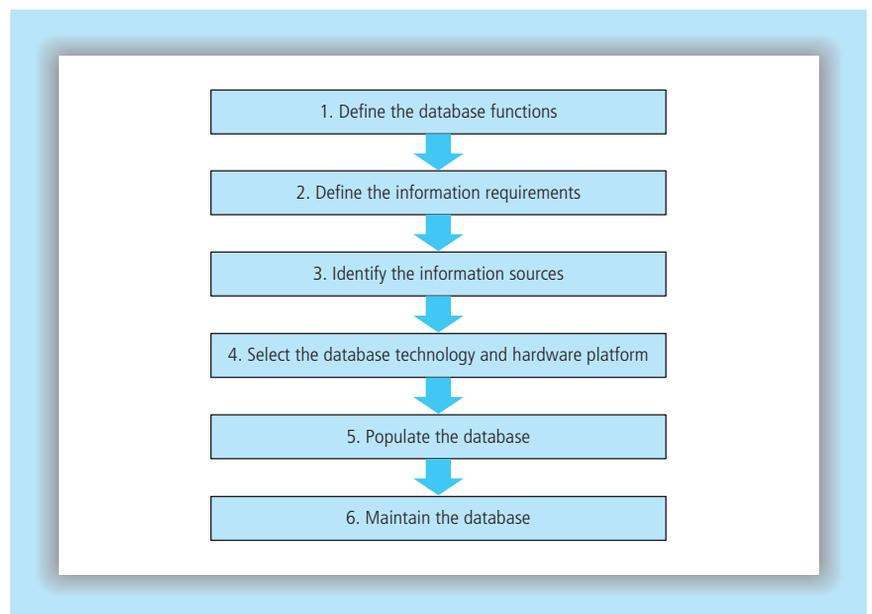


Figure 4.1
Building a customer-related database

Define the database functions

Databases support the four forms of CRM – strategic, operational, analytical and collaborative.

Strategic CRM needs data about markets, market offerings, customers, channels, competitors, performance and potential to be able to identify which customers to target for customer acquisition, retention and development, and what to offer them. Collaborative CRM implementations generally use the operational and analytical data as described below, so that partners in distribution channels can align their efforts to serve end-customers.

Customer-related data is necessary for both operational and analytical CRM purposes.

Operational CRM uses customer-related data to help in the everyday running of the business. For example:

- a telecoms customer service representative (CSR) needs to access a customer record when she receives a telephone query
- a hotel receptionist needs access to a guest's history so that she can reserve the preferred type of room – smoking or non-smoking, standard or de-luxe
- a salesperson needs to check a customer's payment history to find out whether the account has reached the maximum credit limit.

Analytical CRM uses customer-related data to support the marketing, sales and service decisions that aim to enhance the value created for and from customers. For example:

- the telecoms company might want to target a retention offer to customers who are signalling an intention to switch to a different supplier
- the hotel company might want to promote a weekend break to customers who have indicated their complete delight in previous customer satisfaction surveys
- a sales manager might want to compute his sales representatives' customer profitability, given the level of service that is being provided.

Customer-related data are typically organized into two subsets, reflecting these operational and analytical purposes. Operational data resides in an OLTP (online transaction processing) database, and analytical data resides in an OLAP (online analytical processing) database. The information in the OLAP database is normally a summarized, restructured, extract of the OLTP database, sufficient to perform the analytical tasks. The analytical database might also draw in data from a number of internal and external sources. OLTP data needs to be very accurate and up to date. When a customer calls a contact centre to enquire about an invoice, it is no use the CSR telling the customer what the average invoice is for a customer in her postcode. The customer wants personal, accurate, contemporary, information. OLAP databases can perform well with less current data.

Define the information requirements

The people best placed to answer the question ‘what information is needed?’ are those who interact or communicate with customers for sales, marketing and service purposes, and those who have to make strategic CRM decisions.

A direct marketer who is planning an e-mail campaign might want to know open and click-through rates, and click-to-open rates (CTOR) for previous e-campaigns, broken down by target market, offer and execution. She would also want to know e-mail addresses, e-mail preferences (html or plain text), and preferred salutation (first name? Mr? Ms?). Operational and analytical needs like these help define the contents of customer-related databases.

Senior managers reviewing your company’s strategic CRM decisions will require a completely different set of information. They may want to know the following. How is the market segmented? Who are our current customers? What do they buy? Who else do they buy from? What are our customers’ requirements, expectations and preferences across all components of the value proposition, including product, service, channel and communication?

With the advent of packaged CRM applications, much of the database design work has been done by the software vendors. The availability of industry-specific CRM applications, with their corresponding industry-specific data models, allows for a much closer fit with a company’s data needs. Where there is a good fit out of the box, the database design process for both operational and analytical CRM applications becomes one of implementing exceptions that have been overlooked by the generic industry model. Some CRM vendors have also built in the extract, transform and load processes to move information from OLTP to OLAP databases although it is highly likely that a client will need to modify and customize the standard processes.

Customer information fields

Most CRM software has predefined fields in different modules, whether for sales, marketing or service applications. For example, in a sales application, a number of fields (columns) of information about customers are common: contact data, contact history, transactional history, current pipeline, future opportunities, products and communication preferences.

Contact data

Who is the main contact (name) and who else (other names) is involved in buying decisions? What are their roles? Who are the decision-makers, buyers, influencers, initiators and gatekeepers? What are the customer’s invoice addresses, delivery addresses, phone numbers, fax numbers, e-mail addresses, street addresses and postal addresses?

Contact history

Who has communicated with the customer, when, about what, in which medium and with what outcome?

Transactional history

What has the customer bought and when? What has been offered to the customer, but not been purchased?

Current pipeline

What opportunities are currently in the sales pipeline? What is the value of each opportunity? What is the probability of closing? Is there a 10 per cent, 20 per cent ... 90 per cent chance of making a sale? Some CRM applications enable sales people to allocate red, amber or green signals to opportunities according to the probability of success.

Opportunities

Whereas 'transactional history' looks backwards, 'opportunity' looks forwards. This is where opportunities that have not yet been opened or discussed are recorded.

Products

What products does the customer have? When were these products purchased, and when are they due for renewal? Have there been any service issues related to these products in the past?

Communication preferences

What is the preferred medium of communication – mail, telephone, e-mail, face-to-face, etc.? If it is e-mail, is plain text or html preferred? What is the preferred salutation? And the preferred contact time and location? Customers may prefer you to contact them by phone for some communications (e.g. an urgent product recall), by mail for others (e.g. invoicing), by e-mail (e.g. for advice about special offers) and face-to-face for other reasons (e.g. news about new products). These preferences can change over time. When a customer's preferences are used during customer communications, it is evidence that the company is responsive to customer expectations. Many companies allow customers to opt in to, or out of, different forms of communication. Customers may prefer to adjust their own preferences. Amazon.com, for example, allows customers to opt to receive e-mail about six different types of content: terms and conditions of shopping at Amazon; new products; research surveys; magazine subscription renewal notices; information about and from Amazon's partners and special offers.

Identify the information sources

Information for customer-related databases can be sourced internally or externally. Prior to building the database it is necessary to audit the company to find out what data are available. Internal data are the foundation of most CRM programmes, though the amount of information available about customers depends on the degree of contact that the company has with the customer. Some companies sell through partners, agents and distributors and have little knowledge about the demand chain beyond their immediate contact.

100 Customer Relationship Management

Internal data can be found in various functional areas.

- Marketing might have data on market size, market segmentation, customer profiles, customer acquisition channels, marketing campaign records, product registrations and requests for product information.
- Sales might have records on customer purchasing history including recency, frequency and monetary value, buyers' names and contact details, account number, SIC code, important buying criteria, terms of trade such as discounts and payment period, potential customers (prospects), responses to proposals, competitor products and pricing, and customer requirements and preferences.
- Customer service might have records of service histories, service requirements, customer satisfaction levels, customer complaints, resolved and unresolved issues, enquiries, and loyalty programme membership and status.
- Finance may have data on credit ratings, accounts receivable and payment histories.
- Your webmaster may have click-stream data.

Enhancing the data

External data can be used to enhance the internal data and can be imported from a number of sources including market research companies and marketing database companies. The business intelligence company Claritas, for example, offers clients access to their Behaviourbank and Lifestyle Selector databases. These databases are populated with data obtained from many millions of returned questionnaires. Experian, another intelligence company, provides geodemographic data to its clients. External data can be classified into three groups:²

1. compiled list data
2. census data
3. modelled data.

Compiled list data

Compiled list data are individual level data assembled by list bureaux or list vendors. They build their lists from a variety of personal, household and business sources. They might use local or council tax records, questionnaire response data, warranty card registrations or businesses' published annual reports. Lists can be purchased outright or rented for a period of time and a defined number of uses. Once the list or its permitted use has expired, it must be removed from the database.

If you were a retailer thinking of diversifying from leisurewear into dancewear and had little relevant customer data of your own, you might be interested in buying or renting data from an external source. Data could have been compiled by the bureau or vendor from a variety of sources, such as:

- memberships of dance schools
- student enrolments on dance courses at school and college
- recent purchasers of dance equipment

- lifestyle questionnaire respondents who cite dance as an interest
- subscribers to dance magazines
- purchasers of tickets for dance and musical theatre.

Census data

Census data are obtained from government census records. In different parts of the world, different information is available. Some censuses are unreliable; others do not make much data available for non-governmental use.

In the USA, where the census is conducted every ten years, you cannot obtain census data at the household level, but you can at a more aggregated geodemographic level, such as zip code, census tract and block group. Census tracts are subdivisions of counties. Block groups are subdivisions of census tracts, the boundaries of which are generally streets. In the USA there are about 225 000 block groups, with an average of over 1000 persons per group. Census data available at geodemographic level includes:

- median income
- average household size
- average home value
- average monthly mortgage
- percentage ethnic breakdown
- marital status
- percentage college educated.

For the UK census there are 155 000 enumeration districts, each comprising about 150 households and ten postcodes. The enumeration district is the basis for much geodemographic data.

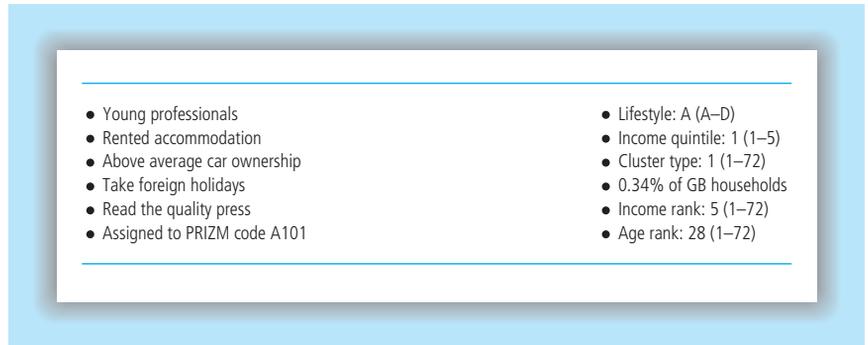
Individual-level data are better predictors of behaviour than aggregated geodemographic data. However, in the absence of individual-level data, census data may be the only option for enhancing your internal data. For example, a car reseller could use census data about median income and average household size to predict who might be prospects for a purchase promotion.

Modelled data

Modelled data are generated by third parties from data that they assemble from a variety of sources. You buy processed, rather than raw, data from these sources. Often they have performed clustering routines on the data. For example, Claritas has developed a customer classification scheme called PRIZM. In Great Britain, PRIZM describes the lifestyles of people living in a particular postcode. Every postcode is assigned to one of 72 different clusters on the basis of their responses to a variety of lifestyle and demographic questions. Eighty per cent of the data used in the clustering process is less than three years old.

Figure 4.2 provides the PRIZM profile of residents of one postcode in the London suburb of Twickenham. They are assigned to PRIZM code A101, which applies to about one-third of one per cent of households in the country. The figure profiles their occupational status, living accommodation, car ownership, vacation choices and media consumption.

Figure 4.2
PRIZM analysis of
TW9 1UU, England



If you want to use external data to enhance your internal data, you'll need to send a copy of the data that you want to enhance to the external data source. The source will match its files to yours using an algorithm that recognizes equivalence between the files (often using names and addresses). The source then attaches the relevant data to your files and returns them to you.

Secondary and primary data

Customer-related data are either secondary or primary. Secondary data are data that have already been collected, perhaps for a purpose that is very different from your CRM requirement. Primary data are that collected for the first time, either for CRM or other purposes.

Primary data collection through traditional means, such as surveys, can be very expensive. Companies have, therefore, had to find relatively low cost ways to generate primary customer data for CRM applications. Among the data-building schemes that have been used are the following:

- **Competition entries:** customers are invited to enter competitions of skill or lotteries. They surrender personal data on the entry forms.
- **Subscriptions:** customers may be invited to subscribe to a newsletter or magazine, again surrendering personal details
- **Registrations:** customers are invited to register their purchase. This may be so that they can be advised on product updates.
- **Loyalty programmes:** many companies run loyalty programmes. These enable companies to link purchasing behaviour to individual customers and segments. When joining a programme, customers complete application forms providing the company with personal, demographic and even lifestyle data.

Select the database technology and hardware platform

Customer-related data can be stored in a database in a number of different ways.

1. hierarchical
2. network
3. relational.

Hierarchical and network databases were the most common form between the 1960s and 1980s. The hierarchical database is the oldest form and not well suited to most CRM applications. You can imagine the hierarchical model as an organization chart or family tree, in which a child can have only one parent, but a parent can have many children. The only way to get access to the lower levels is to start at the top and work downwards. When data is stored in hierarchical format, you may end up working through several layers of higher-level data before getting to the data you need. Product databases are generally hierarchical. A major product category will be subdivided repeatedly until all forms of the product have their own record.

To extend the family tree metaphor, the network database allows children to have one, none or more than one parent. Before the network database had the chance to become popular, the relational database superseded it, eventually becoming an ANSI standard in 1971.³

Relational databases

Relational databases are now the standard architecture for CRM applications (see Figure 4.3). Relational databases store data in two dimensional tables comprised of rows and columns. Relational databases have one or more fields that provide a unique form of identification for each record. This is called the primary key. For sales databases, each customer is generally assigned a unique number which appears in the first column. Therefore, each row has a unique number. Companies also have other databases for marketing, service, inventory, payments and so on. The customer's unique identifying number enables linkages to be made between the various databases.

Let's imagine you are a customer of an online retailer. You buy a book and supply the retailer with your name, address, preferred delivery choice and credit-card details. A record is created for you on the 'Customer' database, with a unique identifying number. An 'Orders received' database records your purchase and preferred delivery choice. An 'Inventory' database records that there has been a reduction in the stock of the item you ordered. This may trigger a re-ordering process when inventory reaches a critical level. A 'Payment' database records your payment by credit-card. There will be one-to-many linkages between your customer record and these other databases. With the advent of enterprise suites from vendors such as Oracle and SAP, all of these databases may reside in the one system and be preintegrated. The choice of hardware platform is influenced by several conditions:

1. The size of the databases. Even standard desktop PCs are capable of storing huge amounts of customer data. However, they are not designed for this data to be shared easily between several users.

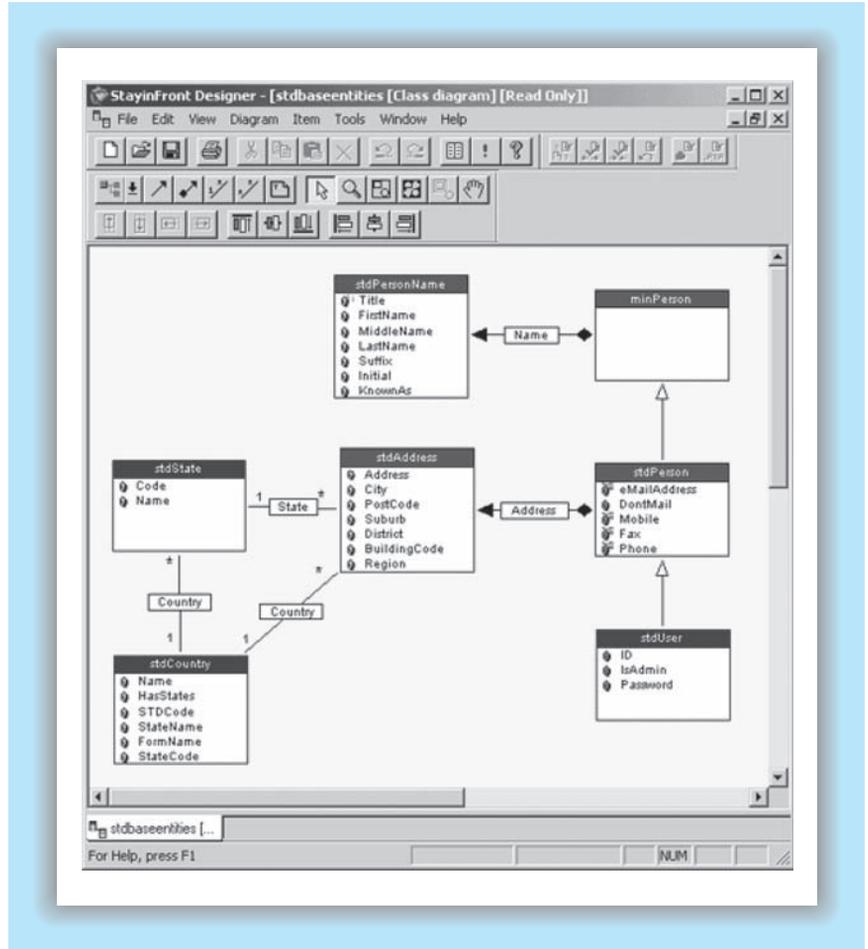


Figure 4.3
Relational database model⁴

- Existing technology. Most companies will already have technology that lends itself to database applications.
- The number and location of users. Many CRM applications are quite simple, but in an increasingly global marketplace the hardware may need very careful specification and periodic review. For example, the hardware might need to enable a geographically dispersed, multilingual, user group to access data for both analytical and operational purposes.

Relational database management system (RDBMS)

A relational database management system can be defined as follows:

An RDBMS is a software programme that allows users to create, update and administer a relational database.

There are a number of relational database management systems available from technology firms that are well suited to CRM applications. Leading RDBMS products are Oracle, DB2 from IBM, and Microsoft's SQL server. Most RDBMS products use SQL to access, update and query the database.

The selection of the CRM database can be done in parallel with the next step in this process, selection of CRM applications. Modern database applications come together with their own database schema, which predetermines the tables and columns in the database structure. Each CRM vendor then supports a specified list of database technologies, for example, Oracle or SQL server.

Indeed, it is possible to buy an entire platform, consisting of integrated hardware, operating system (OS), database and CRM applications. Leading platforms include UNIX, Microsoft and IBM. The UNIX platform offers a number of hardware/OS/database options, such as Hewlett-Packard hardware, Digital UNIX operating system and Oracle database. The IBM platform employs AS/400 hardware, OS/400 operating system and DB2/400 database. Microsoft NT servers are becoming more popular for CRM applications, due to the ease with which they can be scaled and expanded.

Populate the database

Having decided what information is needed and the database and hardware requirements, the next task is to obtain the data and enter it onto the database. CRM applications need data that are appropriately accurate. We use the 'appropriately' because the level of accuracy depends upon the function of the database. As noted earlier, operational CRM applications generally need more accurate and contemporary data than analytical applications.

You may have personally experienced the results of poor quality data. Perhaps you have received a mailed invitation to become a donor to a charity, to which you already donate direct from your salary. This could have happened when a prospecting list that has been bought by the charity was not been checked against current donor lists. Perhaps you have been addressed as Mrs although you prefer Ms. This is caused because the company has either not obtained or not acted or checked your communication preferences.

One of the biggest issues with customer data is not so much incorrect data as missing data. Many organizations find it difficult to obtain even basic customer data, such as e-mail addresses and preferences. The main steps in ensuring that the database is populated with appropriately accurate data are as follows:

1. source the data
2. verify the data
3. validate the data
4. de-duplicate the data
5. merge and purge data from two or more sources.

Sourcing: organizations must develop explicit processes to obtain information from customers, such as on initial sign-up or when concluding a service call. Organizations cannot rely on customer goodwill; data must be collected whenever interaction occurs.

Verification: this task is conducted to ensure that the data has been entered exactly as found in the original source. This can be a very labour-intensive process since it generally involves keying the data in twice with the computer programmed to flag mismatches. An alternative is to check visually that the data entered match the data at the primary source.

Validation: this is concerned with checking the accuracy of the data that are entered. There are a number of common inaccuracies, many associated with name and address fields: misspelt names, incorrect titles, inappropriate salutations. A number of processes can improve data accuracy.

- range validation: does an entry lie outside the possible range for a field?
- missing values: the computer can check for values that are missing in any column.
- check against external sources: you could check postcodes against an authoritative external listing from the mail authorities.

De-duplication: also known as de-duping. Customers become aware that their details appear more than once on a database when they receive identical communications from a company. This might occur when external data is not cross-checked against internal data, when two or more internal lists are used for the mailing or when customers have more than one address on a database. There may be sound cost reasons for this (de-duplication does cost money), but from the customer’s perspective it can look wasteful and unprofessional. De-duplication software is available to help in the process.

The de-duplication process needs to be alert to the possibility of two types of error:

1. Removing a record that should be retained. For example, if a property is divided into unnumbered apartments and you have transactions

DupID	FirstName	LastName	CompanyName	StreetAddress	City	State
483	Stephen	Ayres	US Veterans Affairs Med. Ctr.	1601 SW Archer Rd.	Gainesville	FL
483	Stephen	Aires	US Affairs Medical Center	1601 SW Archer Rd.	Gainesville	FI
573	Karl	Asha	ClearCommerce Corporation	11500 Metric Blvd.	Austen	TX
573	DanieleCarl	Asha	ClearCommerce Corporation	11500 Metric Boulevard	Austen	TX
870	Sherrell	Ballard	Southern Farm Bureau Life Ins.	Box 78	Jackson	MS
870	Cheryl	Ballard	Southern Farm Bur Lf Insur Co	P O Box 78	Jackson	MS
1359	Timothy	Bremere	General Cstly Co of Wlscosin	1 General Dr.	Sun Prairie	VM
1359	Tim	Bremer	General Casualty Companies	One general Drive	Sun Prairie	VM
2101	Mike	Condry	Celina Financial Corp.	One Insurance Sq.	Celina	OH
2101	Mike	Condry	Celina Insurance Group	One Insurance Square	Celina	OH
2800		Carmar	Indiana Lumbermans Insurance	Box 68600	Indianapolis	IN
2800	Patty	Carmar	Indiana Lumbermens Mutl Insur	P O Box 68600	Indianapolis	IN
3363	Robert	Delaney	Principal Mutual Life Insurance	711 High St	Des Moines	IA
3363	B	Delaney	THE PRINCIPAL FINANCIAL GROUP	711 High St	Des Moines	IA
3532	Danny	Teo	Bosley Medical	91 wilsheer	Beverly Hills	CA
3532	Danny	Teo	Bosley	9100 Wlshire	Beverly Hills	CA

Figure 4.4 Output from merge-purge operation⁵

with more than one resident, then it would be a mistake to assume duplication and delete records. Similarly, you may have more than one customer in a household, bearing the same family name or initials.

2. Retaining a record that should be removed. For example, you may have separate records for a customer under different titles such as Mr and Dr.

Merge and purge: also known as merge–purge (see Figure 4.4), this is a process that is performed when two or more databases are merged. This might happen when an external database is merged to an internal database, when two internal databases are merged (e.g. marketing and customer service databases), or when two external lists are bought and merged for a particular purpose such as a campaign. There can be significant costs savings for marketing campaigns when duplications are purged from the combined lists.

Maintain the database

Customer databases need to be updated to keep them useful. Consider the following statistics:

- 19% of managing directors change jobs in any year
- 8% of businesses relocate in any year
- in the UK, 5% of postcodes change in an average year
- in western economies about 1.2% of the population dies each year
- in the USA, over 40 million people change addresses each year.

It does not take long for databases to degrade. Companies can maintain data integrity in a number of ways.

1. Ensure that data from all new transactions, campaigns and communications is inserted into the database immediately. You will need to develop rules and ensure that they are applied.
2. Regularly de-duplicate databases.
3. Audit a subset of the files every year. Measure the amount of degradation. Identify the source of degradation: is it a particular data source or field?
4. Purge customers who have been inactive for a certain period of time. For frequently bought products, the dormant time period might be six months or less. For products with a longer repeat purchase cycle, the period will be longer. It is not always clear what a suitable dormancy period is. Some credit-card users, for example, may have different cards in different currencies. Inactivity for a year only indicates that the owner has not travelled to a country in the previous year. The owner may make several trips in the coming year.
5. Drip-feed the database. Every time there is a customer contact there is an opportunity to add new or verify existing data.
6. Get customers to update their own records. When Amazon customers buy online, they need to confirm or update invoice and delivery details.

7. Remove customers' records when they request this.
8. Insert decoy records. If the database is managed by an external agency, you might want to check the effectiveness of the agency's performance by inserting a few dummy records into the database. If the agency fails to spot the dummies, you may have a problem with their service standards.

Users with administrative rights can update records. Database updating and maintenance is also enabled by database query language. Common languages are SQL (Structured Query Language) and QBE (Query By Example). Maintenance queries available in SQL include UPDATE, INSERT and DELETE commands. You can use the commands to update your customer-related data. INSERT, for example, adds a new record to the database.

Desirable data attributes

Maintaining the database means that users will be more likely to have their need for accurate and relevant data met. Accuracy and relevance are two of six desirable data attributes that have been identified – data should be shareable, transportable, accurate, relevant, timely and secure.⁶ You can remember these desirable data attributes through the mnemonic STARTS.

Data need to be **shareable** because several users may require access to the same data at the same time. For example, profile information about customers who have bought annual travel insurance might need to be made available to customer service agents in several geographic locations simultaneously as they deal with customer enquiries in response to an advertising campaign.

Data need to be **transportable** from storage location to user. Data need to be made available wherever and whenever users require. The user might be a hot-desking customer service representative, a delivery driver en route to a pick-up, an independent mortgage consultant or a salesperson in front of a prospect. Today's international corporations with globally distributed customers, product portfolios across several categories and multiple routes to market face particularly challenging data transportation problems. Electronic customer databases are essential for today's businesses, together with enabling technologies, such as data synchronization, wireless communications and web browsers to make the data fully transportable.

Data **accuracy** is a troublesome issue. In an ideal world it would be wonderful to have 100 per cent accurate data. But data accuracy carries a high costs. Data are captured, entered, integrated and analysed at various moments. Any or all of these processes may be the source of inaccuracy. Keystroke mistakes can cause errors at the point of data entry. Inappropriate analytical processes can lead to ill-founded conclusions. In CRM, data inaccuracy can lead to undue waste in marketing campaigns, inappropriate prospecting by salespeople and

general suboptimal customer experience. It also erodes trust in the CRM system, thus reducing usage. This leads to further degrading of data quality. To counter this, usage volumes and data quality should be monitored. Data need to be entered at source rather than second hand; user buy-in needs to be managed; data quality processes such as de-duplication need to be introduced. Newsagency and book retailer WH Smith attribute high response rates of CRM-enabled direct marketing to the accuracy of their database. For example, an offer of Delia Smith's *How to Cook* book achieved an 8 per cent response rate, significantly more than was the norm before their data quality project was implemented.

Relevant data is pertinent for a given purpose. To check a customer's credit worthiness you need their transaction and payment histories, and their current employment and income status. To flag customers who are hot prospects for a cross-sell campaign, you need their propensity-to-buy scores. In designing a data management system to support a CRM strategy, relevance is a major issue. You need to know what decisions will be made and what information is needed to enable them to be made well.

Timely data is data that is available as and when needed. Data that is retrieved after a decision is made is not helpful. Equally, decision-makers do not want to be burdened with data before the need is felt. Bank tellers need to have propensity-to-buy information available to them at the time a customer is being served.

Data **security** is a hugely important issue for most companies. Data, particularly data about customers, is a major resource and a source of competitive advantage. It provides the foundation for delivery of better solutions to customers. Companies do need to protect their data against loss, sabotage and theft. Many companies regularly back-up their data. Security is enhanced through physical and electronic barriers such as firewalls. Managing data security in a partner environment is particularly challenging, as it is essential that competing partners do not see each other's sales leads and opportunity information, despite being signed into the same CRM system through the same portal.

Data integration

As noted earlier, in most companies there are several customer-related databases, maintained by different functions or channels. There might also be customer data in product or production databases, as well as call centres and websites, as suggested in Figure 4.5. External data from suppliers, business partners, franchisees and others may also need to be integrated.

Failure to integrate databases may lead to inefficiency, duplication and damaged customer relationships. Poor integration is indicated when you have bought an item online, only to be offered the same item at a later time through a different channel of the same company.

Customer data integration relies on standardization of data across databases. An indicator of the magnitude of the problem is that when Dun & Bradstreet was integrating data from several sources to create

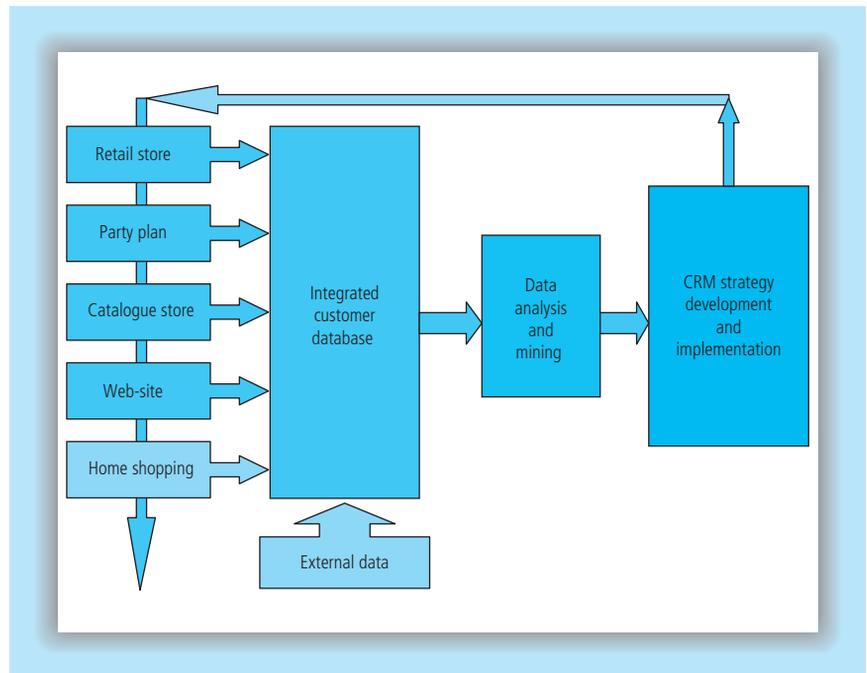


Figure 4.5
A single view of the customer

a marketing database it found 113 different entries for AT&T alone. These included ATT, A.T.T., AT and T and so on.

Companies often face the challenge of integrating data from several sources into a coherent single view of the customer. Sometimes this becomes a significant challenge in a CRM project, and a necessary hurdle to cross before implementing marketing, sales or service CRM applications.

The major on-premise CRM vendors, such as Oracle and SAP, offer solutions to this problem. SAP, for example, offers Master Data Management as part of its NetWeaver business integration platform. This enables companies to capture and consolidate data from different sources into a centralized database.

For companies with older mainframe (legacy) systems, another solution to the problem of database integration is to convert to newer systems with a centralized database that can accept real time inputs from a number of channels.⁷ However, where there is considerable investment in legacy systems and a huge number of records this may not be cost effective. Legacy systems are typically batch-processing systems. In other words, they do not accept real time data. Many technology firms have developed software and systems to allow companies to integrate databases held on different legacy systems. Sometimes middleware has to be written to integrate data from diverse sources. Middleware is a class of software that connects different parts of a system that would not otherwise be able to communicate to each other. Middleware acts as a broker of information between systems, receiving information from source systems, and passing it to destination systems in a format that can be understood. It is often referred to as a kind of 'glue' that holds a network together.

Case 4.1

Data integration at the American Heart Association

The American Heart Association (AHA) is a not-for-profit US health organization dedicated to reducing disability and death from heart attack, stroke and related cardiovascular disorders.

One of the AHA's major goals has been improving its relationships with stakeholders, including many thousands of volunteers conducting unpaid work for the organization, donors, businesses and the media. However, a challenge facing the AHA in achieving this goal was integrating the organization's data, which was previously located in over 150 separate databases, often geographically isolated and specific to certain departments within the organization. These provided a fragmented view of customers' profiles and history of activities.

AHA chose to implement a CRM software system across the organization to integrate all existing databases. Since implementation the AHA has found its staff is far more productive, it is able to respond to customers more quickly and provide more personalized service. Donations from customers have increased by over 20 per cent, using the system to contact potential donors compared to previous activities.

Data warehousing

As companies have grown larger they have become separated both geographically and culturally from the markets and customers they serve. Disney, an American corporation, has operations in Europe, Asia and Australasia, as well as in the USA. Benetton, the French fashion brand has operations across five continents. In retailing alone it operates over 7000 stores and concessions. Companies such as these generate a huge volume of data that needs to be converted into information that can be used for both operational and analytical purposes.

The data warehouse is a solution to that problem. Data warehouses are really no more than repositories of large amounts of operational, historical and other customer-related data. Data volume can reach terabyte levels, i.e. 2^{40} bytes of data. A warehouse is a repository for data imported from other databases. Attached to the front end of the warehouse is a set of analytical procedures for making sense out of the data. Retailers, home shopping companies and banks have been early adopters of data warehouses.

Watson describes a data warehouse as follows:⁸

- **subject-oriented:** the warehouse organizes data around the essential subjects of the business (customers and products) rather than around applications such as inventory management or order processing.
- **integrated:** it is consistent in the way that data from several sources is extracted and transformed. For example, coding conventions are standardized: M = male, F = female.

- **time-variant:** data are organized by various time-periods (e.g. months).
- **non-volatile:** the warehouse's database is not updated in real time. There is periodic bulk uploading of transactional and other data. This makes the data less subject to momentary change.

There are a number of steps and processes in building a warehouse. First, you must identify where the relevant data is stored. This can be a challenge. When the Commonwealth Bank opted to implement CRM in its retail banking business, it found that relevant customer data were resident on over 80 separate systems. Secondly, data must be extracted from those systems. It is possible that when these systems were developed they were not expected to align with other systems.

The data then needs to be transformed into a standardized, consistent and clean format. Data in different systems may have been stored in different forms, as Figure 4.6 indicates. Also, the cleanliness of data from different parts of the business may vary. The culture in sales may be very driven by quarterly performance targets. Getting sales representatives to maintain their customer files may be not straightforward. Much of their information may be in their heads. On the other hand, direct marketers may be very dedicated to keeping their data in good shape.

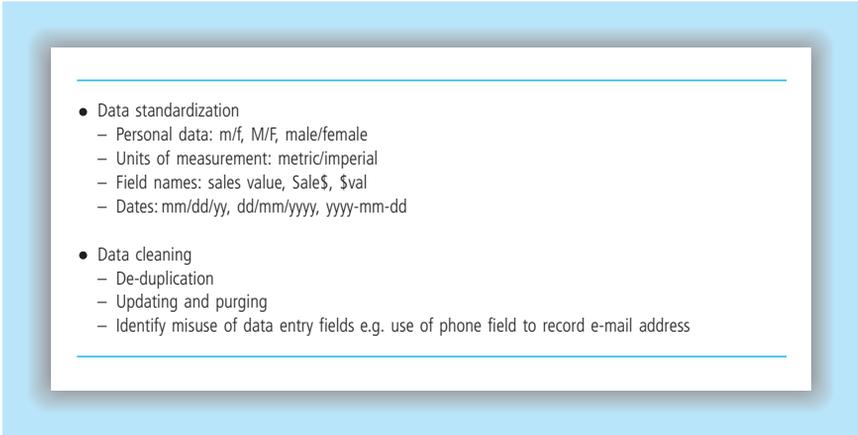
- 
- Data standardization
 - Personal data: m/f, M/F, male/female
 - Units of measurement: metric/imperial
 - Field names: sales value, Sale\$, \$val
 - Dates: mm/dd/yy, dd/mm/yyyy, yyyy-mm-dd
 - Data cleaning
 - De-duplication
 - Updating and purging
 - Identify misuse of data entry fields e.g. use of phone field to record e-mail address

Figure 4.6
Data transformation

After transformation, the data then needs to be uploaded into the warehouse. Archival data that have little relevance to today's operations may be set aside, or only uploaded if there is sufficient space. Recent operational and transactional data from the various functions, channels and touchpoints will most probably be prioritized for uploading.

Refreshing the data in the warehouse is important. This may be done on a daily or weekly basis depending upon the speed of change in the business and its environment.

Data marts

A data mart is a scaled down version, or subset, of the data warehouse, customized for use in a particular business function or department.

Marketing and sales may have their own data marts enabling them to conduct separate analyses and make strategic and tactical decisions. Some large data warehousing projects have taken years to implement and have yielded few measurable benefits. According to a Gartner Inc., 75 per cent of data warehouse implementations will fail to meet their delivery targets. The Meta Group says 20 per cent fail outright and 50 per cent fall short of expectations.⁹ Data mart project costs are lower because the volume of data stored are reduced, the number of users is capped, and the business focus is more precise. Technology requirements are less demanding.

Case 4.2

Data warehousing at Owens & Minor Inc.

Owens & Minor Inc., a Fortune 500 company headquartered in Richmond, VA, is the USA's leading distributor of national name brand medical/surgical supplies. The company's data warehouse project was first implemented in April 1997, starting with a single subject area – sales. Today, the data warehouse environment has grown to integrate over 20 different subject areas with over ten years of history. The size of the warehouse is just under 2 terabytes of total space. Internally there are over 900 users out of a total employee base of 3000, which is a very high percentage of business intelligence users. Externally Owens & Minor has four different extranet user groups that total around 600 users.

Source: The Data Warehousing Institute¹⁰

Data access and interrogation

CRM applications allow users to interact with customer-related databases for operational purposes. Sales representatives add data to customer records after a call is completed; CSRs in call centres log inbound calls on customer records; marketers update online brochures as product specifications change.

In addition, CRM users want to interrogate data for analytical purposes, or receive management reports. There are three main ways of doing this – standard reports, database queries, and data mining.¹¹

Standard reports

Standard reports are automatically generated periodically by the CRM system. Examples include monthly reports to sales management about sales representatives' activity and performance against quota, and daily reports of call centre activity. OLAP technologies allow users to drill down into the data on a screen rather than resorting to a flat, fixed-format, report. Starting with aggregated sales data for a region, a sales

manager can drill down into data about individual sales representatives and their customers, to reveal where causes of underperformance lie. Special reports can also be produced when ad hoc queries are made of a database, data warehouse or data mart. Most database management systems incorporate some reporting capability.

Database queries

A number of different types of query languages are available to CRM users when they want to raise a database query. Some are graphical – users can click and drag the data they want, and then drill down until they reach the level of granularity they require. Database managers may prefer to use SQL, which is now the standard query language for relational databases. SQL queries employing standard commands, such as SELECT, INSERT, DELETE, UPDATE, CREATE, DROP, can be used to access required data.

Data mining

In the CRM context, data mining can be defined as follows:

Data mining is the application of descriptive and predictive analytics to support the marketing, sales and service functions.¹²

Although data mining can be performed on operational databases, it is more commonly applied to the more stable datasets held in data marts or warehouses. Higher processing speeds, reduced storage costs and better software packages have made data mining more attractive and economical.

Data mining can provide answers to questions that are important for both strategic and operational CRM purposes. For example:

1. How can our market and customer base be segmented?
2. Which customers are most valuable?
3. Which customers offer most potential for the future?
4. What types of customers are buying our products? Or not buying?
5. Are there any patterns of purchasing behaviour in our customer base?
6. Should we charge the same price to all these segments?
7. What is the profile of customers who default on payment?
8. What are the costs of customer acquisition?
9. What sorts of customer should be targeted for acquisition?
10. What offers should be made to specific customer groups to increase their value?
11. Which customers should be targeted for customer retention efforts?
12. Which retention tactics work well?

Data mining helps CRM in a number of ways. It can find **associations** between data. For example, the data may reveal that customers who buy low fat desserts are also big buyers of herbal health and beauty aids, or

that consumers of wine enjoy live theatre productions. One analyst at Wal-Mart, the American retailer, noted a correlation between diaper sales and beer sales, which was particularly strong on Fridays. On investigating further he found that fathers were buying the diapers and picking up a six-pack at the same time. The company responded to this information by locating these items closer to each other. Sales of both rose strongly.¹³

Case 4.3

Data mining at Marks & Spencer

Data mining has proven to be a successful strategy for the UK retailer Marks & Spencer (M&S). The company generates large volumes of data from the ten million customers per week it serves in over 300 stores. The organization claims data mining lets it build one-to-one relationships with every customer, to the point that whenever individual customers come into a store the retailer knows exactly what products it should offer in order to build profitability.

Marks & Spencer believes two factors are important in data mining. First is the quality of the data. This is higher when the identity of customers is known, usually as a result of e-commerce tracking or loyalty programme membership. Second is to have clear business goals in mind before starting data mining. For example, M&S uses data mining to identify 'high margin', 'average margin' or 'low margin' customer groups. The company then profiles 'high margin' customers. This is used to guide customer retention activities with appropriate targeted advertising and promotions. This technique can also be used to profile 'average margin' or 'low margin' customers who have the potential to be developed into 'high margin' customers.

Sequential patterns often emerge from data mining. Data miners look for 'if ... then' rules in customer behaviour. For example, they might find a rule such as 'If a customer buys walking shoes in November, then there is a 40 per cent probability that they will buy rainwear within the next six months', or 'If a customer calls a contact centre to request information about interest rates, then there is a 50 per cent probability the customer will churn in the next three months'. Rules such as these enable CRM users to implement timely tactics. In the first instance, there is an opportunity for cross-selling. Secondly, there may be an opportunity to save the customer.

Data-mining also works by **classifying**. Customers can be classified into mutually exclusive groups. For example, you might be able to segment your existing customers into groups according to the value they produce for your company. You can then profile each group. When you identify a potential new customer you can judge which group the prospect most resembles. That will give you an idea of the prospect's potential value.

You could also classify customers into quintiles or deciles in terms of important transactional information such as the recency, frequency and monetary value of the purchases they have made. This is called RFM

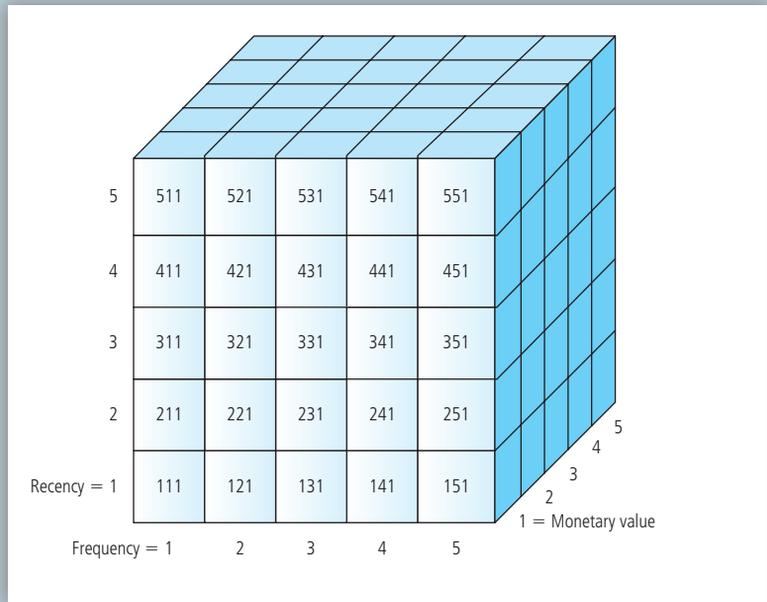


Figure 4.7
A recency–
frequency–
monetary value
matrix

analysis. Then you can experiment with different treatments, making different offers and communicating in different ways to selected cells of the RFM matrix (see Figure 4.7). You can expect to find that customers who have bought most recently, frequently or spend most with you are the most responsive in general terms.

Another approach in data mining is **clustering**. CRM practitioners attempt to cluster customers into groups. The general objective of clustering is to minimize the differences between members of a cluster while also maximizing the differences between clusters. Clustering techniques work by using a defined range of variables to perform the clustering procedure. You might, for example, use all available transaction data to generate customer segments. There are a number of techniques, such as cluster analysis, which find the hidden clusters.¹⁴ Once statistical clusters have been formed they need to be interpreted. Lifestyle market segments are outputs of cluster analysis on large sets of data. Cluster labels such as ‘Young working class families’ or ‘Wealthy suburbanites’ are often used to capture the essence of the cluster.

Finally, data mining can contribute to CRM by **making predictions**. CRM practitioners might use historic purchasing behaviour to predict future purchasing behaviour and customer lifetime value.

These five major approaches to data mining can be used in various sequences. For example, you could use clustering to create customer segments, then within segments use transactional data to predict future purchasing and customer lifetime value.

According to Gartner Inc., market leaders SAS and SPSS offer broad data mining solutions that meet most market needs.¹⁵ There are many

other vendors. Successful vendors of CRM analytics provide the following:

- packaged applications to support common CRM decisions such as cross-sell and customer churn prediction
- a user interface suitable for business users
- the capability to access data from various sources including data warehouses, data marts, call centres, e-commerce or web-tracking systems, as well as third party data sources
- robust data mining statistical tools such as cluster analysis, decision trees and neural networks that can provide reliable insights into different types and volumes of data
- reporting tools that make the results of analysis available to decision-makers such as campaign managers and call centre agents.

Privacy issues

Privacy and data protection are major concerns to legislators around the world. Customers are increasingly concerned about the amount of information commercial organizations have about them, and the uses to which that information is put. In fact, consumers are not aware of just how much information is available to companies. When you use the Internet, small programmes called cookies are downloaded onto your hard disk from the sites you visit. A very small number of websites obtain permission from their site visitors prior to the download; most do not.

There have been two major responses to the privacy concerns of customers. The first is self-regulation by companies and associations. For example, a number of companies publish their privacy policies and make a commercial virtue out of their transparency. Professional bodies in fields such as direct marketing, advertising and market research have adopted codes of practice that members must abide by.

The second response has been legislation. In 1980, the Organization for Economic Cooperation and Development (OECD) developed a set of principles that has served the foundation for personal data protection legislation around the world.¹⁶ These principles are voluntary guidelines that member nations can use when framing laws to protect individuals against abuses by data gatherers. The principles are as follows:

- **Purpose specification:** at the time of data collection, the consumer should be provided with a clear statement of the purposes for which the data is being collected.
- **Data collection processes:** data should be collected only by fair and lawful means.
- **Limited application:** data should be used only for valid business purposes.
- **Data quality:** personal data should be relevant for the purposes used and kept accurate, complete and up to date.
- **Use limitation:** personal data should not be disclosed, sold, made available or otherwise used for purposes than as specified at the time

of collection unless the consumer gives consent or as required by law. Consumer consent can be obtained either through an opt-in or opt-out process. Opt-in means that consumers agree that their data may be used for a particular purpose. Opt-out means that consumers prohibit use for that purpose.

- **Openness:** consumers should be able to receive information about developments, practices and policies with regard to their personal data. They should be able to find out what data has been collected and the uses to which it has been put. Consumers should have access to the data controller.
- **Access:** consumers should be able to access their data in readable form, to challenge the data and, if the challenge is successful, have the data erased, corrected or completed.
- **Data security:** personal data should be protected against risks such as loss, unauthorized access, destruction, use, modification or disclosure.
- **Accountability:** a data controller should be accountable for compliance with these measures.

Legislation has been enacted at a number of levels. In 1995, the Council of the European Union issued Directive 95/46/EC on the 'Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data'. This applies to all forms of data and information processing including e-commerce. It required all member states to upgrade their legislation to a common standard by 1998. Companies are now only allowed to process personal data where the individual has given consent or where, for legal or contractual reasons, processing is necessary. EU countries are not allowed to export personal data to countries where such exacting standards do not apply. Legislation guarantees certain rights to citizens of the EU:

- **notification:** individuals are to be advised without delay about what information is being collected, and the origins of that data, if not from the individual
- **explanation** of the logic behind the results of automated decisions based on customer data (e.g. why a credit application was rejected)
- **correction/deleting/blocking** of data that do not comply with legislation
- **objection:** individuals can object to the way in which their data are processed (opt-out). Where the objection is justified, the data controller must no longer process the information.

Data controllers are also required to comply with certain obligations, including:

- Only collect and process data for legitimate and explicit purposes.
- Only collect personal data when individual consent has been granted, or is required to enter into or fulfil a contract, or is required by law.
- Ensure the data is accurate and up to date.
- At the point of data collection, to advise the individual of the identity of the collector, the reason for data collection, the recipients of the

data, and the individual's rights in respect of data access, correction and deletion.

- Ensure that the data is kept secure and safe from unauthorized access and disclosure.

The USA has not adopted these legislative standards, but in order to enable US companies to do business with EU organizations, the US Commerce Department has devised a set of 'Safe Harbor' principles. US organizations in the Safe Harbor are assumed to adhere to seven principles regarding notice (as in notification, above), choice, onward transfer (disclosure to third parties), security, data integrity, access and enforcement (accountability). US companies obtain Safe Harbor refuge by voluntarily certifying that they adhere to these principles. This enables data transfers to be made to the USA. Two areas of difference between the EU Directive and these Safe Harbor principles are in access and enforcement. The Safe Harbor wording for access is weaker. The Safe Harbor principle states that 'individuals must have *reasonable* access to personal information about them that an organization holds, and to be able to correct or amend the information *where it is inaccurate*'. The enforcement principle is unclear about sanctions should a company breach the standard and it allows no possibility of enforcement by government agencies.

In the USA, there is a tendency to rely on self-regulation by individual or associated companies, rather than legislation at state or federal level. For example, the World-Wide Web Consortium (W3C) has developed a Platform for Privacy Preferences (P3P) standard for improving privacy protection in e-commerce. This comprises three major elements:

1. A personal profile: each Internet user creates a file consisting of personal data and privacy rules for use of that data. Personal data might include demographic, lifestyle, preference and click-stream data. Privacy rules are the rules that the user prescribes for use of the data, e.g. opt-in or opt-out rules, and disclosure to third parties. The profile is stored in encrypted form on the user's hard drive, can be updated at any time by the users and is administered by the user's web browser.
2. A profile of website privacy practices: each website discloses what information has been accessed from the user's personal profile and how it has been used.
3. Automated protocols for accessing and using the user's data: these allow either the user or the user's agent (perhaps the web browser) automatically to ensure that the personal profile and the privacy rules are being complied with. If compliance is assured, then users can enter websites and transact without problems.

This is now being complemented with a more rigorous approach to legislation. In Australia, privacy legislation has been enacted at state and federal levels.

Summary

In this chapter you've read about the development, management and use of customer-related databases. CRM cannot deliver its promised benefits without appropriate customer-related data. Customer-related data are used for strategic, operational, analytical and collaborative CRM purposes. Customer-related databases need to be constructed with a very clear idea of the applications for which the data are needed. These applications range across the full territory of CRM strategy development and implementation. Customer-related data can be used to answer strategic questions such as 'Which customers should we serve?' and tactical questions such as 'What is the best day to communicate with a given customer?'

We described a six-step approach to developing a high quality customer-related database, consisting of defining the database functions, establishing the information requirements, identifying the information sources, selecting the database technology and hardware platform, populating and maintaining the database. We saw how compiled list data, census data and modelled data can be imported to enhance the basic data available in company-maintained databases, most of which adopt the standard relational architecture. Data integration from disparate databases is often a barrier to the delivery of desired CRM outcomes. Attached to the front end of many databases are data mining systems that allow users to make sense of the data. We ended by looking at data warehouses, data marts and privacy issues.

References

1. Based on O'Connor, J. and Galvin, E. (2001) *Marketing in the digital age*, (2nd edn). Harlow, England: Financial Times/Prentice Hall.
2. Drozdenko, R.G. and Drake, P.D. (2002) *Optimal database marketing: strategy, development and data mining*. Thousand Oaks, CA: Sage.
3. ANSI is the American National Standards Institute.
4. Courtesy of StayinFront Inc., www.stayinfront.com Used with permission.
5. Courtesy of Intelligent Search Technology Ltd, www.intelligentsearch.com
6. Based on Watson, R.T. (1999) *Data management: databases and organisations*. New York: John Wiley.
7. Drozdenko, R.G. and Drake, P.D. (2002) *Optimal database marketing: strategy, development and data mining*. Thousand Oaks, CA: Sage.
8. Watson, R.T. (1999) *Data management: databases and organizations*. New York: John Wiley.
9. D'Addario, J. (2002) *The application revolution*. <http://www.tdwi.org/Publications/display.aspx?id=6460&t=y>. Accessed 11 September 2007.
10. The Data Warehousing Institute. <http://www.tdwi.org/display.aspx?ID=7145>. Accessed 11 September 2007.

11. Zikmund, W.G., McLeod, R. Jr. and Gilbert, F.W. (2003) *Customer relationship management: integrating marketing strategy and information technology*. Hoboken, NJ: John Wiley.
12. Gartner Inc. (2006) *Magic quadrant for data mining*, 1Q06. www.gartner.com
13. Dempsey, M. (1995) Customers compartmentalised. *Financial Times*, 1 March.
14. Saunders, J. (1994) Cluster analysis. In: G.J. Hooley and M.K. Hussey (eds). *Quantitative methods in marketing*. London: Dryden Press, pp. 13–28.
15. Gartner Inc. (2006) *Magic quadrant for data mining*, 1Q06. www.gartner.com
16. Swift, R.S. (2001) *Accelerating customer relationships using CRM and relationship technologies*. Upper Saddle River, NJ: Prentice Hall.