1

Business intelligence

The advent of low-cost data storage technologies and the wide availability of Internet connections have made it easier for individuals and organizations to access large amounts of data. Such data are often heterogeneous in origin, content and representation, as they include commercial, financial and administrative transactions, web navigation paths, emails, texts and hypertexts, and the results of clinical tests, to name just a few examples. Their accessibility opens up promising scenarios and opportunities, and raises an enticing question: is it possible to convert such data into information and knowledge that can then be used by decision makers to aid and improve the governance of enterprises and of public administration?

Business intelligence may be defined as a set of mathematical models and analysis methodologies that exploit the available data to generate information and knowledge useful for complex decision-making processes. This opening chapter will describe in general terms the problems entailed in business intelligence, highlighting the interconnections with other disciplines and identifying the primary components typical of a business intelligence environment.

1.1 Effective and timely decisions

In complex organizations, public or private, decisions are made on a continual basis. Such decisions may be more or less critical, have long- or short-term effects and involve people and roles at various hierarchical levels. The ability of these *knowledge workers* to make decisions, both as individuals and as a community, is one of the primary factors that influence the performance and competitive strength of a given organization.

Business Intelligence: Data Mining and Optimization for Decision Making *Carlo Vercellis*
 ISBN: 978-0-470-51138-1

Most knowledge workers reach their decisions primarily using easy and intuitive methodologies, which take into account specific elements such as experience, knowledge of the application domain and the available information. This approach leads to a stagnant decision-making style which is inappropriate for the unstable conditions determined by frequent and rapid changes in the economic environment. Indeed, decision-making processes within today's organizations are often too complex and dynamic to be effectively dealt with through an intuitive approach, and require instead a more rigorous attitude based on analytical methodologies and mathematical models. The importance and strategic value of analytics in determining competitive advantage for enterprises has been recently pointed out by several authors, as described in the references at the end of this chapter. Examples 1.1 and 1.2 illustrate two highly complex decision-making processes in rapidly changing conditions.

Example 1.1 – Retention in the mobile phone industry. The marketing manager of a mobile phone company realizes that a large number of customers are discontinuing their service, leaving her company in favor of some competing provider. As can be imagined, low customer loyalty, also known as customer *attrition* or *churn*, is a critical factor for many companies operating in service industries. Suppose that the marketing manager can rely on a budget adequate to pursue a customer retention campaign aimed at 2000 individuals out of a total customer base of 2 million people. Hence, the question naturally arises of how she should go about choosing those customers to be contacted so as to optimize the effectiveness of the campaign. In other words, how can the probability that each single customer will discontinue the service be estimated so as to target the best group of customers and thus reduce churning and maximize customer retention? By knowing these probabilities, the target group can be chosen as the 2000 people having the highest churn likelihood among the customers of high business value. Without the support of advanced mathematical models and data mining techniques, described in Chapter 5, it would be arduous to derive a reliable estimate of the churn probability and to determine the best recipients of a specific marketing campaign.

Example 1.2 – Logistics planning. The logistics manager of a manufacturing company wishes to develop a medium-term logistic-production plan. This is a decision-making process of high complexity which includes, among other choices, the allocation of the demand originating from different market areas to the production sites, the procurement of raw materials and purchased parts from suppliers, the production planning of the plants and the distribution of end products to market areas. In a typical manufacturing company this could well entail tens of facilities, hundreds of suppliers, and thousands of finished goods and components, over a time span of one year divided into weeks. The magnitude and complexity of the problem suggest that advanced optimization models are required to devise the best logistic plan. As we will see in Chapter 14, optimization models allow highly complex and large-scale problems to be tackled successfully within a business intelligence framework.

The main purpose of business intelligence systems is to provide knowledge workers with tools and methodologies that allow them to make *effective* and *timely* decisions.

Effective decisions. The application of rigorous analytical methods allows decision makers to rely on information and knowledge which are more dependable. As a result, they are able to make better decisions and devise action plans that allow their objectives to be reached in a more effective way. Indeed, turning to formal analytical methods forces decision makers to explicitly describe both the criteria for evaluating alternative choices and the mechanisms regulating the problem under investigation. Furthermore, the ensuing in-depth examination and thought lead to a deeper awareness and comprehension of the underlying logic of the decision-making process.

Timely decisions. Enterprises operate in economic environments characterized by growing levels of competition and high dynamism. As a consequence, the ability to rapidly react to the actions of competitors and to new market conditions is a critical factor in the success or even the survival of a company.

Figure 1.1 illustrates the major benefits that a given organization may draw from the adoption of a business intelligence system. When facing problems such as those described in Examples 1.1 and 1.2 above, decision makers ask themselves a series of questions and develop the corresponding analysis. Hence, they examine and compare several options, selecting among them the best decision, given the conditions at hand.

If decision makers can rely on a business intelligence system facilitating their activity, we can expect that the overall quality of the decision-making process will be greatly improved. With the help of mathematical models and algorithms, it is actually possible to analyze a larger number of alternative

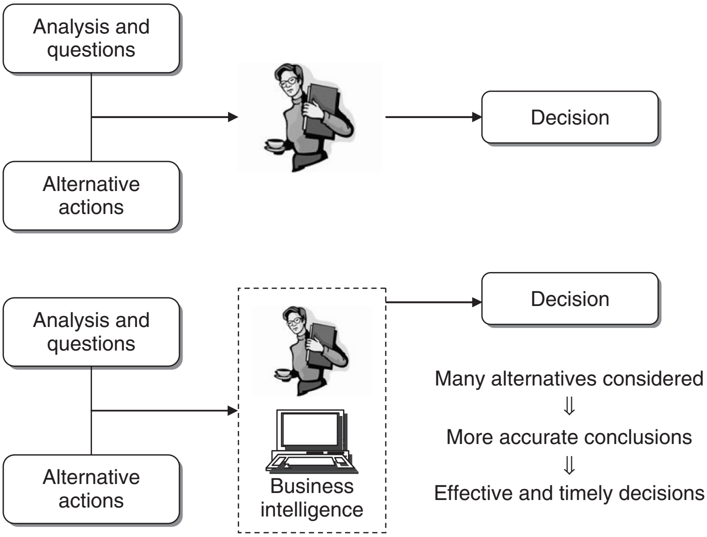

Figure 1.1 Benefits of a business intelligence system

actions, achieve more accurate conclusions and reach effective and timely decisions. We may therefore conclude that the major advantage deriving from the adoption of a business intelligence system is found in the increased *effectiveness* of the decision-making process.

1.2 Data, information and knowledge

As observed above, a vast amount of data has been accumulated within the information systems of public and private organizations. These data originate partly from internal transactions of an administrative, logistical and commercial nature and partly from external sources. However, even if they have been gathered and stored in a systematic and structured way, these data cannot be used directly for decision-making purposes. They need to be processed by means of appropriate extraction tools and analytical methods capable of transforming them into information and knowledge that can be subsequently used by decision makers.

The difference between *data*, *information* and *knowledge* can be better understood through the following remarks.

Data. Generally, data represent a structured codification of single primary entities, as well as of transactions involving two or more primary entities. For example, for a retailer data refer to primary entities such as customers, points of sale and items, while sales receipts represent the commercial transactions.

Information. Information is the outcome of extraction and processing activities carried out on data, and it appears meaningful for those who receive it in a specific domain. For example, to the sales manager of a retail company, the proportion of sales receipts in the amount of over €100 per week, or the number of customers holding a loyalty card who have reduced by more than 50% the monthly amount spent in the last three months, represent meaningful pieces of information that can be extracted from raw stored data.

Knowledge. Information is transformed into knowledge when it is used to make decisions and develop the corresponding actions. Therefore, we can think of knowledge as consisting of information put to work into a specific domain, enhanced by the experience and competence of decision makers in tackling and solving complex problems. For a retail company, a sales analysis may detect that a group of customers, living in an area where a competitor has recently opened a new point of sale, have reduced their usual amount of business. The knowledge extracted in this way will eventually lead to actions aimed at solving the problem detected, for example by introducing a new free home delivery service for the customers residing in that specific area. We wish to point out that knowledge can be extracted from data both in a *passive* way, through the analysis criteria suggested by the decision makers, or through the *active* application of mathematical models, in the form of inductive learning or optimization, as described in the following chapters.

Several public and private enterprises and organizations have developed in recent years formal and systematic mechanisms to gather, store and share their wealth of knowledge, which is now perceived as an invaluable intangible asset. The activity of providing support to knowledge workers through the integration of decision-making processes and enabling information technologies is usually referred to as *knowledge management*.

It is apparent that business intelligence and knowledge management share some degree of similarity in their objectives. The main purpose of both disciplines is to develop environments that can support knowledge workers in decision-making processes and complex problem-solving activities. To draw a boundary between the two approaches, we may observe that knowledge management methodologies primarily focus on the treatment of information that is usually unstructured, at times implicit, contained mostly in documents, conversations and past experience. Conversely, business intelligence systems are based on structured information, most often of a quantitative nature and usually organized in a database. However, this distinction is a somewhat fuzzy one: for example, the ability to analyze emails and web pages through text mining methods progressively induces business intelligence systems to deal with unstructured information.

1.3 The role of mathematical models

A business intelligence system provides decision makers with information and knowledge extracted from data, through the application of mathematical models and algorithms. In some instances, this activity may reduce to calculations of totals and percentages, graphically represented by simple histograms, whereas more elaborate analyses require the development of advanced optimization and learning models.

In general terms, the adoption of a business intelligence system tends to promote a scientific and rational approach to the management of enterprises and complex organizations. Even the use of a spreadsheet to estimate the effects on the budget of fluctuations in interest rates, despite its simplicity, forces decision makers to generate a mental representation of the financial flows process.

Classical scientific disciplines, such as physics, have always resorted to mathematical models for the abstract representation of real systems. Other disciplines, such as operations research, have instead exploited the application of scientific methods and mathematical models to the study of artificial systems, for example public and private organizations. Part II of this book will describe the main mathematical models used in business intelligence architectures and decision support systems, as well as the corresponding solution methods, while Part III will illustrate several related applications.

The rational approach typical of a business intelligence analysis can be summarized schematically in the following main characteristics.

- First, the objectives of the analysis are identified and the performance indicators that will be used to evaluate alternative options are defined.

- Mathematical models are then developed by exploiting the relationships among system control variables, parameters and evaluation metrics.

- Finally, *what-if* analyses are carried out to evaluate the effects on the performance determined by variations in the control variables and changes in the parameters.

Although their primary objective is to enhance the effectiveness of the decision-making process, the adoption of mathematical models also affords other advantages, which can be appreciated particularly in the long term. First, the development of an abstract model forces decision makers to focus on the main features of the analyzed domain, thus inducing a deeper understanding of the phenomenon under investigation. Furthermore, the knowledge about the domain acquired when building a mathematical model can be more easily transferred in the long run to other individuals within the same organization, thus allowing a sharper preservation of knowledge in comparison to empirical decision-making processes. Finally, a mathematical model developed for a specific decision-making task is so general and flexible that in most cases it can be applied to other ensuing situations to solve problems of similar type.

1.4 Business intelligence architectures

The architecture of a business intelligence system, depicted in Figure 1.2, includes three major components.

Data sources. In a first stage, it is necessary to gather and integrate the data stored in the various primary and secondary sources, which are heterogeneous in origin and type. The sources consist for the most part of data belonging to operational systems, but may also include unstructured documents, such as emails and data received from external providers. Generally speaking, a major effort is required to unify and integrate the different data sources, as shown in Chapter 3.

Data warehouses and data marts. Using extraction and transformation tools known as *extract, transform, load* (ETL), the data originating from the different sources are stored in databases intended to support business intelligence analyses. These databases are usually referred to as *data warehouses* and *data marts*, and they will be the subject of Chapter 3.

Business intelligence methodologies. Data are finally extracted and used to feed mathematical models and analysis methodologies intended to support decision makers. In a business intelligence system, several decision support applications may be implemented, most of which will be described in the following chapters:

- multidimensional cube analysis;
- exploratory data analysis;

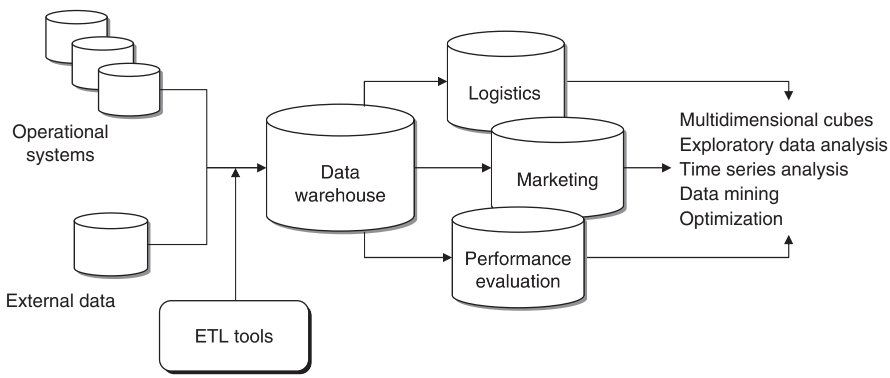

Figure 1.2 A typical business intelligence architecture

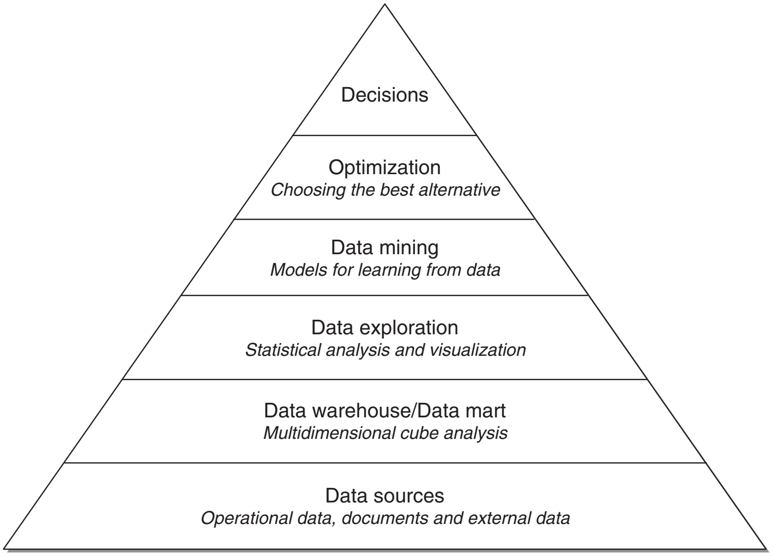

Figure 1.3 The main components of a business intelligence system

- time series analysis;
- inductive learning models for data mining;
- optimization models.

The pyramid in Figure 1.3 shows the building blocks of a business intelligence system. So far, we have seen the components of the first two levels when discussing Figure 1.2. We now turn to the description of the upper tiers.

Data exploration. At the third level of the pyramid we find the tools for performing a *passive* business intelligence analysis, which consist of query and reporting systems, as well as statistical methods. These are referred to as passive methodologies because decision makers are requested to generate prior hypotheses or define data extraction criteria, and then use the analysis tools to find answers and confirm their original insight. For instance, consider the sales manager of a company who notices that revenues in a given geographic area have dropped for a specific group of customers. Hence, she might want to bear out her hypothesis by using extraction and visualization tools, and then apply a statistical test to verify that her conclusions are adequately supported by data. Statistical techniques for exploratory data analysis will be described in Chapters 6 and 7.

Data mining. The fourth level includes *active* business intelligence methodologies, whose purpose is the extraction of information and knowledge from data.

These include mathematical models for pattern recognition, machine learning and data mining techniques, which will be dealt with in Part II of this book. Unlike the tools described at the previous level of the pyramid, the models of an active kind do not require decision makers to formulate any prior hypothesis to be later verified. Their purpose is instead to expand the decision makers' knowledge.

Optimization. By moving up one level in the pyramid we find optimization models that allow us to determine the best solution out of a set of alternative actions, which is usually fairly extensive and sometimes even infinite. Example 1.2 shows a typical field of application of optimization models. Other optimization models applied in marketing and logistics will be described in Chapters 13 and 14.

Decisions. Finally, the top of the pyramid corresponds to the choice and the actual adoption of a specific decision, and in some way represents the natural conclusion of the decision-making process. Even when business intelligence methodologies are available and successfully adopted, the choice of a decision pertains to the decision makers, who may also take advantage of informal and unstructured information available to adapt and modify the recommendations and the conclusions achieved through the use of mathematical models.

As we progress from the bottom to the top of the pyramid, business intelligence systems offer increasingly more advanced support tools of an active type. Even roles and competencies change. At the bottom, the required competencies are provided for the most part by the information systems specialists within the organization, usually referred to as *database administrators*. Analysts and experts in mathematical and statistical models are responsible for the intermediate phases. Finally, the activities of decision makers responsible for the application domain appear dominant at the top.

As described above, business intelligence systems address the needs of different types of complex organizations, including agencies of public administration and associations. However, if we restrict our attention to enterprises, business intelligence methodologies can be found mainly within three departments of a company, as depicted in Figure 1.4: marketing and sales; logistics and production; accounting and control. The applications of business intelligence described in Part III of this volume will be precisely devoted to these topics.

1.4.1 Cycle of a business intelligence analysis

Each business intelligence analysis follows its own path according to the application domain, the personal attitude of the decision makers and the available analytical methodologies. However, it is possible to identify an ideal cyclical

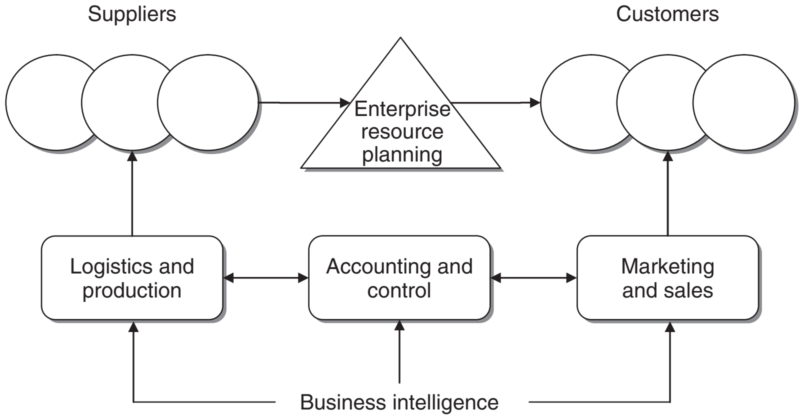

Figure 1.4 Departments of an enterprise concerned with business intelligence systems

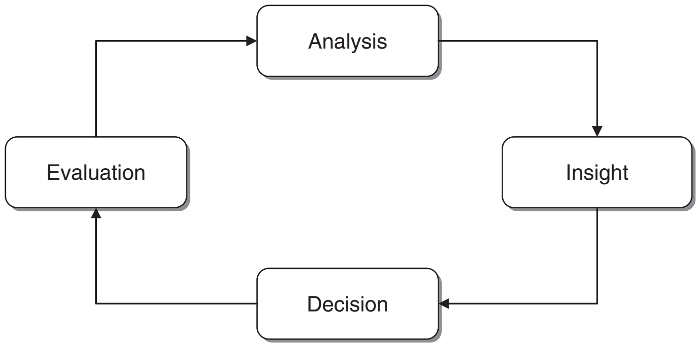

Figure 1.5 Cycle of a business intelligence analysis

path characterizing the evolution of a typical business intelligence analysis, as shown in Figure 1.5, even though differences still exist based upon the peculiarity of each specific context.

Analysis. During the analysis phase, it is necessary to recognize and accurately spell out the problem at hand. Decision makers must then create a mental representation of the phenomenon being analyzed, by identifying the critical factors that are perceived as the most relevant. The availability of business intelligence methodologies may help already in this stage, by permitting decision makers to rapidly develop various paths of investigation. For instance, the exploration of data cubes in a multidimensional analysis, according to different logical views as described in Chapter 3, allows decision makers to modify their hypotheses flexibly and rapidly, until they reach an interpretation scheme that they deem satisfactory. Thus, the first phase in the business intelligence cycle leads decision makers to ask several questions and to obtain quick responses in an interactive way.

Insight. The second phase allows decision makers to better and more deeply understand the problem at hand, often at a causal level. For instance, if the analysis carried out in the first phase shows that a large number of customers are discontinuing an insurance policy upon yearly expiration, in the second phase it will be necessary to identify the profile and characteristics shared by such customers. The information obtained through the analysis phase is then transformed into knowledge during the insight phase. On the one hand, the extraction of knowledge may occur due to the intuition of the decision makers and therefore be based on their experience and possibly on unstructured information available to them. On the other hand, inductive learning models may also prove very useful during this stage of analysis, particularly when applied to structured data.

Decision. During the third phase, knowledge obtained as a result of the insight phase is converted into decisions and subsequently into actions. The availability of business intelligence methodologies allows the analysis and insight phases to be executed more rapidly so that more effective and timely decisions can be made that better suit the strategic priorities of a given organization. This leads to an overall reduction in the execution time of the *analysis–decision–action–revision* cycle, and thus to a decision-making process of better quality.

Evaluation. Finally, the fourth phase of the business intelligence cycle involves performance measurement and evaluation. Extensive metrics should then be devised that are not exclusively limited to the financial aspects but also take into account the major performance indicators defined for the different company departments. Chapter 15 will describe powerful analytical methodologies for performance evaluation.

1.4.2 Enabling factors in business intelligence projects

Some factors are more critical than others to the success of a business intelligence project: *technologies*, *analytics* and *human resources*.

Technologies. Hardware and software technologies are significant enabling factors that have facilitated the development of business intelligence systems within enterprises and complex organizations. On the one hand, the computing capabilities of microprocessors have increased on average by 100% every 18 months during the last two decades, and prices have fallen. This trend has enabled the use of advanced algorithms which are required to employ inductive learning methods and optimization models, keeping the processing times within a reasonable range. Moreover, it permits the adoption of state-of-the-art graphical visualization techniques, featuring real-time animations. A further relevant enabling factor derives from the exponential increase in the capacity of mass storage devices, again at decreasing costs, enabling any organization to store terabytes of data for business intelligence systems. And network connectivity, in the form of *Extranets* or *Intranets*, has played a primary role in the diffusion within organizations of information and knowledge extracted from business intelligence systems. Finally, the easy integration of hardware and software purchased by different suppliers, or developed internally by an organization, is a further relevant factor affecting the diffusion of data analysis tools.

Analytics. As stated above, mathematical models and analytical methodologies play a key role in information enhancement and knowledge extraction from the data available inside most organizations. The mere visualization of the data according to timely and flexible logical views, as described in Chapter 3, plays a relevant role in facilitating the decision-making process, but still represents a passive form of support. Therefore, it is necessary to apply more advanced models of inductive learning and optimization in order to achieve active forms of support for the decision-making process.

Human resources. The human assets of an organization are built up by the competencies of those who operate within its boundaries, whether as individuals or collectively. The overall knowledge possessed and shared by these individuals constitutes the *organizational culture*. The ability of knowledge workers to acquire information and then translate it into practical actions is one of the major assets of any organization, and has a major impact on the quality of the decision-making process. If a given enterprise has implemented an advanced business intelligence system, there still remains much scope to emphasize the personal skills of its knowledge workers, who are required to perform the analyses and to interpret the results, to work out creative solutions and to devise effective action plans. All the available analytical tools being equal, a company employing human resources endowed with a greater mental agility and willing to accept changes in the decision-making style will be at an advantage over its competitors.

1.4.3 Development of a business intelligence system

The development of a business intelligence system can be assimilated to a project, with a specific final objective, expected development times and costs, and the usage and coordination of the resources needed to perform planned

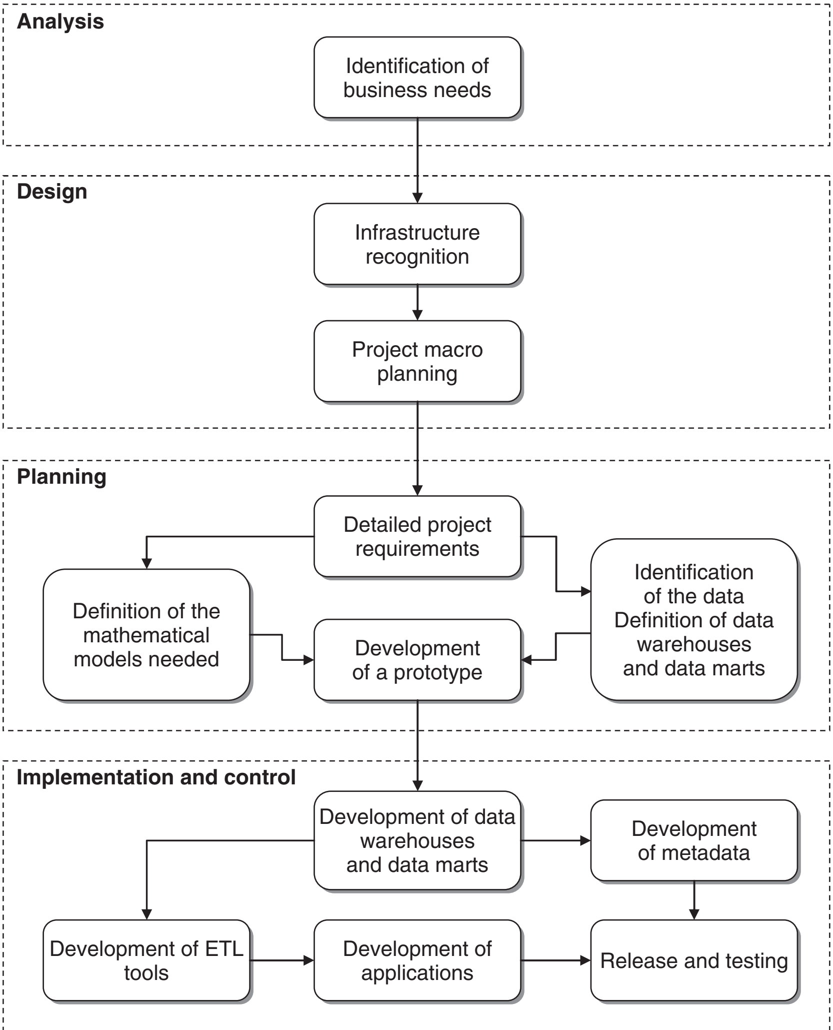

Figure 1.6 Phases in the development of a business intelligence system

activities. Figure 1.6 shows the typical development cycle of a business intelligence architecture. Obviously, the specific path followed by each organization might differ from that outlined in the figure. For instance, if the basic information structures, including the data warehouse and the data marts, are already in place, the corresponding phases indicated in Figure 1.6 will not be required.

Analysis. During the first phase, the needs of the organization relative to the development of a business intelligence system should be carefully identified. This preliminary phase is generally conducted through a series of interviews of knowledge workers performing different roles and activities within the organization. It is necessary to clearly describe the general objectives and priorities of the project, as well as to set out the costs and benefits deriving from the development of the business intelligence system.

Design. The second phase includes two sub-phases and is aimed at deriving a provisional plan of the overall architecture, taking into account any development in the near future and the evolution of the system in the mid term. First, it is necessary to make an assessment of the existing information infrastructures. Moreover, the main decision-making processes that are to be supported by the business intelligence system should be examined, in order to adequately determine the information requirements. Later on, using classical project management methodologies, the project plan will be laid down, identifying development phases, priorities, expected execution times and costs, together with the required roles and resources.

Planning. The planning stage includes a sub-phase where the functions of the business intelligence system are defined and described in greater detail. Subsequently, existing data as well as other data that might be retrieved externally are assessed. This allows the information structures of the business intelligence architecture, which consist of a central data warehouse and possibly some satellite data marts, to be designed. Simultaneously with the recognition of the available data, the mathematical models to be adopted should be defined, ensuring the availability of the data required to feed each model and verifying that the efficiency of the algorithms to be utilized will be adequate for the magnitude of the resulting problems. Finally, it is appropriate to create a system prototype, at low cost and with limited capabilities, in order to uncover beforehand any discrepancy between actual needs and project specifications.

Implementation and control. The last phase consists of five main sub-phases. First, the data warehouse and each specific data mart are developed. These represent the information infrastructures that will feed the business intelligence system. In order to explain the meaning of the data contained in the data warehouse and the transformations applied in advance to the primary data, a *metadata* archive should be created, as described in Chapter 3. Moreover, ETL procedures are set out to extract and transform the data existing in the primary sources, loading them into the data warehouse and the data marts. The next step is aimed at developing the core business intelligence applications that allow the planned analyses to be carried out. Finally, the system is released for test and usage.

Figure 1.7 provides an overview of the main methodologies that may be included in a business intelligence system, most of which will be described

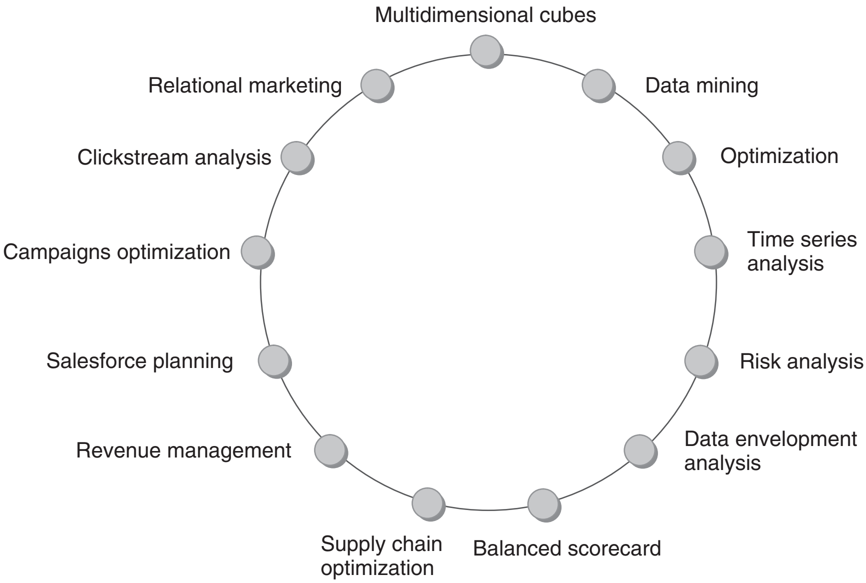

Figure 1.7 Portfolio of available methodologies in a business intelligence system

in the following chapters. Some of them have a methodological nature and can be used across different application domains, while others can only be applied to specific tasks.

1.5 Ethics and business intelligence

The adoption of business intelligence methodologies, data mining methods and decision support systems raises some ethical problems that should not be overlooked. Indeed, the progress toward the information and knowledge society opens up countless opportunities, but may also generate distortions and risks which should be prevented and avoided by using adequate control rules and mechanisms. Usage of data by public and private organizations that is improper and does not respect the individuals' right to privacy should not be tolerated. More generally, we must guard against the excessive growth of the political and economic power of enterprises allowing the transformation processes outlined above to exclusively and unilaterally benefit such enterprises themselves, at the expense of consumers, workers and inhabitants of the Earth ecosystem.

However, even failing specific regulations that would prevent the abuse of data gathering and invasive investigations, it is essential that business intelligence analysts and decision makers abide by the ethical principle of respect for the personal rights of the individuals. The risk of overstepping the boundary between correct and intrusive use of information is particularly high within the relational marketing and web mining fields, described in Chapter 13. For example, even if disguised under apparently inoffensive names such as 'data enrichment', private information on individuals and households does circulate, but that does not mean that it is ethical for decision makers and enterprises to use it.

Respect for the right to privacy is not the only ethical issue concerning the use of business intelligence systems. There has been much discussion in recent years of the social responsibilities of enterprises, leading to the introduction of the new concept of *stakeholders*. This term refers to anyone with any interest in the activities of a given enterprise, such as investors, employees, labor unions and civil society as a whole. There is a diversity of opinion on whether a company should pursue the short-term maximization of profits, acting exclusively in the interest of shareholders, or should instead adopt an approach that takes into account the social consequences of its decisions.

As this is not the right place to discuss a problem of such magnitude, we will confine ourselves to pointing out that analyses based on business intelligence systems are affected by this issue and therefore run the risk of being used to maximize profits even when different considerations should prevail related to the social consequences of the decisions made, according to a logic that we believe should be rejected. For example, is it right to develop an optimization model with the purpose of distributing costs on an international scale in order to circumvent the tax systems of certain countries? Is it legitimate to make a decision on the optimal position of the tank in a vehicle in order to minimize production costs, even if this may cause serious harm to the passengers in the event of a collision? As proven by these examples, analysts developing a mathematical model and those who make the decisions cannot remain neutral, but have the moral obligation to take an ethical stance.

1.6 Notes and readings

As observed above, business intelligence methodologies are interdisciplinary by nature and only recently has the scientific community begun to treat them as a separate subject. As a consequence, most publications in recent years have been released in the form of press or promotional reports, with a few exceptions. The following are some suggested readings: Moss and Atre (2003), offering a description of the guidelines to follow in the development of business intelligence systems; Simon and Shaffer (2001) on business intelligence applications for e-commerce; Kudyba and Hoptroff (2001) for a general introduction to the subject; and finally, Giovinazzo (2002) and Marshall *et al*. (2004) focus on business intelligence applications over the Internet. The strategic role of analytical methods, in the form of predictive and optimization mathematical models, has been pointed out recently by a number of authors, among them Davenport and Harris (2007) and Ayres (2007).

The integration of business intelligence architectures, decision support systems and knowledge management is examined by Bolloju *et al*. (2002), Nemati *et al*. (2002) and Malone *et al*. (2003). The volume by Rasmussen *et al*. (2002) describes the role of business intelligence methodologies for financial applications, which are not covered in this text. For considerations of a general nature on the ethical implications of corporate decisions, see Bakan (2005). Snapper (1998) examines the ethical aspects involved in the application of business intelligence methodologies in the medical sector.