Bedryfsingenieurswese
Industrial Engineering

University of Stellenbosch

Department of Industrial Engineering

Intelligent Risk Profiling for Project Management

Kennith Loftus

12708828

Assignment presented in partial fulfilment of the requirements for the degree of Masters of Industrial Engineering at the University of Stellenbosch.

Study leader: Mr CJ Fourie

19 November 2003

Declaration

I, the undersigned, hereby declare that the work contained in this assignment is my own original work and has not previously in its entirety or in part been submitted at any university for a degree. Ek, die ondergetekende verklaar hiermee dat die werk gedoen in hierdie werkstuk my eie oorspronklike werk is wat nog nie voorheen gedeeltelik of volledig by enige universiteit vir 'n graad aangebied is nie.

Signature: Date:

Synopsis

Whenever projects fail, analysis of the causes has shown that risks were present from day one. Often individuals at some level in the project team have knowledge of these risks and they could have been identified and appropriate remedial action taken. Risk, whether identified or not, generally results in some increase in financial exposure on behalf of the organisation, but, if managed well, offers a potential that could lead to increased profits.

There has been a tremendous explosion regarding the amount of data that organisations generate, collect and store. Managers are beginning to recognize the value of this asset and are increasingly relying on intelligent systems to access, analyse, summarise and interpret information from large and multiple data sources. These systems help them to make critical decisions at a faster rate or with a greater degree of confidence. Data mining is a promising new technology that helps bring intelligence into these systems.

The purpose of this thesis is to present a methodology that integrates a data mining technique with a decision support system in order to form an intelligent decision support system. The implementation of such an intelligent decision support system will enable project and project risk managers to improve the management of and reduce risk within a project.

This thesis consists of two sections. The first section describes the processes and characteristics of project management, project risk management, data mining and decision support systems. The aim is to provide the reader with a background about these four management methodologies. The second section describes the methodology of how the processes of project and project risk management can benefit from the integration of a data mining technique and a decision support system.

An application that uses the case-based reasoning approach as a data mining technique to intelligently profile a project according to its risks is demonstrated.

Opsomming

Wanneer projekte misluk, toon 'n analise van die oorsake dat risiko's vanuit die staanspoor daar teenwoordig was. Individuele persone op verskillende vlakke in die projekspan is dikwels daarvan bewus. Hierdie risiko's kon geïdentifiseer gewees het en regstellende stappe kon geneem gewees het. Risiko, hetsy geïdentifiseer of nie, loop gewoonlik uit op 'n sekere mate van toename in finansiële blootstelling namens die organisasie, maar wanneer dit goed bestuur word, bied dit 'n potensiaal vir verhoogde wins.

Daar is 'n geweldige vermeerdering in die hoeveelheid data wat organisasies genereer, versamel en berg. Bestuurders begin alreeds die onskatbare waarde van hierdie bate besef en steun toenemend op intelligensiestelsels vir toegang, analise, opsomming en interpretasie van inligting van omvangryke en veelsoortige databronne. Hierdie sisteme stel hulle in staat om kritieke besluite vinniger of met 'n groter mate van vertroue te neem. Dataontginning is 'n belowende nuwe tegnologie wat daartoe bydra dat intelligensie in hierdie sisteme ingebring word.

Die doel van hierdie tesis is om 'n metodologie wat 'n dataontginningstegniek met 'n besluitnemingsondersteuningsisteem integreer sodat 'n intelligente besluitnemings-ondersteuningsisteem gevorm kan word. Die implementering van so 'n intelligensie besluitnemingsondersteuningsisteem sal projekbestuurders en projekrisikobestuurders in staat stel om die bestuur van 'n projek te verbeter en die risiko binne die projek te verminder.

Hierdie tesis word in twee dele aangebied. Die eerste deel beskryf die prosesse en karakteristieke van projekbestuur, projekrisikobestuur, dataontginning en besluitondersteuningsisteme. Sodoende word aan die leser agtergrondinligting van hierdie vier bestuursmetodologieë verskaf. Die tweede deel beskryf die metodologie en hoe die prosesse van projekbestuur en projekrisikobestuur voordeel kan trek uit die integrasie van 'n dataontginningstegniek en 'n besluitondersteuningsisteem.

'n Toepassing is ontwikkel wat die gevallebasis beredeneringsbenadering as 'n dataontginingstegniek gebruik om 'n projek op 'n intelligente wyse volgens sy risiko's uit te beeld.

Acknowledgements

I would like to thank the following people for their contribution in making this study possible:

- Dr C.J. Fourie, for his guidance and support throughout the course of this study.
- Mr J. Bekker for his invaluable assistance with programming.
- The National Research Foundation (NRF), for their financial assistance during this study.
- The staff of the Department of Industrial Engineering, for their support.
- To my parents, Malcolm and Lidia Loftus, for their persistent encouragement until the completion of this thesis.
- To the Lord, for providing me with the determination to complete this study.

Table of Contents

List of figures

List of tables

Glossary

AI	Artificial intelligence
CBR	Case-based reasoning
DBMS	Database management system
DM	Data mining
DP	Data processing
DSS	Decision support system
EERD	Extended entity relationship diagram
EIS	Executive information system
IRR	Internal rate of return
IS	Information system
MBMS	Model base management system
MIS	Management information system
MRP	Materials requirements planning
NPV	Net present value
PM	Project management
PMBOK	Project management body of knowledge
PMI	Project Management Institute (United States of America)
PMIS	Project management information systems
PRM	Project risk management
RAMP	Risk analysis and management for projects
RDMS	Relational database management system
RM	Risk management
SCERT	Synergistic contingency evaluation and review technique

1. Project management

Project management offers a structured approach to managing projects. The purpose of this chapter is to present an overview of project management and the processes involved within project management. Various definitions for a project and project management are given and associated characteristics discussed. A summary of the nine knowledge areas of project management as defined by the Project Management Institute's 'Guide to the PMBOK' (1996) and which describe best practices and their application in project management is also presented. The chapter continues with an overview of the interface between the project and its parent organisation (i.e., how the project is organised as part of its host). A look at the facilitative role that computers play in project management concludes the chapter.

1.1 What is a project

Organisations perform work. Work generally involves either operations or projects, although the two may overlap. Operations and projects may share many characteristics; for example:

- Performed by people
- Constrained by limited resources
- Planned, executed and controlled

Operations and projects differ primarily in that operations are ongoing and repetitive while projects are temporary and unique. The Project Management Institute's (PMI) 'Guide to the PMBOK' (1996) defines a project as "... a temporary endeavour undertaken to create a unique product or service. Temporary means that every project has a definite end. Unique means that the product or service is different in some distinguishing way from all similar products or services."

Turner (1993) on the other hand defines a project as: "... an endeavour in which human, (or machine), material and financial resources are organised in a novel way, to undertake a unique scope of work, of given specification, within constraints of cost and time, so as to deliver beneficial change defined by quantitative and qualitative objectives."

According to Burke (1999), traditionally work in the construction industry and defence procurement were seen as projects, but in recent years most pro-active companies are structuring their work as projects (management-by-projects) and using project management techniques to ensure successful completion.

Projects are undertaken at all levels of the organisation and range in size, scope, cost and time. They may involve a single person or many thousands. Projects may involve a single unit of an organisation, cross units of the same organisation, or even cross-organisational boundaries as in joint ventures or partnering. Examples of projects include (Burke, 1999):

- The transition period during which change occurs.
- Designing and constructing a building, a house or a yacht.
- Designing and testing a new prototype (a car or a washing machine).
- The launch of a new product (advertising and marketing project).
- Implementing a new computer system (IT project, or upgrade).
- Designing and implementing a new organisational structure (human resource project).
- Planning and conducting an audit (quality management project).
- Improving productivity within a target period.
- Disaster recovery (limiting the damage of fires, floods or any type of accident).
- Olympics or Springboks' tour of New Zealand (a sports project).
- Rolling Stones' world tour (an entertainment project).
- Moving house or going on holiday (a domestic project).

In the broadest sense, a project is an autonomous, finite task to be accomplished, whether on large or small scale or on long or short term, is not particularly relevant. What is relevant is that the project be seen as a delimited deliverable. Other primary features of a project include:

- **Purpose**. A project is usually a one-off activity with a well-defined set of desired end results.

- **Start and finish**. Every project has a definite beginning and a definite end. The end is reached when the project's objectives have been achieved or when it becomes clear that the project objectives will not or cannot be met and the project is terminated.

- **Life cycle**. Projects have life cycles. From a beginning the level of effort builds up, peaks, begins to decline and finally terminates.

- **Budget**. Every project has a budget with an associated cash flow.

- **Interdependencies**. Projects often interact with other projects being carried out simultaneously by their parent organisation but projects always interact with the parent organisation's standard, ongoing operations.

- **Uniqueness**. Every project has some elements or activities that are essentially unique and non-repetitive.

- **Resources**. Every project uses resources which may be from different departments and need co-ordinating.

- **Conflict**. Projects compete with functional departments for resources and personnel.

- A **single point of responsibility** (i.e. the project manager).

- **Team roles**. Team roles and relationships that are subject to change and need to be developed, defined and established (team building).

1.2 Project management

PMI's 'Guide to the PMBOK' (1996) defines project management as: "…the application of knowledge, skills, tools and techniques to project activities in order to meet stakeholders needs and expectations from a project." In other words the project manager must do whatever is required to make the project happen.

The definition clearly identifies that the purpose of the project is to meet the stakeholders' needs and expectations. It is therefore important for the project manager to establish who the stakeholders are and analyse their needs and expectations to define at the outset, the project's scope of work and objectives.

Kerzner (2001) on the other hand describes project management as the involvement of project planning and project monitoring and includes such items as:

- Project planning
  - Definition of work requirements
  - Definition of quantity and quality of work
  - Definition of resources needed
- Project monitoring
  - Tracking progress
  - Comparing actual outcome to predicted outcome
  - Analysing impact
  - Making adjustments

He then defines successful project management as having achieved the project objectives:

- Within time
- Within cost
- At the desired performance/technology or specification level
- While utilizing the assigned resources effectively and efficiently
- Within acceptance by the customer/user
- With minimum or mutually agreed upon scope changes

- Without disturbing the main work flow of the organisation
- Without changing the corporate culture

Figure 1 is a pictorial representation of project management. The objective of the figure is to show that project management is designed to manage or control company resources on a given activity, within time, within cost and within performance. Integrated, time, cost and performance are the constraints on the project. If the project is to be accomplished for an outside customer, the project has a fourth constraint: good customer relations.

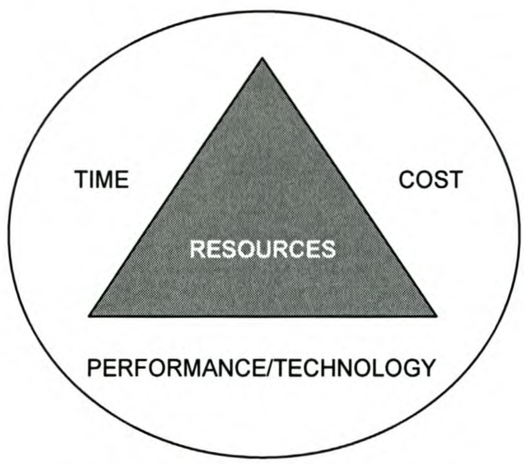

Figure 1: Overview of project management

Source: Kerzner, 2001

The term *project management* is sometimes used to describe an organisational approach to the management of ongoing operations. This approach, more properly called management-by-products, treats many aspects of ongoing operations as projects in order to apply project management to them. Burke (1999) argues that the management-by-products approach encourages:

- Organisational flexibility
- Decentralised management responsibility
- A holistic view of problems
- Goal orientated problem solution processes

1.3 Project management environment

The project environment directly influences the project and how it should be managed. Consider the following project environmental factors (Burke, 1999):

- Stakeholders (all interested parties)
- Client/sponsor's requirements
- The company's organisation structure
- Market requirements
- Competitors
- New technology
- Rules and regulations
- Economic cycle

Other environmental factors include:

- Physical (Green) factors
- Politics
- Community impacts
- Infrastructure
- Logistics

Burke argues that for project managers to be effective they must have a thorough understanding of the project environment which may well be changing, and so continually shift the goal posts. The project environment consists of among other factors the numerous stakeholders and players that have an input or are affected by the project. All must be managed as any one person could derail the project.

1.4 Project Management Body of Knowledge

Over the last fifty years a considerable body of knowledge has been built up around project management tools, skills and techniques. This database of information has grown into what is now called the project management body of knowledge (PMBOK).

The purpose of the body of knowledge is to identify and describe best practices that are applicable to most projects most of the time. Burke (1999) summarises the nine knowledge areas as defined by the Project Management Institute's 'A Guide to the PMBOK' (1996):

- **Project integration:** integrates the three main project management processes of planning, execution and control – where inputs from several knowledge areas are brought together.
- **Project scope management:** includes the process required to ensure that the project includes all the work required to complete the project successfully. It is primarily concerned with defining and controlling what is or is not included in the project, to meet the sponsors' and stakeholders' goals and objectives. It consists of authorisation, scope planning, scope definition, scope change management and scope verification.
- **Project time management:** includes the process required to ensure timely performance of the project. It consists of activity definition, activity sequencing, duration estimating, establishing the calendar, schedule development and time control.
- **Project cost management:** includes the process required to ensure that the project is completed within the approved budget. It consists of resource planning, cost estimating, cost budgeting, cash flow and cost control.
- **Project quality management:** includes the process required to ensure that the project will satisfy the needs for which it was undertaken. It consists of determining the required condition, quality planning, quality assurance and quality control.
- **Project human resource management:** includes the process required to make the most effective use of people involved with the project. It consists of organisation planning, staff acquisition and team development.
- **Project communications management:** includes the process required to ensure proper collection and dissemination of project information. It consists of

communication planning, information distribution, project meetings, progress reporting and administrative closure.

- **Project risk management:** includes the process concerned with identifying, analysing and responding to project risk. It consists of risk identification, risk quantification and impact, response development and risk control.
- **Project procurement management:** includes the process required to acquire goods and services from outside the performing project team or organisation. It consists of procurement planning, solicitation, source selection, contract administration and contract closeout.

The knowledge areas can be sub-divided into four primary elements that determine the deliverable objectives of the project:

- Scope
- Time
- Cost
- Quality

The other knowledge areas provide the means of achieving the deliverable objectives, namely:

- Integration
- Human resources
- Communication
- Risk
- Procurement and contract

1.5 The project life cycle

The PMI's 'Guide to the PMBOK' (1996) states that because projects are unique undertakings, they involve a certain degree of uncertainty. Organisations performing projects will usually divide each project into several project phases to provide better management control and appropriate links to the ongoing operations of the performing organisation. Collectively, the project phases are known as the project life cycle.

There is general agreement that projects pass through a four-phase life cycle under the following headings (Burke, 1999):

- **Conception and initiation phase:** the first phase starts the project by establishing a need or opportunity for the product, facility or service. The feasibility of proceeding with the project is investigated and on acceptance moves to the next phase.
- **Design and development phase:** the second phase uses the guidelines set by the feasibility study to design the product and develop detailed schedules and plans for making or implementing the product.
- **Implementation or construction phase:** the third phase implements the project as per baseline plan developed in the previous phase.
- **Commissioning and handover phase:** the fourth phase confirms the project has been implemented or built to the design and terminates the project.

The project life cycle phases are illustrated in Figure 2.

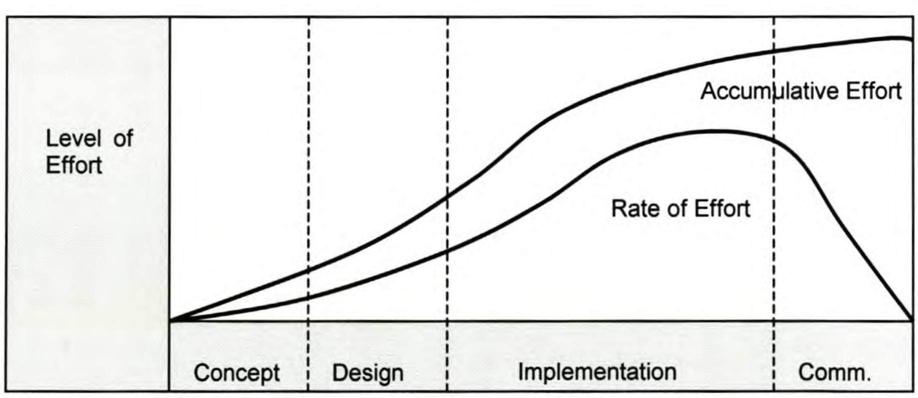

Figure 2: Project life cycle

1.5.1 Characteristics of the project phases

Each project phase is marked by the completion of one or more deliverables e.g. feasibility study, a detail design or a working prototype. The end of a project phase is generally marked by a review of both the deliverables and performance in order to determine if the project should continue into the next phase and detect and correct errors cost effectively. Each phase can be planned and controlled as a mini project and performed.

1.5.2 Characteristics of the project life cycle

The project life cycle serves to define the beginning and the end of a project. The project life cycle definition will determine whether the feasibility study is treated as the first project phase or as a separate stand alone-project. The PMI's 'Guide to the PMBOK' (1996) states that the phase sequence defined by most project life cycles generally involves some form of technology transfer or hand off, such as requirements to design, construction to operations or design to manufacturing. Deliverables from the proceeding phase are usually approved before work starts on the next phase.

According to the PMI's 'Guide to the PMBOK' (1996), project life cycles generally define:

- What technical work should be done in each phase.
- Who should be involved in each phase.

The 'Guide to the PMBOK' further identifies 3 common characteristics most project life cycle descriptions share:

- Cost and staffing levels are low at the start, higher towards the end and drop rapidly as the project draws to a conclusion. This pattern is illustrated in Figure 3.
- The probability of successfully completing the project is at its lowest, and hence risk and uncertainty are at its highest, at the start of a project. The probability of successful completion generally gets progressively higher as the project continues.
- The ability of the stakeholders to influence the final characteristics of the project product and the final cost of the project are at its highest at the start and become progressively lower as the project continues.

Although many project life cycles have similar phase names with similar work product requirements, few are identical. Most have four or five phases but some can have nine or more.

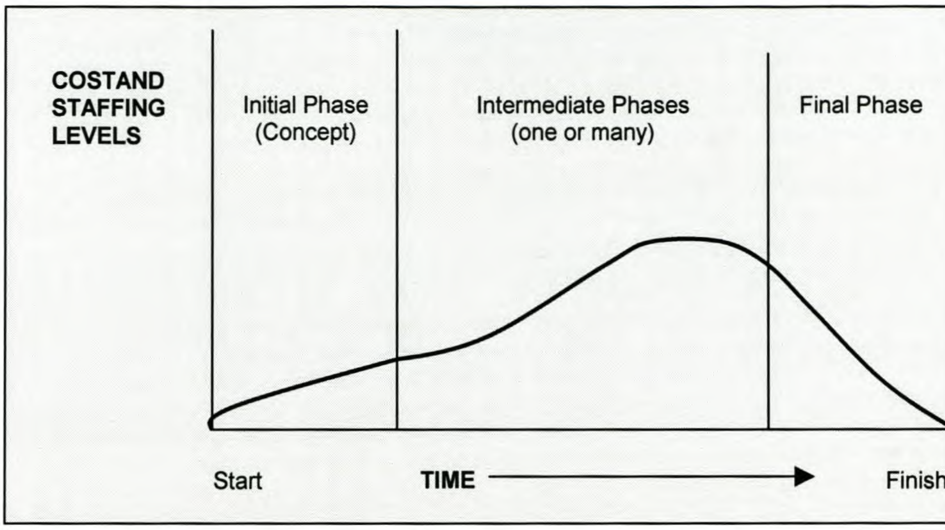

Figure 3: Cost and staffing levels vs time

Source: PMI's 'Guide to the PMBOK', 1996

1.6 *Project organisation*

Organisation structures are often quantified by their lines of responsibility and lines of authority. On projects this relationship can be presented as a continuum of organisation structures from functional to pure project with a range of matrix organisation structures in between. For simplicity the three main types of project organisation structure will be discussed:

- Functional
- Pure project
- Matrix

1.6.1 Functional organisation structure

This traditional organisation structure (Figure 4) is based on the sub-division of product lines or disciplines into separate departments together with a vertical hierarchy where each employee has a clear superior. Functional organisations are common in companies dominated by marketing or manufacturing departments (whenever there is a large amount of repetitive work). Staff are grouped by speciality, such as human resources, production, marketing, engineering and accounting at the top level with engineering further subdivided into electrical and mechanical. Functional organisations still have projects but the perceived scope of the project is limited to within the boundaries of the function. A project in a functional organisation structure is most likely to be successful when all project resources are located under one functional group.

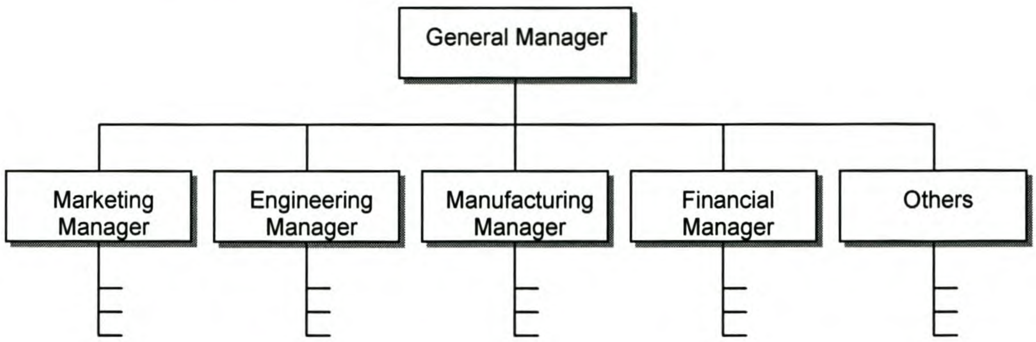

Figure 4: Functional organisation structure

1.6.2 Pure project organisation structure

A pure project organisation structure (Figure 5) is similar in shape to the functional organisational structure except now all the departments and company resources are involved in project work. Project managers have a great deal of independence and authority. This type of structure is typical of large projects of long duration. A major problem with this kind of organisation is the uncertainty of the employees about where they will go when the project is completed. This terminal anxiety can impede the project completion.

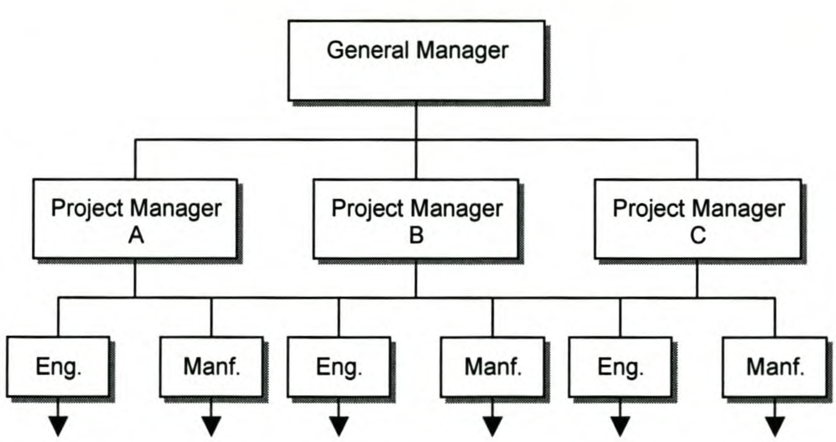

Figure 5: Pure project organisation structure

1.6.3 Matrix organisation structure

The matrix structure presents how the project structure overlays the functional structure and outlines the relationship between the project manager, functional manager and their subordinates (Figure 6). The topology of the matrix structure has the same format as a mathematical matrix. In this case the vertical lines represent the functional department's responsibility and authority while the horizontal lines represent the project's responsibility and authority (Burke, 1999). According to Rosenau (1998), the matrix organisation recognizes that both full-time and part-time assignment of personnel is required and simplifies the allocation and shifting of project priorities in response to management needs. Functional departments are responsible for staffing, developing personnel, and assuring the technical quality of the work done by the personnel. The project managers are responsible for defining the work to be done and establishing a reasonable plan (including schedule and budget) for accomplishing it.

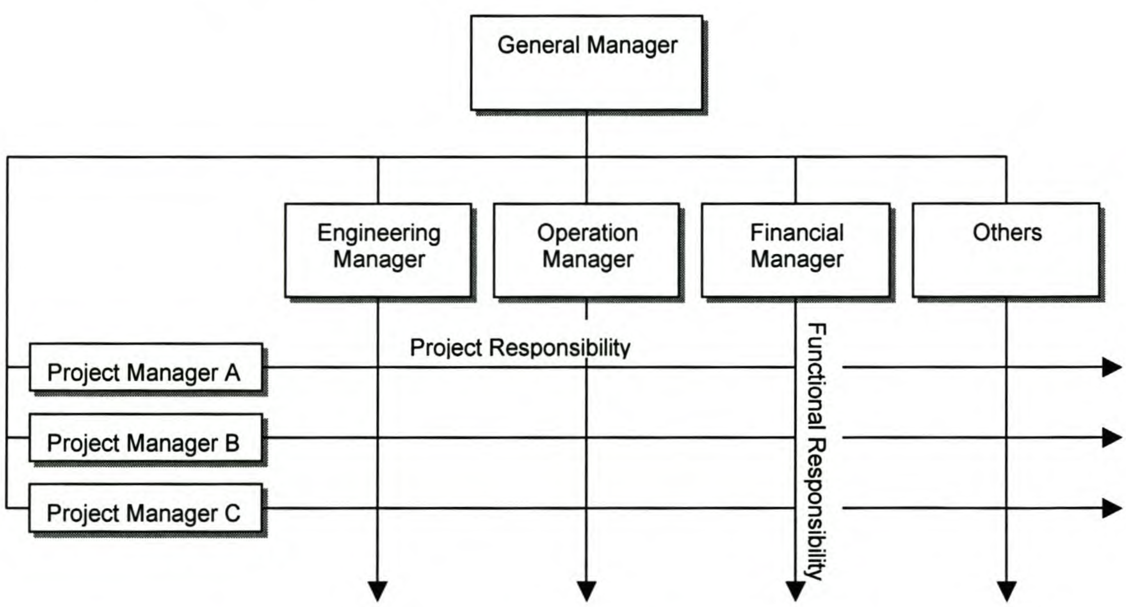

Figure 6: Matrix organisation structure

The matrix structure is considered by many practitioners to be the natural project organisation structure (Burke, 1999) and can be applied at different levels ranging from weak to strong. According to PMI's 'Guide to the PMBOK' (1996), weak matrices retain many of the characteristics of a functional organisation. The project manager's role is more of a co-ordinator or expediter than that of a manager. In similar fashion strong matrices have many of the characteristics of a project organisation—full-time project managers with considerable authority and full-time administrative staff.

Rosenau (1998) concludes that from a project management point of view, the functional organisation is least desirable. The project organisation is most useful on large projects of long duration and the matrix organisation is probably the best organisational option if one has many projects.

1.7 Project management processes

Projects are composed of processes performed by people and generally fall into one of two major categories as defined by PMI's 'Guide to the PMBOK' (1996):

- Project management processes concerned with describing and organising the work of the project.
- Product-oriented processes concerned with specifying and creating the project product.

Product management processes and product-orientated processes overlap and interact throughout the project.

PMI's PMBOK (1996) organises project management processes into five groups of one or more processes each:

- **Initiating processes**—recognising that a project or phase should begin and committing to do so.
- **Planning processes**—devising and maintaining a workable scheme to accomplish the business requirements that the project was undertaken to address.
- **Executing processes**—coordinating people and other resources to carry out the plan.
- **Controlling processes**—ensuring that project objectives are met by monitoring and measuring progress and taking corrective action when necessary.
- **Closing processes**—formalising acceptance of the project or phase and bringing it to an orderly end.

The process groups are furthermore linked by the results they produce. Among the central process groups the links are iterated—planning provides executing with a documented project plan early on, and then provides documented updates to the plan as the project progresses. These connections are illustrated in Figure 7.

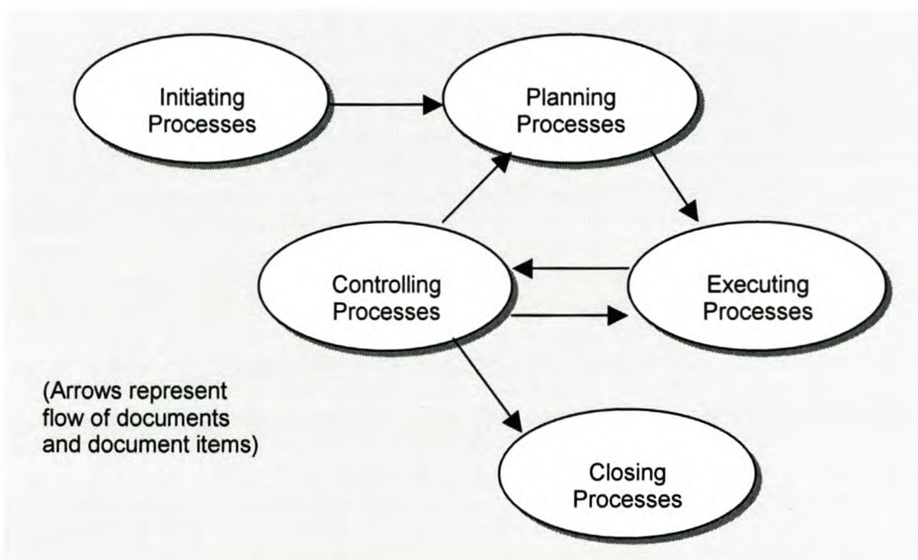

Figure 7: Links among process groups in a phase

Source: PMI's 'Guide to the PMBOK', 1996

1.8 Project management information systems

Organisations typically have numerous projects running simultaneously. The challenge is to not only manage and report on them effectively but to accurately predict resource needs in the near future and track the outcomes and their progress. Information is essential to the design and execution of decisions of allocating resources in the management of a project. Decisions agreed on during planning, organisation, directing and controlling the project must be based on timely and relevant information.

Computer technology is an integral part of modern project management. Project Management Information Systems (PMIS*) is a critical element of the project management infrastructure related to information creation, collection, storage and dissemination and on how to design and implement processes and tools, specific to their project needs, in order to be more efficient. The objective of a PMIS is to provide the basis to plan, to monitor, to do integrated project evaluation and to show the interrelationships among cost, schedule and technical performance for the entire project and for the strategic direction of the organisation. With the development of and proliferation of microcomputers and the corresponding availability of a wide variety of project management software, project managers use one or more PMIS.

The latest microcomputer-based PMIS are considerably more sophisticated than earlier systems and use the microcomputer's graphics, colour and other features more extensively. Many systems can handle almost any size project, limited only by the memory available by the computer. According to Meredith and Mantel (2000), the PMIS trend of the early 1990s has been to integrate the project management software with spreadsheets, databases, word processors, communication, and the other capabilities of Windows-based software packages. The current trend is to facilitate the global sharing of project information, including complete status reporting, through local networks as well as the Internet. Project information systems are becoming increasingly important in the overall management of projects.

**The acronym PMIS is used as both singular and plural throughout the document.*

1.9 Chapter summary

A project is defined as a temporary endeavour undertaken to create a unique product or service. Project management is the application of skills, tools and techniques to project activities in order to meet stakeholders' needs and expectations from a project. Project activities fall into the two categories of project planning and project monitoring, while the success of a project is usually measured against the three constraints of time, cost and user requirements.

Projects undertaken are usually divided into several project phases to provide better management control. It is generally accepted that projects fall into the following four phases:

- Concept and initiation phase
- Design and development phase
- Implementation or construction phase
- Commissioning and handover phase

Collectively, the project phases are known as the project life cycle.

Projects are typically part of an organisation larger than the project. The project organisation structure identifies the relationship between the project participants together with defining their duties, responsibilities and authority. Because of the dynamic nature of projects, it is possible to have a number of organisation structures running concurrently and during the project all the organisation structures may be used. These structures outline the relationships between the various participants, lines of authority and the lines of communication. The matrix structure is considered to be the natural project organisation structure.

Projects are composed of processes performed by people and generally fall into one of two major categories:

- Project management processes concerned with describing and organising the work of the project.

- Product-oriented processes concerned with specifying and creating the project product.

Product management processes and product-orientated processes overlap and interact throughout the project.

Project Management Information Systems (PMIS) is a critical element of project management. The objective of a PMIS is to provide the basis to define the project requirements and elements in a standardised way to communicate these requirements and elements to the project team and to the project owner as a basis for evaluating and measuring performance and directing and reprogramming the effort required to realise the project purposes.

2. Project risk management

Project success does not happen automatically. There are many risks that can cause the project to miss its key objectives of cost, time and scope. Project risk management is practiced to provide the project manager with an early warning on the risks of importance so that they can be addressed and improve the chance for project success.

Sadly, many organisations do not know much about risk management and do not even attempt to practice it (Hullet, 2001).

This chapter presents an overview of project risk management. Risk is defined and a look is taken at how risk changes as the project passes through the different life cycle phases. An overview of different project risk management approaches and a more detailed look at the various processes involved within project risk management concludes the chapter.

2.1 What is risk

Projects inherently contain risks. There are many definitions for describing risk. A simple definition of a "risk" is a problem that could cause some loss or threaten the success of a project but which has not happened yet (Wiegers,1998). These potential problems might have an adverse impact on the cost, schedule, technical success of the project, the quality of the product or project team morale. Wiegers further states that risks, as potential problems, need to be differentiated from the current problems facing a project because different approaches are taken for addressing the two issues. Current, real problems require prompt corrective action. Whereas looming risks can be dealt with in several different ways. A risk could be avoided totally by changing the project approach or even cancelling the project or the risk can be absorbed and no specific action taken to avoid or minimise it.

Chapman and Ward (1997) give a broad definition of project risk as "...the implications of the existence of significant uncertainty about the level of project performance achievable." They continue to say that a source of risk is any factor that can affect project performance and risk arises when this effect is both uncertain and significant in its impact on project performance. Setting tight cost or time targets makes a project more cost or time risky by definition Since achievement of targets are uncertain if targets are tight. Conversely, setting slack time or low quality requirements implies low time or quality risk. However, inappropriate targets are themselves a source of risk and a failure to acknowledge the need for a minimum level of performance against criteria automatically generates risk on those dimensions.

It is the writer's view that project risk may be defined as any event with a probability of occurring that prevents or limits project objectives as defined at the outset of the project, and these objectives may be revised or changed as the project progresses through the project life cycle.

2.1.1 Risk and uncertainty

Many risks are attributable to uncertainty about the things that are believed to be under control. It is what is *not* known that can prevent or limit the success of a project. Uncertainty is a normal and unavoidable characteristic of most projects. It can result from the continuously increasing complexity of the product created and the haste with which it is created. Living on the edge of rapidly changing technology or business conditions is a very real source of uncertainty. Lack of practical knowledge about the project and product development techniques and tools being used presents an additional source of uncertainty.

Risk can also be seen as a combination of constraint and uncertainty as illustrated in Figure 8 (Tusler, 1996). The illustration plots uncertainty against constraint. The curved line indicates the 'acceptable level of risk' whatever that may be in the individual case. Reducing either the uncertainty or constraint may reduce the risk to an acceptable level. Tusler states that in practice, few people have the opportunity to reduce constraint therefore most focus on the reduction of uncertainty. It is also worth noting from the diagram that total elimination of risk can rarely be achieved. Consideration must therefore be given to how the remaining risk can most effectively be managed.

Controlling risk partly means reducing uncertainty or the probable impact. Of course reducing uncertainty has a cost. Such costs need to be balanced against the potential cost incurred if the risk is not addressed and does indeed occur. It may not always be cost-effective to reduce uncertainty too much.

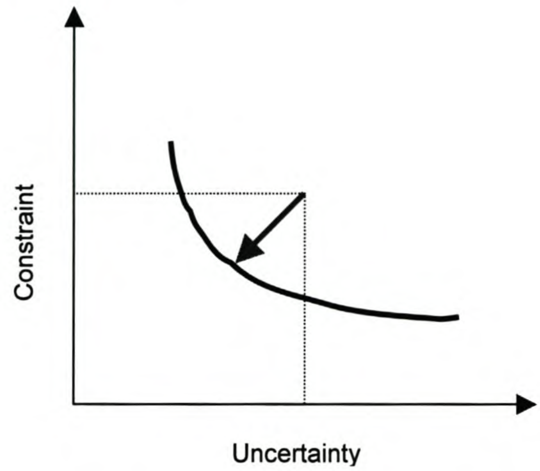

Figure 8: Minimising risk in projects

Source: Tusler, 1996

2.1.2 Risk versus opportunity

Risk and opportunity go hand in hand. Many development projects strive to advance current capabilities and achieve something that has not been done before. The opportunity for advancement cannot be achieved without taking risk.

Risk in itself is not bad. Risk is essential to progress and failure is often a key part of learning. Learning must take place to balance the possible negative consequences of risk against the potential benefits of its associated opportunity.

According to Chapman and Ward (1997), if risk is seen as a 'bad thing', a source of fear to be avoided, people develop blinkers as a natural defence mechanism. If risk is seen as a 'good thing', an opportunity and source of satisfaction to be seized, people take off their blinkers. They start to look for opportunities.

2.2 Risk through the project life cycle

The project life cycle provides an overview of the project phases viz. concept, design, implement and commission. Figure 9 outlines how risk and the amount at stake change as the project progresses.

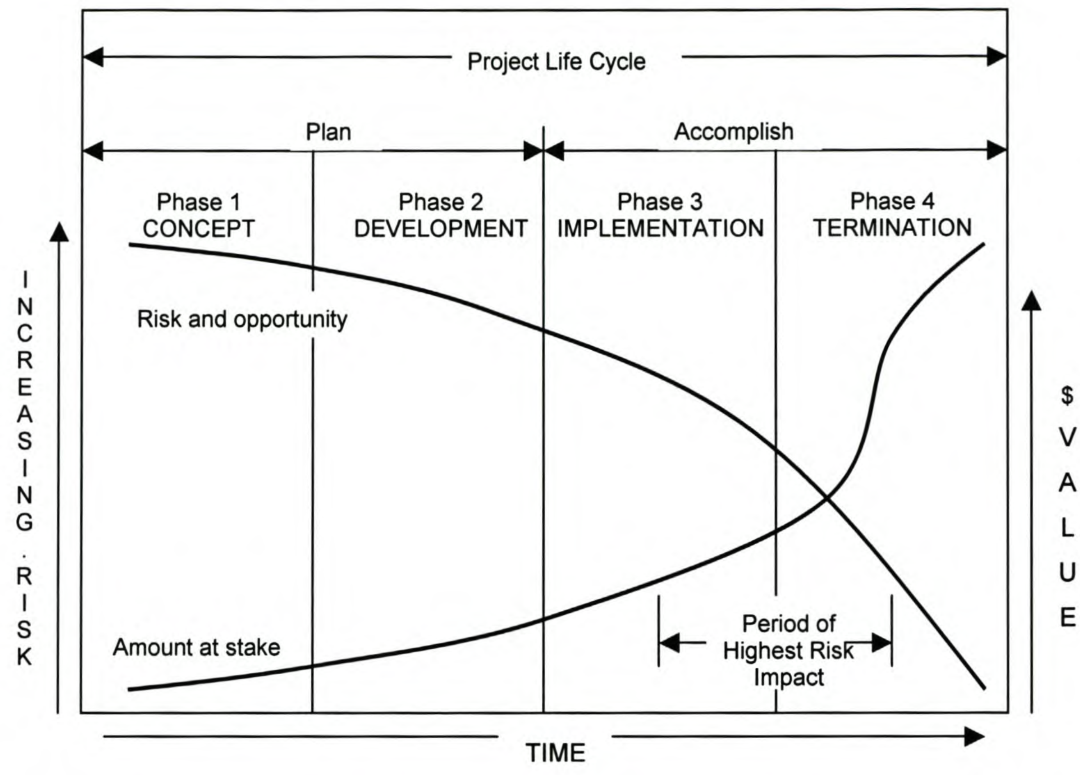

Figure 9: Risk through the project life cycle

Source: Burke, 1999

Risk and opportunity are high at the outset of the project (during the concept and design phases) when there is the greatest degree of uncertainty about the future. As the project progresses, these parameters reduce as decisions are made, design freezes are implemented and the remaining unknowns are translated into knowns. These unknowns are eventually zero when the project has been successfully completed. On the other hand, the amount at stake starts low and steadily rises as capital is invested to complete the project.

In Figure 9 the period of highest vulnerability to risk occurring is during the last two phases (implementation and commissioning). During these phases, adverse conditions may be discovered, particularly during commissioning and start-up. This is also when the level of influence is low and the cost to change is high.

2.3 Risk management overview

A key component of project management is making decisions. Ideally these should be based on complete information with a high degree of certainty of the outcome. However, in the real world most decisions are based on incomplete information with an associated level of uncertainty about the outcome.

Company success is achieved by pursuing opportunities to gain a competitive advantage and projects have typically been set up to take advantage of these opportunities viz. to make something new or change an existing facility. So risk has always been an intrinsic part of project management. With increasing market competition, improved technology and an increasing rate of change, risk management is gaining in significance and importance.

Risk management is a very broad topic. It is entirely a matter of choice where the line is drawn between project risk management and general project management. Grey (1995) describes three views, illustrated in Figures 10 – 12 of the relationship between risk management and project management. He argues that they are all perfectly valid and each one might be a useful way of thinking about the subject in different circumstances.

Figure 10 is the traditional view of risk management, a part of the project management function carried out by the project manager or delegated to a member of his or her team.

Figure 11 is almost the opposite. It is based on the idea that if there were no risks in a project, the need for project management would fade away. It would become an administrative task. Expressed differently, the main purpose of project management is to manage the risks in a project. This view is sometimes summed up in the term 'Risk-driven project management'.

The third view (in Figure 12) depicts closely the picture of risk management set out in this chapter. It illustrates the fact that risk management has to be considered in all aspects of project management but there are also some tasks that most project managers would expect to delegate to consultants or external specialists.

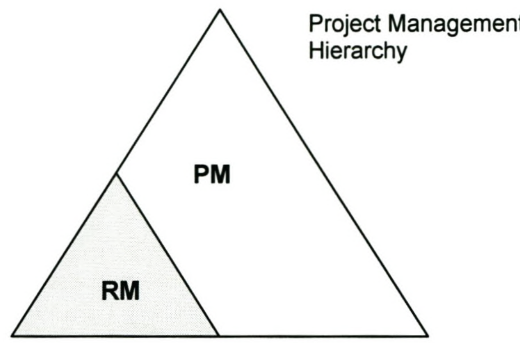

Figure 10: Risk management supporting project management

Source: Grey, 1995

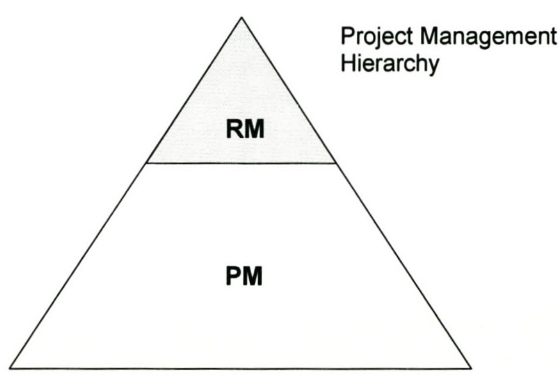

Figure 11: Risk management is *raison d'être* of project management

Source: Grey, 1995

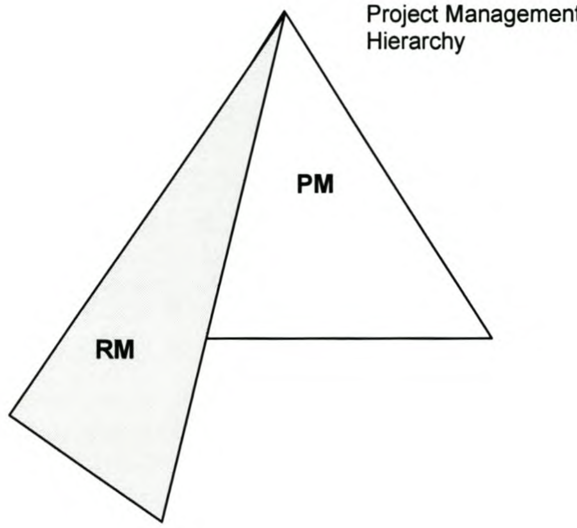

Figure 12: Risk management permeates all of project management

Source: Grey, 1995

PMI's 'Guide to the PMBOK' (1996) formally defines project risk management as "...the process concerned with identifying, analysing and responding to project risk (or uncertainty). It includes the maximising of the results of positive events and minimising the consequences of adverse events". It argues that project risk management consists of four processes:

- **Risk identification**—determining which risks are likely to affect the project and documenting the characteristics of each.
- **Risk quantification**—evaluating risks and risk interactions to assess the range of possible project outcomes.
- **Risk response development**—defining enhancement steps for opportunities and responses to threats.
- **Risk response control**—responding to changes in risk over the course of the project.

These processes interact with each other and with the processes in the other knowledge areas as well. Each process generally occurs at least once in every project phase.

There are many different descriptions for the process described here. Later in the chapter other project risk management processes are discussed and their characteristics summarised into a general model for project risk management.

2.3.1 Why risk management should be applied formally

A formal risk management process provides a number of benefits to the project team. According to Wiegers (1998), it gives a structured mechanism to provide visibility into threats to project success. By considering the potential impact of each risk event, the most severe risks and ways of controlling them can be focused on first. A team approach allows the various project stakeholders to collaboratively address their shared risks and to assign responsibility for risk mitigation to the most appropriate individuals. Risk assessment can be combined with project estimation to quantify possible schedule slippage if certain risks materialise into problems. A formal approach is required to ensure that risk management actions will be initiated in a timely fashion, completed as planned and will be effective. The net result of these activities is to help avoid preventable surprises late in the project and therefore improve the chance of meeting the organisation's and project commitments.

The entire organisation also enjoys benefits from risk management. Sharing what does and does not work to control risks across multiple projects helps projects avoid repeating the mistakes of the past. Members of the organisation can pool their experience and identify opportunities to control common risks through education, process improvement and application of improved management techniques. Over a period of time, a checklist can be built of risk items and mitigation strategies from multiple projects that can help future projects manage their risk effectively.

2.3.2 Role of risk management

The essential purpose of risk management is to improve project performance via systematic identification, appraisal and management of project related risk. Chapman and Ward (1997) argue that focusing on reducing threats or adverse outcomes, what they refer to as 'downside' risk, misses a key part of the overall picture. They say that the aim of improving performance implies wide perspective that seeks to exploit opportunities or favourable possibilities. They refer to this as an 'upside' risk.

To realise in practical terms the advantages of this wide perspective, it is essential to see project risk management as an important extension of conventional project planning with the potential to influence project design and base plans on a routine basis.

2.3.3 When to apply risk management

Risk management is most valuable early in a project proposal while there is still the flexibility in design and planning to consider how the serious risks may be avoided. Not all risks can be avoided e.g. changes in the predictions of the demand for the service or product. Risk management should therefore be continued throughout the life cycle of a project.

The greatest uncertainty is encountered early in the life cycle of a new project. Consequently decisions taken during the earliest stages of a project can have a very large impact on its final cost and duration.

Applying risk management at the project appraisal stage will give the client a much clearer idea of the project and will enable decisions to be made with far more confidence.

2.4 Project risk management processes

Methods for project risk management fall into three broad groups:

- Issue-based methods which invite consideration of technical, commercial, management and other types of risk or else provide a checklist of things that might go wrong.
- Scoring techniques based on a questionnaire enquiring if a certain factor is applicable to a project and assigning project points, depending how bad it is with the total number of points as a measure of the overall riskiness of the project.
- Quantitative techniques that aim to represent the likelihood and impact of risk in terms of the usual planning measures such as time and money.

Some methods are a mixture of all these approaches. Tables 1 to 4 summarise different project risk management models as used in different application areas.

Table 1: Project risk management processes as described by the PMI's 'Guide to the PMBOK' (1996)

PROCESS	OUTPUT
Risk identification	Sources of risk
	Potential risk events
	Risk symptoms
Risk quantification	Opportunities to pursue, threats to respond to
	Opportunities to ignore, threats to accept
Risk response development	Risk management plan
	Contingency plans
	Reserves
	Contractual agreements
Risk response control	Corrective action
	Updates to risk management plan

Table 2: Project risk management processes for construction projects

Source: Thompson and Perry, 1992

Qualitative analysis	Risk identification
	Initial risk analysis
Quantitative analysis	Estimates of activity cost and duration uncertainty
	Probabilistic combination of individual uncertainties
Risk management	Identifying preventive measures
	Reducing uncertainty in project stages
	Reduce transfer contract strategy
	Risk transfer to insurers
	Provide risk allowances
	Establish contingency plans

Table 3: Risk analysis and management for projects (RAMP)

Source: Institute of Civil Engineers and the Faculty and Institute of Actuaries

Process launch	Defining the RAMP strategy
	Establishing the baseline
Risk review	Plan and initiate risk review
	Identify risks
	Evaluate risks
	Devise measures for mitigating risks
	Asses residual risks
	Plan response to residual risk
	Communicate mitigation strategy and response plan
Risk management	Implement strategy and plans
	Control risks
Process close-down	Asses investment outturn
	Review RAMP process

RAMP (risk analysis and management for projects) is a process that has been developed by a joint working party of the actuarial and civil engineering professions for the purpose of evaluating and controlling risk in major projects (see Table 3). RAMP demonstrates how to identify, analyse and mitigate risks and how to place financial values on them. RAMP was developed with major projects in mind. Its principles are also applicable to smaller projects.

SCERT (Synergistic Contingency Evaluation and Review Technique) was initially developed for planning North Sea oil products and subsequently adopted by BP worldwide for planning and costing (see Table 4). The SCERT approach was later developed to be applied by other organisations for the determination of availability, reliability, maintainability and technical choices (Turner, 1995).

Table 4: Project risk management using SCERT

Source: Turner, 1995

STAGE	PHASE	STEPS
Qualitative analysis	Scope phase	Compose breakdown
		Identification of sources of risk
		Identification of responses
		Identification of secondary risk
		Identification of secondary responses
	Structuring phase	Risk and response links
		Major and minor links
		Focus of risks and responses
		Sequence of decision making
Quantitative analysis	Parameter phase	Deciding when to quantify
		Quantifying uncertainty
		Combining risks
Risk management		Identifying data and response needs
		Managing the needs as part of project risk
		Interpretation of results
		Developing risk reduction strategies and responses

Different application areas often use different names for the processes described here. For example:

- Risk identification and risk quantification are sometimes treated as a single process and the combined process may be called risk analysis or risk assessment.
- Risk response development is sometimes called response planning or risk mitigation.
- Risk response development and risk control are sometimes treated as a single process and the combined process may be called risk management.

From the four project risk management models presented above, a generic project risk management model, shown in Table 5, can be derived. The proposed model is shown below.

Table 5: Proposed model of project risk management

Risk analysis	Qualitative	Identification
		Assessment
	Quantitative	Estimating
		Probabilities
Risk management	Risk response development	Avoidance
		Mitigation
		Acceptance
	Implementing risk strategy	
	Risk response control	
	Documentation	

The model is subsequently discussed.

The project risk management model is divided into risk analysis and risk management.

Risk analysis can be qualitative and quantitative. Firstly the sources of risk must be identified. Secondly their effects must be assessed or analysed.

Risk management requires management responses and policies to reduce and control the main risks identified in the analysis.

2.4.1 Qualitative analysis

This process has two aims viz. risk identification and risk assessment which are discussed below.

2.4.1.1 Risk identification and assessment

This is clearly a crucial phase. If risk is not identified it cannot be evaluated and managed. The objective is to compile and document a list of the main risk sources and a description of their likely consequences, perhaps including a first approximation of their potential effect on estimates of cost and time (Thompson and Perry, 1992). Risk identification is not a one-time event. It should be performed on a regular basis throughout the project.

Risk identification should address both internal and external risks. Internal risks are things that the project team can control and influence. External risks are beyond the control or influence of the project team such as market shifts or government action.

Four techniques are commonly used:

1. Check lists of risks compiled from previous experience.
2. Flow charting to help understand the causes and effects of risks.
3. Interviews with key project participants.
4. Brainstorming with the project team.

Outputs from risk identification and assessment include:

- Sources of risk. Descriptions of the sources of risk should generally include estimates of:
    - The probability that a risk event from a source should occur.
    - The impact of the risk event.
    - The range of outcomes.
    - Expected time of the risk event.
    - Anticipated frequency of the risk.
- Potential risk events such as a natural disaster.
- Risk symptoms or triggers. For example, poor morale may be an early warning signal of an impending schedule delay.

The benefit from this first step is the understanding that is created of the project and its potential problems as well as the provision of thought about the management responses to the risks.

2.4.2 Quantitative analysis

Quantitative analysis usually involves more sophisticated analysis techniques, often requiring computer programs and is primarily concerned with determining which risk events warrant response. This is the most formal aspect and requires the following:

- Estimates of uncertainty in predicting the cost and duration of activities
- Probabilistic combination of individual uncertainties.

According to Thompson and Perry (1992), the most useful techniques are sensitivity analysis and probability analysis. The choice of technique should however depend on many factors, principally:

- The type and size of project.
- The information available.
- The cost of the analysis and time available to carry it out.
- The experience and expertise of the analyst.

2.4.2.1 Tools and techniques for risk quantification

1. **Sensitivity analysis.** This is a technique used to consider the effect on the whole project of changes in the value of each variable that is considered to be a potentially serious risk to the project. A sensitivity analysis should be performed for all the risks and uncertainties that may affect a project in order to identify those that have a large impact on the project outcome. Project outcome is usually considered in terms of speed of completion, final cost or an economic criterion such as Net Present Value (NPV) or Internal Rate of Return (IRR). It can also be used to identify the variables that need to be considered for carrying out a probability analysis.

2. **Probability analysis.** This analysis overcomes many of the limitations of sensitivity analysis by specifying the probability distribution for each risk and then considering the effects on the risks in combination. The result of the analysis is a range of values in which the final outcome could be found.

3. **Simulation.** Simulation uses a representation or model of a system to analyse the behaviour or performance of the system. Schedule simulation can be used to quantify

the risk of various schedule alternatives, different project strategies or individual activities. Simulation can also be used to assess the range of possible cost outcomes.

4. **Decision trees.** A decision tree is a diagram that depicts key interactions amongst decisions and associated chance events as they are understood by the decision maker. They are commonly used to study alternative projects and the effects of a design and other choices on project costs. The addition of estimated costs, values of outcomes and probabilities provides a basis for analysing complex problems.

5. **Expert judgement.** Expert judgement can often be applied instead of or in addition to the mathematical techniques described above. For example, risk events could be described as having a high, medium or low probability of occurrence and a severe, moderate or limited impact.

2.4.3 Risk response development

Risk response development involves defining enhancement steps for opportunities and responses to threats. Responses generally fall into one of three categories (PMI's 'Guide to the PMBOK', 1996):

- Avoidance—eliminating a specific threat usually by eliminating the cause. The project team can never eliminate all risk but specific risk events can often be eliminated.
- Mitigation—reducing the expected monetary value of a risk by reducing the probability of occurrence (e.g. using proven technology to reduce the probability that the product of the project will not work) reducing the risk event value (e.g. buying insurance) or both.
- Acceptance—accepting the consequences. Acceptance can be active e.g. by developing a contingency plan to execute should the risk event occur or passive e.g. by accepting a lower profit if some activities overrun.

2.4.3.1 Tools and techniques for risk response development

1. **Procurement.** Procurement viz. acquiring goods or services from outside the immediate project organisation, is often an appropriate response to some types of risk. For example, risks associated with using a particular technology may be mitigated by contracting with an organisation that has experience with that technology.

2. **Contingency planning.** Contingency planning involves defining action steps to be taken if an identified risk event should occur.

3. **Alternative strategies.** Risk events can often be prevented or avoided by changing the planned approach.

4. **Insurance.** Insurance is often used to deal with some categories of risk. A third party accepts insurable risks for the payment of a premium. The premium is now the quantified impact of the risk on the project.

Outputs from risk response development include:

- **Risk management plan.** The risk management plan should document the procedures that will be used to manage risk throughout the project.
- **Contingency plans.** Contingency plans are pre-defined action steps to be taken if an identified risk event should occur.
- **Reserves.** A reserve is a provision in the project plan to mitigate cost and/ or schedule risk.

A risk response plan is needed to minimise the probability of, and contain the impact of all the remaining risks that cannot economically or practically be avoided.

2.4.4 Implementing risk strategy

A precise and comprehensive structure for implementing the risk strategy is essential if the strategy is to be carried out. A lack of attention to this could result in the failure of the project. The results of the risk identification, the risk quantification analysis and risk response plans are used to manage risks as part of the mainstream management of the project. It is essential that the risk analysis, strategies and plans continue to be monitored and updated regularly as risk exposures change and risk events occur in between risk reviews. Any significant changes or developments during the implementation of the risk response strategy should be reported promptly to the project or project risk manager.

2.4.5 Risk response control

When changes occur, the basic cycle of identify, quantify and respond is repeated. It is important to understand that even the most thorough and comprehensive analysis cannot identify all risks and probabilities correctly. Control and iteration are required.

2.4.5.1 Tools and techniques for risk response control

PMI's body of knowledge describes two techniques for risk response control:

1. **Workarounds.** Workarounds are unplanned responses to negative risk events. Workarounds are unplanned only in the sense that the response was not defined in advance of the risk event occurring.
2. **Additional risk response development.** If the risk event was not anticipated or the effect is greater than expected, the planned response may not be adequate and it will be necessary to repeat the response development process and perhaps quantification process as well.

2.4.6 Documentation

It is important to document the project risk management process. Documentation should not only occur at the end but throughout the entire project risk management process. Documentation is important for five reasons (Turner, 1995):

- It clarifies thinking.
- It aids communication.
- It can be used to brief new staff.
- It captures expertise for future projects.
- It provides an audit trail if decisions prove unsuccessful.

2.4.6.1 Learning from the past

While it cannot exactly be predicted which of the many threats to projects might come to pass, learning from previous experiences can avoid the same pain and suffering on future projects. As risk management strategies are implemented on projects, records should be kept of risk management activities for future reference. Historical information about what actually happened on previous projects can be especially helpful in identifying potential risks. Information on previous projects can be collected in the following ways:

- Results of even informal risk assessments should be recorded to capture the thinking of the participants on each project.
- Records of the mitigation strategies attempted for each of the risks pursued should be kept, noting which approaches worked well and which did not pay off.

- Post-project reviews should be conducted to identify the unanticipated problems that arose and they should be added to a checklist of potential risk factors that the next project team can think about.

Anything done to improve the ability to avoid or minimise previous problems on future projects will improve a company's business success.

2.5 Chapter summary

Risk can be defined as a problem that could cause some loss or threaten the success of a project but which has not happened yet. Risk can also be seen as a combination of constraint and uncertainty. Reducing either the uncertainty or constraint may reduce the risk to an acceptable level.

Project risk management is defined as the process concerned with identifying, analysing and responding to project risk (or uncertainty). It includes the maximising of the results of positive events and minimising the consequences of adverse events. The essential purpose of project risk management is to improve project performance via systematic identification, appraisal and management of project related risk.

Risk and the amount at stake change as the project progresses. Project risk management is most valuable in the early stages of the project life cycle but should be continued throughout the life cycle of the project.

Risk management is a process consisting of well-defined steps which, when taken in sequence, support better decision making by contributing to a greater insight into risks and their impacts. It is as much about identifying opportunities as it is about avoiding losses. By adopting effective risk management techniques one can help to improve the cost, schedule and performance of the project.

Risk management will not remove all risks from projects Its principal aim is to ensure that risks are kept to manageable levels.

3. Data mining

The technologies for generating and collecting data have been advancing rapidly. At the current stage, lack of data is no longer a problem; however, the inability to generate useful information from data is. The explosive growth in data and databases results in the need to develop new technologies and tools to process data into useful information and knowledge, intelligently and automatically. Data mining, therefore, has become a research area with increasing importance.

The purpose of this chapter is to present an overview of data mining. Definitions for data mining as well as other terms used to describe data mining are discussed. Different types and techniques of data mining are described. The chapter continues with an overview of the data mining process. A description of problems and issues experienced in data mining concludes the chapter.

3.1 What is data mining

Data mining (DM) is the process of posing various queries and extracting useful information, patterns and trends often previously unknown from large quantities of data possibly stored in databases. For many organisations the goals of data mining include improving marketing capabilities, detecting abnormal patterns and predicting the future based on past experiences and current trends.

Various terms have been used to refer to data mining as shown in Figure 13. These include knowledge/data/information discovery and knowledge/data/information extraction (Thuraisingham, 1999). Many people see data mining as the process of extracting previously unknown information while knowledge discovery is seen as the process of making sense of the extracted data. Alternatively, others view data mining as simply an essential step in the process of knowledge discovery in databases.

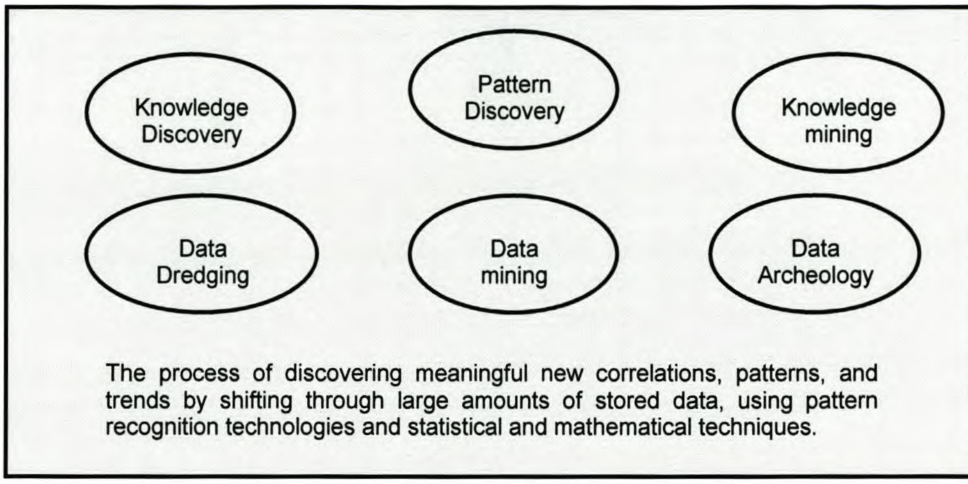

Figure 13: Different definitions of data mining

Source: Thuraisingham, 1999

Data mining is about discovering facts. Some seemingly irrelevant at first but when viewed in a broader context and applied with human intellect and supportive technology, they turn out to hold profound meaning and knowledge. Knowledge is power. The power to compete and win. It drives decisions and progress.

3.2 Data mining technologies

Data mining is an integration of multiple technologies as illustrated in Figure 14. These include data management such as database management, data warehousing, statistics, machine learning, decision support and others such as visualisation and computer technology.

3.2.1 Data mining and decision support systems

Decision support is a broad term for the entire information technology infrastructure that organisations use to make informed decisions. Decision support systems are a collection of tools and processes to help managers make decisions and support them in management. Tools for scheduling meetings, organising events, spreadsheets and performance evaluation tools are examples of decision support systems.

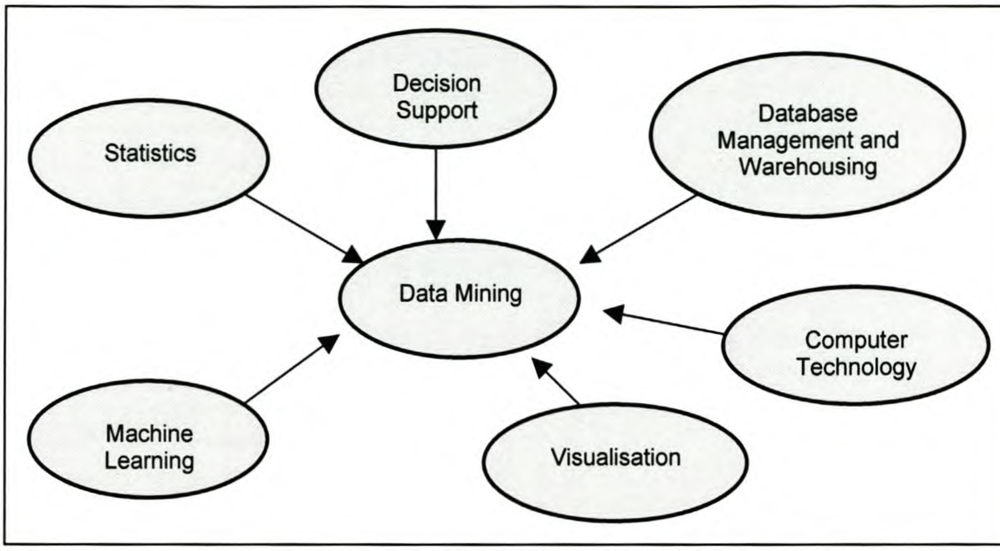

Figure 14: Data mining technologies

Source: Thuraisingham, 1999

3.2.2 Data mining and data warehousing

Data mining potential can be enhanced if the appropriate data has been collected and stored in a data warehouse. A data warehouse is a relational database management system (RDMS) designed specifically to meet the needs of transaction processing systems. The data warehouse is a key data management technology for integrating the various data sources and organising the data so that it can be effectively mined.

3.2.3 Data mining and statistics

Statistics is another important technology that supports data mining. Researchers in statistical analysis are integrating their techniques with machine learning techniques to develop more sophisticated statistical techniques in data mining.

3.2.4 Data mining and machine learning

Machine learning is the automation of a learning process. The machine learns various rules from the patterns observed and then applies these rules to solve the problems. The principles used in machine learning and data mining are similar but data mining is that part of machine learning which is concerned with finding understandable knowledge in large sets of data. Therefore integration of database management and machine learning techniques are needed for data mining.

3.2.5 Data mining and computer technology

Data mining requires complex calculations to be applied to vast amounts of data. The advances in computer power, performance and data input and output speeds have made large-scale data mining practical and profitable.

3.2.6 Data mining and visualisation

Visualisation is the graphical presentation of data. Using visualisation techniques in the data mining process is one of the areas of focus of researches in the computer visualisation field. Often representing data graphically brings out points that are not normally seen.

3.3 Two approaches to data mining

According to Berry and Linoff (2000), there are two approaches to data mining viz. directed data mining and undirected data mining also known as supervised and unsupervised data mining respectively. These two approaches are not mutually exclusive. Data mining efforts often include a combination of both.

3.3.1 Directed data mining

Berry and Linoff (2000) describe directed data mining as a top-down approach. It is used when it is known what to look for when the data mining effort can be directed to a particular goal. Information gathered from known examples is applied to unknown examples. Data mining models based on this approach are called predictive models because they are making predictions about unknown examples. The predictive models use experience to assign scores to some relevant outcome in the future and are usually very accurate. The goal in making predictions is to learn from the past and to learn in such a way that the knowledge can be applied in the future.

3.3.2 Undirected data mining

Berry and Linoff (2000) describe undirected data mining as a bottom-up approach that lets the data speak for itself. Undirected data mining finds patterns in the data and leaves it up to the user to determine whether or not these patterns are important. These patterns provide insight that may even prove very informative. Undirected data mining is necessarily interactive. Advanced algorithms can find patterns in the data but only people can determine whether the patterns have any significance and what the patterns might mean.

3.4 Data mining outcomes

The outcomes of data mining are also referred to as the data mining functions, tasks or types. These are the results that one can expect to see as a result of data mining. Different papers and text use different terms to describe the outcomes. Berry and Linoff (2000) identify six outcomes:

- Classification
- Estimation
- Prediction
- Affinity grouping or association rules
- Clustering
- Description and visualisation

The first three tasks, classification, estimation and prediction are all examples of directed data mining. The next three tasks, affinity grouping, clustering and visualisation are examples of undirected data mining. These outcomes are discussed briefly.

3.4.1 Classification

Classification consists of examining the features of a newly presented object and assigning to it a predefined class. The objects to be classified are generally represented by records in a database. Classification is carried out by developing training sets with preclassified examples and then building up a model that fits the description of the classes. This model is then applied to the data not yet classified in order to classify it. Examples of classification tasks include (Berry and Linoff, 2000):

- Assigning keywords to articles as they come in off the news wire.
- Classifying credit applicants as low, medium or high risk.
- Determining which home telephone lines are used for Internet access.
- Assigning customers to predefined customer segments.

3.4.2 Estimation

Classification deals with discrete outcomes like yes or no, debit card, mortgage or car loan. Estimation deals with continuously valued outcomes. Given some input data, estimation is used to come up with a value for some unknown continuous variable such as an income, height or credit card balance. For example, based on the spending patterns of a person and his age, one can estimate his salary or the number of children he has. In practice estimation is often used to perform a classification task. Other examples of estimation tasks include (Berry and Linoff, 2000):

- Estimating the number of children in a family.
- Estimating a family's total household income.
- Estimating the value of a piece of real estate.

3.4.3 Prediction

Prediction tasks predict the future value of some variable. For example, based on the education of a person, his current job and the trends in the industry, one can predict that his salary will be a certain amount by the end of the year.

Berry and Linoff (2000) state that prediction can be thought of as classification or estimation. The difference is one of emphasis. Predictive tasks feel different because the records are classified according to some predicted future behaviour or estimated future value. With prediction, the only way to check the accuracy of the classification is to wait and see. Other examples of prediction tasks include (Berry and Linoff, 2000):

- Predicting the size of the balance that will be transferred if a credit card prospect accepts a balance transfer offer.
- Predicting which customers will leave within the next six months.
- Predicting which telephone subscribers will order a value-added service such as a three-way calling or voice mail.

Berry and Linoff (2000) conclude that the techniques used for classification and estimation can be adapted for use in prediction by using training examples where the value to be predicted is already known along with historical data for those examples. The historical data is used to build a model that explains the current observed behaviour. When this model is applied to current inputs, the result is a prediction of future behaviour.

3.4.4 Affinity grouping or association rule

The task of affinity grouping is to determine which items go together. Affinity grouping is used to answer questions like: Who are the people that travel together? What are the items that are purchased together? While prediction is some future value and estimation is an estimated value, affinity grouping makes associations between current values.

3.4.5 Clustering

Clustering is the task of creating a partition so that all the members of each set of the partition are similar according to some metric. A cluster is a set of objects grouped together because of their similarity or proximity. Clustering is often confused with classification. While classification classifies an entity based on some predefined values or classes of attributes, clustering groups similar records not based on some predefined values or classes. In clustering there are no predefined classes and no examples. The records are grouped together on the basis of self-similarity. It is up to the miner to determine what meaning, if any, to attach to the resulting clusters.

3.4.6 Description and visualisation

Berry and Linoff (2000) argue that sometimes the purpose of data mining is simply to describe what is going on in a complicated database in a way that increases our understanding of the people, products or processes that produced the data in the first place. They say that a good enough description of a behaviour will often suggest an explanation for it as well.

Data visualisation is one powerful form of descriptive data mining. It makes it possible for the analyst to gain a deeper, more intuitive understanding of the data and as such can work well alongside data mining.

3.5 Data mining techniques

Four data mining techniques are discussed in this chapter:

- Clustering analysis
- Decision tree analysis
- Neural networks
- Case-based reasoning

Data mining techniques are numerous. The writer has chosen to describe these four techniques as they occur most often in literature on data mining and together they cover a wide range of data mining situations.

3.5.1 Clustering

Cluster analysis is an exploratory data analysis technique for solving classification problems. Its object is to sort cases (people, things, events, etc) into groups or clusters so that the degree of association is strong between members of the same cluster and weak between members of different clusters. In an unsupervised learning environment the system has to discover its own classes. One way in which it does this is to cluster the data in the database as shown in Figure 15. The first step is to discover subsets of related objects and then find descriptions (indicated in the figure as D1, D2, D3) that describe each of these subsets. Each class or cluster thus describes in terms of the data collected, the class to which its members belong and this description may be abstracted through use from the particular to the general class or type.

Clustering basically partitions the database so that each partition or group is similar according to some criteria or metric. Clustering according to similarity is a concept that appears in many disciplines. If a measure of similarity is available there are a number of techniques for forming clusters. Membership of groups can be based on the level of similarity between members and from this the rules of membership can be defined.

Clustering in databases are the processes of separating a data set into components that reflect a consistent pattern of behaviour. Once the patterns have been established they can then be used to "deconstruct" data into more understandable subsets. They also

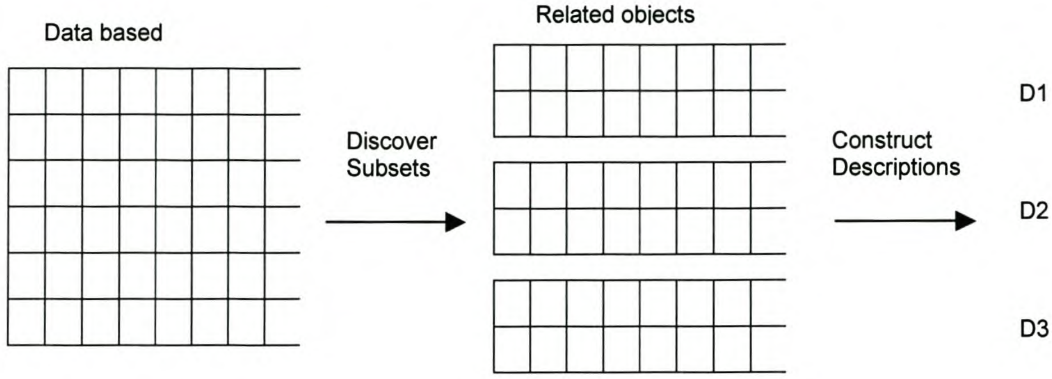

Figure 15: Discovering clusters and descriptions in databases

Source: Dilly, 1995

provide sub-groups of a population for further analysis or action which is important when dealing with very large databases.

Because cluster detection is an undirected technique, it can be applied without prior knowledge of the structure to be discovered. Cluster analysis is thus a technique of discovery. It may reveal associations and structure in data which, though not previously evident, nevertheless are sensible and useful once found.

3.5.1.1 When to use cluster analysis

Berry and Linoff (2000) suggest that cluster analysis should be used when it is suspected that there are natural groupings of data that have a lot in common with each other.

In general, clustering is often useful when there are many competing patterns in the data, making it hard to spot any single pattern. Creating clusters of similar records reduces the complexity within clusters so that other data mining techniques are more likely to succeed.

3.5.2 Decision tree analysis

Decision trees are used to predict and/or classify. They are based on simple tree models where at each branch during tree growth, the data set is strategically partitioned into different classes and subclasses. At each split, the most effective way of partitioning and classifying the data set is accomplished by using the most distinguishing feature encountered at that step of the algorithm.

During the training phase, the data set is partitioned iteratively. During each pass the data set is split on that feature (or attribute) that produces the most effective classification. Only those factors most significant to the effective partitioning are used. The implementation phase then produces decision rules which are equivalent to the partitions (or branches) created during the training phase. When the model is applied to data, each record flows through the tree along a path determined by a series of tests such as “is field 3 greater than 27?” or “is field 4 red, green or blue?” until the record reaches a leaf or terminal node of the tree. There it is given a class label based on the class of records that reached that node in the training set.

The following is an example of objects that describe the weather at a given time (Dilly, 1995). The objects contain information on the outlook, humidity etc. Some objects are positive examples denoted by P and others are negative i.e. N. Classification is in this case the construction of a tree structure, illustrated in the following diagram which can be used to classify all the objects correctly.

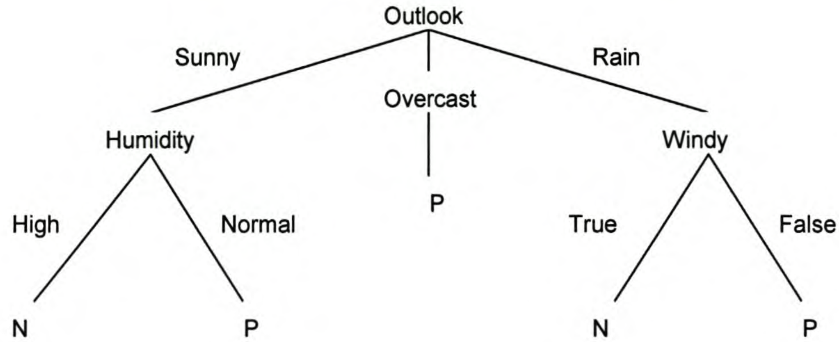

Figure 16: Decision tree structure

Source: Dilly, 1995

3.5.2.1 When to use decision trees

Decision tree methods are a good choice when the data mining task is classification of records or prediction of outcomes. According to Berry and Linoff (2000) decision trees should be used when the goal is to assign each record to one of a few categories. Decision trees are also a natural choice when the goal is to generate rules that can be easily understood, explained and translated into SQL or a natural language.

3.5.3 Neural networks

Neural networks are probably the most widely known and the least understood of the major data mining techniques. Much of the confusion stems from over reliance on the metaphor of the brain that gives the technique its name.

Neural networks are an approach to computing that involves developing mathematical structures with the ability to learn. The methods are the result of academic investigations to model nervous system learning. Neural networks have the ability to derive meaning from complicated or imprecise data and can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques. A trained neural network can be thought of as an "expert" in the category of information it has been given to analyse. This expert can then be used to provide projections given new situations of interest and answer "what if" questions.

Neural networks use a set of processing elements (or nodes) analogous to neurons in the brain. These processing elements are interconnected in a network that can then identify patterns in data once it is exposed to the data i.e. the network learns from experience just as people do. This distinguishes neural networks from traditional computing programs that simply follow instructions in a fixed sequential order.

Figure 17 illustrates a neural network. The model contains an input layer, hidden layer and output layer. Each of the nodes in the hidden layer takes many inputs and generates an output that is a non-linear function of the weighted sum of the inputs. The weights assigned to each of the inputs are obtained during a training process in which the outputs generated by the net are compared with target outputs. The required answers to be produced by the network are compared with the generated outputs and the deviation between them is used as feedback to adjust the weights.

The number of inputs, hidden nodes, outputs and the weighting algorithms for the connections between the nodes determine the complexity of a neural network. There are many approaches for finding the right number of hidden nodes and readjusting weights by a training process.

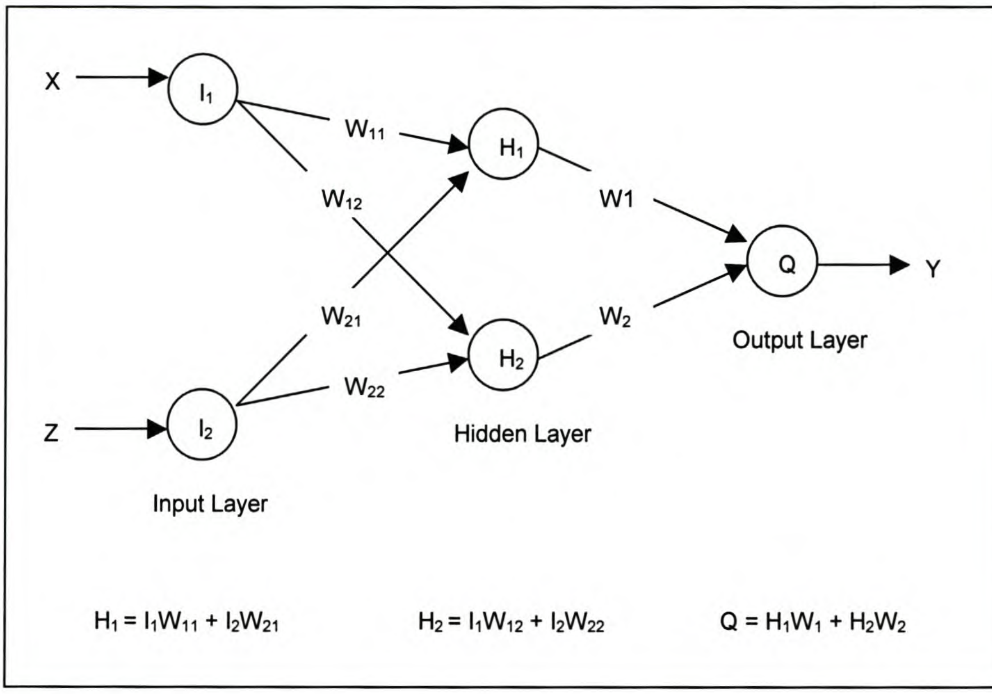

Figure 17: A neural network model

The network illustrated in Figure 17 is a feed-forward network with a hidden layer. By feed-forward we mean that data enters at the input nodes and exits the output nodes without ever looping back on itself.

3.5.3.1 When to use neural networks

According to Berry and Linoff (2000), neural networks are a good choice for most classification and prediction tasks when the results of the model are more important than understanding how the model works.

Neural networks do not work well when there are hundreds of thousands of input features. Large numbers of features make it more difficult for the network to find patterns and can result in long training phases that never converge to a good solution.

3.5.4 Case-based reasoning

Case-based reasoning (CBR) is a problem-solving approach that takes advantage of the knowledge gained from previous attempts to solve a particular problem. A record of each past attempt is stored as a case. The collection of historical cases, the case base then becomes the model. When a CBR system solves a problem, rather than starting from scratch, it searches its case base for cases whose attributes are similar to the problem that it is being asked to solve. The CBR system then creates a solution by synthesizing the similar cases and adjusting the final answer for differences between the current situation and the ones described in the cases. As the case base grows, the accuracy of the system should improve (Dhar and Stein, 1997).

A case can be described as a collection of attributes. Together the attributes describe a scenario involving a situation and an action or solution.

Figure 18 illustrates fragments from two cases. Three attributes are used to describe the problem and are represented by the sticks. The lengths of the sticks are proportional to the value of the numeric attribute they represent. Collectively, the attributes specify a situation.

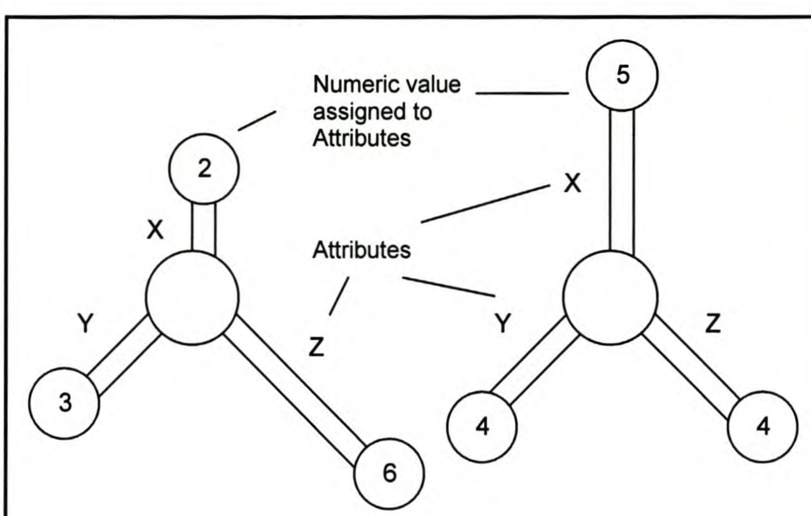

Figure 18: Fragments of two cases

Figure 19 illustrates the "solution" or "action" part of a case. The outcome attributes represent the solution part of the case.

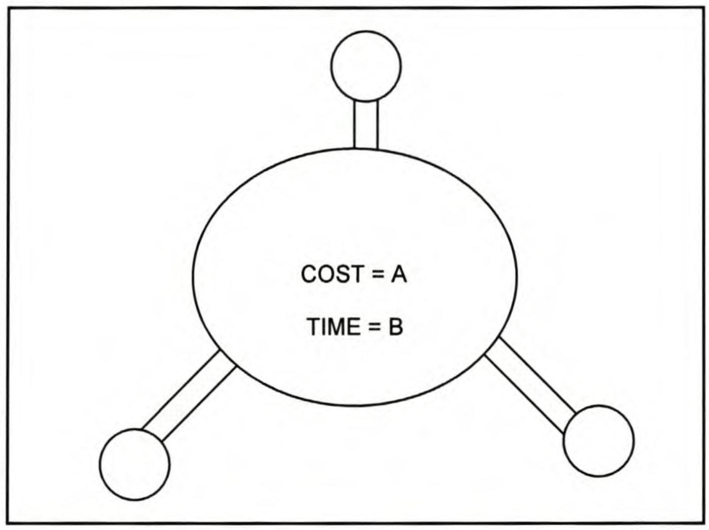

Figure 19: The "solution" or "action" part of the case

The process of finding similar cases and making adjustments to account for the differences between them is called case-based reasoning. In order to be able to deal with a wide variety of situations, it is useful to collect many different cases. A collection of cases is called a case base. Each case is a snapshot of the solution space. It represents a combination of attributes and an outcome. The more cases in the case base, the more complete the cover of the solution space.

Storing cases is good and well, but how can one use all the information stored in the cases? To find useful information in the cases, a CBR system sends a probe into the case-base. A probe is a kind of query that is used to find "close" cases. The probe describes the situation that one is interested in matching. A probe consists of attributes, either situation or outcome, that is used to describe a state of affairs. For example, if a probe is sent with certain "situation" attributes, the CBR system finds cases that are similar to the specified probe. Instead of just presenting the cases to the user, the system automatically makes adjustments to the "solution" component of the closest cases found. These adjustments take into account differences between the situation component of the current problem (the probe) and the situation component of the retrieved cases. The system synthesizes the results into a new solution.

3.5.5 Why use case-based reasoning?

To understand why CBR makes sense, it is worth asking why cases are used so widely in instruction.

Cases transform abstract concepts into real images. Having seen cases, one gets a better sense of how solutions work under different conditions. Future problems can be formulated better and faster. A case provides some inherently useful information because of how it organises information and accesses it. It provides a good starting point from where to begin the search for a better or more detailed solution. It forces one to consider explicitly the similarities and differences between different situations in a structured way, along with the attributes that are used to describe the cases.

3.6 Data mining process in overview

When people talk about data mining, they focus primarily on the actual mining and discovery aspects. However, mining the data is only one step in the overall process. The specific approaches to data mining differ from company to company or researcher to researcher. Cabena, Hadjinian, Stadler, Verhees and Zanasi (1998) propose the following steps in the data mining process:

- Business objective determination
- Data preparation
- Data transformation
- Data mining
- Analysis and results
- Assimilation of knowledge

These steps are illustrated in Figure 20. A brief discussion of each follows after the figure.

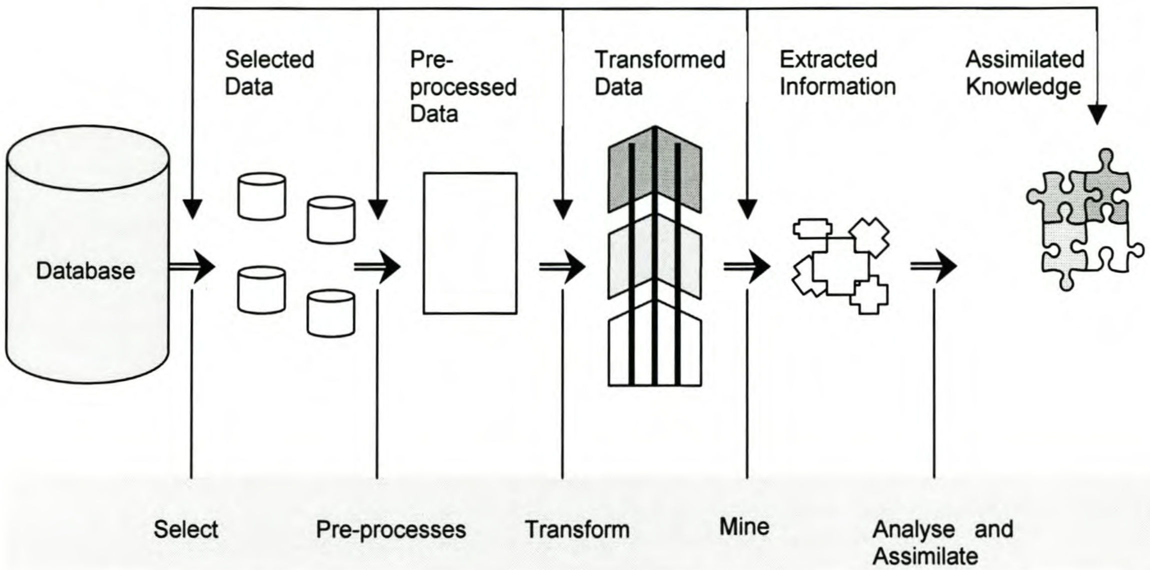

Figure 20: The data mining process

Source: Cabena, Hadjinian, Stadler, Verhees and Zanasi, 1998

3.6.1 Business objective determination

The business problem or challenge must be clearly defined. This is an essential ingredient in any data mining project as the business objectives drive the entire data mining process. They are the basis on which the initial project is established. Data mining for the sake of data mining is rarely successful.

3.6.2 Data preparation

Data preparation can further be divided into 2 sub processes viz. data selection and data pre-processing:

Data selection. Identify all internal and external sources of data and select which subset of the data is needed for the data mining application.

Data pre-processing. Study the quality of the data to pave the way for further analysis and to determine the kind of mining operation that will be possible and worth performing.

3.6.3 Data transformation

To transform the data into an analytical model. Model the data to suite the intended analysis and the data formats required by the data mining algorithms. Presentation of a sound analytical data model of the data to the data mining algorithms is critical for success.

3.6.4 Data mining

To mine the data transformed in the previous step with the selected data mining algorithms. This step is quick and automated.

3.6.5 Analysis of results

To interpret and evaluate the output from the previous step. Typically a visualisation technique is used.

3.6.6 Assimilation of knowledge

To incorporate the business insights gained from the previous step into the organisation's business and information system.

Not all steps in the process are of equal weight in terms of time and effort spent.

Figure 21 presents a broad outline of the steps in the process and the relative effort typically associated with each of them. As shown, 60% of the time goes into preparing the data for data mining. The actual mining step typically constitutes about 10% of the overall effort.

3.7 Data mining problems/issues

Data mining systems rely on databases to supply the raw data for input and this raises problems in that databases tend be dynamic, incomplete, noisy, and large. Other problems arise as a result of the adequacy and relevance of the information stored. Dilly (1995) identifies and discusses the following problems/issues in data mining.

3.7.1 Limited information

A database is often designed for purposes different from data mining and sometimes the properties or attributes that would simplify the learning task are not present nor can they be requested from the real world. Inconclusive data causes problems. If some attributes essential to knowledge about the application domain are not present in the data, it may be impossible to discover significant knowledge about a given domain.

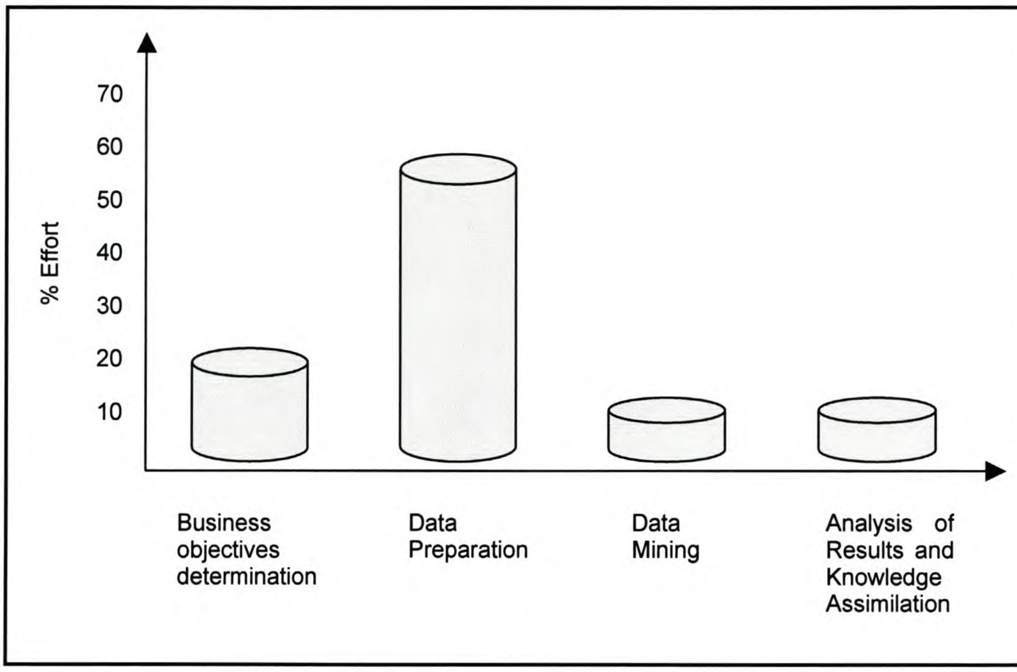

Figure 21: Effort required for each data mining process step

Source: Cabena, Hadjinian, Stadler, Verhees and Zanasi, 1998

3.7.2 Noise and missing values

Databases are usually contaminated by errors so it cannot be assumed that the data they contain is entirely correct. Attributes that rely on subjective or measurement judgements can give rise to errors such as some examples being mis-classified. Error in either the values of attributes or class information are known as noise. Obviously, where possible, it is desirable to eliminate noise from the classification information as this affects the overall accuracy of the generated rules.

3.7.3 Uncertainty

Uncertainty refers to the severity of the error and the degree of noise in the data. Data precision is an important consideration in a discovery system.

3.7.4 Size, updates, and irrelevant fields

Databases tend to be large and dynamic in that their contents are ever-changing as information is added, modified or removed. The problem with this from the data mining perspective is how to ensure that the rules are up-to-date and consistent with the most current information. Also the learning system has to be time-sensitive as some data values vary over a period of time and the discovery system is affected by the 'timeliness' of the data.

Another issue is the relevance or irrelevance of the fields in the database to the current focus of discovery.

3.8 Example of data mining applied in market management

Retail stores are increasingly using loyalty cards to reward their frequent buyers. The underlying concept is that because card holders get special treatment, such as exclusive discounts on selected items, they are thus encouraged to do more of their shopping at that store and are less likely to visit the competition.

COIN, one of the most customer orientated non-food retail chains in Italy, wanted to analyse the data obtained from the loyalty card programme to better understand the kinds of customers who were shopping in their stores, with a view to better targeting store promotions. COIN chose IBM to perform this challenging service.

The output gained by COIN and IBM specialists could allow to group the different clients following the demographic characteristics and purchasing behaviour. But allowed also to understand what were the shop sectors more visited, in what period of the day, season or year and what is the best way to understand the COIN clients, that have different characteristics from other chains clients.

Source: Cabena, Hadjinian, Stadler, Verhees, Zanansi (1998).

3.9 Chapter summary

Data mining is the process of posing various queries and extracting useful information, patterns and trends previously unknown from large quantities of data possibly stored in databases. It is also described as knowledge/data/information discovery and knowledge/data/information extraction.

Data mining is an integration of multiple technologies. These include data management, visualisation and computer technology.

Directed data mining and undirected data mining, also known as supervised and unsupervised data mining respectively, are two approaches to data mining.

The outcomes of data mining are classified according to the results that one can expect to see as a result of data mining. Different papers and text use different terms to describe the outcomes. Six outcomes are identified:

- Classification
- Estimation
- Prediction
- Affinity grouping or association rules
- Clustering
- Description and visualisation

Four data mining techniques were discussed in this chapter. They are:

- Clustering analysis
- Decision tree analysis
- Neural networks
- Case-based reasoning

Data mining does not constitute the only step in the overall process and it is not the most time consuming step either. The largest part of the process is in the data preparation with the actual mining step only playing a minor role in the overall effort.

4. Decision support systems

Decision support systems (DSS*) couple the intellectual sources of individuals with the capabilities of the computer to improve the quality of decisions. It is a computer-based support system for management decision makers who deal with semi-structured problems.

The purpose of this chapter is to present an overview of decision support systems. Definitions of a DSS as well as the purposes of DSS are discussed. DSS are compared to those of data processing and information management systems. The chapter continues describing the characteristics and components of DSS. A look at the different classifications of DSS concludes the chapter.

The terms *procedural knowledge* and *descriptive knowledge* are frequently mentioned in the section relating to DSS. A short description of these two terms is given:

Procedural knowledge is generic knowledge regarding problem-solving procedures. In contrast, descriptive or declarative knowledge relates to the specific knowledge domain of the problem to be solved.

**The acronym DSS is used as both singular and plural throughout the document.*

4.1 What is a DSS

Traditional definitions of decision support systems identify it as a system intended to support decision makers in addressing unstructured or semi-structured decisions. Unstructured decisions are those that are resistant to computerisation and depend primarily on intuition. Semi-structured decisions are those that are partially programmable but still require human judgement. DSS are meant to be an adjunct to decision makers to extend their capabilities but not to replace their judgement.

DSS is sometimes used as an umbrella term to describe any and every computerised system used to support decision making in an organisation. An organisation might have an executive information system (EIS) for its top executives, separate DSS for marketing, finance, accounting, an MRP system for production and several expert systems for product repair diagnostics and help desks (Aronson and Turban, 1998). DSS encompasses them all. In spite of the differences in style and substance, all DSS share similar purposes and exhibit a common set of characteristics. Henceforth, one can say that the term *decision support system* refers to a computerised, supporting participant in decision making. A DSS is also descriptive instead of prescriptive compared to exact analytical methods.

Aronson and Turban (1998) give a working definition that defines a range, from a basic to an ideal DSS:

> "A DSS is an interactive, flexible, and adaptable computer-based information system specially developed for supporting the solution of a non-structured management problem for improved decision making. It uses data, provides easy user interface, and can incorporate the decision maker's own insights.
>
> In addition, a DSS may use models, is built by an interactive process (often by end users), supports all phases of decision making, and may include a knowledge component."

4.2 Purposes of decision support systems

Unlike information systems or data processing systems, a decision support system is not primarily concerned with record keeping, production of standard reports or processing transactions. While a DSS engages in such activities to varying degrees, they do not constitute its focus.

In very broad terms the purpose of DSS is to improve decision-making ability by allowing more or better decisions within the constraints of cognitive, temporal and economic limits. Its purpose is to increase a decision maker's productivity. DSS provide support for decision makers mainly in semi-structured and unstructured situations by bringing together human judgment and computerised information. Such problems cannot be solved (or cannot be solved conveniently) by other computerised systems or by standard quantitative methods or tools (Aronson and Turban, 1998).

More specifically the purpose of a DSS is to facilitate one or more of a decision maker's abilities. Its intention could be to make tasks of knowledge collection easier by providing access to a variety of data sources, formats and types ranging from geographic information systems to object-orientated ones. The purpose of a DSS could be to help a decision maker's formulation ability. It may be designed to actually formulate plans for analysis or action. The purpose of a DSS could be to supplement the decision maker's ability to recognise problems. Similarly, DSS could facilitate a decision maker's evaluation, implementation and governing abilities.

Regarding decision making as a flow of problem-solving episodes, the purpose of a DSS could be to help problem-solving flows go more smoothly or rapidly. Because each problem-solving episode begins with the recognition that the problem exists, a DSS could fulfil this purpose by stimulating the user to perceive problems needing to be solved. It could go so far as to break a problem posed by the user into sub-problems that, when solved, allow the user's problem to be solved.

One other way to assess the purpose of a DSS stems from the knowledge management perspective. When decision making is thought of as a knowledge–manufacturing process, the purpose of a DSS is clearly one of helping the user manage knowledge. A DSS can be used to enhance the user's competence in representing and processing knowledge. It supplements human knowledge management skills with computer-based means for managing knowledge. A DSS accepts, stores, uses, derives and presents knowledge pertinent to the decisions being made. Its capabilities are very much defined by the types of knowledge with which it can work, the ways in which it can represent these various types of knowledge and its capabilities for processing these representations.

4.3 Distinguishing DSS from DP systems and MIS

One way to appreciate the characteristics of a DSS is to compare and contrast them with the traits of two other major types of business computer systems. These are the data processing (DP) systems and management information systems (MIS). Both predate the advent of the DSS. All three share the trait of being concerned with record keeping. On the other hand the three kinds of business computing systems differ in various ways because each serves a different purpose in the management of an organisation's knowledge resources.

The main purpose of DP systems is to automate the handling of a large number of transactions. For example, a bank must deal with large volumes of deposit and withdrawal transactions every day. It must track each transaction's effect on one or more accounts correctly. It must maintain a history of all the transactions that have occurred in order to give a basis for auditing its operations. At the heart of a DP system lies a body of descriptive knowledge i.e. data. This is the computerised record of what is known as a result of various transactions having transpired.

Unlike a DP system, the central purpose of the MIS is to provide managers with periodic reports that recap certain predetermined aspects of an organisation's past operations. The point of such computer-generated reports is to give managers at various organisational levels, regular snapshots of what has been happening in the organisation. Such descriptive knowledge can help them in controlling the organisation's operations. Whereas a DP is concerned with transforming transactions into records and generating transactions from records, the MIS concern with record keeping focuses on using this stored descriptive knowledge as a base for generating recurring standard reports. A MIS can be regarded as an adaptation or extension of the DP idea to emphasize the generation of standard reports for managers rather than the generation of voluminous transactions for customers, suppliers, employees or regulators. The information contained in standard reports that a MIS furnishes to managers certainly can be factored into their decision-making activities. When this is the case, a MIS could be fairly regarded as a kind of DSS.

4.3.1 Comparison between DSS and MIS

Aronson and Turban (1998) describe a few comparisons between DSS and MIS. They mention the following:

MIS can be viewed as an IS infrastructure that can generate standard and exception reports and summaries, provide answers to queries and help in monitoring and tracking. Thus, there are marketing MIS, accounting MIS etc. A DSS on the other hand is basically a problem-solving tool and it is often used to address ad hoc and unexpected problems. MIS is usually developed by the IS department because of its permanent nature. DSS is usually an end-user tool. It can provide decision support within a short time. An MIS can provide quick decision support only to situations for which the models and software were pre-written.

Because of its unstructured nature, DSS is usually developed by a prototype approach. MIS on the other hand, is often developed by a structured methodology such as the system development life cycle.

A DSS can evolve as the decision maker learns more about the problem. Often managers cannot specify in advance what they want from computer programmers and model builders. Many computerised applications are developed in a way that requires detailed specifications to be formalised in advance. This requirement is not reasonable in many semi-structured and unstructured decision-making tasks.

4.4 Characteristics of DSS

Ideally a decision maker should have immediate, focused and clear access to whatever pieces of whatever types of knowledge required on the spur of the moment in coping with semi-structured or unstructured decisions. The pursuit of this ideal separates decision support systems from their DP and MIS ancestors. It also suggests the characteristics expected to be observed in a DSS.

There is no agreement on standard characteristics of DSS. A few of the characteristics of DSS as mentioned by Aronson and Turban (1998) are given below:

- DSS provide support for decision makers mainly in semi-structured and unstructured situations by bringing together human judgement and computerised information.
- DSS support all phases of the decision-making process i.e. intelligence, design, choice and implementation.
- DSS support a variety of decision-making processes and styles.
- DSS are adaptive over a period of time. They are flexible so that users can add, delete, combine change or rearrange basic elements.
- DSS attempt to improve the effectiveness of decision-making (accuracy timeliness, quality), rather than its efficiency (cost) of making decisions.
- A DSS usually utilises models for analysing decision-making situations.
- A DSS should provide access to a variety of data sources, formats and types ranging from geographic information systems to object-orientated ones.

Salvendy (1995) on the other hand, identifies five characteristics to be observed in a DSS:

- A DSS includes a body of knowledge that describes some aspects of the decision maker's world, specifies how to accomplish various tasks and indicates what conclusions are valid in various circumstances and so forth.
- It has an ability to acquire and maintain descriptive knowledge i.e. record keeping and other kinds of knowledge such as i.e. procedure keeping, rule keeping, etc.
- It has an ability to present knowledge on an ad hoc basis in various customised ways as well as in standard reports.
- A DSS has an ability to select any desired subset of stored knowledge for either presentation or deriving new knowledge in the course of problem recognition and/or problem solving.
- A DSS can interact directly with a decision maker or a participant in decision making in such a way that the user has flexibility in the choice and sequencing of knowledge management activities.

There are variations among DSS with respect to each of these characteristics. For example, one DSS may possess descriptive and procedural knowledge while another holds only descriptive and reasoning knowledge. Yet another may store only descriptive knowledge. One DSS may ask its user a series of questions to find out what the user wants. Another DSS allows its user to state a command telling a DSS what is desired. Regardless of such variations, these five characteristics combine to amplify a decision maker's knowledge management capabilities.

4.5 Components of a DSS

Aronson and Turban (1998) identify the following subsystems of which a DSS is composed:

- **Data management sub-system.** The data management subsystem includes the database which contains relevant data for the situation and is managed by software called the database management system (DBMS).
- **Model management sub-system.** A software package that includes financial, statistical, management science or other quantitative models that provide the system's analytical capabilities and appropriate software management. Modelling languages for building custom models are also included. This software is often called a model base management system (MBMS).

- **Knowledge management system.** This sub-system can support any of the other sub-systems or act as an independent component. It provides intelligence to augment the decision maker's own.
- **User interface sub-system.** The user communicates with and commands the DSS through this sub-system.
- The user is considered to be part of the system. Some of the unique contributions of the DSS are derived from the intense interaction between the computer and the decision maker.

A schematic view of a DSS and the above components is shown in Figure 22.

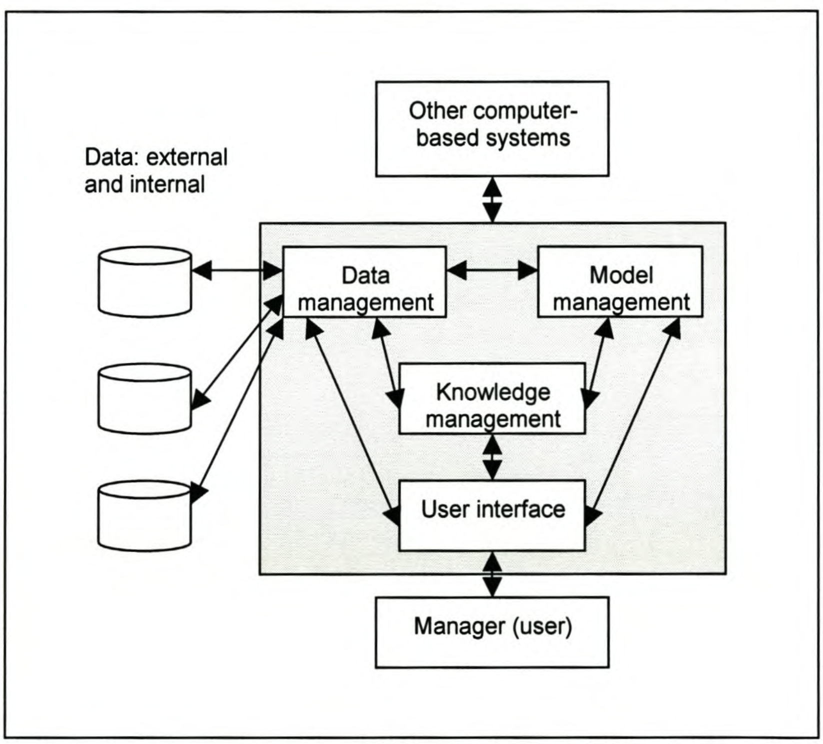

Figure 22: A schematic view of DSS

Source: Aronson and Turban (1998)

4.6 Classifications of DSS

DSS applications can be classified in several ways. The design process as well as the operation and implementation of the DSS depends in many cases on the types of DSS involved.

4.6.1 Text-orientated DSS

Information (including data and knowledge) is often stored in a textual format and must be accessed by the decision makers. A text-orientated DSS supports a decision maker by electronically keeping track of textually represented information that could have a bearing on decisions. Documents can be electronically created, revised and viewed as needed. Information technologies such as document imaging, hypertext and intelligent agents can be incorporated into the text-orientated DSS applications.

4.6.2 Database-orientated DSS

Here the database plays a major role in the DSS structure. Rather than being treated as streams of text, data are organised in a highly structured format (relational or object orientated). The database-orientated DSS features strong report generation and query capabilities.

4.6.3 Spreadsheet-orientated DSS

A spreadsheet is a modelling language that allows the user to write models to execute DSS analysis. These do not only create, view and modify procedural knowledge but also instruct the system to execute their self-contained instructions. The most popular tools for developing DSS are Microsoft Excel and Lotus 1-2-3 both of which are spreadsheet applications. Packages such as Excel can include a rudimentary DBMS or can readily interface with one e.g. Microsoft Access, allowing the manipulation of descriptive knowledge, a property of a database-orientated DSS.

4.6.4 Solver-orientated DSS

A solver is an algorithm or procedure written as a computer programme for performing certain computations for solving a particular problem type. For example, an economic order quantity procedure for calculating an optimal ordering quantity or a linear regression routine for calculating trend. Spreadsheet tools such as Excel and Lotus have solvers, called functions. The solver-orientated DSS can be flexible allowing solvers to be changed, added or deleted as needed.

4.6.5 Rule-orientated DSS

In a rule-orientated DSS, knowledge is represented in the guise of both procedural and inferential (reasoning) rules, often in an expert system. These rules can be qualitative or quantitative or both.

4.6.6 Compound DSS

A compound DSS is a hybrid system that includes two or more of the five basic structures described above.

4.7 Chapter summary

Traditional definitions of decision support systems identify it as a system intended to support decision makers in addressing unstructured or semi-structured decisions. DSS are meant to be an adjunct to decision makers to extend their capabilities but not to replace their judgement. DSS is sometimes used as an umbrella term to describe any and every computerised system used to support decision making in an organisation.

The main purpose of a DSS is to improve the decision-making ability of the decision maker. It also gives the decision maker access to data sources, formats and types and facilitates a decision maker's evaluation, implementation and governing abilities. A DSS accepts, stores, uses, derives and presents knowledge pertinent to the decisions being made.

A DSS is similar to DP and MIS in that all three share the trait of being concerned with record keeping. It differs in that it is basically a problem-solving tool and can evolve as the decision maker learns more about the problem.

Arguably the most important characteristic of DSS is to provide support for decision makers by bringing together human judgement and computerised information.

The following subsystems comprise a DSS:

- Data management subsystem
- Model management subsystem
- Knowledge management system
- User interface subsystem

Five basic structures of DSS were described in this chapter. They are the text, database, spreadsheet, solver and rule-orientated DSS. A compound DSS is a hybrid system that includes two or more of the above-mentioned structures.

5. Integration of the methodologies

This chapter identifies the need to combine the processes of the management methodologies presented in the previous chapters in order to augment the project manager's ability to manage risk in projects. The four methodologies are:

- Project management
- Project risk management
- Data mining
- Decision support systems

5.1 Project management

Projects are undertaken at all levels of the organisation and range in size, scope, cost and time and involves the application of knowledge, skills, tools and techniques to project activities in order to meet stakeholders needs and expectations from a project. They are a one-time endeavour with a set of desired results. Because projects are unique undertakings, they involve a certain degree of uncertainty.

Project management is a process that offers a structured approach to managing projects. A key component of project management is making decisions. Ideally these would be based on complete information with a high degree of certainty of the outcome. However, in the real world most decisions are based on incomplete information with an associated level of uncertainty about the outcome. Project management is therefore a risky business on its own, emphasising the need for a formal project risk management process to ensure project success. Project risk management is subsequently discussed.

5.2 Project risk management

Projects are launched to take advantage of opportunities. In the process one is, however, sometimes at the mercy of uncertainties that have risks attached. Project success does not just happen. There are many risks that can cause the project to miss its key objectives of cost, time and scope.

Project risk management is a process consisting of well-defined steps which, when taken in sequence, support better decision making by contributing to a greater insight into risks and their impacts. It provides a structured mechanism to provide visibility into threats and is practiced to give the project manager an early warning on the risks of importance so they can be addressed appropriately and improve the chance for project success.

Risks occur throughout the project life cycle but risks are greatest at the outset of the project when there is a great degree of uncertainty about the future. At this point, decisions taken, can have a large impact on the cost and duration of the project. An effective risk management strategy that is applied at the early stages of the project can save cost and time and can influence the performance capability of the project. But what is needed for the risk management strategy to be effective? If risk managers can have access to a database of historical data on the risks of previous projects, they can use this data to profile projects according to their risks. This information can be used to improve the decisions made when planning risk strategies of new projects with the result of savings in terms of cost and time and overall better performance of the project. Extracting the information from a database and presenting it in a useful form, from where knowledge can be discovered to predict the impact of risks on a project, can be accomplished through data mining, which is discussed next.

5.3 Data mining

Data mining is the process of applying artificial intelligent techniques to a large data set in order to determine patterns in the data. Data mining, from a business perspective, can be defined as the process of scanning a large data set to glean information. Data mining is critical to the enterprise that wants to exploit operational data and other available data to improve the quality of decision-making.

Data mining is a useful technique for discovering knowledge in data and can be applied to project risk management. It can be used to extract comprehensible and valid information from large relational databases to support decision making. Data on previous or completed projects can be entered into a database. Through predictive modelling, information gathered about the known projects can be used to solve problems of future projects.

An appropriate data mining technique for making predictions on new projects is the case-based reasoning technique. Case-based reasoning (CBR) is a sub-field of Artificial Intelligence that is based on the idea that knowledge gained form past problem solving experiences can be re-used and learned from in solving new problems. In project risk management, records containing data on risks involved in past projects can be stored as a case. The collection of historical cases forms the case base. When risks and certain attributes of a new project have been identified, the case base is searched for projects with similar risks and attributes. The knowledge gained from these projects can then be applied to the risk management strategy of the new project. As the case base grows, the accuracy of the system improves.

The strategic impact of data mining is only beginning to be felt on the processes of decision-making and learning through information discovery. There is a bright, exciting future for knowledge discovery in data warehouses and the central role data mining can play in its evolution, as well as in decision-making. This is discussed in 5.4.

5.4 Decision support systems

Strategic planning is one of the most difficult tasks of modern management. It involves all functional areas in an organisation and several relevant outside factors, a fact that complicates the planning process especially in dealing with long-run uncertainties. Thus, strategic planning is clearly not a structured decision situation and is therefore a potential candidate for decision support system (DSS) applications.

The purpose of a DSS is to improve decision-making ability. DSS provide support for decision makers mainly in semi-structured and unstructured situations by bringing together human judgement and computerised information.

Project information provides the intelligence for managing the project. Information must be processed into knowledge so that decisions can be made and executed. A database of records containing data of projects and their risks, together with data mining as a knowledge discovery technique can be integrated to form a DSS for project and particularly project risk management. The DSS could facilitate a decision maker's evaluation and implementation of risk management strategies within project management.

5.5 Chapter summary

Project management involves making decisions. These decisions are not always based on complete information and therefore a certain amount of risk is involved. Formal risk management greatly improves the likelihood of successful project completion and it reduces the otherwise negative consequences of those risks that cannot be avoided.

Knowledge on risks from previous similar projects and how they were managed, can be used to improve project risk management strategies and therefore the chances for project success. By keeping records of past experiences of risk encountered on projects, data can be mined to present solutions to risks encountered on current projects. Case-based reasoning is a data mining technique identified to transform this data into information so that the knowledge derived from the information can be used to better manage the risks.

The integration of this data mining technique with a decision support system will result in a DSS that can facilitate the decision maker in the evaluation and implementation of risk management strategies within the project management environment.

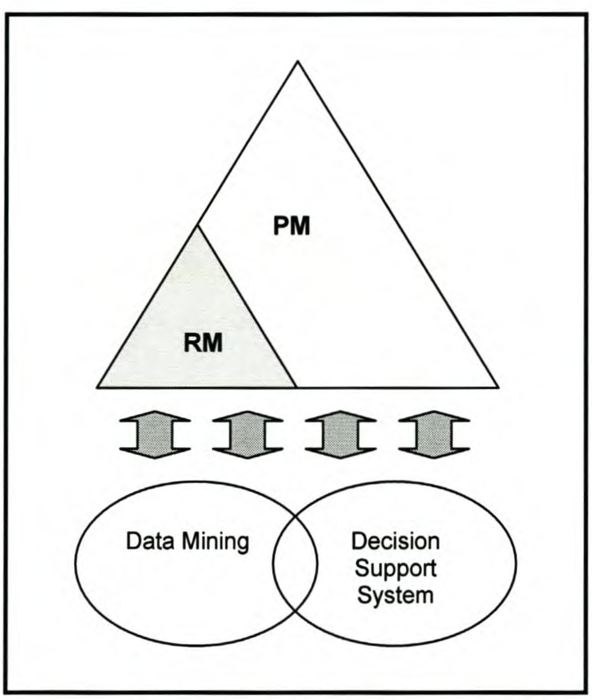

Figure 23: Integration of methodologies

6. Decision support system for risk analysis and management

This chapter presents a proposed architecture for a decision support system that supports project and projects risk managers in the assessment and management of risks encountered in projects. The DSS uses a data mining technique to intelligently profile a project according to identified risk factors.

The chapter also describes why a DSS is needed and how it can support the project risk manager. A break down of the system into its components and a classification of the system is given. Discussing how project risk management can benefit from the system concludes the chapter.

6.1 Why a DSS for project risk management?

To explain the need for a DSS to perform risk analysis, the phenomenon that occurred in the financial industry with the advent of the electronic spreadsheet must be examined. Prior to this innovation, a financial analyst who wanted to, say, make projections about the impact of different sales growth scenarios on a particular firm would (a) copy by hand the income statement, balance sheet, and cash flow data into a ledger; then (b) perform the appropriate arithmetic and accounting operations for a scenario; (c) repeat steps a and b for each scenario; and, finally, (d) perform the analysis. The frustrating thing was that the calculations required to do the work in a, b, and c took a large share of the time, but the actual analysis of the end results, d, which took a much shorter time, was what the analyst was really getting paid for!

With the advent of electronic spreadsheets, though, things changed. Now an analyst can experiment with new scenarios as fast as they can be typed. Once the initial data are entered and the spreadsheet is set up (once only), the analyst is free to experiment extensively. All of the extra time that the analyst would have spent calculating ratios and adding columns by hand can now be spent doing analyses and making decisions.

A spreadsheet-orientated DSS aided the financial analyst in his work. Similarly, a DSS can aid the project risk manager by performing time consuming, repetitive calculations and experimenting with influencing variables and their effect on the project outcome. Over a period of time, a checklist can be built of risk items and mitigation strategies from multiple projects that can help future projects manage their risk effectively. Realisations of decision support entities can also be stored over time thus providing information for future decision making.

6.2 Components of the proposed DSS

The DSS is built up through the integration of the following subsystems:

- Data management subsystem
- Model management subsystem
- User interface
- User (Decision maker)

These components are now discussed in more detail.

6.2.1 Data management subsystem

The data management subsystem is composed of the following elements:

- DSS database and knowledge base
- Data management system
- Query facility

The database is a collection of organised and interrelated data pertaining to the organisation, projects and possible risks within projects. The relational database structure is used to organize the relationships among the data records stored in the databases. The knowledge base is a collection of knowledge related to projects and their risks, for example, successful risk mitigation strategies for specific risks and how they were implemented. With a knowledge base in place, inferences can be drawn based on the facts and relationships contained in the data knowledge base.

The function of the database management system is to create and update the data and knowledge base and to support activities such as navigation among records and report generation.

The query facility's function is to access, manipulate and query the data and knowledge.

6.2.2 Model management subsystem

This is a software package that includes quantitative models that provide the system's analytical capabilities and appropriate software management. A spreadsheet is an example of a modelling language that allows the user to write models to execute DSS analysis. Case-based reasoning, a data mining technique, is to be included in the software package. An advantage of using case-based reasoning is the fact that retrieving and adapting old cases is faster than deriving the answers from scratch.

6.2.3 User interface

The user communicates with and commands the DSS through this subsystem. It includes not only the hardware and software, but also factors that deal with ease of use, accessibility and human-machine interactions. The user interface component may use standard objects such as pull-down menus and buttons through a graphical user interface to make the DSS easy to use. It presents the summarised knowledge for the user.

6.2.4 User

The person faced with performing the risk analysis and making decisions that the DSS is designed to support is called the user or the decision maker (the project risk manager or someone on the PRM team).

Figure 24 illustrates the composition of the proposed DSS.

6.3 Classification of the proposed DSS

The DSS can be classified as a "model driven" support system. The value of such a system is largely in the quality of its model. Its analysis capabilities are based on a strong theory or model, coupled with a good user interface to make the model easy to use.

The integration of a DSS and a data mining technique transforms the DSS into an intelligent DSS. The data mining technique, case-based reasoning, can be used to intelligently profile a project according to its risks and suggest ways to reduce or eliminate the risks. The system contains historical data about projects and their associated risks. Data of a new project can be entered into the system and through the case-based reasoning technique, the model compares and profiles the project against historical data and develops graphs of variables like an expected risk of the project and how this risk varies in response to certain defined parameters.

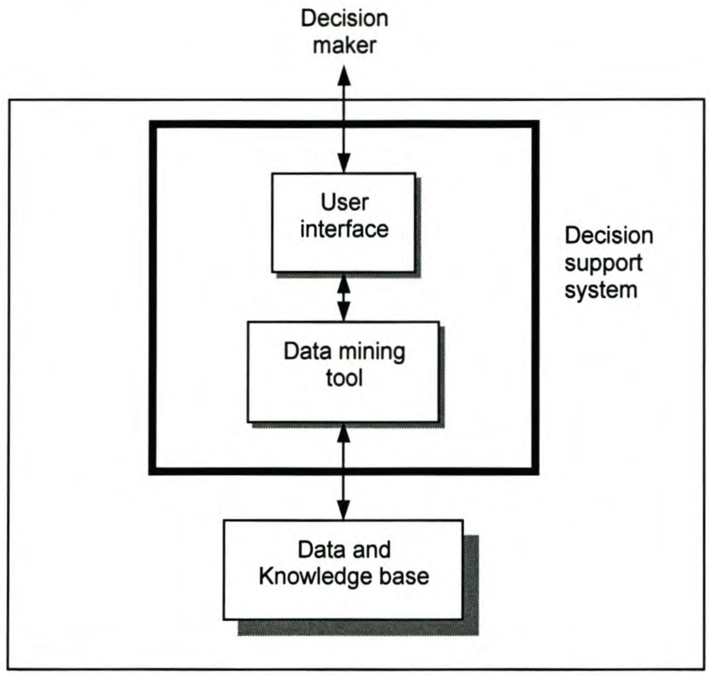

Figure 24: Composition of the proposed project risk management DSS

6.4 More on case-based reasoning and the proposed DSS

Case-based reasoning is a sub-field of Artificial Intelligence that is based on the idea that past problem solving experiences can be re-used, learned from and applied to solving new problems. The intelligent component of the system contains a history of past problems and the (successful) solutions applied. Future problems can then be considered through analogy with these past cases to hone in rapidly on the most promising type of solutions. A further step is incorporating machine learning through the updating of the "case base" with those cases for which the suggested solutions proved successful.

One of the benefits of cased-based reasoning is the fact that retrieving and adapting old cases is a lot faster than deriving new solutions from scratch. It encourages the decision maker to make good use of past experiences. By looking at previously solved problems, one avoids having to encode rules describing how the details of each problem affect the proposed solution.

Another benefit of CBR is that it "learns" over a period of time. After a new problem is solved, the case base is changed by the addition of the new case. By adding new cases, the system retains more and more knowledge along with problem-solving augmentation and achieves learning.

6.5 Benefits of the proposed DSS

The proposed DSS can assist managerial decision making by presenting information and interpretations for various alternatives. Such a system can help the decision maker to make more effective and efficient decisions. Decisions made with such a system not only helps to reduce the impact of time and money of risks on projects, but also impact of time, money and effort in making these decisions.

The insight created of the project and its potential problems as well as providing understanding about management responses to risk helps the risk manager to capitalise on, estimate and organise risks to improve the performance of the project.

The DSS allows documentation and the generation of reports. This has its own advantages:

- It clarifies thinking.
- Aids communication.
- Can be used to brief new staff.
- Captures expertise for future projects.
- Provides an audit trail if decisions are proven unsuccessful.

Sharing and documenting, what does and does not work to control risks across multiple projects, helps project managers avoid repeating the mistakes of the past. Members of the organisation can pool their experience and identify opportunities to control common risks through education, process improvement and application of improved management techniques.

With a knowledge base and the ability to draw inferences from it, the computer (as an element of the decision support system) running an AI application can be put to practical use as a problem solver and decision maker. By searching the knowledge base for relevant facts and relationships, the computer can aid the project risk manager by presenting one or more alternative solutions to the case in question. The DSS knowledge base and inferencing capability augment those of the user. This concept is illustrated in Figure 25.

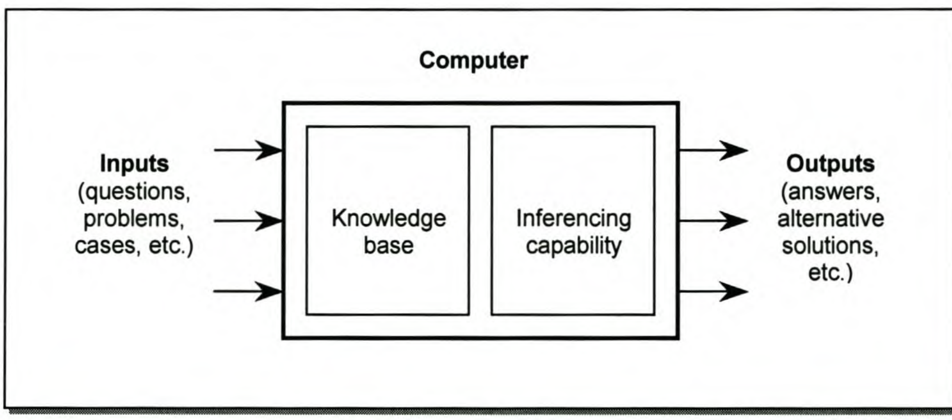

Figure 25: Applying AI concepts with a computer to form a DSS

Source: Aronson and Turban, 1998

6.5.1 Benefits to project risk management

There are many benefits of the proposed DSS in implementing risk management procedures. Some of these include:

- More effective strategic planning.
- Better cost control.
- Better schedule control.
- Enhancing shareholder value by minimising losses and maximising opportunities.
- Increased knowledge and understanding of exposure to risk.
- A systematic, well-informed and thorough method of decision making.
- Increased preparedness for outside review.
- Minimised disruptions.

6.6 Chapter summary

The architecture for a DSS for risk analysis and management is proposed in this chapter. A model driven decision support system, consisting of a data and knowledge base integrated with a data mining technique, can facilitate the project risk manager in assessing and managing risk. The database contains data pertaining to the organisation, past projects and the risks within the projects. The knowledge base contains knowledge of risk mitigation and management strategies that can be applied to risks under varying circumstances.

By using case-based reasoning, a data mining technique, the decision support system is transformed into an intelligent decision support system that profiles a project according to its risks and suggest ways to reduce or eliminate the risks.

7. Application

This chapter describes an application that would form part of the decision support system described in chapter 6.

Appendix A demonstrates how the system can support the project risk manager during the risk analysis phase in profiling a new project according to the impact its risks can have on the cost and time of the project and ultimately the performance of the project.

7.1 Description of the system

The application was developed and implemented in Microsoft Access. The model is simple, computing information like averages and totals. Records of cases containing data on previous projects and their risks form a case base. The case-based reasoning approach is used to match previous cases that are similar to the problem at hand. The result of the matching is cases that fit the description of the problem.

7.1.1 Primary function of system

The system is used as a tool when the user or decision maker requires information of how the risks in a project can influence the project in terms of time and cost. The risks that could occur in the project must first be identified. This tool can be used as part of the risk assessment phase.

7.1.1.1 How it works

The user enters data that describes the project in question in terms of its:

- Estimated cost
- Estimated duration
- Risks identified

These specifications can be represented as a probe that searches the case base for cases that match the attributes. In CBR, for probing purposes, all attributes are created equally.

7.1.1.2 Output

Once the probe finds similar cases, the cases are retrieved from the case base and presented in a report. The report gives the details of the cases and their risks. A summary is given of the expected impact the risks identified will have on the projects in terms of cost and time.

This process is illustrated in Figure 26.

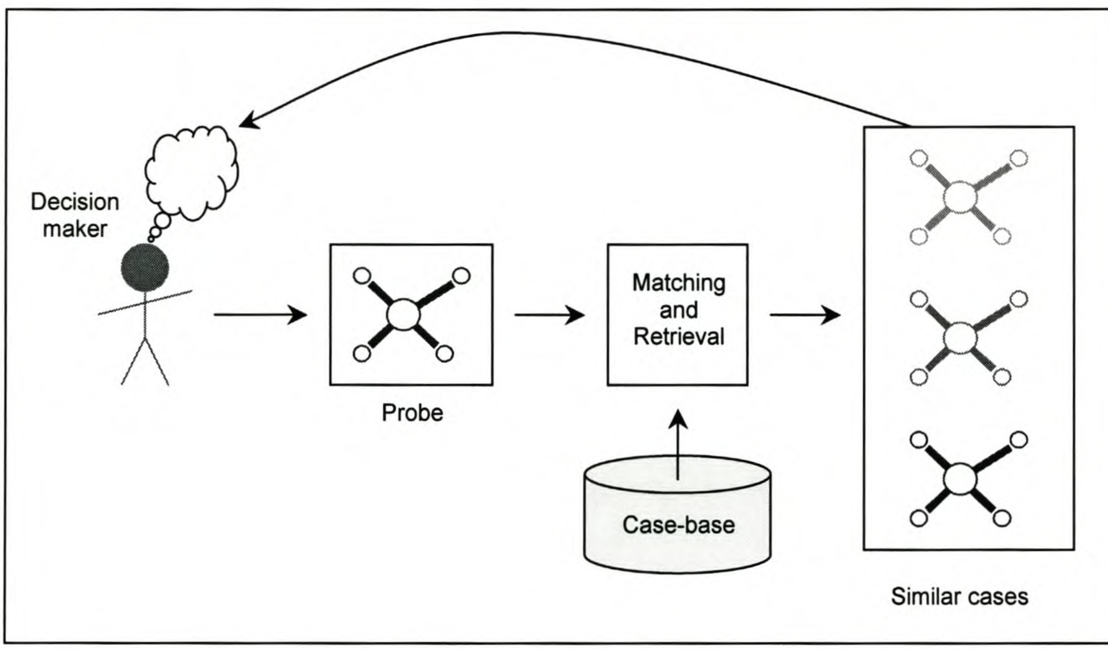

Figure 26: A probe is used to find similar cases in a case base

7.1.2 Other functionalities

Apart from being able to search for matching cases the system also allows the user to enter new cases and scroll through the case base in search for specific projects.

7.2 Learning capability

Learning is the last step in this system. After successful completion of a project, the case base is changed by the addition of the new case. The system retains more and more data from which information and knowledge can be derived and achieves learning.

An example of the model is shown in Appendix A.

7.3 Outcome of tests

The model was tested on cases with the required project data and was found to be useful in giving the decision maker a good estimate on the impact, in terms of cost and time, the identified project risks could have on the project.

The accuracy of the model will increase as the database is populated with more cases.

The model was found to be easy to navigate through and user friendly in entering and retrieving data.

8. Conclusion and recommendations

Different management methodologies i.e. project management, project risk management, data mining and decision support systems were studied in an attempt to present a new methodology for improving the management of risks in projects and improving the overall performance of a project.

A structure for an intelligent decision support system is presented. The implementation of such an intelligent decision support system will support project managers and assist project risk managers to improve the management of and reduce risk within a project.

An application was developed and modelled that demonstrates how the outcomes of past projects can be used to give the decision maker a better understanding of the impact identified risks can have on a project. On testing the model, it was found to be satisfactory and can be used to improve the overall performance of a project.

For further study it is suggested that the development of the intelligent DSS be completed and implemented.

9. References

1. "A Guide to the Project Management Body of Knowledge", Project Management Institute, Four Campus Boulevard, Newton Square, PA 19073–3299 USA, 1996.

2. Aronson, Jaye and Turban Efraim, "Decision Support Systems and Intelligent Systems", Fifth Edition, Prentice Hall, Upper Saddle River, NJ 07458, 1998.

3. Berry, Michael J. A., and Linoff, Gordon, "Mastering Data Mining: The Art and Science of Customer Relationship Management", Wiley, New York, N. Y., 2000.

4. Burke, Rory, "Project Management Planning & Control Techniques", Third Edition, John Wiley & Sons Ltd, 1999.

5. Cabena, Hadjinian, Stadler, Verhees, Zanasi, "Discovering Data Mining from Concept to Implementation", Prentice Hall, Upper Saddle River, NJ 07458, 1998.

6. Chapman, Chris and Ward, Stephen, "Project Risk Management - Processes, Techniques and Insights", John Wiley & Sons Ltd, 1997.

7. Dhar, Vasant and Stein, Roger, "Intelligent Decision Support Methods - The Science of Knowledge Work", Prentice Hall, Upper Saddle River, NJ 07458, 2000.

8. Grey, Stephen, "Practical Risk Assessment for Project Management", John Wiley & Sons Ltd, 1995.

9. Hullet, David T., "Key Characteristics of a mature Risk Management Process", International Institute for Learning, Inc., New York, NY USA, 2001.

10. Kerzner, Harold, "Project Management – A Systems Approach to Planning, Scheduling, and Controlling", Seventh Edition, John Wiley & Sons Ltd, 2001.

11. Mantel, Samuel J. JR. and Meredith, Jack R., "Project Management – A Managerial Approach", University of Cincinnati, John Wiley & Sons, Inc., New York , USA, 2000.

12. "RAMP—Risk Analysis and Management for Projects", Institute of Civil Engineers and the Faculty and Institute of Actuaries, Thomas Telford Ltd, 1 Heron Quay, London E14 4JD.

13. Rosenau, Milton, D., "Successful Project Management", Third Edition, John Wiley & Sons Ltd, 1998.

14. Salvendy, Gavriel (Editor), "Handbook of Industrial Engineering", Second Edition, John Wiley & Sons Ltd, 1995.

15. Thuraisingham, Bhavani M., "Data Mining: Technologies, Techniques, Tools and Trends", CRC Press, Boca Raton, London, New York, Washington, DC, 1999.

16. Thompson, P.A. and Perry, J.G (Editors), "Engineering Construction Risks: A guide to Project Risk Analysis and Assessment Implications for Project Clients and Project Managers", Centre for Research in the Management of Projects, University of Manchester Institute of Science and Technology, Thomas Telford Ltd, 1 Heron Quay, London E14 4JD.

17. Turner, Rodney, "The Commercial Project Manager", McGraw-Hill Book Company Europe, 1995.

Internet references:

18. Dilly, Ruth, "Data Mining – An Introduction", Parallel Computer Centre, Queens University Belfast, Version 2, December 1995, http://www.pcc.qub.ac.uk/tec/courses/datamining/stu_notes/dm_book_1.html

19. Tusler, Robert, "Project Risk Management Principles", Coblands Consulting, September 1998, http://www.netcomuk.co.uk/~rtusler/project/principl.html

20. Wiegers, Karl E., "Know Your Enemy: Software Risk Management", Software Development, vol. 6, no. 10, October 1998. Process Impact, http://www.processimpact.com//articles/risk_mgmt.html

Appendix A Demonstration of application

This Appendix assists the user through the various functions of the model and how to initiate them. An example is given that assists the user through the process of searching for cases that match a probe.

Functions of the model

The model is a Microsoft Access (Office 2000 version) application. Open the model by selecting the Microsoft Access file; Risk Profiler, found on the CD that accompanies this document. On opening this file, the "Switchboard" form opens to present the user with various options.

The Switchboard

The "Switchboard" form is illustrated in Figure 27.

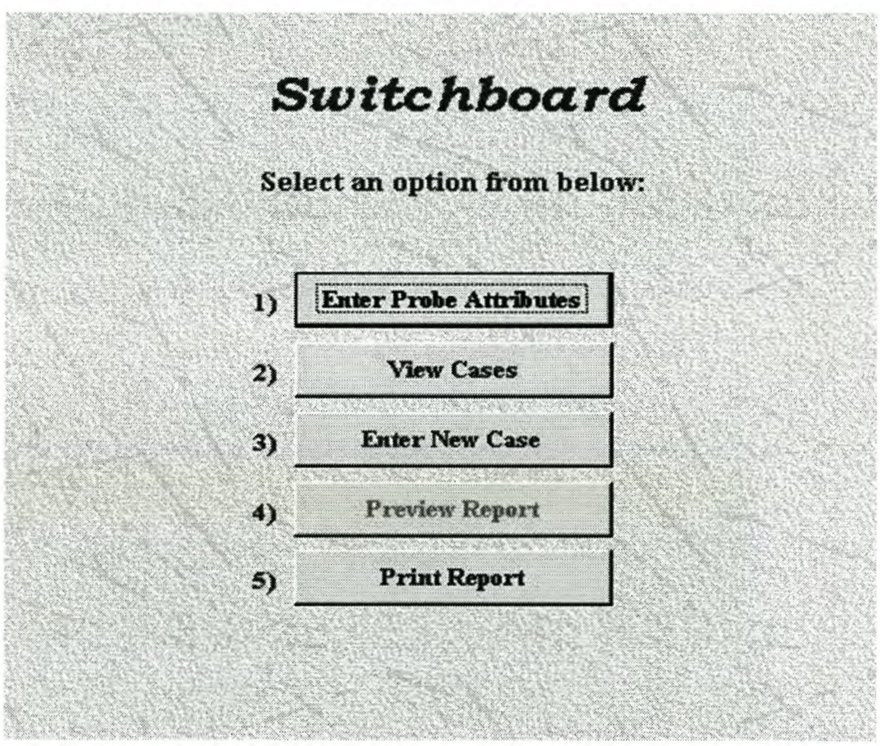

Figure 27: Switchboard form

Links to the various functions are presented on the "Switchboard" form. The links to "Preview Report" and "Print Report" should only be activated after a probe has been run. The "Print Report" button prints the report to the default printer.

Entering the probe attributes

Figure 28 illustrates the form for entering the probe attributes.

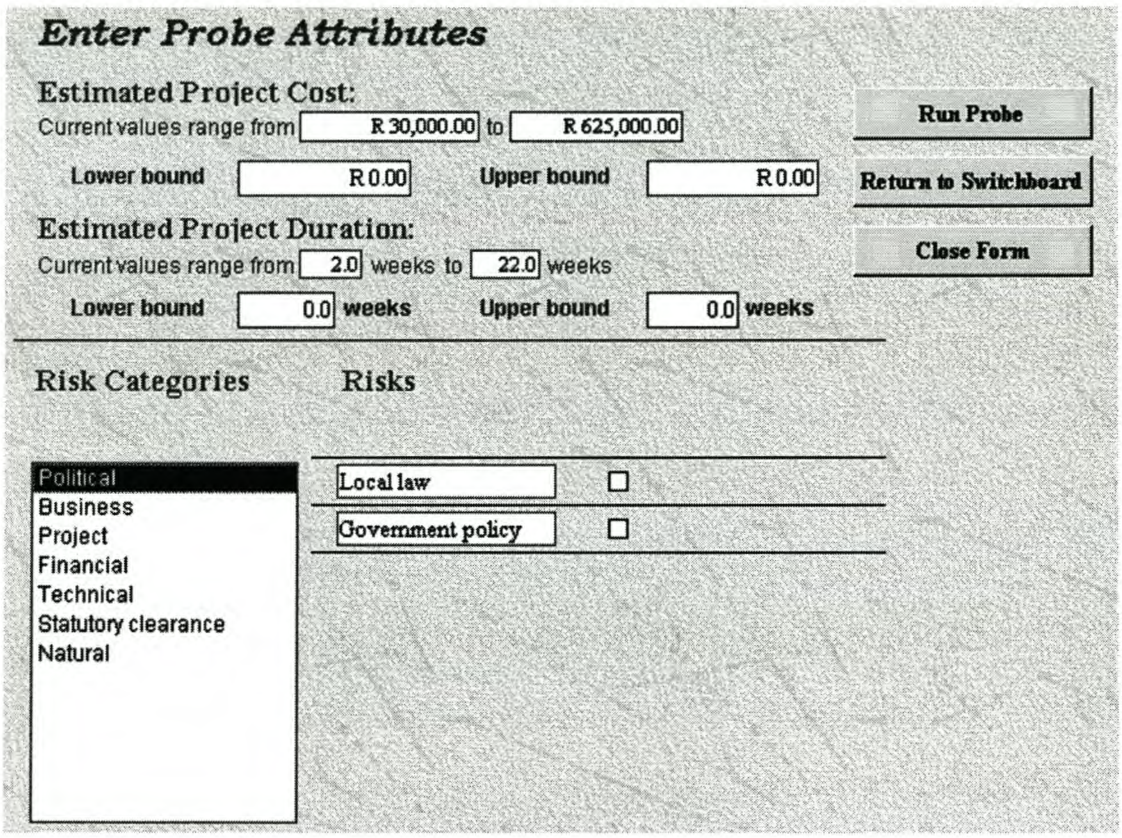

Figure 28: Probe attributes form

On this form the user enters the probe attributes. As shown on the form the attributes are:

- Estimated project cost
- Estimated project duration
- Risks

The user is required to enter a range for the estimated project cost and duration. This is done by entering values into lower and upper bound fields. For guidance, the lowest and highest values contained in the database for project cost and duration are provided.

A list of the risk categories is given to make the selection of the risks simpler. The user simply clicks on the appropriate risk category and the associated risks are listed. To select a risk, the user must check the box that appears next to the risk.

The "Run Probe" button activates the probe and a report that contains the matching cases is automatically compiled.

Finding the matching cases

A series of queries are run to extract from the database matching projects or cases. The following criteria must be fulfilled for the cases to be a match:

- The cases 'project cost' and 'project duration' must fall within the range as specified in the lower and upper bound fields on the 'Enter Probe Attributes' form.

AND

- The cases must have an exact match with the risks identified on the 'Enter Probe Attributes' form.

Viewing cases

Figure 29 illustrates the "Project Information" form that allows the user to scroll through the case base.

This form presents the project information as stored in the case base. A summary of the estimated cost and time, actual cost and time and estimated deviation of cost and time as a result of the risks occurring, is also given. The actual cost is calculated by summing up the actual costs of each risk. The estimated cost and estimated time is calculated by summing up the product of the impact with its corresponding probability of occurrence for each risk. The estimated deviation is the difference between the actual and estimated values. A record selector allows the user to scroll through the case base.

Calculations for the above-mentioned values are shown below:

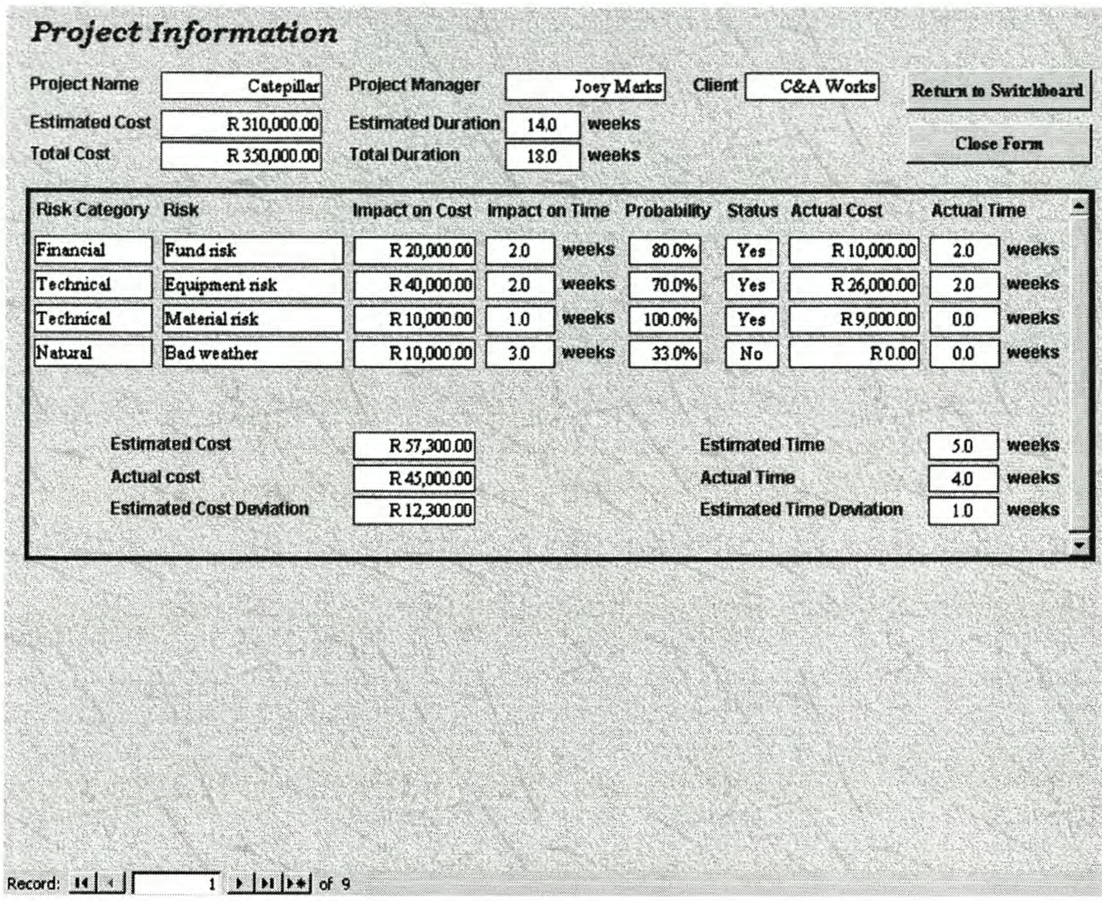

Figure 29: Project information form

Risk	Impact on Cost		Probability		Estimated Cost
Fund risk	R 20,000.00	X	80.0%	=	R 16,000.00
Equipment Risk	R 40,000.00	X	70.0%	=	R 28,000.00
Material Risk	R 10,000.00	X	100.0%	=	R 10,000.00
Bad weather	R 10,000.00	X	33.0%	=	R 3,300.00
					R 57,300.00

Risk	Impact on Time		Probability		Estimated Time
Fund risk	2.0 weeks	X	80.0%	=	1.6 weeks
Equipment Risk	2.0 weeks	X	70.0%	=	1.4 weeks
Material Risk	1.0 weeks	X	100.0%	=	1.0 weeks
Bad weather	3.0 weeks	X	33.0%	=	0.99 weeks
					4.99 weeks ≈ **5 weeks**

Risk	Actual Cost
Fund risk	R 10,000.00
Equipment Risk	R 26,000.00
Material Risk	R 9,000.00
Bad weather	R 0.00
	R 45,000.00

Risk	Actual Time
Fund risk	2.0 weeks
Equipment Risk	2.0 weeks
Material Risk	0.0 weeks
Bad weather	0.0 weeks
	4 weeks

Estimated cost		**Actual cost**		**|Estimated cost deviation|**
R 57,000.00	—	R 45,000.00	⇒	R 12,300.00
Estimated time		**Actual cost**		**|Estimated time deviation|**
5 weeks	—	4 weeks	⇒	1 week

Entering a new case

Figure 30 illustrates the "Enter Project Data" form.

The "Enter Project Data" form allows the user to enter a new case (a successfully completed project) into the case base. A drop-down box contains a list of risks to choose from. After entering the data, the form can be closed or one can return to the switchboard.

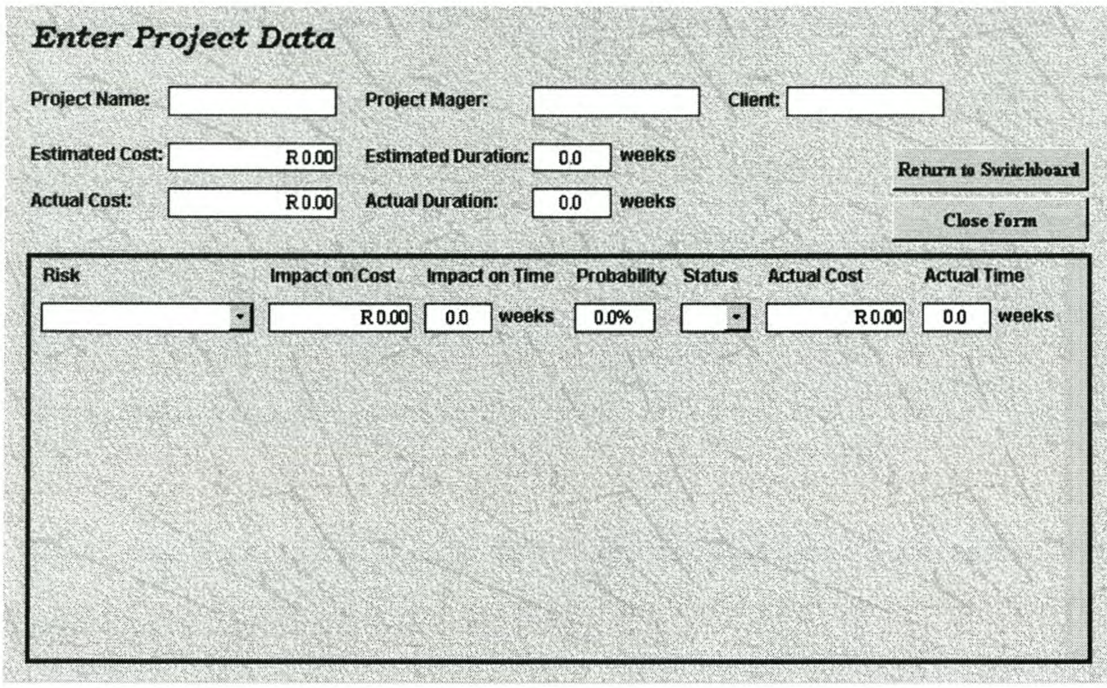

Figure 30: Form for entering a new case (project)

Example

In this example it is shown how the system profiles a new project according to the impact its risks may have in terms of cost and time.

The project in this example is fictitious and is referred to as "Project X".

Probe attributes

In the project planning phase it was estimated that the total project cost would be R300,000.00 and would take an approximate 14 weeks to complete.

During the risk identification phase it was realised that the following risks could influence the outcome of Project X:

- Financial:
    - Fund risk
- Technical:
    - Equipment risk
    - Material risk
- Natural
    - Bad weather

The following attributes concerning the estimated cost and duration of the project were chosen in the search for similar cases in the case base:

- Estimated cost:
    - Lower bound = R280,000.00
    - Upper bound = R320,000.00
- Estimated duration:
    - Lower bound = 12 weeks
    - Upper bound = 16 weeks

Figure 31 illustrates the attributes entered in on the "Enter Probe Attributes" form.

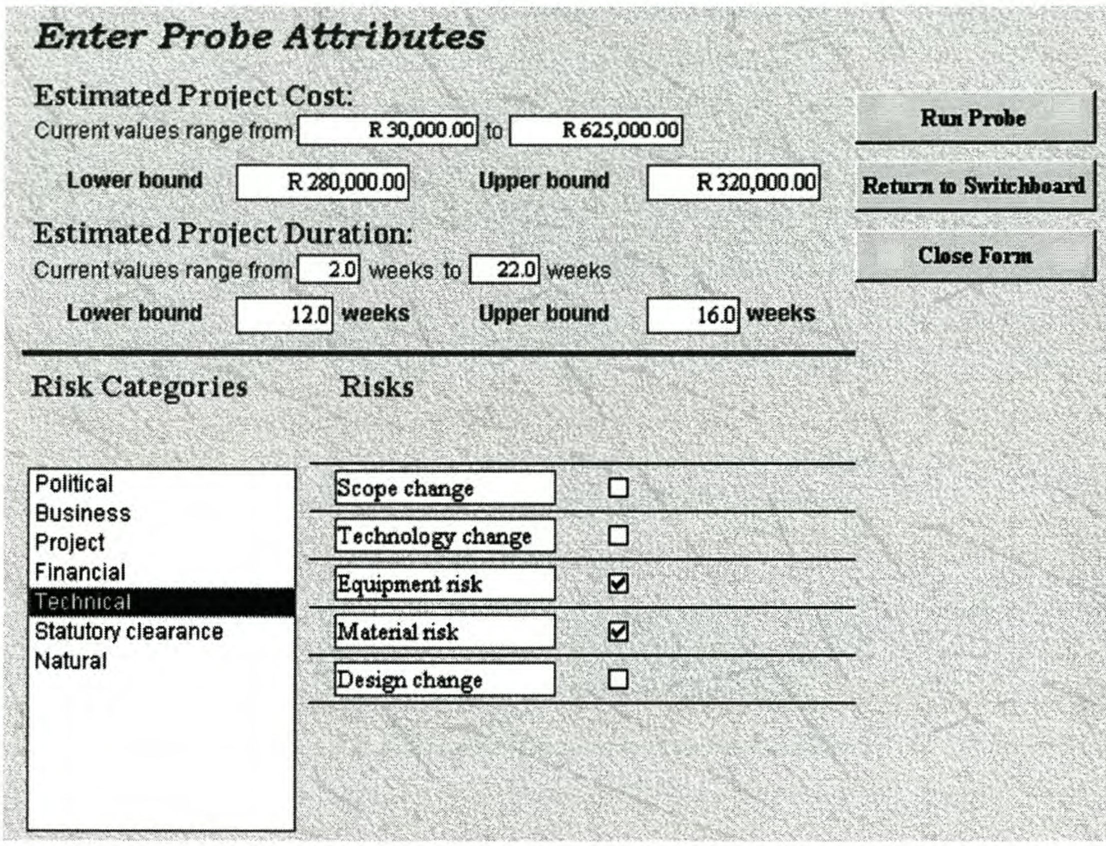

Figure 31: Probe attributes are entered

Report summary

The report compiled containing the matching cases is illustrated in Figure 32 through to Figure 33.

The report indicates that 2 cases from the case base matched the probe viz. projects Caterpillar and MalCo. The report gives the details of the projects and their risks. The risks listed were identified in the risk analysis phases of the particular project. Their impact, in terms of cost, time and probability of occurrence were also determined. The status of the risk indicates whether the risk did or did not occur. The risks' actual impact in terms of cost and time is also given. A summary of the estimated cost and time, actual cost and time and estimated deviation of cost and time as a result of the risks, is included.

Matching Cases

Project Name	Catepillar	**Project Manager**	Joey Marks	**Client**	C&A Works
Estimated Cost	R 310,000.00	**Estimated Duration**	14.0 **weeks**		
Total Cost	R 350,000.00	**Total Duration**	18.0 **weeks**		

Risk	Impact on Cost	Impact on Time	Prob.	Status	Actual Cost	Actual Time
Fund risk	R 20,000.00	2.0 weeks	80.0%	Yes	R 10,000.00	2.0 weeks
Equipment risk	R 40,000.00	2.0 weeks	70.0%	Yes	R 26,000.00	2.0 weeks
Material risk	R 10,000.00	1.0 weeks	100.0%	Yes	R 9,000.00	0.0 weeks
Bad weather	R 10,000.00	3.0 weeks	33.0%	No	R 0.00	0.0 weeks

Estimated Cost	**R 57,300.00**	**Estimated Time**	**5.0 weeks**
Actual Cost	**R 45,000.00**	**Actual Time**	**4.0 weeks**
Estimated Cost Deviation	**R 12,300.00**	**Estimated Time Deviation**	**1.0 weeks**

Project Name	MalCo	**Project Manager**	Joey Marks	**Client**	CT Municilality
Estimated Cost	R 285,000.00	**Estimated Duration**	12.0 **weeks**		
Total Cost	R 320,000.00	**Total Duration**	15.0 **weeks**		

Risk	Impact on Cost	Impact on Time	Prob.	Status	Actual Cost	Actual Time
Fund risk	R 7,000.00	2.0 weeks	50.0%	No	R 0.00	0.0 weeks
Equipment risk	R 20,000.00	3.0 weeks	80.0%	Yes	R 23,000.00	2.0 weeks
Material risk	R 9,000.00	2.0 weeks	40.0%	No	R 0.00	0.0 weeks
Bad weather	R 5,000.00	0.5 weeks	50.0%	Yes	R 6,000.00	0.5 weeks

Estimated Cost	**R 25,600.00**	**Estimated Time**	**4.5 weeks**
Actual Cost	**R 29,000.00**	**Actual Time**	**2.5 weeks**
Estimated Cost Deviation	**R 3,400.00**	**Estimated Time Deviation**	**2.0 weeks**

Figure 32: Report showing retrieved cases

Summary of impact identified risks could have on project:

	Estimated Cost	Actual Cost	Estimated Cost Deviation
Ave	R 41,450.00	R 37,000.00	R 7,850.00
STDEVP	R 15,850.00	R 8,000.00	R 4,450.00
% Err of Mean	38.2%	21.6%	56.7%

	Estimated Time	Actual Time	Estimated Time Deviation
Ave	4.7 weeks	3.3 weeks	1.5 weeks
STDEVP	0.27 weeks	0.75 weeks	0.48 weeks
% Err of Mean	5.7%	23.1%	32.7%

Figure 33: Report showing retrieved cases continued

The summary of the report gives the averages, standard deviation and percentage error of mean for the estimated cost and time, actual cost and time and estimated deviation of the matching cases. This information gives the decision maker an idea of the impact the risks can have on the new project.

'Ave' Returns the average (arithmetic mean) of the arguments. For example:

Estimated cost

Ave = R 57,300.00 + R 25,600.00 / 2 = R 41,450.00

'STDEVP' is the standard deviation based on the entire population given as arguments. The standard deviation is a measure of how widely the values are dispersed from the average value (the mean). The standard deviation is calculated using the "biased" or "n" method. The following formula was used to calculate the STDEVP:

$$s^2 = \frac{\sum x^2 - n\bar{x}^2}{n}$$

For example:

Estimated cost

$$s^2 = \frac{\left(57300^2 + 25600^2\right) - 2(41450)^2}{2} = 15850$$

STDEVP = R 15,850.00

'% Err of Mean' (percentage error of mean) also called the coefficient of variation is a measure of how much the average represents the set of values it came from. The formula is:

% Err of Mean = STDEVP / Ave

For Example:

Estimated cost

% Err of Mean R 15,850.00 / R 41,450.00 = 38.2%

Appendix B Design of application

This Appendix defines the various entities and their relationships in the design of the model applied in Microsoft Access. Definitions of the attributes of the entities are also given.

Data Model (EERD)

Any object or event about which data is collected is an entity. An entity relationship model is used to set out system boundaries. The elements that make up a system, in this case the model the application is based on, can be referred to as the entities.The following entities were identified for the model:

- Projects
- Projects_Risks (an associative entity to link the 'Projects' and 'Risks' entities)
- Risks
- Risk Categories

The relationships between the entities in the extended entity relationship diagram (EERD) and more descriptive Microsoft Access relationship diagram are depicted in Figure 34 and Figure 35 respectively.

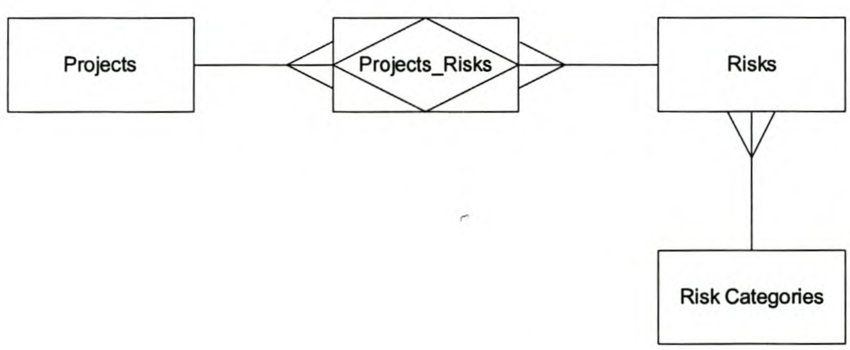

Figure 34: EERD of application

As depicted by the 'one to many' relationships, a project can have many risks and a risk can be assigned to more than one project. A risk only falls under one category while a category may have many risks defined.

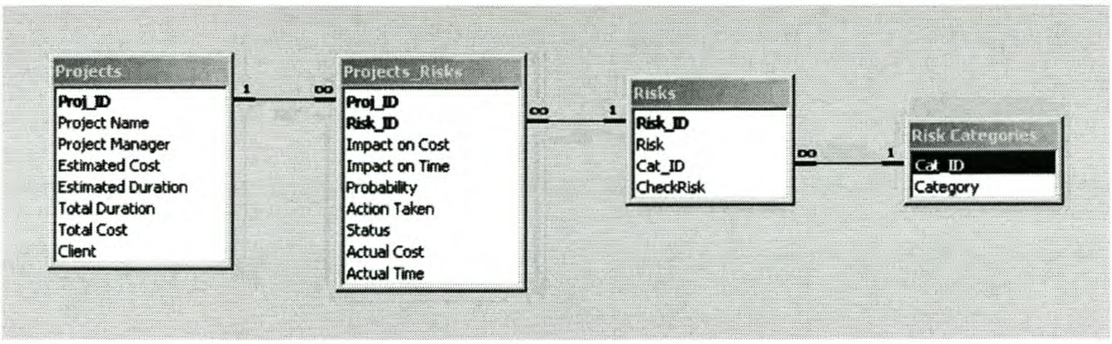

Figure 35: EERD in Microsoft Access

Attribute definitions

An attribute or data item is a characteristic of an entity and determines the contents of data stored in an entity. Typical values assigned to data items may be numbers, alphabetic characters, special characters or combinations of all three. A short description of each attribute used in the model is given below:

Projects

Proj_ID	:	This attribute is used to establish distinction between different projects with the use of a number/integer which is specific for each project.
Project Name	:	Stores the name of a specific project.
Project Manager	:	Stores the name of the project manager who is incharge of the project.
Estimated Cost	:	Stores the estimated cost (in Rand) of the project as caculated during the scoping or design stage.
Estimated Duration	:	Stores the estimated duration (in weeks) of the project as caculated during the scoping or design stage.
Total Cost	:	Stores the actual project costs (in Rand) incured.
Total Duration	:	Stores the actual duration (in weeks) of the project.
Client	:	Stores the name of the client.

Projects_Risks

Proj_ID	:	This attribute is used to establish distinction between different projects with the use of a number/integer which is specific for each project.
Risk_ID	:	This attribute is used to establish distinction between different risks with the use of a number/integer which is specific for each type of risk.
Impact on Cost	:	Stores the estimated value (in Rand) of the financial impact the risk will have on the project if the risk occurs.
Impact on Time	:	Stores the estimated time (in weeks) the project will lose if the risk occurs.
Probability	:	Stores the value (%) of the probability that the risk will occur.
Status	:	Used to indicate whether or not the risk took place.
Actual Cost	:	Stores the value (in Rand) of the financial impact the occurrence of the risk had on the project.
Actual Time	:	Stores the time (in weeks) lost to the project due to the occurrence of the risk.

Risks

Risk_ID	:	This attribute is used to establish distinction between different risks with the use of a number/integer which is specific for each type of risk.
Risk	:	Store the name of the risk.
Cat_ID	:	This attribute is used to establish distinction between the different risk categories with the use of a number/integer which is specific for each type of risk.
Check Risk	:	Used to indicate whether or not a risk has been selected.

Risk Categories

Cat_ID	:	This attribute is used to establish distinction between the different risk categories with the use of a number/integer, which is specific for each type of risk.
Category	:	Stores the name of the risk category.