Data Quality Assurance for Strategic Decision Making in Abu Dhabi’s Public Organisations

Omar Alketbi

DATA QUALITY ASSURANCE FOR STRATEGIC DECISION MAKING IN ABU DHABI'S PUBLIC ORGANISATIONS

By

Omar Alketbi

Abstract

Data quality is an important aspect of an organisation's strategies for supporting decision makers in reaching the best decisions possible and consequently attaining the organisation's objectives. In the case of public organisations, decisions ultimately concern the public and hence further diligence is required to make sure that these decisions do, for instance, preserve economic resources, maintain public health, and provide national security. The decision making process requires a wealth of information in order to achieve efficient results. Public organisations typically acquire great amounts of data generated by public services. However, the vast amount of data stored in public organisations' databases may be one of the main reasons for inefficient decisions made by public organisations. Processing vast amounts of data and extracting accurate information are not easy tasks. Although technology helps in this respect, for example, the use of decision support systems, it is not sufficient for improving decisions to a significant level of assurance. The research proposed using data mining to improve results obtained by decision support systems. However, more considerations are needed than the mere technological aspects. The research argues that a complete data quality framework is needed in order to improve data quality and consequently the decision making process in public organisations. A series of surveys conducted in seven public organisations in Abu Dhabi Emirate of the United Arab Emirates contributed to the design of a data quality framework. The framework comprises elements found necessary to attain the quality of data reaching decision makers. The framework comprises seven elements ranging from technical to human-based found important to attain data quality in public organisations taking Abu Dhabi public organisations as the case. The interaction and integration of these elements contributes to the quality of data reaching decision makers and hence to the efficiency of decisions made by public organisations. The framework suggests that public organisations may need to adopt a methodological basis to support the decision making process. This includes more training courses and supportive bodies of the organisational units, such as decision support centres, information security and strategic management. The framework also underscores the importance of acknowledging human and cultural factors involved in the decision making process. Such factors have implications for how training and raising awareness are implemented to lead to effective methods of system development.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1. Introduction

Retaining data is no longer a primary concern of an organisation as it is common that a numerous number of transactions are regularly stored in data warehouses. Only a few of these transactions hold real relevance to the organisation's decision making process. The amounts of data stored in databases increase daily and go beyond the technical skills and human capacity to interpret that into valuable information. Database management systems have advanced at a faster rate than the techniques used for extracting and utilising information to be used in making decisions (Power, 2007), in the strategic sense of using the trend of the past endeavours to anticipate the future tendency (Lv & Li, 2009). Obtaining, storing and managing information in larger organisations are now ordinary business operations and are usually performed automatically by electronic data repositories (Saxena & Rajpoot, 2009).

Decision support systems (DSSs) are one of the techniques used in the process of making decisions. Decision support systems are widely used in organisations around the world, including some public organisations, for the main aim of helping executives to make more accurate decisions based on advanced levels of data refined and presented to them. A decision support system refers to a class of computer-based systems that help in the process of decision making (Hardin & Chhieng, 2007). Padhy et al. (2012) argue that the value of strategic information systems is easily recognised yet efficiency and speed are not the only factors of competitiveness. The vast amounts of data have called for new methods to analyse and understand them. Conclusions and inferences from these data need special tools and techniques that are able to delve deeper than traditional decision support systems can. Public administrations face challenges of using the correct analysis of huge amounts of data. These data are used for producing statistical analyses and forecasts on economic, social, health and education issues, which are very important for government planning in aspects where such as development of interest rates and inflation, economic growth, education standards, household income, crime trends, and climate change are major inputs. The ability to utilise DSS in public organisations usually collides with the difficulty of attaining refined and quality data that such systems need. Furthermore, DSS systems incur high costs to maintain and operate due to a large number of variables involved in the process whereby educated decisions need as many variables as can be included.

Artificial intelligence systems such as neural networks, data miming, and fuzzy inference algorithms are common techniques used to extract knowledge and hence support decision makers. Data mining currently receives a high level of interest in different disciplines. This interest is based on the proven ability of data mining techniques to contribute to knowledge discovery and consequently make better use of the stored data (Han et al., 2006). However, a significant amount of the organisation's information may be in a textual format described in natural languages or does not have a structure like the one present in data tables and structured relational databases. This type of information, found mainly in the form of paper-based and electronic documents and emails, particularly in public organisations, cannot be used with traditional data mining tools, and thus minimises their potentials use with data mining tools. It is common that data miners prepare the data for mining. This operation, called pre-processing, is complex and may take a long time to complete, depending on the size of the project, and requires a significant part of the organisational resources. It is usually performed only by large enterprises (Calderon et al., 2003). Only a few examples of using data mining to support management decisions are found in the literature. Mladenic et al. (2003) maintain that until now there has been no systematic attempt to integrate data mining in the decision support process. Reasons behind that are many but mainly include the nature of data mining processes that combine computer science and statistics, which create some confusion about what implementation aspects may be suitable for managerial decisions.

Abu Dhabi Emirate is undergoing a complete overhaul of public organisational systems. The overhaul aims at bringing together a range of processes, including the human side as well as software and hardware deployment, in order to improve the results and quality of the undertaken decisions. For that, the Emirate has already established a quality assessment and control department. However, results have not been as expected.

Seven public organisations in Abu Dhabi Emirate were chosen as the study's focus for the aim of providing a data quality framework for the Abu Dhabi's public sector. The study comprised interviews conducted with senior managers at the organisations in Abu Dhabi Emirate. The organisations are: the Abu Dhabi Police Organisation

(ADPO), Al Ain Hospital, the Department of Economic Development of Abu Dhabi, the General Authority of Youth and Sports Welfare, Zayed Foundation, Tawam Hospital, and the Department of Municipal Affairs. Two senior managers from each organisation were interviewed. Decision making of these organisations is supported by quantitative and qualitative quality data sent by different departments and organisational bodies countrywide to support decision making. The decision process is a little complicated and comprises several factors. The organisations aim to realise the full and potential value of the data that they acquire. Hence, challenges associated with data formats, content, validity and reliability need to be addressed. Moreover, there is a social link too, which is an important aspect to consider regarding the special attributes of Arab culture, for example collectivism and high-power distance. Training and awareness are also some aspects of development. The choice of the organisations was based on several considerations but most importantly the involvement of these organisations in different federal bodies of the public sector in the country, such as law enforcement, traffic regulations, the healthcare sector, and civil defence, among others. Furthermore, public organisations in Abu Dhabi undergo continuous development and innovation to pursue best practices in digital information management. Therefore, the suggested framework may serve public organisations and agencies at different levels, including government bodies, hospitals, police forces, and others.

1.2. Problem Definition

Strategic decision making in organisations is a complex process that has many aspects and involves many variables. In the case of public organisations, the decision making process may be even more complex due to the bureaucratic and non-profit nature of these organisations. The complexity of the process often results in inefficient outcomes in terms of the decisions taken by decision makers in public organisations. Inefficient decisions in public organisations may lead to economic loss and social deficiencies, and may even be disastrous. Generally speaking, inefficient strategic decisions of public organisations result in waste of public resources and squandering. The decision making process requires a wealth of information in order to generate efficient results. In the same respect, public organisations typically acquire great amounts of data generated by public services. However, the vast amounts of data stored in public organisations' databases may be one of the main reasons for inefficient decisions made by public organisations. Processing vast amounts of data and extracting accurate information is not an easy task. Although technology helps in this respect, it is not sufficient for improving decisions to a significant level of assurance, as many experiences have shown. Accordingly, these organisations need "good" information to make good decisions. The goodness of the information is controversial and is subject to many, subjective and objective, considerations. Generally speaking, there are certain measures that can be adopted to assess the quality of the acquired information. The loose nature of definitions of standards and measures for the quality of information (or data), as well as the fact that most data quality measures are devised on an ad hoc basis (Pipino et al., 2002), renders those measures ever more individually determined. For example, data accuracy, reliability, timeliness, completeness and relevance, among others, are common dimensions in many considerations for data quality (Pipino et al., 2002).

1.3. Aim and Objectives

The research attempts to understand data quality issues in public organisations in Abu Dhabi Emirate and accordingly suggest appropriate ways for overcoming these issues up to improving data quality used for strategic decisions of these organisations.

For the above aim, a framework that incorporates information systems, data quality standards, human resources and other elements deemed necessary is proposed.

Using the framework, data that are highly utilised will eventually improve quality. A data quality framework will achieve consistent direction towards optimal decision making in an organisation. The improved data quality will then ensure that the organisation is more able to make informed and accurate decisions on policies and strategies.

The study will also provide special insights into data mining techniques as part of the framework solution to data quality issues. This includes investigating such techniques suitable for pre-processing data and hence reducing costs and improving data quality.

The objectives to be attained are the following:

- Indicate the areas of improvement for a better decision making process, with prioritisation decision tools;

- Improve the relationship between data providers and decision makers of and closely monitor them;
- Attain awareness of a data quality culture throughout the organisation;
- Educate and support data providers;
- Improve the understanding of the processes of communicating information;
- Set data quality requirements for different departments;
- Provide best practice guidelines;
- Devise a feedback system aimed at improving data quality and rectifying data quality errors.

The aim of the research can be broken down into the following questions that the study attempts to answer:

1. Is data quality an issue in public organisations in general and in Abu Dhabi's public organisations in particular? How is this issue manifested?
2. What are the main problems facing improving data quality in public organisations?
3. How can data quality contribute to better decision making?
4. Are there specific methods to be followed for improving data quality in Abu Dhabi's public organisations? If so, what are these methods? How can these methods assure better data quality?

1.4. Thesis Structure

The thesis is divided into nine chapters:

Chapter 1: Introduction

This chapter provides an introductory overview of the thesis, which includes an introduction and background, motivation, problem definition, aim and objectives, and the research question.

Chapter 2: Literature Review

The literature review explores the main themes of the research aim and objectives, namely, data quality, information systems, decision support systems and data mining, with a main focus on public organisations as opposed to private sector organisations. The chapter attempts to provide definitions as well as other relevant details such as uses and applications of the main concepts involved. The review helps acquire knowledge of the main themes of the thesis, which is used in later chapters.

Chapter 3: Methodology

This chapter describes the research methods used for collecting data used for devising the data quality framework for Abu Dhabi public organisations. It provides an insight into the choice of the research strategy and the advantages of using it over other ones. The chapter highlights the importance of the choice of the adopted research and provides details of how it was conducted.

Chapter 4: Data Mining for Data Quality

This chapter describes some of the main state-of-the-art data mining techniques and their possible uses for knowledge extraction and data quality improvement. The chapter also highlights the uses of data mining by public sector organisations in some countries.

Chapter 5: Data Mining based Experiments for Data Quality: Using Classifiers to Predict Missing Data Values

This chapter provides an implementation of data mining methods, namely classification, on datasets provided by one of the case study organisations, Abu Dhabi Police. The aim is to use classification to predict missing data values in the datasets. The experiments show the various success rates attained by using different classifiers.

Chapter 6: Integration of Decision Support Systems and Data Mining for Improving Data Quality

In this chapter, a data mining decision support integrated (DM-DS) system is suggested to improve the performance of DSSs by feeding them quality data. The chapter provides an insight into some DM-DS integrated systems found in the literature and uses the results of the investigation as a basis for the proposed system.

Chapter 7: Findings

This chapter details the findings obtained from the interviews conducted with senior managers at seven public organisations in Abu Dhabi Emirate.

Chapter 8: Discussion

This chapter provides a discussion of the findings obtained from the interviews. The outcome is a framework aiming at improving data quality for Abu Dhabi public organisations. The proposed framework is evaluated and compared with other existing data quality frameworks.

Chapter 9: Conclusion

This chapter provides a conclusion of the work conducted. It provides details about how the aim and objectives were satisfied by the completion of the work and highlights the limitations and aspects of future work. It has an allocated section on the work's contribution to knowledge.

Chapter 2: Literature Review

2.1. Introduction

This chapter explores the main themes of the research aim and objectives, namely: data quality, information systems, decision support systems and data mining, with a main focus of public organisations as opposed to private sector organisations. Data quality is a measure used by an organisation of the extent to which the data are consistent, comprehensible and relevant to particular situations. Modern information systems are complex systems that include several elements other than the data and the processing and storing machine. This includes the human element, the organisational structure, and the classification criteria of the stored information, among others (O'Brien & Marakas, 2010). Decision support systems are computer-based information systems designed to aid organisations regarding their decision-making activities and are usually used in the mid and upper levels of corporations to aid in management functions and planning (March & Hevner, 2007). Data mining currently receives a high level of interest in different disciplines. This interest is based on the proven ability of data mining techniques to contribute to knowledge discovery and consequently make better use of the stored data (Han et al., 2006). Therefore, those four concepts are of paramount importance to the research and comprise the core means of managing and improving data. Mainly, the chapter attempts to provide definitions for the main concepts involved, as well as other relevant details such as uses and applications. In essence, the chapter provides aspects of uses in a public sector setting, mainly in Abu Dhabi Emirate. The review helps acquire knowledge of the main themes of the thesis, which will be used in later chapters to experiment with data mining techniques on datasets provided by case study organisations and to suggest optimisation of decision support systems by data mining integration.

2.2. Data Quality: Issues, Processes and Importance

This section introduces the concept of data quality and discusses several aspects which relate to the concept. A number of issues and problems that affect the quality of data are discussed. Moreover, different ways in which the issues related to data quality as identified in the discussion can be resolved are presented. Particular emphasis is on how solving data-associated problems can help improve the overall quality of data. The relevant processes, tools and technologies used to improve the quality of data are presented. Lastly, the various ways in which quality data is important to the public sector are discussed. The knowledge gained from this section will help understanding of the concept of data quality and consequently, suggest methods of overcoming and anticipating data quality issues.

2.2.1. Data quality: inherent characteristics

Generally speaking, the quality of data that is used by an organisation can be understood in terms of the extent to which the data are consistent, comprehensible and relevant to particular situations (Singh & Singh, 2010). The extent to which data is of quality with regard to particular situations depends on the particular organisation that is using the data. For example, educational institutions have different requirements for data as compared to organisations operating in other sectors. It is only when data adheres to these needs that it can be said to be of high quality. Further, an organisation that uses high quality data stands to benefit a lot in terms of the quality of decisions made and their overall outcomes (Rhind, n.d.). This implies that the issue of data quality is of great importance to organisations.

2.2.1.1. Accuracy

There are several issues that are inherent to data quality. The first one is accuracy. For data to be of high quality, it must represent the actual values that it stands for (Watson et al., 2006). This implies that since data is supposed to represent actual reality, it must not fail to do so. There are several ways in which data may fail to be accurate, thus losing its overall quality. For example, when data lacks links that allow two or more separate systems to access and edit it, it usually fails to reflect reality (Watson et al., 2006). Furthermore, issues of accuracy of data may arise as a result of large-scale errors in the data itself. The result of this is that the overall quality of the data is compromised because of lack of accuracy in individual values.

2.2.1.2. Completeness

The second issue related to quality of data is its completeness. Generally, completeness of data can be understood in terms of whether or not the data represents all the available information and covers it. In order for data to be termed as complete, all the values must be available and should be in a usable state (Singh & Singh, 2010).

The importance of completeness in data cannot be overemphasised. According to Chapman (2005), organisations can guarantee the quality of their data by prioritising the use of small but complete sets of data over large amounts of data that are not complete. Therefore, completeness is an important aspect of data quality.

2.2.1.3. Consistency

The third aspect of data quality is consistency. For data to be of high quality, it has to be completely consistent. This means that all values that are similar should relate to the same kind of information. There are two types of consistency of data: structural and semantic (Chapman, 2005). For structural consistency, the entities, attributes and types used in the presentation of data should have a uniform format. This is usually achieved by ensuring that the database is well-designed and has good attributes throughout. On the other hand, data that is said to be semantically consistent is presented in such a manner that it is clear and completely unambiguous (Chapman, 2005).

2.2.1.4. Validity

The fourth aspect of quality of data is validity. Typically, data should be correct and completely reasonable for it to be regarded as of quality (Singh & Singh, 2010). This suggests that the data collected, processed and stored must be able to support the analysis that it is meant to facilitate.

2.2.1.5. Timeliness

Another issue that is inherent in data quality is the timeliness of the data. Generally, data that is of high quality is able to capture the actual and constant changes that happen in the real world. This is applicable to different scenarios. For instance, Watson et al. (2006) point out that flexibility of data used within the context of a learning institution should reflect the rapid changes that take place in the day-to-day operations within such an institution. Chapman (2005) argues that timeliness is required even in collecting and processing data about scientific phenomena. Therefore, the level to which data enjoys timeliness determines its overall quality.

The issues inherent in data quality can be illustrated as shown in Figure 1.

Figure 1: Issues inherent in data quality

2.2.2. Potential solutions to data quality issues

Several solutions have been recommended to tackle the issues of accuracy, completeness, consistency, validity and timeliness of data as detailed above. This is because the different types of errors that arise from these issues compromise the overall quality of the data. It is therefore necessary for organisations and other stakeholders to solve these issues as a way of avoiding the consequences of using data that is of poor quality.

First of all, it is important to note that solutions to the issues of data quality are meant to be applied to the entire process that data is subjected to between its reception and final use. This is due to the point that errors in data can occur at any of the stages of the data handling process, including reception, entry, integration, maintenance and extraction (Singh & Singh, 2010).Theoretically, organisations seek to eliminate these issues by ensuring that the correct processes and procedures are used during all stages of the data management process. The data management process that is commonly used by organisations is represented in Figure 2.

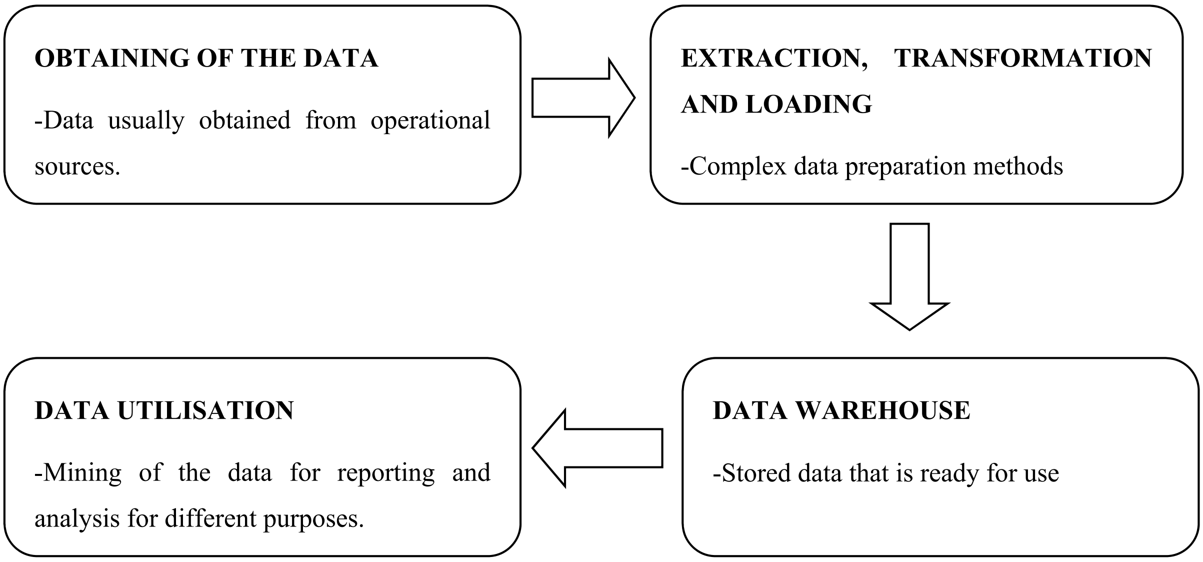

Figure 2: The data management process used by organisations

In addition to this, there are other specific methods that have been developed to correct different forms of error in data. Although all these methods fall under the data cleansing process, the generic data handling approach, the methods and the tools that they employ to seek to address specific issues. For instance, Cong et al. (2007) developed a complicated model for solving the problems of accuracy and consistency in data. In this model, it is observed that since the problems of inconsistencies and inaccuracy in data result from failure to follow the right procedures in preparing the data, the problems can be resolved by three primary methods: repairing individual values in the dataset, use of incremental methods on the values and ensuring that the repair process is accurately done (Cuong et al., 2007). Such an approach ensures that the data cleansing process takes matters of consistency and accuracy into consideration.

According to Rahm and Do (2000), in order for the data cleansing process to be effective in removing the errors that arise from issues of consistency, accuracy and others, the process must take into consideration several requirements. To begin with, it is stated that the cleansing process should be able to detect all manner of errors in individuals and integrated sources (Rahm & Do, 2000). This is necessary for all the errors to be removed from the data. Secondly, the process should be performed together with other specific operations that are meant to improve the overall quality of the data. This denotes that data cleansing works best when it is performed in conjunction with other procedures that are meant to reduce the level of errors in data.

One of the other procedures that should be used with data cleansing is data validation. Through data validation procedures, potential errors that may not be detected by the data cleansing procedures are pointed out, analysed and resolved. The importance of this procedure is that it is completely dependent on how well the data cleansing process is carried out. Therefore, by ensuring that data cleansing is carried out in conjunction with data validation, all the different types of error that have been identified can be resolved. This in turn ensures that the data that is stored and utilised is of the highest possible quality.

2.2.3. Relevant processes and the latest tools and technologies

There are several processes that are used to ensure that data is of high quality. Moreover, there are several tools and technologies that are usually used to ensure that data is of the highest quality possible. All the processes, tools and technologies that are used are usually based on basic data handling process. This essentially covers all the steps that are involved from accessing data to storing it in a form that can be easily used by an organisation. According to Rahm and Do (2000), such a process is usually divided into three main phases: *operational sources of the data*, the *extraction*, *transformation and loading* phase and the last phase is the *storage of the data*. In all these phases, complex processes of extraction, translation, matching and integration are applied to transform the data to a form that is ready for use.

In order to enhance this process and improve the overall quality of the data that is obtained from it, organisations use the Total Quality Data Management (TQDM) approach. In this approach, a four-stage cycle involving defining, measuring, analysing and improving data is applied to the entire data handling process for continuous improvement (Wang et al., 2001). With regard to tools and technologies, Rahm and Do (2000) identify three areas in data quality for which special tools have been developed, as follows: *data analysis*, *cleansing* and the phase of *extraction, transformation and loading*. Such tools include *Copy Manager*, *Quality Management (QM) Software*, *QuickAddress* and *Integrity*, among others. The following is a list of data management tools that can be used by organisations.

a) Copy Manager

b) QM Software

c) QuickAddress

d) Trillium Software

e) Integrity

f) Data QUALITY Tool

g) Dataflux.

2.2.4. Importance of data quality in the public sector

There are several ways in which data quality is of high importance in the public sector. Generally speaking, quality data increases the productivity of organisations. Improvement of the level of productivity of an organisation as a result of availability of high quality data occurs in several ways and across different sectors of the economy. For example, availability of data of good quality can be used to help organisations reallocate resources appropriately (Yiu, 2012). This leads to attention being given to areas of great concern in the operations of the organisation.

Secondly, data that is of good quality can be used to enhance the productivity of organisations by being used to make improvements in the way the organisations offer their services (Bujak et al., 2012). For example, organisations operating in the public health sector can considerably improve their service delivery by relying on more accurate and complete data about the health condition of their patients. It is by relying on continuous improvements in service delivery, which in turn is dependent on the use of quality data, that organisations are able to increase their productivity to the general public. The end result is that the performance of the organisation is improved over time and the benefits arising from increased productivity are passed down to the general public.

The most obvious way in which data quality is important to the public sector is with regard to the role that such data plays in the decision-making processes of organisations and the impact of this to the general public. Usually, the decision-making process plays a very important role in the way organisations carry out their responsibilities. For example, organisations rely on well-developed predictions to plan for how they can respond to possible future scenarios (Yiu, 2012). By forecasting their future operations, organisations are able to respond to changes in policy quite effectively. In order for organisations to make such decisions, it is important that they have access to high quality data. Therefore, quality data is important for organisations in the public sector in that it helps them make decisions about their future courses of action. This is important because by doing so, organisations are able to plan on how to provide high quality services to the public (Health Information and Quality Authority HIQA, 2011). This is applicable to different public sector organisations in that they can use quality data to predict trends in the market and respond by instituting the most appropriate risk management strategies.

Another way in which high quality data is important to the public sector relates to the benefits that members of the public and organisations can derive from the process of sharing information on the operations of organisations that operate in the sector. In general, the public is in support of the practice of organisations sharing important data on how they operate (Shakespeare, 2012). For this process to be successful, organisations need to share with the general public data that is of the highest possible quality. When this happens, individuals are able to understand the way organisations conduct their operations. This enhances the accountability of organisations in the public sector. This is helpful not only to individuals but also to different organisations. For example, when one is accessing specific documents, the input of several public organisations is required at different stages of the process. When such organisations share high quality data among themselves and the general public, this makes the process much easier and more transparent.

According to SOA & LL Global (2011), poor quality data usually leads to undesirable consequences for organisations, which include the risk of losing profitability, the manner in which the organisation manages its capital and its overall rating by the public. When applied to the public sector, it can be seen that organisations require high quality data as a way of avoiding the consequences of using data that is of poor quality. For example, an organisation that fails to use accurate data to present its performance to the public risks losing its overall rating and profitability over the long-term.

Data quality is important in the public sector because it helps the public understand the way organisations conduct their operations. Further, by using high quality data, organisations are able to reallocate their resources and make general improvements to the way they conduct their operations. In order to ensure that they have and use high quality data, organisations apply the TQDM approach to the data handling process. Data quality in public organisations will be investigated later in this chapter, and primary data from seven organisations in Abu Dhabi Emirate are explored in later chapters.

2.3. Information Systems

This section provides insights into information systems, in particular their applications in the public sector. In the general sense, an information system is a systematic approach followed to manage, sort, organise, retrieve, access, and modify data or information stored in one central place or several places connected in a certain manner. This definition may embed any type of information management system, including manual and legacy systems. However, modern information systems are based on computer processing of information which is stored mainly in digital formats.

The rapid development of information systems is mainly based on the rapid growth of information generation and communication techniques. Modern technology has allowed not only better ways of generating and appreciating information, but also methods of efficient communication of this information. Information and Communication Technologies (ICT) offer a multitude of ways to manipulate and communicate information among users and across different systems. Technology has a key role in modern information systems. Therefore, continuous and rapid changes and updates to these systems occur based on the rapid development of the underlying technologies (O'Brien & Marakas, 2010).

Modern forms of information systems have become a necessity for organisations in different fields due to their abilities to manage large amounts of information in organisations, nowadays. Public organisations are one type of organisation that usually acquire vast amounts of data. Management of these amounts may pose a challenge to these organisations given the sensitivity of the information they hold, as well as the choice of efficient information management systems. Furthermore, information systems can be very involved and complicated, and may require a significant amount of skill and knowledge. The vast amount of data handled by information systems may be significantly benefited from, given the right choice of tools.

2.3.1. Development of information systems (IS)

This section provides an overview of the development of information systems as they are currently utilised. The study focuses on modern views of information systems rather that involving legacy concepts.

2.3.1.1. History of information systems

The history of information systems can be traced with the history of computer science. The four major trends are outlined below (Lucas & Henry, 1994):

(i) the use of information processing technology as a part of corporate strategy;

(ii) technology as a pervasive part of the work environment;

(iii) the use of technology to transform the organisation; and

(iv) the use of personal computers as managerial workstations in the development of Information Technology.

The historical evolvement of information systems went through major changes up to their current state. Some of the major historical focuses of the development of Information Systems (IS) are as follows (O'Brien & Marakas, 2010).

2.3.1.1.1. Data Processing: 1950s - 1960s

Between the 1950s and 1960s, IS were used for processing electronic data, including a range of business transaction processing, digital record keeping, and storing accounting data in a classical way.

2.3.1.1.2. Management Reporting: 1960s - 1970s

Managers started to employ IS to support their decision-making processes in the 1960s to 1970s. The period is also known as the era of Management Information Systems (MIS). In this period, IS usually produced various business reports – financial reports and information about products and services, which assisted managers in making strategic decisions.

2.3.1.1.3. Decision Support: 1970s - 1980s

IS started to work as a decision support system in the 1970s to 1980s. In this period, information was used to assist managers through interactive management support of the managerial decision-making process.

2.3.1.1.4. Strategic and End-User Support: 1980s - 1990s

From the 1980s to the 1990s, IS took a new shape and went through a profound development as end-user computing systems, executive information systems (EIS), expert systems, and strategic information systems. In this period, end-user computing systems facilitated direct computing support for end-user productivity and work group collaboration. Executive information systems provided the important documents for different groups of people in an organisation to assist in their decision-making processes. Expert systems provided the knowledge-based expert instruction for end-users. Strategic information systems provided knowledge and information about the products and services of rivals to help in making strategic decisions to gain competitive advantages.

2.3.1.1.5. Electronic Business and Commerce: 1990s - 2000s

There was a phenomenal growth of Internet-based e-business and e-commerce with the advancement of IS between the 1990s and 2000s. With the advantages of IT and IS, there was an emergence of the Internet, intranets, extranets and other networks in this period.

2.3.1.1.6. Enterprise Resource Planning and Business Intelligence: 2000s - 2010s

Information Systems became more dynamic and were used for enterprise resource planning (ERP), including enterprise-wide common-interface applications, data mining and data visualisation, customer relationship management and supply chain management.

2.3.1.2. Types of information system

Information systems can be categorised on the basis of purpose dimension and scope dimension (Gordon & Gordon, 2004). To perform in the purpose dimension, IS should have the elements of Automation Systems (AS), Transaction Processing Systems (TPS), Management Support Systems (MSS), Decision Support Systems (DSS), Groupware, and Executive Information Systems (EIS). To perform IS in the scope dimension, the information systems should hold the attributes of serving among individuals, be departmental/functional, allow enterprise, and act as inter-organisational systems. There are different types of Information Systems that are used by modern organisations (Rainer & Cegielski, 2013). Some of the IS are Transaction Processing Systems (TPS), Management Information Systems (MIS), and Enterprise

Resource Planning (ERP) systems, Customer Relationship Management (CRM) systems, and Supply Chain Management (SCM) systems. IS may be classified into two different categories – operations support systems (OSS) and management information systems (MIS) (O'Brien & Marakas, 2010).

2.3.1.2.1. Operations Support Systems

Information systems have always processed data produced by, and practised in, business operations. Operations support systems generate different types of information on products and services to use internally and externally, although the information is not focused on a particular group of people (for example, managers) in an organisation. Operations systems are used "*to process business transactions, control industrial processes, support enterprise communications and collaborations, and update corporate databases efficiently*" (O'Brien & Marakas, 2010). A brief discussion of the above process follows.

2.3.1.2.2. Transactional Processing Systems

Rainer and Cegielski (2013) and O'Brien and Marakas (2010) define Transactional Processing Systems as "*inputs for functional area information systems and business intelligence systems, as well as being used for business operations such as customer relationship management, knowledge management, and e-commerce*". For instance, Point of Sale (POS) is a prime example of a transactional process system that is used in various retail stores.

2.3.1.2.3. Process Control Systems

O'Brien and Marakas (2010), Jessup and Valacich (2003), and Rainer and Cegielski (2013) assert that process control systems are operations support systems that are used for monitoring and controlling industrial processes.

2.3.1.2.4. Enterprise Collaboration Systems

Today's organisations use enterprise collaboration systems to assist their teams and workgroups, and for communication within and outside of the organisation (O'Brien & Marakas, 2010). Email, chat rooms and video conferencing are some of the examples of enterprise collaboration systems that are used by modern organisations.

2.3.1.2.5. Management Support Systems

O'Brien and Marakas (2010, p.14) stated that "*when information system applications focus on providing information and support for effective decision making by managers, they are called management support systems*". A manager cannot perform all tasks independently. Therefore, different information and support systems are needed for making decisions within an organisational framework. Categorised management support systems include (1) management information systems, (2) information systems for supporting decisions, and (3) information systems for executives. A brief explanation of the above management support systems are discussed as follows.

2.3.1.2.6. Management Information Systems

Management Information Systems (MIS) are used for providing information to the managers whilst making strategic decisions for the business. Some of the MIS are used for sales analysis, analysing performance of the production, and cost trend reporting systems.

2.3.1.2.7. Decision Support Systems

Decision support systems (DSS) are designed to deliver interactive ad hoc support for managers and other business professionals. DSS is used for setting a competitive price for products and services, cost-benefit analysis and analysing risks in business.

2.3.1.2.8. Executive Information Systems

Executive information systems (EIS) deliver important data and information from MIS and DSS and other sources and blend them together to assist executives in making decisions. Some of the emerging executive information systems are systems for analysing business performance, tracking different strategies and activities of competitors, and planning economic growth and development of the business, which can assist management in making strategic business decisions.

2.3.1.2.9. Information Systems for Strategy Making

Technology has become the de facto strategic tool in the knowledge economy (Dwivedi et al., 2012). As a business asset, technology has captured its position in the area of business makers and has become a strategic tool as no business can survive without technology in this age of information technology. The impacts of IT on any organisation, regardless of being public or private, are enormous and no other organisation can escape from its power, especially in the age of IT and globalisation (Dwivedi et al., 2012; Pearlson & Saunders, 2009; Clarke, 2001; Remenyi, 1991). Managers or decision-makers in private or public organisations must take into account the impacts of technological investments in terms of changes in organisational culture, work-life balance and employee resistance towards changes in management, besides business benefits and return on investment (ROI) from the adoption of IT (Rockart & Scott, 1984). In addition, to meet changes and challenges of IT, managers must consider how information technology shapes their organisations internally and externally with a view to providing competitive advantages (Dwivedi et al. 2012). Managers and decision-makers need to address different resource constraints, including money, to prioritise investment decisions and to assure stakeholders that their decision to adopt IT is to achieve the long-term strategic objectives of the business (Collin, 2008). IT is more than just a computer or combination of hardware and software as some may assume; it is in fact an effective design and use of information systems that the organisation will have to apply in order to give customers value for money and deliver value to all stakeholders (McLean & Turban, 2005).

Using Porter's model for analysis, the strategic impact of IT can reveal three generic tiers (Haag & Cummings, 2013). The three tiers for withstanding the competition in any type of organisation are (i) overall cost leadership, (ii) differentiation, and (iii) focus (Figure 3).

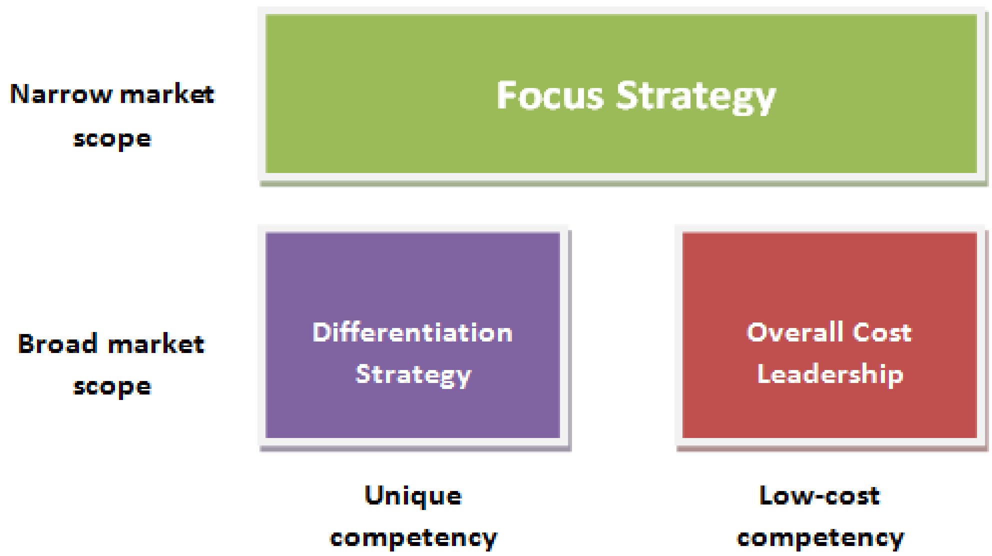

Figure 3: Porter's model of generic strategies applied to IT [*Source: Haag and Cummings, 2013, p.23*]

Porter defines overall cost leadership as offering the same or better quality products or services at lower prices than any of the competitors is able to offer. Differentiation is clarified as a business proposition or service that has a unique quality in the market place. Focus is predefined by Porter as a strategy which usually concentrates on both products and services to (i) a targeted business or businesses or individual or group of people, (ii) within a segment of product line, and (iii) to a particular domicile (Haag & Cummings, 2013).

There are three fundamental challenges posed by information technology in organisations. Firstly, with the consistent innovation in technology, business models are bound to be reshaped; secondly, information technology has increased the demands of stakeholders along with customers; finally, adoption of technology provides a competitive edge in any business (Dwivedi et al., 2012). One of the prime concepts that have brought enormous changes to business through IT is "*providing value of stakeholders such as customers, investors, employees, supplier and environments*" (Porter, 2001). The five forces in an IS strategy based on Porter's five forces model and are as follows (Dwivedi et al., 2012):

1. IT and Buying Power
2. IT and Entry Barriers
3. IT and Threats of Substitutes
4. IT and Industry Rivalry
5. IT and Selling Power.

The alignment of business strategies has to be coherent with the IS strategy to meet the changes and challenges of the IT. The rationale behind the alignment of corporate strategy and IS strategy is that today's business technology cannot be viewed as being different from business (Clarke, 2011). By taking into account the ultimate influence of the technology in business, a strategic model that includes all the elements of corporate and information systems strategy may be as follows (Clarke, 2011) (Figure 4).

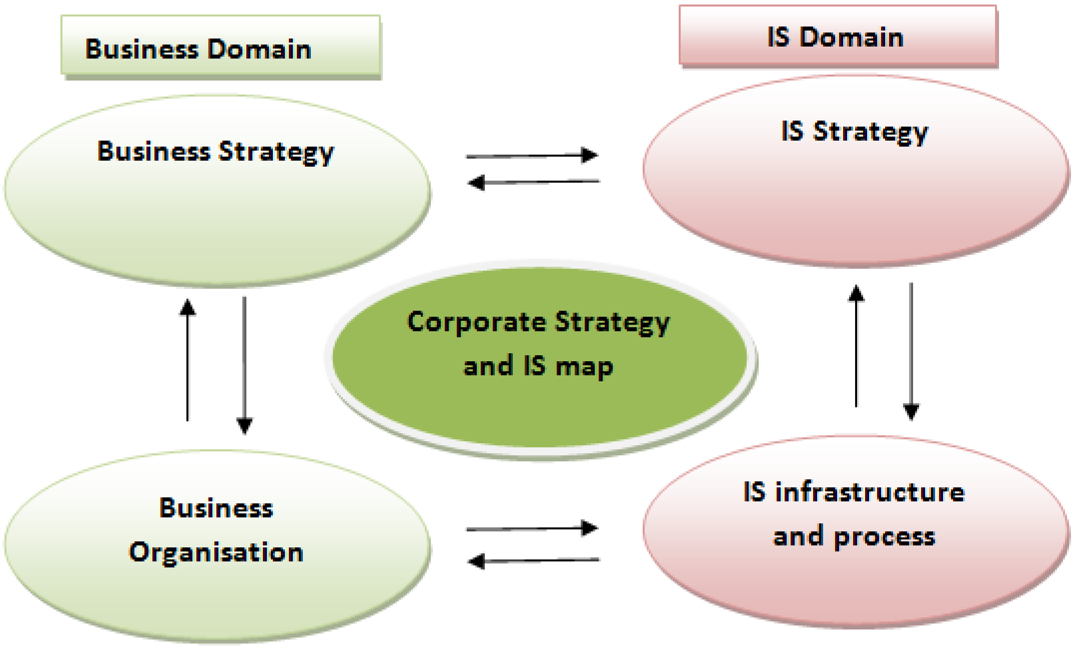

Figure 4: Extended Strategic Alignment Model [*Source: Clarke (2001), p.92*]

It is mainly information that drives the process, whilst central to that process is the organisation's corporate strategy, together with the IS map. Information needs are met through the business domain and IS domain interacting to support the operations of an organisation. Functional focus is an important aspect to consider whilst developing the strategy. Functional focus is fundamentally based on the maximising of resources that are available to the organisation (Hofer & Schendel, 1978).

2.3.2. Information systems in the public sector

In a research study conducted by Heeks (2000), it was shown that information has different meanings to different governments of different countries. For example, the U.S. government believes that general information should be made public because information is collected from everyone and they should have free access to that information. However, the European Commission (EC) (2000) has a mixed view on information systems and outlines an electronic government model to share the information within a transaction service. In further clarification of transaction service, the EC (2000) states that, "*transaction services, such as electronic forms, are perceived as the future of electronic government*". The notion of transactional services attempts to help the state's citizens to submit electronic documents in the form of communication with the public authority to obtain the appropriate service. The data received from the citizen is then transmitted to the public authority into the information system followed by the business rules.

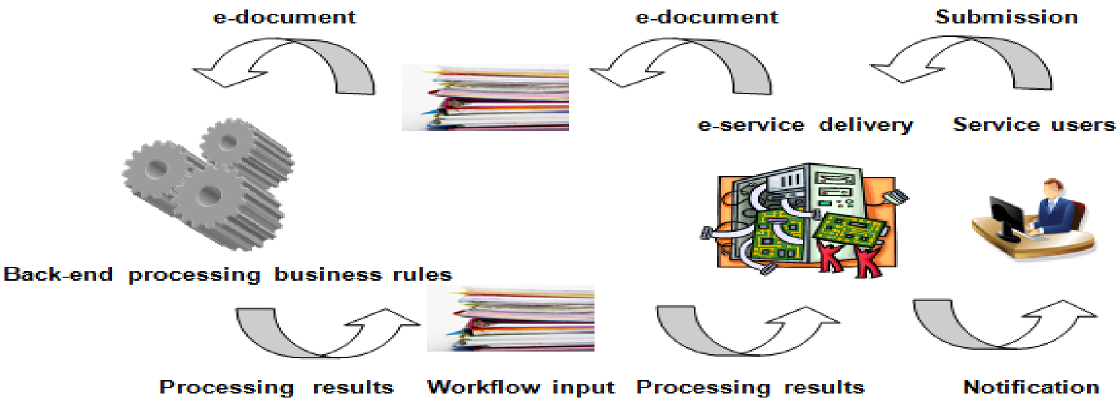

Figure 5 depicts the method of transactional services in e-governance and how this permits users to submit their documents electronically, which fulfils the desire of the service provider while customers gain the most satisfactory service from the public authority.

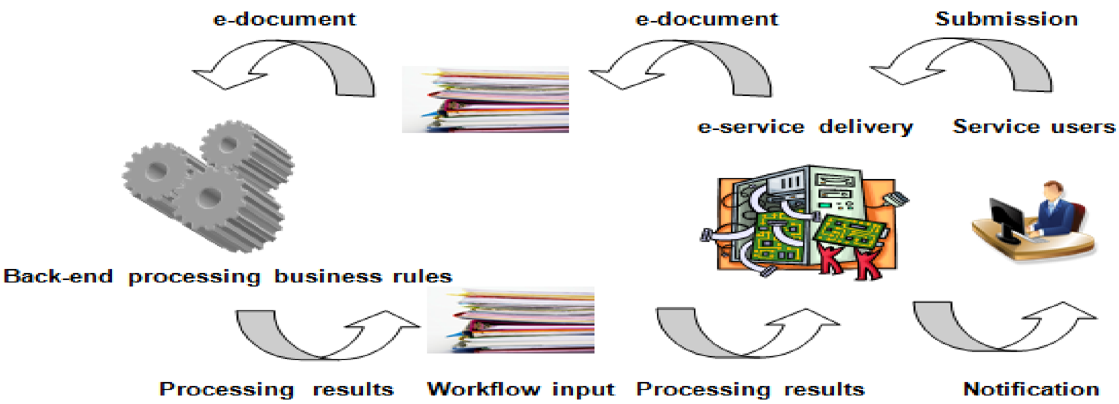

Figure 5: Full Processing Cycle for Transaction Services [Source: Vassilakis et al. (2004), p.50]

Information systems are not mature in the public sector, however. Vassilakis et al. (2004) have argued this by providing a number of points about why thorough information systems do not exist in the public sector. These are:

- Limited security
- Platform diversity
- Inflexibility of legacy systems
- Task scheduling complexity

- Change management.

These points and others may require further understanding before a thorough information system can be utilised in the public sector as a central hub for information.

2.3.3. Use of information systems in the public sector of Abu Dhabi

Most of the services that are found in bricks-and-mortar localities in developed countries are also found online. That even extends to government services in those countries. The expansion of the Internet has also urged governments to be committed to serving their citizens through online facilities. Countries in the developed world have strong presence online under the so-called e-governance umbrella. Some of the developing countries have started introducing e-governance into their service mechanisms. Industrialised countries are taking the zest of e-governance to achieve benefits. Such benefits include providing citizens and organisations with more convenient access to government information and resources, conducting transactions with businesses and with those working in the public sector, and delivering public services to citizens (Davidrajuh, 2004). Further to these benefits, the main objective of any country from e-governance is to utilise e-governance for the development of the nation's economy and improvement of the quality of life and opportunities of the citizens. In Abu Dhabi, this aim is set forth. Following the successful endeavour of the neighbouring emirate, Dubai, the government of Abu Dhabi has introduced a policy to deliver most of its services to customers electronically and set a vision to build e-Abu Dhabi. The vision of e-Abu Dhabi is explained and clarified by the Director General of the Abu Dhabi Systems and Information Centre (ADSIC), Mr. Rashed Lahej Al Mansoori, who asserts that

> "*the e-Governance site is designed to provide services to all the government departments, authorities and administrations and ensure transparency of the services so that most of the officials use it as a vital tool of providing quality and effective service to the diverse customers*" (ADSIC, 2015).

To acknowledge the fact and advantages of e-governance, ADPO sets a vision in their mission statement and strategy. ADPO is, hence, committed to executing the vision of central government by implementing and providing most of its services online.

According to the official documentation, the vision of e-Abu Dhabi is to develop into a "*high performance government delivering world class services to the benefits of all its customers*" (Abu Dhabi e-Government, 2015). The Government of Abu Dhabi believes that the vision will benefit them in various ways with their customers and also transform the way government works electronically. The Government of Abu Dhabi defines the customers as including individuals and every business. The idea behind this vision includes four focus and design themes for the e-government strategy. The focus concentrates on end-users and increasing the efficiency of the service. The design themes are comprehensive and cross-governmental in design. The vision is also described in Figure 6.

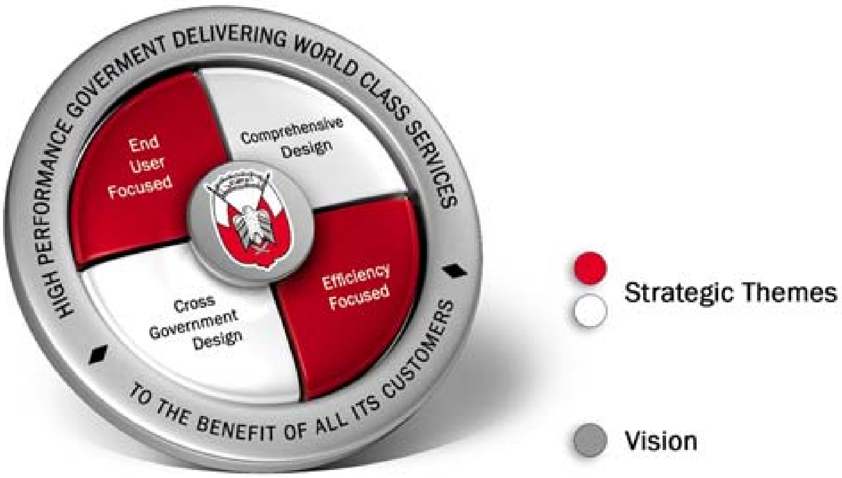

Figure 6: The Vision of e-Abu Dhabi [Source: e-Abu Dhabi: The Abu Dhabi Government Modernization Initiative (abudhabi.ae)]

2.3.3.1. E-Governance in Abu Dhabi

2.3.3.1.1. Real Time Video Surveillance

ADPO introduced a new real-time video surveillance system on 24 February 2009 in the process of e-governance. The new technology, called MOTOA4, provides a critical portfolio delivering innovation and reliable wireless solutions. The installation of MOTOA4 is a part of the accomplishment of the vision, which is to build a safe and secure society. This new system improves communication and amalgamates services among police personnel, police vehicles and control room operators. MOTOA4 provides the most comprehensive video services with high-resolution compared to other systems. Now, ADPO can record, save and analyse video from the MOTOA4 system which can also help in further investigation of any crime. This system is used for operating and delivering a high-efficiency real-time streaming of video from police vehicles and personnel. The service will enable management and control room operators to have real-time knowledge of what is going on in the area and allow them to immediately respond on that basis. The vehicles will be shown on a live map. MOTOA4 is easy to use alongside other operating systems used by ADPO such as automated speed tracking of vehicles, fixed camera installations, facial recognition and automatic number plate recognition. In the future, electronic passports, national IDs and fingerprint reading capabilities will be added to this system. This integrated system has made ADPO the most modernised and sophisticated police force in the region.

2.3.3.1.2. Advanced Radio Communication

ADPO launched a unique advanced radio communication system called NOKIA TETRA to upgrade its communication system. The implementation of the turnkey project is part of the five-year plan and one step forward in the execution of the e-Abu Dhabi. The main platform of the new communication network, Policom, is based on NOKIA TETRA and Policom became the backbone of the ADPO. The Lieutenant General Shaikh Saif Bin Zayed Al Nahyan also, Minister of the Interior, confirmed that,

> "*this advanced radio technology is able to communicate instantly, provide reliability and maintain high security of the commutation. Choosing a qualified company to carry out such a project was challenging. Selecting Nokia to implement the Policom network was the result of lengthy technical investigation and field tours conducted by our police communications experts. We also conducted a prolonged technical study on Nokia's sophisticated TETRAS system*".

2.3.3.1.3. State Police Security System

An agreement was signed between ADPO and EADS Defense and Security (DS) to prepare a plan for an integrated security system solution for the ADPO on 23rd February 2009. This project included analysis of business needs and security gap analysis, comprehensive threat assessment, system design concept and complete implementation plan.

2.3.3.1.4. *E-recruitment*

The ADPO launched a scheme for recruiting employees through an e-recruitment system on 26 September 2005 (us.oneworld.net). Now, people can see their dream jobs online and download application forms from the ADPO website. For the applicant, this e-recruitment provides more flexibility and they can also save their valuable time. On the other hand, ADPO can seek the best talent from the job market and also save huge recruitment administration costs. The Director of Human Resources Management, Colonel Mohammed Al Awadi Al Menhali, stated that there will be four phases in the new recruitment system and they include the latest vacancies being immediately posted on the ADPO website, collecting applications, evaluation of applications and e-appointment.

2.3.3.1.5. *E-services*

In the process of implementation of e-governance, ADPO provides most of the services electronically. The people of Abu Dhabi benefit from the e-services in various ways. They can pay their bills, fines and get information from the electronic services available in most public places and can also use online e-payment and information services. Recently, ADPO installed a number of automated machines in shopping malls and this network, named Sahel (facile), is a part of implementing the vision of e-Abu Dhabi. All machines are sophisticated and have touch screens that make the machine more user-friendly. Now, people not only use these services for their principal purpose but also for making complaints and proposals or simply for writing communication messages. The machines can also produce a receipt for every transaction so that people can keep the slip for their records. The Department of Traffic and Parking of the ADPO has recently introduced the mParking system so that people can pay their parking fees via mobile SMS. This system is easy to use for any registered vehicle owners. If any vehicle is not registered the vehicle owner can still go to the website mpark.rta.ae and pay their parking fees via mobile SMS. E-services also allow people to open a file for the driving licence and book a test.

2.4. Decision Support Systems

2.4.1. Definition

The definition of Decision Support System (DSS) is roughly that it is a computer-based information system which is designed to aid organisations regarding their decision-making activities (March & Hevner, 2007). Decision support systems are used in the mid and upper levels of corporations to aid in management functions and planning. These areas usually change rapidly as a result of external factors such as competitor innovations or inroads, along with rapidly changing customer tastes (Arnott & Pervan, 2008). Understanding of decision support systems is central to this study, thus an additional definition shall be used to ensure understanding. Laudon and Laudon (2010) define a DSS as a system which supports either one or a group of managers working on problem-solving situations to solve problems or issues. This is accomplished by providing management with information scenarios arising from differing ways to look at or ascertain potential outcomes (Laudon & Laudon, 2010). DSS includes varied potential solutions or suggestions which can be analysed (Laudon & Laudon, 2010). In terms of a further explanation of DSSs, Shim et al. (2002) explore many definitions resulting in synthesising the explanation to one that describes a DSS as an interactive computer based decision system that (Figure 8):

i. Is used to provide support to decision-makers as opposed to replacing them.

ii. That the process uses models and data in the performance of its varied functions.

iii. DSS solves problems in varied structure modes. These can include non-structured and semi-structured.

iv. The DSS focuses on the effectiveness of outcomes as opposed to the efficiency or facilitation of the decision process because it seeks the end result as being the purpose.

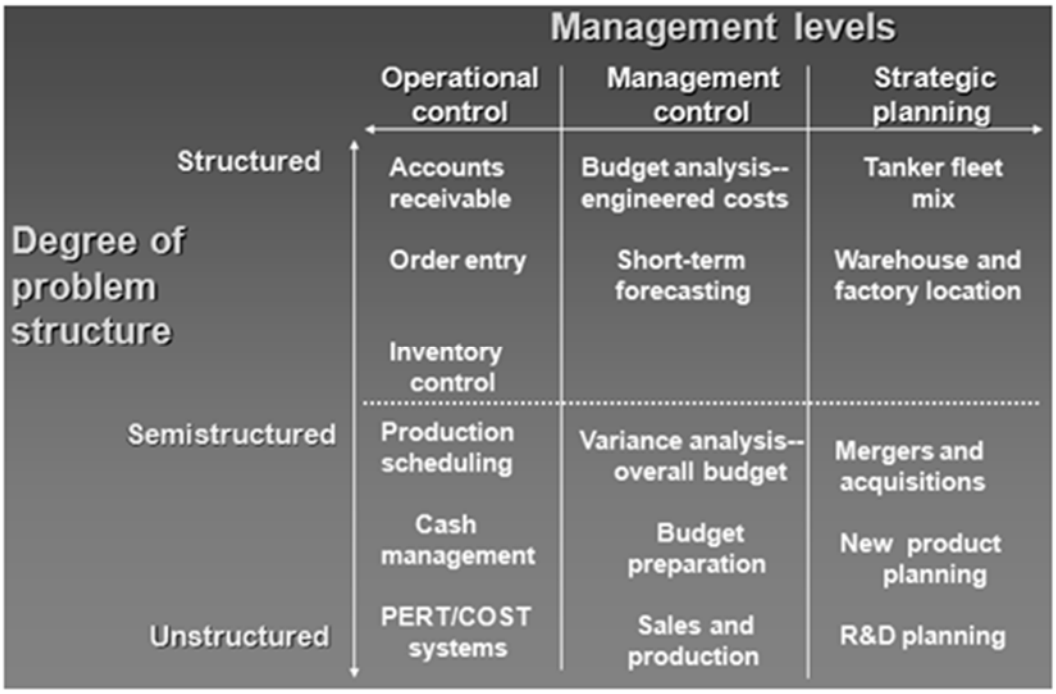

Figure 7: DSS structured, semi structured and unstructured examples (McLeod & Schell, 2007)

As a means to further understand DDS, the levels in terms of problem-solving support from its lowest function to its highest are ranked as Laudon & Laudon (2010) state:

i. The retrieval of information elements is the lowest-ranked function.

ii. The next function is the retrieval of information files.

iii. This is followed by the use of multiple files to create reports.

iv. The estimation concerning the consequences of decisions ranks fourth.

v. The prior function provides the means for the DSS process to propose decisions.

vi. All of the prior areas are aimed at decision-making, which is the top function.

DSS is a concept that leads to the development of systems to address particular types of decision. The use of DSS is not confined or restricted to any one type of application (Laudon & Laudon, 2010). Visually, the DSS process in terms of its general approach to solving problems or posing solutions looks as in Figure 8.

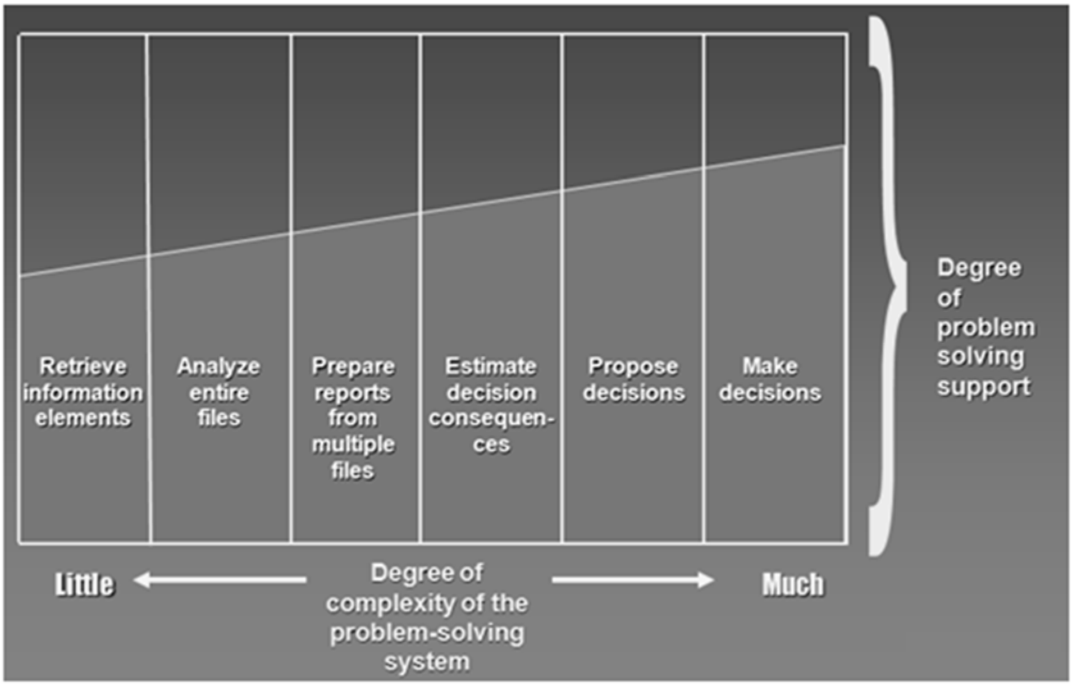

Figure 8: DSS path to solutions (McLeod & Schell, 2007)

The three main objectives of DSS are to provide assistance in helping to solve semi-structured issues or problems to support managers, and to contribute to the effectiveness of the decision process (McLeod & Schell, 2007). Figure 9 helps to visually understand the manner DSS works (McLeod & Schell, 2007):

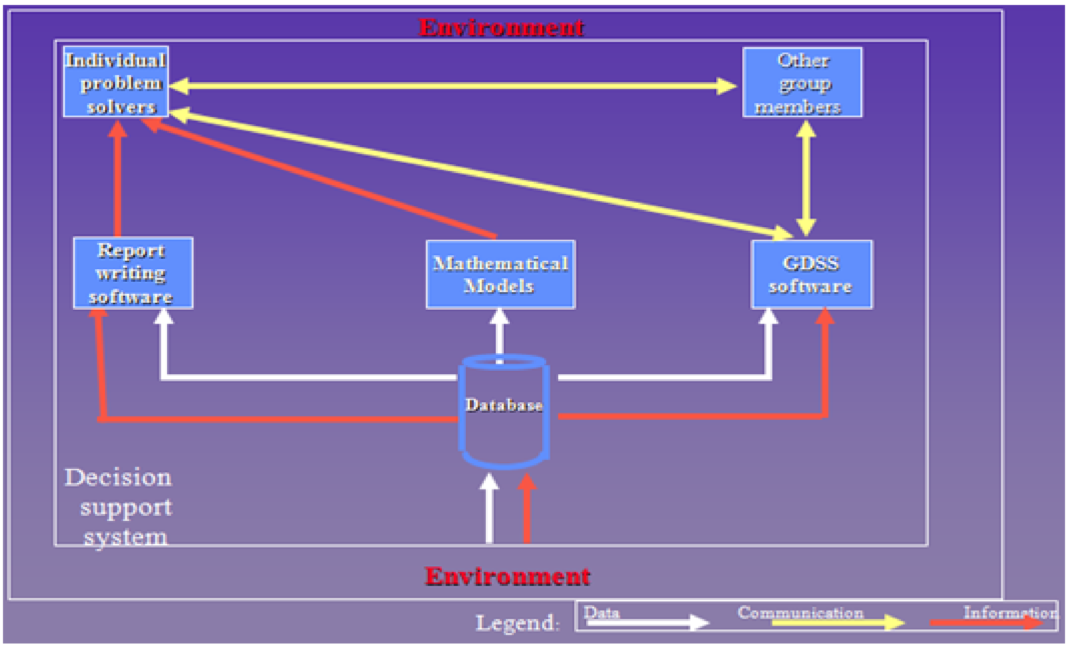

Figure 9: A Model of the DSS Process (McLeod and Schell, 2007)

DSS aids in solving problems by providing improved communications and enhanced focus concerning discussions, along with reducing time needed for decisions as it allows for the exploration of scenarios prior to actual application. The following provides an understanding of how the process (DSS) accomplishes the above (McLeod & Schell, 2007):

- Decision rooms
    - Small groups face-to-face
    - Parallel communication
    - Anonymity
- Local area decision network
    - Members interact on a LAN
- Legislative session
    - Large group of interaction
- Computer-mediated conference
    - Permits large, geographically dispersed group interaction.

2.4.2. History of development

As a subset of information and information management systems, DSS evolved in the late 1950s and early part of the 1960s starting with work conducted at the Carnegie Mellon Institute of Technology (Power, 2002). It evolved out of research undertaken to apply quantitative models in computer software to aid in decision-making (Power, 2002). The first rudimentary example concerning the beginnings of a DSS-type of application was run on an IBM 7094 system in the latter part of the 1960s (Power, 2002). As the power and utility of computer systems improved, the type of software applications that could be run was enhanced (Buchanan & Connell, 2006).

As the capability of computers increased, the theory of DSS also advanced as researchers were able to think about broader application possibilities (Buchanan and Connell, 2006). The book that is said to have influenced the direction and development of DSS was written by Steven Alter in 1980 (Power, 2007). Alter's work advanced and expanded the framework concerning the thinking and approaches to DSS as he explored and provided examples on the process (Power, 2007). Alter's work resulted in classifying DSS into seven types (Power 2002):

1. *File drawer systems.*

   These provide users with access to items in the data system.

2. *Data analysis systems*

   The systems that support data manipulation using computerised tools are tailored to specific tasks and settings.

3. *Analysis information systems*

   This area provides access to decision-oriented data and models.

4. *Accounting and financial models*

   This fiscal aspect is utilised to determine the monetary consequences of various actions and outcomes.

5. *Representational models*

   The consequences concerning actions taken are estimated using simulation models.

6. *Optimisation models*

   These provide the guidelines for the generation of optimal solutions taking into account the constraints that accompany them.

7. *Suggestion models*

   These are models that perform logic processing that lead to a suggested decision for either a task that is understood or one which is structured.

The building blocks offered by Alter (Power, 2007) provided the framework and foundation to be studied and built upon in universities and by organisations. The prior developmental inputs concerning DSS provided the identity that guided later areas as it was recognised that the system would support decision-makers (Buchanan and Connell, 2006). The understanding concerning the direction and purpose of DSS provided needed clarity in the development of its approach (Buchanan and Connell, 2006). The above means researchers looked at the manipulation of quantitative models, and the analysis and accessing of large databases as a key input component, along with the support of decision making by groups (Power, 2007). The two main categories are model driven (which emphasises individual use) and data driven (which is organisational or institutional DSS) (Power, 2007). The following sub-sections provide their historical development inside of the broad DSS time structure provided between the late 1960s and late 1980s.

2.4.2.1. Model-driven DSS

It is broadly recognised that Scott-Morton in 1972 designed the first model-driven DSS system (Hedgebeth, 2007). The model-driven version of DSS is based on emphasis concerning the access and manipulation of models (Hedgebeth, 2007). The above represent optimisation, along with financial and simulation (models) using in most instances a quantitative foundation (Hedgebeth, 2007). In terms of development, Power (2008) argues that in general, quantitative models offer an elementary base of functionality. Model-driven Decision Support Systems utilise the parameters and data that are provided to aid in the analysis of situations as generally large databases are not needed. The above is the case due to the fact that model-driven DSS scenarios are based on specifics as opposed to generalities (Power, 2008).

Gerald R. Wagner and his students at the University of Texas created the first commercial model-driven tool for quantitative financial models, which was termed Interactive Financial Planning System(IFPS) (Power, 2007). In 1982, Ernest Forman in close collaboration with Thomas Saaty designed Expert Choice (Mahdi & Alreshaid, 2005). It represented a DSS generator used in the building of specific systems that was based on the Analytic Hierarchy Process (Mahdi & Alreshaid, 2005). In the late 1970s, the financial program called VisiCalc was commercialised. This program was the first financial software application that could perform multiple functions and set the stage for today's Excel and other financial software. VisiCalc was built on a model-oriented personal adaptation of DSS (Grad, 2007).

2.4.2.2. Data-driven DSS

Hedgebeth (2007) argues the data-driven category of DSS is based on emphasising access and the manipulation of time-based series of data in a company. This can also include in some instances real-time and external data (Hedgebeth, 2007). Power and Sharda (2007) argue that under data-driven DSS file systems, they are accessed using retrieval and query tools. Whilst these provide elementary functionality, data warehouse systems are the base used to supply information (Power & Sharda, 2007). This differs from model-driven DSS as the data-driven mode draws on information it needs to construct its formulations (Power & Sharda, 2007).

The first data-driven DSS applications were developed for American Airlines by Richard Klaas and Charles Weiss. It was called An Analytical Information Management System (AAIMS) and was used to perform what-if simulations (Power, 2008). The late 1970s and early 1980s saw the development of executive support systems (ESS) and executive information systems (EIS) that evolved as a result of what are termed as rational database utilisations (Hung, 2003). These systems were first used by Lockheed Aircraft and Northwest Industries, and aided these companies in the performance of data importation, along with the access to and use of news service feeds and user-friendly screen designs (Hung, 2003).

The relational database approach is a collection of tables as opposed to hierarchies which means adding and accessing information is faster. Relational databases do not use a hierarchy system, thus the rows of items and columns of fields can easily be added to and accessed. This differs from the hierarchy system that organises information based on importance, which can change based on the application or end use.

2.4.3. Uses

The above description and definition of DSSs did not make a distinction in terms of if it is more suitable for business or governmental use. From the definition in the above section, it is clear that Decision Support Systems are used by management to find solutions to a broad range of areas.

A white paper concerning modernising government in the UK provided an example of the uses of Decision Support Systems (UK Government Cabinet Office, 1999). In this white paper, the challenges identified concerning the use of DSS were as follows:

a. Implementing a process and systems where the staffs in the varied bureaucratic governmental departments work together to streamline operations and eliminate task duplication.

b. A focus and commitment to the processes and tasks undertaken by public service staffs to improve individual and departmental work quality internally and for the public.
c. To create a culture in governmental service that is more innovative and risk aversive.
d. To develop a management system in civil service that provides the systems and processes to accomplish the areas and objectives set forth above.

As the scope of these undertakings was so broad, a specific case study example involving the UK is used. Watson (2001) used the UK's INFOSHOP programme as an example of the application of Decision Support Systems by government. It was used to mesh the complex regulations for food safety, health and safety, building planning, and building control into a more efficient system. The differences in regulations for various councils, boroughs, towns, villages and cities is organised under a decision tree framework that uses a specially designed Internet application that can be accessed by users. Inquiries over the phone are handled by operators using a similar system (Watson, 2001).

The operational cost savings result from the fact that before the system was implemented many questions had to be referred to an expert for handling which created inquiry backlogs and costly delays (Watson, 2001). Development of the INFOSHOP system represented a collaborative project that meshed complex regulations and updated the system regarding new ones across a broad range of areas (Watson, 2001). The INFOSHOP example is a form of DSS used in the UK which embraced web technologies under the indicated government initiative described in the 1999 white paper (Gilbert et al., 2004).

A similar system to the UK's has also been adopted in the U.S. and other countries (Gil-Garcia & Pardo, 2005). France offers another example of Decision Support Systems in government under the country's French Association of Management Control in French Local Government (AFIGESE). The system was devised to reduce bureaucracy and to lower administrative costs in the collection of data, along with the evaluation and implementation of public policies. The process entails meshing a highly complex web system of laws, regulations, considerations, and allied areas in terms of updating or amending current public policy. The issue entails taking into consideration the potential ramifications, impacts and effects that changes would have, and how policy modifications need to be weighed against costs and impacts (Peignot et al., 2012).

In terms of the business arena, the application of Decision Support Systems has an extremely broad base of uses as it entails information management, financial and administration areas (which are the primary usage modes for government) (Moss & Atre, 2003). Other areas include production, procurement, engineering, customer information areas, and various combinations of the above (Moss & Atre, 2003). In the business arena, Decision Support Systems have moved primarily to web-based applications under data-driven, and model-driven approaches (Bhargava et al., 2007) (Table 1).

Table 1: Decision Support Systems in Business Application Modes (Bhargava et al., 2007)

Tasks for Model-driven only	Model instantiation Model execution Analysis and reports	Model definition Analysis definition User Interface definition
Tasks for both Data-driven and Model-driven	Data visualization Query and retrieval Data analysis	Data definition Analysis definition User Interface definition
	Tasks for both Application-specific and DSS Generator	**Tasks for DSS Generator Only**

In explaining the functions and uses of DSS in business and government, Shim et al. (2002) maintain that it is founded on three areas. These are data quantification, information management and model manipulation. Data quantification represents the process where large volumes of data are first condensed and then manipulated analytically into what are termed as core indicators so that important data can be extracted. Information management represents the process used to store, retrieve and report information under a structured format. The third area, model manipulation, is where varied scenarios constructed under what-if questions are explored. As indicated, the uses and applications of Decision Support Systems in business are broad. Some usage examples entail supply chain management, financial forecasting, estimation and projection, project management, customer relationship management, resource planning and a host of other areas (Yan et al., 2003).

2.4.4. DSS types

In terms of the types of DSS application, there are five broad categories (Power, 2007). These categories represent Decision Support Systems that are data-driven, communications-driven, knowledge-driven, document-driven and model-driven (Power, 2007). Alter's classified 56 DSS areas may be divided into seven distinct categories or types (Power, 2007), as follows:

2.4.4.1. File drawer systems

These are systems which permit access to data items. In discussing examples, Power (2008, p. 1) states these include "*real-time equipment monitoring, inventory reorder and monitoring systems*". He adds, "*Simple query and reporting tools that access OLTP or a data mart fall into this category*" and that "*Examples include budget analysis and variance monitoring and analysis of investment opportunities*".

2.4.4.2. Data analysis systems

These provide support data manipulation using computerised tools that are tailored to varied tasks or purposes. (Power, 2008, p. 1) states that "*Examples include budget analysis and variance monitoring and analysis of investment opportunities*".

2.4.4.3. Analysis information systems

This is used to provide access to various small models and decision-oriented databases. Examples concerning this area represent "*sales forecasting based on a marketing database, competitor analyses, product planning and analysis*" (Power, 2008, p. 1).

2.4.4.4. Accounting and financial model-based DSS

This model calculates the consequences resulting from possible actions that might be undertaken. These include the estimation of the potential profit from the introduction of a new product, the analysis of operational areas concerning plans, an analysis of break-even costs, along with producing estimates regarding balance sheet projections and income statements. The above are termed as "what if …" scenarios that management can use to determine different operational, production and marketing scenarios (Power, 2007, p. 1).

2.4.4.5. Representational model-based DSS

Management can utilise this model to perform estimates concerning the possible consequences arising from various actions they plan to take using simulation models. These include causal areas arising from reactions to competitive situations which might cause a change in the manner a company operates or needs to respond. Power (2008) cites that such areas can consist of models that analyse risk, market response models, along with simulations regarding production runs and the use of different types of equipment or procedures or processes.

2.4.4.6. Optimisation model-based DSS

Under this usage mode management can look at varied optimal solutions for operations, production, processes and other areas by introducing varied constraints to assess which approach provides flexibility. In the fast-changing environment of business, having systems that are able to adapt to differing conditions is important in terms of meeting unforeseen situations and circumstances. Examples concerning the use of this area include functions such as the allocation of resources, scheduling systems and the optimisation of material use.

2.4.4.7. Suggestion DSS Based on logic models

This function represents a means to undertake what is termed as logical processing that leads to specific decision suggestions for different tasks. As can be seen from the above, DSS types can use external and internal information. These areas use differing approaches in their application (Arnott & Pervan, 2005). They generally are comprised of analytical models used in complex decision processes, and the extraction of useful information from large data sources. In many instances the use of DSS calls for the combination of the above. In today’s knowledge-driven business arena and society (government mode) the experience factors gained from human actions and outcomes has become an increasing source of valuable information. These are termed as expert systems that capture human expertise in applicable knowledge domains. The capture of the expertise that resides in employees is transferred to systems that correlate and allow for this to be shared and used by others and is an important aspect in the use of DSS.

2.4.5. Importance and need

The highly competitive nature of today's business environment calls for increased government efficiencies and service to its populace that has created the need for increased efficiencies (Srivardhana & Pawlowski, 2007; Carter & Belanger, 2005). The vast repositories of information and complexities that are a part of operating huge governmental systems and business operations means the use of past experience factors, information databases, tendencies, and associated facets are needed to aid in decision-making and planning (Srivardhana & Pawlowski, 2007). The above summary analysis has been used to provide an understanding of the complex environment that government and business operate in. It is a part of a broad range of information management and analytical tools that have become an integral part of the knowledge management aspect of society. These represent the services and administration provided by government along with the broad range of business uses.

The complexities that accompany today's competitive business arena and the broad range of services, regulations and governance means information needs to be processed effectively (Heeks, 2002; Edmunds & Morris, 2000). These two arenas need to be able to harness the power of databases and information for solutions that can be accessed for varied uses. In discussing governmental needs Heeks (2002) states that one of the more important functions is defining and meeting specific areas concerning the populace. Therefore, by understanding the differing needs of constituents, governments can craft policies and services. Moreover, the problem of information overload requires insights. Information overload represents the volumes of information businesses and managers receive, generate, gather, use, and need in order to conduct operations, deal with competitors and the expectations of consumers (Edmunds & Morris, 2000).

An important revelation at the core of the need equation is the fact individuals have what is known as human information behaviour. This is described "*as the totality of behaviours (active or passive) that people engage in to gain access to, organise and use, information*" (Huotari & Wilson, 2001, p. 23). "*Thus, it will include not only pro-active steps to gain access but also the passive reception of information, which then, or later, turns out to be of use*". The above provides the basis for understanding where and how organisational informational behaviour evolved. The need for information is not new, but rather represents a human and organisational trait that has always been the case. Hence,

> "*organisational (or corporate) information behaviour embraces not only the formal systems set up to manage internal information flows, but also the systems, including libraries and information centres designed to access external information as well as the organisational and personal communication systems through which information reaches the organisation and is disseminated*" (Huotari & Wilson, 2001, p. 23).

The above areas have brought to light that information is a critical driver of the decision-making process in companies. Decision-making in companies is the basis for success or failure as it is dependent on the quality and quantity of information used (Saaty, 2004). Information overload in business can be just as damaging to a company as too little information (Edmunds & Morris, 2000). In terms of understanding the importance and need of a decision support system, it is necessary to recognise it as a function of leadership, management and organisation (Saaty, 2004). Leadership is significant in this process as it is management that makes decisions and is responsible for establishing systems to gather, correlate and put information to use. The decisions made by management impact the success of the enterprise. This includes employee morale, performance, competitive decisions, marketing, financial and other areas. The above means the fates of all organisations (in business or government) are determined by the quality of their decision-making capabilities.

2.4.6. Advantages and drawbacks

The above section provided an understanding of the need and importance of the use of information by management. Some of the benefits a DSS provides are as follows (Delone, 2003):

a. It improves personal efficiencies.

b. The process of decision-making is improved.

c. Organisational control is increased.

d. Decision-makers are encouraged to explore options.

e. Problem-solving in organisations is enhanced.

f. Interpersonal communications are facilitated.

g. Learning and training is promoted.
h. New information is generated to aid in decision-making.
i. Through enhanced decision-making capabilities a company can gain a competitive advantage.
j. New approaches can be uncovered concerning ways to think about problems and challenges.
k. DSS provides a way to automate management processes.
l. It helps to develop innovative approaches and ideas to increase performance.

In order to understand how DSS aids in causing or facilitating the above points, the points can be divided as follows (Lam & Schaubroeck, 2000):

a. DSS solves unstructured and semi-structured problems.
b. The DSS process provides support for managers in the different levels of an organisation
c. The use of DSS supports groups and individuals.
d. Its use also aids in understanding the interdependence and the sequencing processes of decisions.
e. DSS supports the improvement of intelligence, design functions and choices.
f. Its use (DSS) provides organisations with flexibility and adaptability.
g. DSS is an interactive process that results in ease of use for management and staff
h. The process (DSS) enhances efficiencies.
i. DSS also provides for increased user (human) control over processes.
j. Some of the functions of DSS, depending on its application, use modelling.
k. The purpose of DSS seeks to increase use by management and staff.
l. It provides enhanced data access.
m. DSS can be utilised as a standalone application or in a web-based mode.
n. The use of DSS aids in supporting a variety of decision processes.

o. It also supports various decision tree approaches.

The above understandings provide insight into the advantages, capabilities and characteristics of DSSs. As is the case with almost any process, there are drawbacks that have to be considered. The following represent the different areas of possible drawbacks of DSSs (Filip, 2008; Aguilar-Saven, 2004; Wen et al., 2005):

2.4.6.1. Cost

In order to implement a Decision Support System, management needs to make a monetary commitment that, depending upon the size of the undertaking, can be considerable. The process (implementing DSS) means a company needs to invest in the research processes needed to understand what is required. Inherent in the above is determining what data is needed, from what and how many sources and the manner in which they are to be analysed. The pre-system set-up areas which determine what is needed are based on the specifications, and require specialists in most cases (Aguilar-Saven, 2004). Management needs to weigh the end use of the process versus its costs to ascertain its payback potentials and time horizon (Wen et al., 2005).

2.4.6.2. Decision making facets can be over-emphasised

The use and implementation of management and organisations concerning reliance on computerised decision-making can be over-emphasised in many cases. The implementation of DSS may reinforce a rationale or perspective overemphasising the decision process. Before implementing a DSS, process managers need to be educated concerning the broader implications and context. The political, social and emotional aspects are important in helping to determine the success of an organisation (Filip, 2008).

Prior to utilising DSS, a company needs to undergo continuous examination concerning the circumstances in which a DSS should be used and designed (Aguilar-Saven, 2004). In addition, the question as to whether the situation being considered is appropriate to the use of DSS and what type would be appropriate also needs to be ascertained (Aguilar-Saven, 2004).

2.4.6.3. Relevance

A DSS needs to be designed to deal with relevant areas (Wen et al., 2005). This means that unless a proper and though analysis of ramifications, end use and application are conducted beforehand, the system's relevance after the fact may be limited. One of the key drawbacks is that in some cases after a DSS has been installed in an organisation managers use it in an inappropriate manner (Wen et al., 2005). Training may provide a suitable approach to avoid this (Filip, 2008).

2.4.6.4. Power transfer

A misconception that occurs in some situations is that when a DSS system is built the perception is that the authority for decision-making is vested in the computer or software (Aguilar-Saven, 2004). The purpose of a DSS is to improve the decision-making process as opposed to making decisions (Aguilar-Saven, 2004). The human element is still the key in the process using DSS (Wen et al., 2005).

2.4.6.5. Unanticipated effects

Another drawback that sometimes accompanies the implementation of a Decision Support System is the consequences that are not anticipated (Poley et al., 2008). In explaining this, Poley et al. maintain it is possible for a DSS system to actually overload managers with information thereby potentially inhibiting their effectiveness in making decisions. This is associated with the forward planning, thinking and expertise used to devise the system which needs to be carefully planned and understood in terms of end use (Poley et al., 2008).

2.4.6.6. Obscured responsibility

In terms of understanding the use of DSS, managers and people using the system need to be aware of its end use. Computers and systems do not make bad decisions; it is rather people who provide bad or insufficient information to the system (Poley et al., 2008). This represents a key drawback in using a DSS system as management needs to understand and ensure the data input is current and relevant to the situation at hand (Harrisonet al., 2007). In using a DSS system, management and those using the system sometimes forget that it (DSS) represents an intermediary between those who build the system and those using it (Filip, 2008).

2.4.6.7. Objectivity

Another factor that can be a drawback in the use or implementation of a DSS system is what is termed as a false belief in objectivity (Poley et al., 2008). This is explained as the people using the system need to be trained and educated in terms of understanding the nuances and philosophical aspects of objectivity (Harrisonet al., 2007). Whilst software can provide managers with encouragement to take or engage in more rational action, managers can mistakenly use the system to rationalise their actions, which is a mistake (Harrisonet al., 2007).

2.4.7. Presence in companies

In terms of equating the use of DSS in companies, the two categories described in the above section (History of development) are used. These represent model-driven and data-driven examples. Under the model-driven approach. some of the main approaches used under DSS entail decision analysis, intelligence management, predictive modelling and decision management (Kopackova & Skrobackova, 2003). In a report by Oracle (2014), it is shown that in excess of 85% of U.S.-based Fortune 500 companies use what are termed as Crystal Ball products. This is one of the leading spreadsheet applications used for forecasting, predictive modelling, optimisation and simulation. It is a DSS software type that provides businesses with insight concerning varied factors that affect risk that permit management to make tactical decisions important to achieving objectives (Oracle, 2014). Another DSS application was developed by Frontline Systems that uses a system of solvers and optimisers used in Quattro Pro, Excel and Lotus 1-2-3 which are used by consumers, accounting professionals and companies internationally.

Another type of model-driven DSS software is MeetingWorks. It is a software application that aids company managers in organising and conducting meetings that are more productive and streamlined (Antunes & Carrio, 2003).

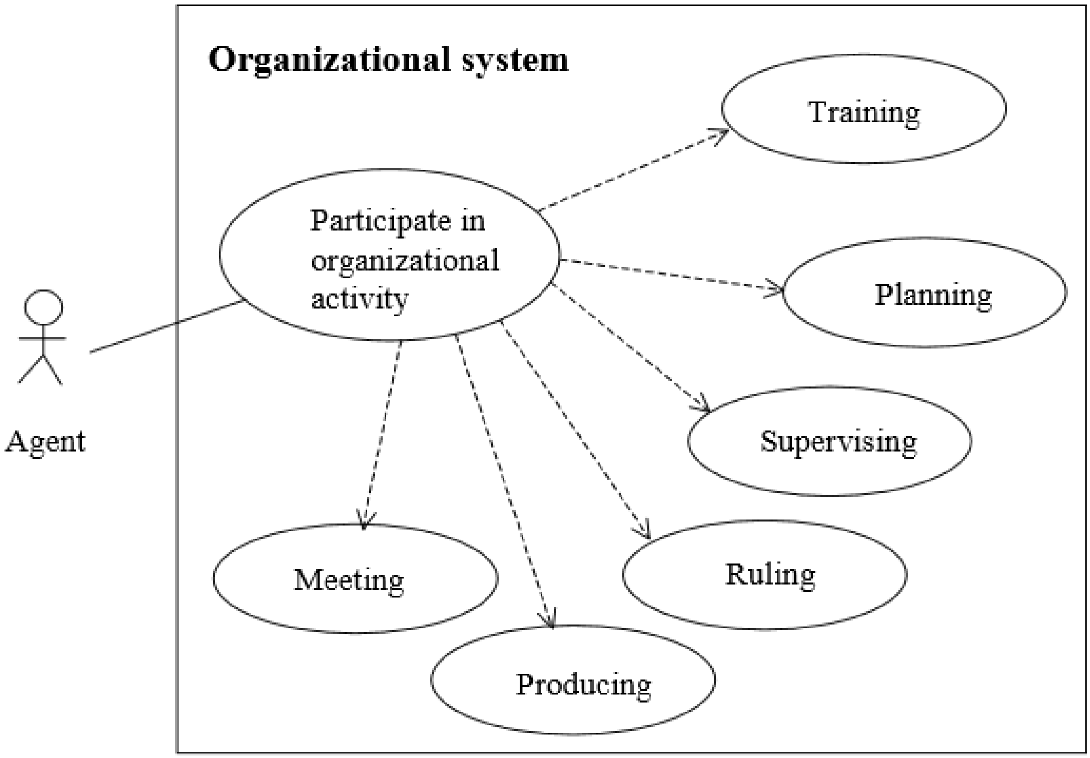

Figure 10: Organisational systems application diagram (Antunes & Carrio, 2003)

The list of users of MeetingWare is too extensive to be listed here as the application is broadly distributed in various types of applications (Antunes & Carrio, 2003).

Data-driven DSS applications are used in companies and large corporations as (Hedgebeth, 2007). One of the largest vendors of data-driven DSS systems is Cognos (Power, 2008). The company's BI/OLAP (Online Analytical Processing) application is a business intelligence (BI) software tied to OLAP, which is a technology used to conduct queries in statistical databases (Jovanovic et al., 2012). The success of OLAP is dependent on the design of multidimensional databases (Jovanovic et al., 2012). Cognos' BI/OLAP application serves over 22,000 corporate customers in virtually all countries (Power, 2008). The business solutions offered by Cognos aid businesses in understanding, managing and monitoring their performance. The software also includes business analysis and reporting, the measurement of profitability, optimisation of forecasting and cost management. Cognos provides an efficient means to aid in the delivery of intelligence data for business. It uses data warehouse information to construct replies to data queries.

Another DSS software application is Brio BI that is distributed by Applix. It has an excess of 2,600 customers (Power, 2008). In elaborating on various data-driven DSS systems, Power (2008) adds:

> *"ESRI is a leading developer of GIS software with more than 300,000 clients worldwide. Hyperion products are used by more than 6,000 customers around the world to enable financial, organisational, customer relationship, supply chain and channel performance management. Information Builders has more than 11,000 customers, including most of the Fortune 100 and U.S. federal government agencies".*

The above examples of DSS model-driven and data-driven applications used in the consumer and business arenas provide extensive proof concerning its broad use.

2.4.8. Presence in public organisations

The examples concerning the use of Decision Support System applications in public organisations is widespread (National Forum for Educational Statistics, 2006). In an extensive report on the use of DSS in public organisations, the National Forum for Educational Statistics (2006) states that various forms are being used in the United States' educational system. The uses range from the administration of personnel, accounting functions, course curriculum design to teacher performance comparisons in districts, states and regions (National Forum for Educational Statistics, 2006). In another report, it is argued that the use of DSS in the public sector aids organisations to utilise data along with models to uncover and identify problems as well as solving them (Bencina, 2006).

The typical use in public organisations entails aiding administrators to use and manipulate data, use checklists, along with building and using models. The application of DSS in the public sector entails cooperative measures to build and form a consensus as opposed to the authority driven approach used in the private sector (Bencina, 2006). This is accomplished by the following steps and stages (Bencina, 2006):

2.4.8.1. First order effects

These aspects entail the following:

1. Social Capital - This represents the building of trust and relationships.

2. Intellectual Capital - The consensus environment of the public sector means arriving at mutual understandings, agreed upon approaches, along with problem frameworks that are shared.
3. Political Capital - This entails forming alliances to work together to arrive at or reach ends mutually agreed upon.
4. High-quality agreements.
5. The use of innovative strategies.

2.4.8.2. Second order effects

These aspects represent the following:

1. Building of new partnerships.
2. Joint action and coordination.
3. The implementation of agreements.
4. Undergoing changes in practice.
5. Understanding that changes in perceptions will be needed.

2.4.8.3. Third order effects

1. New collaborations.
2. Co-evolution represents a core operative aspect as opposed to conflict.
3. New institutions.
4. New norms.

The above building blocks are critical aspects concerning making changes in the public sector (Bencina, 2006). DSS in the public sector can be found in all segments. Whilst the terminology will differ for various countries, the general parameters are the same in terms of application.

2.4.8.4. Departmental digital boards

This represents the use of DSS applications by executive boards to handle transactions under various agencies (finance, health, procurement, etc.).

2.4.8.5. Civil service applications

In the instance of Australia, this represents a government-wide initiative to ensure that all civil service departments have the needed levels of DSS software capabilities.

2.4.8.6. DSS applications in government departments

The following represent the specific Australian government departments or agencies DSS is operational in:

1. HM Revenue and Customs
2. Department for Transport
3. Department for Work and Pensions
4. Ministry of Justice
5. Department for Business Innovation and Skills
6. Department for Environment Food and Rural Affairs
7. Home Office.

The above represents the use of DSS in the Australian government whose use examples are similar to those of France, the United States, the UK and other countries. In order to understand the differences in the application of DSS in private and public organisations, the following points are relevant (Bencina, 2006).

2.4.8.7. Decision making

1. *Private Sector*

Decisions in this sector are made either by an individual or management teams under the authority of the organisational structure.

2. *Public Sector*

Interestingly, decisions in the public sector happen due to a series of complex interactions that occur between trade unions, administrators, pressure groups or other inputs.

2.4.8.8. Decision interests

1. *Private Sector*

In the private sector, decisions are usually dominated by a singular interest that in general is represented by the position of the company competitively.

2. Public Sector

As governmental agencies exist to serve the needs of the public, the interests of society represent the factor.

2.4.8.9. Decision alternatives

1. Private Sector

These are approaches that are evaluated based on sets of quantitative criteria (economic) represented by aspects such as bottom line performance, profits, market share and allied considerations.

2. Public Sector

The evaluative process differs in the public sector as the considerations are broad and comprised of quantitative and qualitative criteria. The values concerning the above are considered as being difficult to establish.

2.4.8.10. Decision horizons

1. Private Sector

The typical planning horizon tends to be months to a number of years. The differences are based on the area, such as marketing decisions (which change rapidly) and new product introductions or entering new markets (which can extend to years).

2. Public Sector

The factors representing changes in the public sector are very slow to evolve. The typical planning horizons are years and typically can represent decades.

The above aspects have been included to offer insights on the use of DSS in the public sector.

2.4.9. Presence in the United Arab Emirates public sector

In terms of identifying the use of DSS in the public sector of the United Arab Emirates (UAE), e-government may be the place to uncover such a use. The UAE e-government represents *"… activities that take place over electronic communications among all levels of government, citizens, and the business community, including: acquiring and providing products and services*; *placing and receiving orders; providing and obtaining information; and completing financial transactions*" (Riad et al., 2011, p. 124). The key element in the above definition is that e-government utilises DSS as its structural foundations. E-government offers the most cost-effective and efficient use of computer aided systems as it drastically reduces installation, communication and information exchange costs. Whilst Western developed countries were able to convert long-standing computer systems to e-government platforms, countries in the Middle East lacked the infrastructures for e-government and have embraced it as the means to introduce the efficiencies and cost saving measures of computer aided systems employing DSS. In the Middle East, most governments are using the installation of DSS service-oriented architecture as the platform to link and integrate services along with applications between different agencies and ministries (Riad et al., 2010).

The use of DSS provides Middle East countries with support for unstructured and semi-structured uses (Riad et al., 2011). This includes the combination of computerised information outputs and human judgment factors where DSS provides support for varied governmental levels (Figure 11).

Figure 11: Figure 8 - DSS Government Support (Riad et al., 2011)

DSS components used in the Middle East usually consist of "*…(i) database management subsystem (DBMS), (ii) model base management subsystem (MBMS), (iii) knowledge-based (Management) Subsystem, and (iv) User interface subsystem (Dialogue)* …" (Riad et al., 2011, p. 126).

The above understanding is critical in understanding the use of DSS in the United Arab Emirates. A recent study conducted by the United Nations (2010) ranked the United Arab Emirates as 49th internationally concerning its use and sophistication in e-government development (Riad et al., 2011, p. 126). Despite this, the country was rated as 86th in terms of e-participation and 99th in its use of online services (Westland

& Al-Khouri, 2010). In providing a base to correlate the standing and development of e-government in the United Arab Emirates, Westland and Al-Khouri (2010) used the developed system in the United Kingdom to illustrate the potential benefits:

Table 2: E-Government Benefits / UK Example (Westland & Al-Khouri, 2010 p.3)

Public Sector *transaction with*	***Examples***	***Benefits***
Citizens	Information Culture Health Education Benefits transactions Taxation	Wider choice of channels, Convenience, lower transaction costs, more personal service, greater awareness of services and policies, greater democratic participation and openness
Business	Support programs Advice and guidance Regulation Taxation	Quicker, faster interactions, reducing transaction costs, and the regulatory burden
Suppliers	e-procurement	Reduced transaction costs, better inventory management, shared data environments
Other public sector bodies	Communication between departments and agencies between central and local government Policy making	Greater accuracy and efficiency, reduced transaction costs. Better use of the knowledge base. More nimble, flexible working arrangements.

The UAE has implemented a card system under its Identity Management Infrastructure (IMI) to aid in access and control of the system (Westland and Al-Khouri, 2010) (Figure 12).

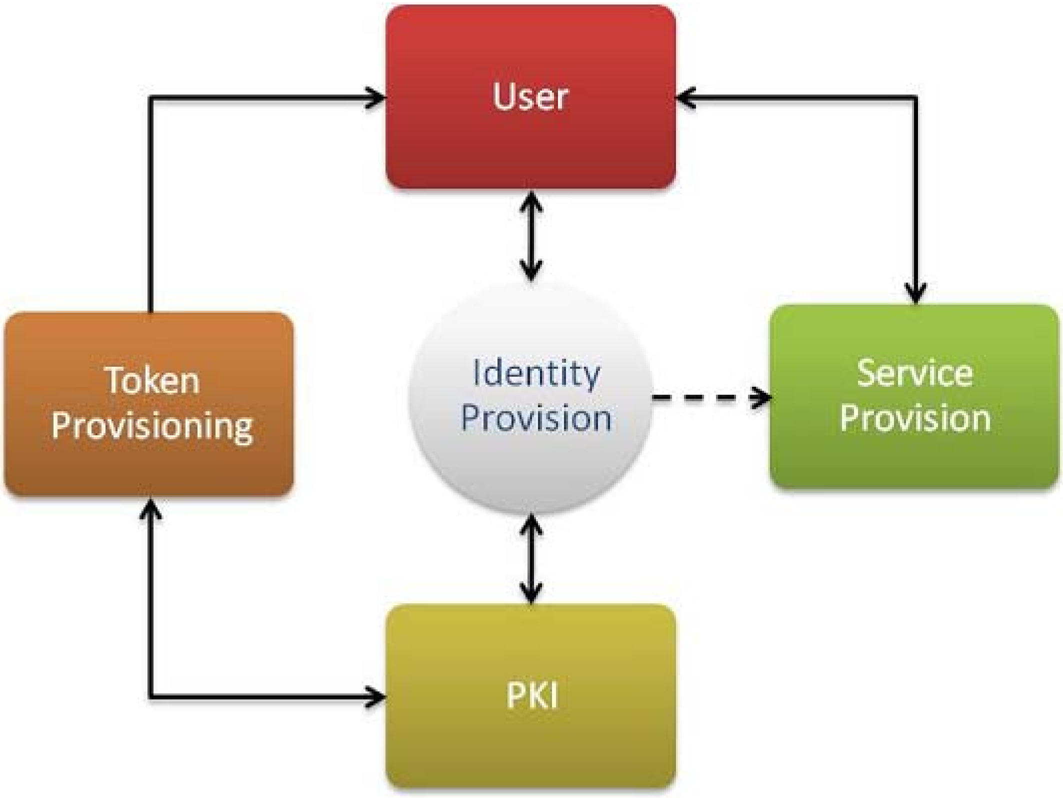

Figure 12: The UAE Identity Card System (Westland & Al-Khoruri, 2010)

Citizens use the government system to enable tracing user interests, inquiries and allied aspects that are aided by DSS applications (Figure 13):

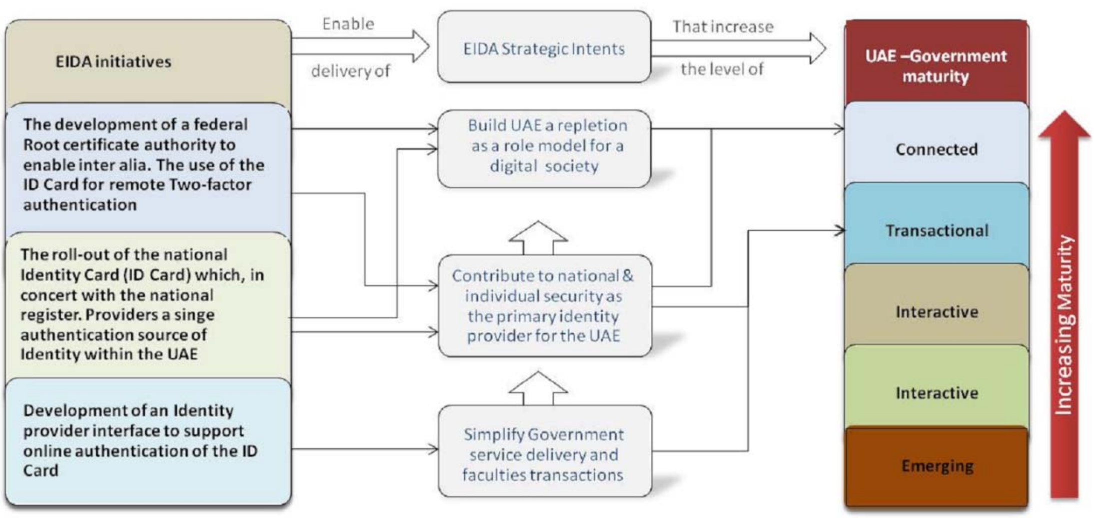

Figure 13: The UAE e-Government System (Westland & Al-Khouri, 2010)

In terms of services under e-government in the UAE, they mainly consist of a broad range of services for citizens, and businesses across varied federal and local governments. These include the following (Al-Hujran, 2012).

2.4.9.1. eServices

The services offered through the UAE government portal are available for businesses, citizens and visitors.

2.4.9.2. mServices

This represents services available through mobile devices.

2.4.9.3. eParticipation

In order to make government services interactive, users can participate in various forums such as chats, blogs, social media and polls.

2.4.9.4. Open data

This provides information on economic areas, various statistics and other information.

As shown above, the UAE has a sophisticated and developed public sector e-government service range that is still improving.

2.4.10. Examples of uses of DSS by organisations worldwide

The examples of DSS usage in the private sector are many. In a case study of DSS use by the East of England Observatory, a website is maintained by its Policy Unit which is available for organisations to use that are in partnership with the agency (Databeacon Staff, 2004). The DSS solution used in this instance represents Databeacon which was developed by Matraxis Ltd (Databeacon Staff, 2004). The application provides the East of England Observatory with a consultation, support and training solution for regional economic indicators (Databeacon Staff, 2004). It offers government agencies and businesses a data architecture that provides analysis and reporting tools that can be used online (Databeacon Staff, 2004). The system has in excess of 1,000 users and provides information to conduct business activities, secure data on the latest regulations, planning and other information to aid in decision making (Databeacon Staff, 2004).

The California Department of Motor Vehicles uses a DSS application that consolidated and streamlined its fee collection and registration system (Fair Isaac, 2006). The DSS application used is a legacy system that is aligned with the state's e-government solution which operates over the Internet (Fair Isaac, 2006). The use of a Business Rules Management Software (BRMS) enabled the Bureau to devise a vehicle registration fee system that updates vehicle fee rates and penalties, along with corresponding rules *"… for autos, commercial vehicles, trailers, motorcycles and off-highway vehicles*" (Fair Isaac, 2006, p. 1). The system has provided the Bureau with precision in implementing new policies and legislation along with the appropriate fees across the wide range of vehicle types (Fair Isaac, 2006). More importantly, the system is highly cost effective as the DSS architecture delivers solutions to all offices and branches (Fair Isaac, 2006).

In examining a business use of DSS, the General Electric (GE) Real Estate division employed it to aid in driving global growth across the division's offices in twenty countries (Power, 2007). The DSS solution was used to enable risk, sales and property management divisions to assess loan profitability, and the likelihood of loan approval (Power, 2007). The property management division uses the system to understand the risks on outstanding loans (Power, 2007).

In another example, Schwartz (2005) discusses the use of a DSS application for the Coca-Cola West Japan Company. The firm uses a key business indicator approach that measures vending machine operational performance (Schwartz, 2005). Other areas and aspects it uses include the replenishment rate, the planning of replenishment visits, and out-of-stock incidences (Schwartz, 2005). These indicators are utilised to analyse data for the headquarters staff at the Coca-Cola West Japan Company concerning vending machines along all of the company's routes (Schwartz, 2005). By being able to analyse vending machine route data, along with increased or declining usage the Coca-Cola West Japan Company was able to optimise route placement concerning its vending machines and reduce operational costs (Schwartz, 2005).

These examples represent a few of the multitude of business and government uses for DSS. As indicated under the section 'Presence in companies', there were 22,000 corporate applications of BI/OLAP used in virtually all countries and over 6,000 uses of GIS (Geographical Information System) software (Power, 2007). These represented a small number of software examples in use, and provide an indication of the extent and exposure of various DSS applications.

2.4.11. Examples of success and failure of DSS

The 1986 space shuttle disaster is a drastic example of a failure of a DSS application. The space shuttle Challenger exploded shortly after it was launched; the investigation uncovered a failure in the use of NASA's Group Decision Support System (GDSS)

(Forest, 2005). The above was traced to a flawed GDSS database that was mismanaged (Forest, 2005). Whilst it was known that there was a flawed 'O' ring in the solid state booster of the shuttle, the decision to launch was a human decision based on inaccurate information. The operational decision to launch was based on a default system that ignored the known fact that the 'O' rings were indicated to have issues operating in cold temperatures. This environmental input was not considered in the launch decision that used the GDSS which did not seek or require a complete engineering sector input approval (Forest, 2005). As a result, the decision to launch was attributed to human error that failed to take into account all variables (Forest, 2005).

A subsequent investigation, that took years, indicated the space shuttle program suffered from budget cutbacks that affected the development and implementation of an effective GDSS mechanism. As indicated in prior segments of this study, the effectiveness of any DSS application is dependent on effective planning and understanding of the contributing variables along with end-use purposes. The failure to adequately account for all of the needed inputs and outcomes along with their consequences was not a part of the abbreviated NASA budget for developing its GDSS software (Forest, 2005).

In an example of the successful uses of DSS, the prior section discussed examples in the private and public sectors. Other success examples include the University of Alberta that used a document-driven version of DSS to post its policies and procedures online (Stellent, 2004). The benefits of this approach enabled real-time information to be posted without the delays of the printed versions used in the past (Stellent, 2004). In another business example, Briggs and Stratton used an SAS version of a DSS application to increase operational efficiencies across its varied departments. The changeover to a DSS-based system significantly reduced the company's global reporting costs and increased the speed of data availability and use (Nelson & Wright, 2005).

Another example of a failed DSS application is the case study of the use of a DSS in the University of Pittsburgh Medical Centre. The system was developed to become a clinical decision support system (CDSS) model designed to ensure compliance with Congestive Heart Failure Core Measures. The system developed what are termed as identifying rules and triggers for treatment measures (Wadhwa et al., 2008) (Table 3):

Table 3: Congestive Heart Failure Core Measures – Rules and Triggers (Wadhwa et al., 2008)

History Rule	IF a past history of CHF was recorded in the Problem List or CHF was a diagnosis in a prior Discharge List
Chest X-ray Rule	IF a Chest X-ray was ordered within 28 days of an echocardiogram (ECHO)
ECHO Rule	IF an ECHO was ordered within 7 days of a Chest X-ray
IV Lasix Rule	IF a single dose of intravenous furosemide or bumetanide was ordered
AICD Rule	If an order for AICD was entered.

The problems encountered in the use of the clinical decision support system (CDSS) model were as follows (Wadhwa et al., 2008).

2.4.11.1.False negatives

There were instances where patients were not entered into the clinical decision support system (CDSS) model, thus the rules or triggers that were used could not be identified.

2.4.11.2.Excessive alerts

One of the issues with the clinical decision support system (CDSS) model that physicians identified is that it generated too many alerts. This caused some staffers to stop using the system.

2.4.11.3.Incomplete physician alert responses

An examination uncovered that in some instances the Congestive Heart Failure Core Measures forms generated as a result of alerts were completed, but the corresponding order sets were not.

The failure in researching and planning the system using test runs and result correlations to work out issues and problems was the reason the system failed (Wadhwa et al., 2008).

The use of DSSs in organisations in general and in public organisations in particular is further investigated in later chapters. In particular, later chapters look into ways for improving decision support systems learning from the success and failure aspects introduced in this section.

2.5. Data Mining

2.5.1. Data mining

Data is the raw form of information. They are numbers, figures, texts, etc. that do not usually hold meaning. Data attributes are the description values of data. They are also known as dimensions, features or variables, e.g., name, address, phone number, etc. The different types of data attributes are as follows:

- Nominal: categories, states, e.g., hair colour: {blond, red, black, white, brown}
- Ordinals: the order matters, but the value that represents each category is not known, e.g., Height: {short, medium, tall, very tall}
- Numerical, e.g., salary, temperature
- Discrete (finite number of states) versus continuous (usually represented by a real number).

Data mining, also known as Knowledge Discovery in Databases (KDD), involves the extraction of interesting patterns or knowledge, usually from large volumes of data. An interesting pattern is non-trivial, which is an implied relationship or association between variables or data items previously unknown. Data mining is also defined as the process of selection, exploration, and modelling of large quantities of data to discover regularities or relations that are at first unknown, with the aim of obtaining clear and useful results for the owner of the database. Data mining provides an increasing potential to support business decisions, end users, business analysts, decision-making, and data presentation techniques.

Data mining currently receives a topical interest not only from computer professionals but also from academics and business managers. Storage media can store more data, and therefore volumes of stored data, whether useful or not, are constantly increasing. For instance, chain stores save large amounts of data about their customers in databases. Among these amounts are usually large volumes of interesting marketing information, such as data available about items purchased, the time of purchase, the customer’s age, etc. Digitisation of analogue phone networks has made it possible to store phone calls for future improvements and quality results, as well as anticipating future incidents based on trends. This can be of great help for police emergency call centres, fire brigades, and other related departments. The sizes of databases for banking transactions are enormous, especially after introducing smart cards and online wallets. This huge amount of data needs to continuously maintained and filtered. These are just a few examples from practice of the increasing volumes of stored data. Rich sources of information for decision-makers are usually dormant and hidden under an avalanche of data. A company would be managed more effectively, if they knew their customers' buying habits, their preferences, their ages, their contact details, and much more information. Therefore, there is a missing link that can connect the existence of large amounts of data and the relevant information that exists in this data.

Data mining is precisely the coupling element, which allows users to filter their huge amounts of data. The purpose of data mining is to discover and display certain patterns in data that can be used to solve the problem (Han etal., 2006). In this respect, showing only the results of computer analysis of large amounts of data is just something like the tip of an iceberg. Analytical data extraction combined with powerful visualisation techniques generally leads to different views of management of the business problem. Examples can be discovering a new correlation between attributes, predicting the future of the data collected in the past, and conducting binary operations on the differentiated datasets.

Organisational Data Mining (ODM) is another type of data mining. Nemati and Barko (2003) define ODM as being "*used to leverage data mining tools and technologies to enhance the decision- making process by transforming data into valuable and actionable knowledge to gain a strategic competitive advantage*." ODM covers a wide array of technologies, including but not limited to e-business intelligence, data analysis, SQL, customer relationship management (CRM), eCRM, EIS, digital dashboards, information portals among others.

In conventional database queries OLAP, a relationship between data elements must be specified. Data mining can discover relationships that the user does not see or did not even suspect existed (Tan et al., 2005). Sometimes the user knows about a relationship yet cannot properly formulate a query and consequently look for an answer, such as searching the Internet by knowing only two keywords among hundreds of thousands of pages.

2.5.2. Types of data mining problems and associated techniques

In general, the triggers of a data mining project involve a combination of different types of problem, which their solutions together solve a certain business problem. Different techniques of data mining exist; each has different applications. It is often the case, however, that more than one technique is used in combination on a dataset. The following is an overview of the common techniques used for data mining, addressing different problems.

2.5.2.1. Visualisation

Data warehouses and tools to use them exist at different levels. Data mining can be done from a simple query on a database through the creation of tables from the stored data to the visual analysis of data from several databases. The first stage starts by setting simple questions, short statements, small tables and conducting analysis (Keim, 2002).In a simple form, a typical computer can display a table or 3D chart with simple analysis. The highest degree is a 2D or 3D visualisation of stored data. Visualisation is important because the graphic design of data is intuitive, more acceptable, faster understood and better remembered by humans. Human senses perceive similarities and anomalies much faster in data that is displayed in a graphical form than in a tabular form, for example, it is easier to remember the faces of people than their names. Visualisation does not mean fancy views of what is written in a table. Generally, it is a reflection of a multi-dimensional problem space into three-dimensional space with computationally intensive data analysis, which provides a view from another angle. In a 3D model, it is naturally easier to efficiently analyse a very complex relationship. To facilitate analysis of trends in a three-dimensional space, it is can be extended by a fourth dimension to the possibilities of animating objects over time (Kohavi, 2000).

Visualisation methods are used because people need to make decisions quickly and efficiently. Here, just one user conducting advanced computing and visualisation operations using high-quality programs for extracting data could find very interesting and effective business results.

2.5.2.2. Data description and summary

The description and summary of data usually point to the concise description of its characteristics, typically in a basic and aggregate form. This provides a description of the data structure. Sometimes, a description and summary of the data is the only target of a data mining project. For example, a retailer might be interested in the sales volume of all outputs split into categories. The differences and changes occurring in the past would be highlighted and summarised. This type of problem would be a low-scale data mining problem.

In almost all data mining projects, however, the description and summary of the data is a minor goal in the process, usually at an early stage. When the data mining process has started, users are not usually well aware of the precise purpose of analysis or exact nature of the data. Initial insights into of data analysis may help users to understand the nature of the data and assume possible hidden information. Simple descriptive statistics and visualisation techniques provide the first ideas about the data. For example, the distribution of clients by age and geographical region suggests that parts of a group of player need to be addressed for future marketing strategies (Hand et. al., 2001).

The description and summary of data typically occur in combination with other data mining problems. For example, the description of data can indicate the nomination (presumption) of important data segments throughout. When these segments are identified and specified, a summarised description of them will be useful for later analysis. It is desirable to conduct a summary and description of data prior to specifying any data mining problem. This is due to the fact that data summary and description is a task in the pre-phase of data compression (Han et al., 2006;Chapman, et al., 2000).

Summary also plays a key role in displaying the final results. Other data mining problem results (for example prediction models or descriptions of concepts) may also be considered, but at a higher conceptual level.

Many statistical packages and reporting systems, such as the online analytical processing (OLAP) system, may cover summary and description of data, but often do not provide methods for advanced modelling. Such tools may be appropriate to carry out data mining if summary and description of data is considered as a separate type of problem and future modelling is needed (Chapman et al., 2000).

2.5.2.3. Segmentation

This technique of data mining mainly aims to split the data into significant classes or subgroups. The members of the same subgroup share common characteristics. An example of segmentation is defining segments of shopping baskets based on the items in these baskets. Segmentation may be performed semi-automatically or manually. The data mining specialist may consider certain subgroups relevant to the business problem based on prior knowledge or experience or according to the results of the summary and description operations on data. Automatic segmentation, known as clustering, can discover hidden and unsuspected structures in data that allow segmentation (Tan et al., 2005).

Segmentation may be a target of data mining sometimes. In this case, segment detection is the main aim of the data mining project. An example of this is when all addresses in certain areas of elderly people and average income would be selected to send advertisement for home care. However, segmentation is often a step to solutions to other problems. Hence, the main aim is to maintain a manageable data size or find homogeneous data subsets that are relatively easy to analyse. In large datasets, the scope of each interesting pattern varies. Choosing the most suitable segmentation hence facilitates the task. For instance, analysing the dependencies among items in millions of shopping carts is an extremely difficult task. However, the task becomes easier if instead, dependencies and interesting segments of shopping carts are identified, such as baskets containing assets of comfort, high-value baskets or baskets of a particular period (Chapman et al., 2000).

Example of automatic segmentation or clustering

A property agency regularly collects information about its customers regarding socioeconomic status, such as income, occupation, sex, age, and others. By cluster analysis, the agency may divide its customer group into more understandable subgroups and analyse the structure of each subgroup. Hence, controls of marketing strategies can be developed separately according to each studied subgroup (Chapman et al., 2000).

2.5.2.4. Description concept

The description concept relates to understandable description of classes or concepts. The aim is to attain high accuracy of the development of prediction models and gain new ideas. For instance, a company may want to know its customer loyalty and disloyalty. A description concept based on these attributes, loyal and disloyal, will help the company establish how to transform disloyal customers into loyal ones (Lin et al., 2008).

A description concept is closely related to both classification and segmentation. Concept descriptions may be used as classification. Furthermore, some classification techniques generate understandable classification models, which may be considered as descriptions of the concept. The main difference, however, is that classification aims to be complete. Classification has to be applied to all cases in the target population. However, the concept description model does not necessarily have to be complete. It is enough to describe only important parts of the concepts or classes. In the above example, it may be sufficient to get the descriptions of clearly loyal customers.

Segmentation on the other hand can lead to a list of objects belonging to a class or concept without providing any understandable description. Typically, segmentation is conducted before description is made. Some techniques, for example, conceptual clustering, may run concept descriptions and segmentation at the same time.

2.5.2.5. Classification

The classification technique considers a group of objects characterised by certain characteristics as belonging to the same class. Each object has a class label of discrete value. The aim of is to construct a classification model (also known as a classifier) which assigns to each object a correct class label (Fayyad et al., 1996). The class labels may be previously defined by the analyst or may have arisen from segmentation. Classification is one of the most significant types of data mining techniques used in many applications. In fact, many data mining problems may be transformed into classification problems. For example, the problem of evaluating the credit risk of a new customer can be transformed into a classification problem of two classes: good and bad credit customers. The classification model can be generated based on credit risk data of existing customers. The generated model can then be used to allocate new customers to the two classes and accordingly reject or accept them.

Classification is connected to almost all other types of data mining technique. For example, a prediction problem can be transformed into a classification problem by discretisation of continuous class labels. Discretisation transforms continuous intervals into discrete values. Discrete intervals produce approximate numerical values and can be used as class labels, hence leading to a classification problem. Description may be also transformed in a classification problem by producing an understandable class or concept. Classification also relates to the dependency analysis by clarifying the dependencies among attributes. Segmentation may also produce class labels or restrict datasets to classification models. It is also beneficial to analyse deviations with classification. Deviations can clarify patterns which would allow a good classification model. Conversely, a classification model can also be used to identify deviations.

Example of classification

A bank can evaluate credit risk of a new customer by classifying two classes: good and bad credit customers. The classification model can be generated based on credit risk data of existing customers. The generated model can then be used to allocate new customers to the two classes and accordingly reject or accept them (Chapman et al., 2000).

2.5.2.6. Prediction

Prediction is an important technique that occurs in a many data mining applications. Prediction is similar to classification with the only difference that the target attribute of prediction is continuous not discrete (Witten, 1999).

The aim of prediction is to find the numerical value of unseen objects with the target attribute. This type of problem is sometimes referred to as regression. It is also called forecasting if it is associated with time series data.

Example of prediction

The annual income of a company is correlated with other attributes such as exchange rate, promotion, inflation rate, and others. Having these values (or accurate estimates), the company can predict its expected revenue for the next year.

2.5.2.7. Dependency analysis

Dependency analysis is used to find a model which describes significant dependencies (or associations) among data items. Associations or dependencies may be used to predict the value of one data item based on information on other data items. Associations describe the affinities of data items (i.e., data items that often occur together). Although units can be used for dependency analysis, units are mostly for understanding only. A typical application scenario of dependency analysis or association is the analysis of shopping baskets. A rule like "in 25% of purchases, peanuts and beer have been bought together" is a typical example for association. There are relatively efficient algorithms for detecting different associations in a set of data. However, selecting the most interesting association is the real challenge. (Shannon, 2002).

Dependency analysis has a close relationship to classification and prediction, as associations are implicitly used to formulate classification and prediction models. Dependency analysis is also linked to concept descriptions, which often highlight dependencies. In applications, dependency analysis often co-occurs with segmentation. In large datasets, results of dependency analysis may not be significant because much noise occurs among the data items. In such cases, it is desirable to perform dependency analysis after segmentation has taken place.

The sequential dependency model is a special type of dependency where the order of events is relevant. For instance, in the shopping basket example, the associations describe dependencies between items at a given point of time, whereas in the sequential pattern model, the purchase of a particular customer or a group of customers is described over a period of time.

2.5.2.8. Clustering

Clustering is an advanced data mining technique that is used to separate objects or observations into groups, where objects in one group are similar to one another and different from objects in different groups.

Many data mining applications require partitioning data into homogeneous groups (or clusters) in order to discover interesting information from which to produce policy, for example, partitioning the customers of a bank into a fixed number of clusters.

However, there are at least two reasons why it is not possible to extend the traditional methods of cluster analysis to data mining applications (Kerby, 2009):

1. Traditional methods of cluster analysis are not efficient when databases have huge dimensions of terabytes size characterised by thousands of records and tens or hundreds of attributes.
2. Traditional methods of cluster analysis do not require the presence of categorical variables with the subsequent special treatment of them and therefore the traditional way of dealing with categorical variables as numeric

variables does not always produce significant results, including the fact that the modes of categorical variables are not ordered.

Even a method that allows the solution of one of the above problems may not suggest the possibility to extend to the other. For example, methods based on the concept of *k*–means are efficient for large datasets but have the disadvantage of being limited to numeric data. They, in fact, consist of algorithms that aim to identify homogeneous clusters by minimising a particular cost function defined by the average Euclidean distance between points in the dataset, where *k* are the points in a *k*–dimensional space.

There are clear limits to the possibility of using these algorithms for categorical variables. However, the algorithms based on the *k*–means concept are a good starting point for building algorithms that also handle categorical data, while maintaining efficiency. In this context, an algorithm can be developed based on the *k*–medians algorithm, capable of handling common categorical variables in data mining, which leads to the algorithm of *k*–means when considering numeric variables only.

Example: Data mining clustering with categorical variables

Let $X = \{X_1, X_2, \ldots, X_n\}$ be a set of n objects and $x_i = \{x_{i1}, x_{i2}, \ldots, x_{im}\}$ represent the attributes of object X_i.

A clustering operation denoted by a positive integer *k* is to find a partition that divides the objects in *X* into *k* disjoint groups. One possible way to achieve this is to analyse all possible partitions in order to find the one best suited for a classification problem. This will result in a very large number of possible options. If *n* objects have to be divided into *k* groups with n_1 objects in the first group, n_2 objects in the second group, and so on, up to n_k objects in the k^{th} group, where $n_1 + n_2 + \ldots\ldots + n_k = n$, then there are $\frac{n!}{n_1!\, n_2! \ldots\ldots n_k!}$ ways to do that.

However, the common solution in these cases is the choice of a grouping criterion to guide the user in searching in the set and then the groups that stem from it. The grouping criterion used to define a cost function, for example, can be as follows (Huang, 1997):

$E = \sum_{l=1}^{k} \sum_{i=1}^{n} y_{il} d(X_i, Q_l)$ where $Q_l = [q_{l_1}, q_{l_2}, \ldots\ldots, q_{l_m}]$ is a representative vector or prototype for cluster l, y_{il} is a generic element of a partition matrix $Y_{n \times l}$, and d is a measure of similarity, often defined as the Euclidean distance, in this case, m-dimensional space. d is defined as: $d(X_i, Q_l) = \sum_{j=1}^{m_r} \left(x_{ij}^r - q_{lj}^r \right)^2 + \gamma_1 \sum_{j=1}^{m_c} \delta\left(x_{ij}^c, q_{lj}^c \right)$. The partition matrix Y has the following properties:

i. $0 \leq y_{il} \leq 1$

ii. $\sum_{l=1}^{k} y_{il} = 1$

If $y_{il} \in \{0, 1\}$ Y is called a hard partition or a fuzzy partition otherwise. m_r and m_c are the number of numeric and categorical attributes respectively. x_{ij}^c and q_{lj}^r are the categorical values, γ_1 is a weight on the categorical values and $\delta(p,q) = \begin{cases} 0 & if \quad p = q \\ 1 & if \quad p \neq q \end{cases}$

.

The matrix Y of size $n \times k$ and any element y_{il} is the matrix of partition, since each matrix is defined as single partition of n objects into k groups.

Clustering high-dimensional data, encountered in many applications such as mining textual data resulting dictionary-size dimensions, is believed to be a computationally hard problem (Zeng & Cheung, 2008). In spite of the existence of a large number of proposed solutions to clustering high-dimensional data, most of these solutions compare a new proposition with one or two competitors, or even with a so-called “naïve” ad hoc solution, but fail to clarify the exact problem definition. As a result of that, it is often unclear whether the compared solutions tackle the same problem (Kriegel et al., 2009).

A major challenge in the clustering area is dealing with high dimensional data. Dimensionality reduction is one of the widely used techniques for dealing with high dimensional data. It avoids the curse of dimensionality: The more data we have, the sparser the database is and consequently it is more difficult to cluster. Dimensionality reduction helps reduce the number of attributes, remove noise, reduce the time required for mining and facilitate data visualisation. Among the used dimensionality reduction techniques are: Wavelet Transform – PCA (Principal Component Analysis), Attribute Selection, Greedy Selection, Greedy Elimination, Breadth-based Selection and Evolutionary Algorithms.

Dimensionality reduction comprises the possibility of information loss. Dimensionality reduction can discard useful instead of irrelevant information (Law, 2006). Sometimes as little as less than 1% of features are used to discriminate data with high dimensionality (De Oliveira & Pedrycz, 2007). Dimensionality reduction is not the only approach to handling high dimensional data. The naïve Bayes classifier has found empirical success in classifying high dimensional datasets (Law, 2006). However, the naïve Bayes classifier may result in skewed probability estimates due to the class conditional assumption. Moreover, dependencies among variables cannot be modelled with the naïve Bayes classifier (Neville, 2010).

2.5.3. Features innovative programs for data mining

Modern operating programs include parallel algorithms for data mining, which significantly accelerate the computationally demanding process. Other components are analytical tools such as regression, clustering (data division into similar groups, such as identifying customer segments for targeted marketing for quick understanding of the different segments and hence mediating visualisation tools) and decision tables for intuitive analysis of data. The modern element is the so-called boosting, which is to increase the accuracy of the model by repeated adjustment, to compensate for errors. Another useful function is to support the prediction of continuous (indiscreet) attributes, such as profit, turnover or market share, which enables suggesting business and marketing strategies.

A very popular feature of Return-On-Investment (ROI) is their curves. They illustrate the costs associated with recommended actions and decisions. They are now a very important piece of information (SPSS, 2006).

The tools also include faster error correction and more accurate business decisions for major progress. The quality tools may be integrated in the decision tables for customers with technical and business focus, which allows qualified managers to manage the operations of OLAP using a visual environment for displaying critical factors and their mutual relationships.

Other interesting elements of a simplistic decision-making process are given in Table 4.

Table 4: Elements of a simplistic decision-making process

1. Navigating through the 3D objects
2. Zooming in and out of data objects
3. Graphic querying of databases
4. Visual grading of data
5. Repeated playback and animation
6. Trend analysis using animation of more than two independent variables
7. Visual filtering of data and information
8. The possibility of global perspectives for details
9. Custom steering angle on the data

2.5.4. Hardware requirements

For the purposes of data mining, it is necessary to deploy computers with high-throughput, and high performance with a large volume of data. An important feature, too, is potential parallelism in data processing to speed up calculations. The requirement is, necessarily, good visualisation graphics hardware. Overall, it is necessary for visual processing of large amounts of data to use powerful hardware with a stable operating system and, if possible, with guaranteed throughput of the system.

Consistent and effective use of data mining techniques for the management of a company gives an obvious advantage to companies that are or anticipate that they will be, for example, under great competitive pressure. Data mining has already found its place in industrial markets, commerce, telecommunications, government and the financial sector, being a powerful tool to support decision-making processes. It will be interesting to monitor developments in this field and practical applications.

2.6. Data Mining in the Public Sector

2.6.1. Overview

Data mining is a deductive query processing – expert systems or small Machine Learning (ML)or statistical programs. O'Brien and Marakas (2010) mentioned that "*data mining is a major use of data warehouse databases and the static data they contain*". Witten et al. (2011) stated that data mining is about providing solutions to problems by analysing data that already exists in database (data warehouse). Managers usually employ data mining in making their strategic decisions with a view to achieving competitive advantages. The decisions include those about the target market, customer relations management, market basket analysis, cross-selling and market segmentation. By data mining, managers can also analyse budget forecasting, analyse risk in the business and monitor and control quality. There is no doubt that data mining can reveal new correlations, patterns, and trends in vast amounts of business data stored in a data warehouse. Figure X shows how data mining drives business knowledge from the data warehouse.

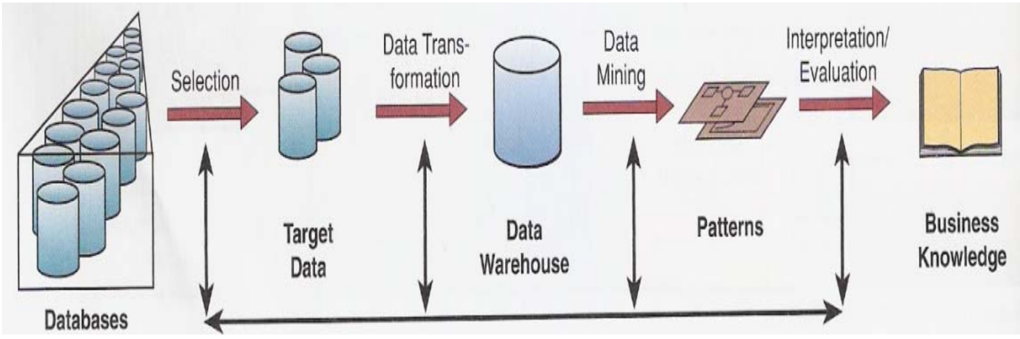

Figure 14: How Data Mining Extracts Business Knowledge from a Data Warehouse [*Source: O'Brien & Marakas (2010: p. 192)*]

2.6.2. Data mining process

Du (2010) outlined three fundamental steps in the data mining process: (i) preparation of input data, (ii) mining of data, and (iii) post processing of output patterns. A brief discussion of the above follows ahead.

2.6.3. Data preparation

In this first step, data preparation needs to select appropriate data and collect them efficiently. This stage also involves pre-processing and formatting of the data and information. Figure X shows how data preparation works in the data mining process.

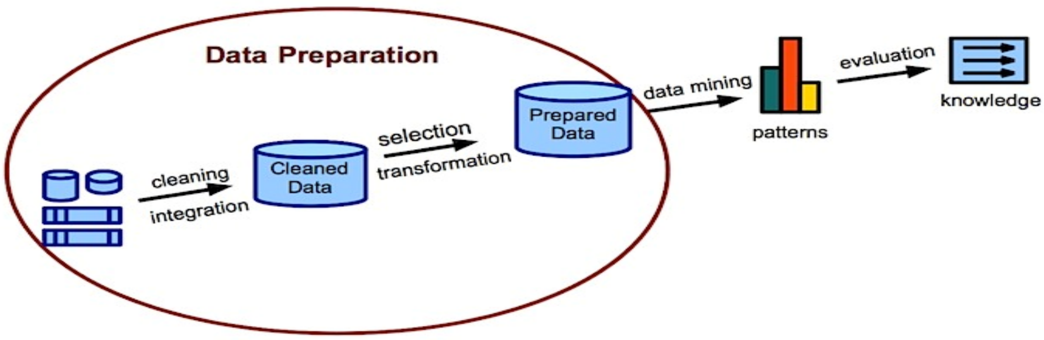

Figure 15: Data Preparation in the Data Mining Process [*Source: datapreparatior.com*]

2.6.4. The mining process

The second stage, mining of data, involves deriving real data from the input data to find patterns. In this process, a practical data mining task is required to comply with the objectives of the investigation. Figure X depicts how data mining selects from the input to find patterns in a business organisation.

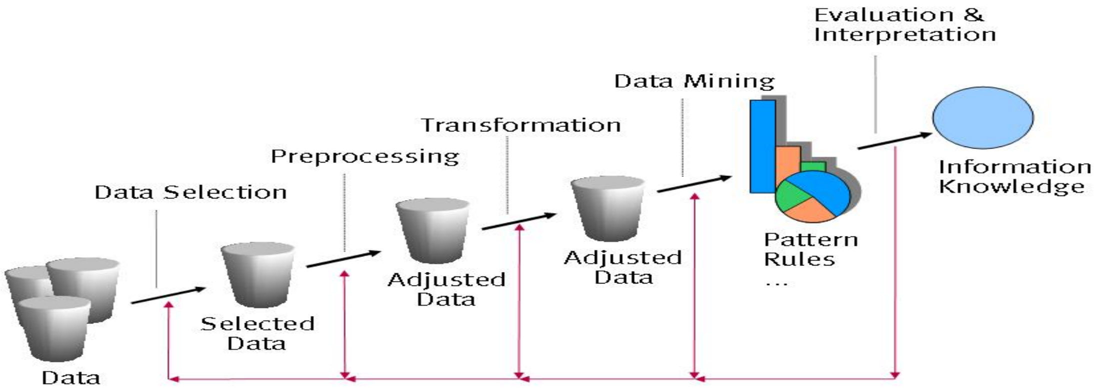

Figure 16: Mining of Data [Source: http://www.wi.hs-wismar.de]

2.6.5. Post-processing of output patterns

This stage covers any further processing of the discovered patterns after mining. The post-processing of output patterns includes pattern evaluation, pattern selection and pattern interpretation.

2.6.6. Use of data mining in the public sector

It has been found from research conducted by Pan and Li (2007) that data mining has been used by the governments of different countries for the last two decades. The authors further mentioned that the then president of the USA, Bill Clinton, introduced the technology of e-government in 1993, and it was approved by the National Performance Review Committee (NPRC). The aim of adopting technology in the government was to reduce the malpractice that existed in U.S. government management and services.

According to Junhua (2002), China employed data mining for implementing e-government in April 1998 under the government online project. The aim of introducing data mining was to improve services within the government departments and divisions.

Malaysia introduced e-government in mid-1996 to implement the policy forthe Multimedia Super Corridor (MSC).

According to Balutis (2001), public organisations perform three different types of activities: Government-to-Customer (G2C), Government-to-Business (G2B), and Government-to-Government (G2G) through information systems. In G2C government usually provides information about different products and services including tourism and recreation, research and education, and also different downloadable forms. G2B provides the services including e-procurement, tax submission and management to different businesses and business professionals. G2G provides services within different departments and divisions of the government and is also used for communicating with different governments for different bilateral and multilateral issues.

2.7. Chapter Summary

This chapter provided a literature review of the main themes of the research. The chapter conducted an insight into the four main concepts of data, including data handling and improvement at the strategic level, namely, data quality, information systems, decision support systems and data mining. Section 2.2.4 showed data quality is important in the public sector, particularly because it helps the public understand the way organisations conduct their operations. Furthermore, it was shown in Section 2.3 that the development of information systems has increased their complexity and added several extra elements other than data, hardware and software. This includes among others the human element, the organisational structure, and the classification criteria of the stored information. Modern forms of information systems have become a necessity for organisations in different fields due to their abilities to manage large amounts of information in organisations, nowadays. Public organisations are one type of organisation that usually acquire vast amounts of data. Management of these amounts may pose a challenge to these organisations given the sensitivity of the information they hold, as well as the choice of efficient information management systems. Furthermore, information systems can be very involved and complicated, and may require a significant amount of skill and knowledge. The vast amount of data handled by information systems may be significantly benefited from, given the right choice of tools. The chapter also provided details of the development of information systems, some of their main uses and applications, and aspects of utilisation in the public sector. It also included an overview of the uses of information systems in Abu Dhabi in general and the Abu Dhabi Police Organisation in particular.

It was also shown in section 2.4 that although decision support systems are important for improving data quality for decision-making in different aspects, they may be insufficient for the sole purpose. On the other hand, data mining techniques provide promising results when deployed on large amounts of data in extracting knowledge and hence helping decision-making. However, the choice of the data mining tools and techniques for a certain deployment is decisive for attaining good results. The chapter also highlighted the use of each of the explored concepts in the public and private sectors. It was shown in Section 2.6 that although there are common points of deployment of both concepts in each sector, there are distinguishing factors between the two sectors that should be highly regarded when deploying any solution. The chapter also provided aspects of established or potential uses of data mining and decision support systems in the public sector of the United Arab Emirates.

Chapter 3: Methodology

3.1. Introduction

This research used the case study research strategy. This strategy is advocated in the literature when the research questions seek to explain how and why a certain social phenomenon works (Yin, 2013). As per the questions set by this research, the case study strategy was used. The data from the case study organisations were collected by the means of interviews. This approach has several advantages over others. For example, the case study strategy is important to allow a rich understanding of the context of the research and the processes being enacted, and is mostly used in exploratory and explanatory research. The case study strategy also allows triangulation, which is a method of cross-referencing and validation of data collected from different sources (Saunders et al., 2009).

Seven public (i.e. state) organisations from Abu Dhabi Emirate were chosen to represent the Abu Dhabi's public sector. The choice of the organisations was based on several considerations but most importantly the involvement of these organisations in different federal bodies of the public sector in the country, such as law enforcement, traffic regulations, healthcare, and civil defence, among others. Furthermore, these organisations undergo continuous development and innovation to pursue best practices in digital information management. The organisations are: Abu Dhabi Police Organisation (ADPO), Al Ain Hospital, the Department of Economic Development of Abu Dhabi, the General Authority of Youth and Sports Welfare, Zayed Foundation, Tawam Hospital, and the Department of Municipal Affairs. ADPO is a major public organisation in the United Arab Emirates.

Considering ADPO as one of the case study organisation has a significant outcome on the research. Particularly, it allows better generalisability of the attained results to other public organisations in Abu Dhabi Emirate and even the UAE. This is because ADPO is as a hub for many other government bodies and services in the entire country. Furthermore, research showed that ADPO is well-known for adopting best practices and international standards for information management, including decision support and strategic information systems.

This work argues that the quality of the data reaching decision makers is not sufficient relative to the impact of the decisions on the general public. Research shows that there is a positive relationship between data quality and efficiency of decisions: higher data quality promotes better decisions (Raghunathan, 1999). Therefore, in order to improve the decision making process, it is necessary to provide decision makers with quality data.

A series of interviews and questionnaires were conducted at various levels in the chosen organisations in order to understand the decision making process undertaken in public organisations in Abu Dhabi, and hence identify data quality issues in the process. The study aimed not only to identify potential issues related to data quality in these organisations, but also to understand these issues and know their causes in order to suggest suitable remedies.

The research question put forward by the study is: *"What are the main challenges to effective decisions in public organisations in Abu Dhabi Emirate?*" Based on this question, the study's aim was to identify how much data quality is an issue in the decision making process in public organisations in Abu Dhabi by analysing the challenges to effective decisions in these organisations, and thus to explore solutions to data quality issues. The solutions would be organised in a normative framework that would allow a systematic approach to addressing these issues. The framework would provide quality measures for data flowing in the various channels in an organisation, which could be used to sift data according to criteria established by the framework and eventually improve data quality. The data quality framework would provide a data quality solution to the public sector based on an extensive study into data quality requirements and data problems encountered in the abovementioned public organisations in Abu Dhabi Emirate. The framework would also benefit from research on established solutions suggested by studies in the literature or adopted by relevant bodies worldwide.

3.2. Data Collection

Data collection started by observations while working closely with officials at ADPO. The observations were substantiated by informal meetings with managers at the Information and Systems department, and questionnaires with to 30 staff members at the department. The researcher spoke to staff members, senior officials and decision makers over the phone and personally prior to conducting the formal interviews. It was noticed the data reaching decision makers underwent a series of processes and across a number of departments, which might affect its quality and relevance for the intended purpose. For example, ADPO has called for continuous data quality audits as per the large databases maintained by the organisation, yet these audits have not been undertaken. Furthermore, the central systems witness frequent lags and failures. The IT department reported that the information systems require extensive inspection and rework. An IT manager met face to face referred the systems' issues to redundant databases and data inconsistency. Manager of Information and Systems at ADPO maintained that the issues are related to inability to implement appropriate data handling strategies, which cause issues with data quality and data suppliers, thereby affecting the IT systems.

The above-cited issues induced from the empirical observations were all indicators of data quality problems, which led to the study's hypothesis. Given at the time of the study that the data reaching decision makers went through a number of processes and across various organisational levels (i.e. transactional, managerial and strategic), the hypothesis put forward was that: the quality of the data is not sufficient for effective decisions in ADPO. In order to move forward to test this hypothesis, semi-structured interviews were conducted in ADPO with two senior managers in the period between 26th May and 10th June 2013.

The initial analysis of the results obtained from interviews, combined with the results from the observations, informal meetings and questionnaires, revealed that the decision making process was negatively influenced mainly by the poor quality of data reaching decision makers from the various sources. In order to answer the research question set earlier, the issue had to be investigated further by considering other public organisations in Abu Dhabi. The organisations considered were: Abu Dhabi Police Organisation (ADPO), Al Ain Hospital, the Department of Economic Development of Abu Dhabi, the General Authority of Youth and Sports Welfare, Zayed Foundation, Tawam Hospital, and Abu Dhabi's Department of Municipal Affairs. The main interest the study scrutinised the decision making process in these organisations in order to learn of any data quality issues associated with the process.

The theoretical sampling approach, that is, collecting data in order to build a theory, was used and data was collected in the form of interviews with decision makers. That helped focus on issues with the quality of data reaching decision makers. All interviews were recorded, transcribed and coded. Grounded Theory was used as the theoretical framework to analyse and identify categories in the collected data.

3.2.1. Overview of the case study organisations

The information in this section was collected directly from the seven organisations through informal interviews with officials, supplied by the organisations upon request, and based on information publically available. The results were verified by documentary analysis from the literature review.

3.2.1.1. Abu Dhabi Police Organisation (ADPO)

ADPO operates with other UAE police departments through the Ministry of Interior to achieve a safer society. ADPO serves four major UAE districts: Abu Dhabi, Al-Ain City, the External Region, and the Western Region. ADPO has several units which include police patrol, emergency response, crime investigation, and traffic control. The primary objective of the organisation is to become an intelligence-led, proactive police force that reacts to the needs of society with the highest level of integrity and training. For this aim, ADPO has constantly undergone development of its information systems in order to integrate a range of processes that include the human side along with the software and hardware deployment with the ultimate goal of improving the accuracy and quality of their undertaken decisions. ADPO has established the "Decision-Making Support Centre" to help the organisation explore future challenges rather than just conduct research on current phenomena. The centre also helps in quality assessment and control. In 2011, ADPO implemented a Geographic Information System (GIS) in order to integrate people and processes for making better decisions (*source: Direct contact with ADPO officials and published documentation by ADPO*).

3.2.1.2. Al Ain Hospital

Al Ain Hospital is a highly specialised government hospital in health care and emergency. The hospital is one of the two main hospitals in the Al Ain area of Abu Dhabi Emirate. The hospital provides a wide range of general and specialised clinical services. Al Ain Hospital is also the base for medical education and training for two teaching institutes from Europe and the UAE. Al Ain Hospital is aiming at improving its performance, expanding its services, enhancing the skills of workers and raising the efficiency of its existing medical facilities to provide an ideal medical care model in the region (*source: The hospital's vision and objectives available on its website*).

Al Ain Hospital gives great importance to interdisciplinary cooperation and team work, and seeks the highest standards of professional training and qualifications of all its advisers and staff. As such, the hospital deploys medical knowledge and entrepreneurial technology as a basis for the decision making process (*source: direct contact with the hospital's official. Details are provided in the discussion.*).

3.2.1.3. Department of Economic Development of Abu Dhabi

The department aims to achieve economic development of Abu Dhabi, regulate its economic and business affairs, prepare economic studies of the overall social and economic issues, and analyse the elements that affect these variables. The relevant decisions taken by the department include the following endeavours:

- Propose economic policies for the Emirate
- Seek optimum utilisation of available resources
- Design policies and procedures that contribute to increasing rates of economic growth, and to identifying trends, priorities and actions that enhance the competitiveness of Abu Dhabi Emirate
- Identify areas of activities and services that make Abu Dhabi Emirate a distinguished local, regional and global centre.

(*Source: the above information is available on the department's website*).

For the decision making process, the department collects and analyses statistics and figures, conducts studies, and prepares feasibility studies of different projects (*source: direct contact with the department's official. Details are provided in the discussion.*).

3.2.1.4. General Authority of Youth and Sports Welfare (GAYSW)

GAYSW is the supreme governmental authority responsible for the welfare of youth and sport sector in the UAE. GAYSW plays a leading role in providing a positive attractive environment that enables youth (including the gifted and talented) to develop physical and mental abilities, utilise free time, improve creative skills, deepen national identity principles, and promote loyalty, belonging and voluntary work sense (*source: the GAYSW website*).

GAYSW uses a set of statistical and decision support tools in its decision process, which range from human insights based on experience and knowledge, to scientific and statistical methods to make decisions. The aim is to provide competitive edge and improve results at the national and international levels (*source: direct contact with the GAYSW officials. Details are provided in the discussion*).

3.2.1.5. Zayed Foundation

Zayed Foundation is a charitable organisation that aims to contribute to society and support cultural and humanitarian centres and scientific research institutions involved in raising awareness of national traditions, customs, and scientific efforts for the development of human civilisation (*source: the Foundation website*).

The foundation believes that the main aspect of achieving efficient decision making is developing the skills of staff. For that, the foundation participates in national and international training projects and programmes and coordinates with the competent authorities in national initiatives for best practices. It also establishes and promotes corporate values and applies the system of governance for the sole aim of achieving the best return possible on investment (*source: direct contact with the Foundation officials. Details are provided in the discussion*).

3.2.1.6. Tawam Hospital

Tawam Hospital is a semi government hospital that provides healthcare services to the community of Al Ain and referral services to the UAE and surrounding GCC countries. One of the largest hospitals in the UAE, Tawam Hospital aims to provide quality health care services that meet the needs and expectations of the UAE population and the surrounding GCC countries (*source: the hospital's website*).

The decision making process at the hospital is undertaken by a team of managerial and medical experts. They make use of the wealth of information provided by the hospital's database and decision making systems. Decision makers at the hospital also liaise and cooperate in some cases with Ministry of Health officials and state policy makers to ensure that the hospital's decisions align with the country's overall healthcare policies (*source: direct contact with the hospital's officials. Details are provided in the discussion.*).

3.2.1.7. Department of Municipal Affairs (DMA)

The Department of Municipal Affairs (DMA) acts as a focal point of all municipal planning and oversees public projects in Abu Dhabi Emirate. DMA aims to produce efficiencies and higher customer satisfaction in accordance with the national policy agenda, which represents a new era in municipal services to the general public.

As a regulatory body, DMA supervises three regional municipal councils and municipal administrations: Abu Dhabi Municipality, Al Ain Municipality and Western Region Municipality.

The decision making process in DMA is based on its organisational structure:

- Local Governance: responsible for the municipal regulations, strategic support, municipal council management and customer complaints
- Municipal Support: responsible for municipal operations support, inter-department coordination, and training
- Property Registrar: responsible for handling complex property requests and transactions, and managing the Property Registrar database
- DMA Support Services: responsible for providing administrative and legal support to the DMA.

(*Source: the department's website and direct contact with officials*).

3.2.2. Interview procedure

At this stage, the study aimed to particularly scrutinise the data at the decision makers' side. The interviews were set to assess how data arrived and how was used in decision making. The hypothesis proposed after the earlier surveys/observations detailed above was that the quality of the data reaching decision makers is not sufficient for effective decisions. In order to test the hypothesis, semi-structured interviews were conducted face to face with two senior staff members in each organisation (between 1st and 10th June 2013 in ADPO and in October 2014 and November 2014 in the other organisations). The theoretical sampling approach was used to obtain data from the interviews. The theoretical sampling approach enables comprehensive comparisons across the conducted interviews and arriving at insights that are well-established in the data (Laperrouze et al., 2010). Furthermore, according to Charmaz (1990, cited in

Calman, 2012), theoretical sampling is used to generate further data to confirm or refute original categories.

Two senior managers from each organisation were interviewed as shown in Table 5.

Table 5: The interviewees' posts in the organisations

Organisation	Interviewee Post
Abu Dhabi Police Organisation (ADPO)	Interviewee 1: Director of Strategy and Performance Development Interviewee 2: Head of IT and Systems
Al Ain Hospital	Interviewee 3: Operations Director Interviewee 4: Quality Assurance Manager
Department of Economic Development of Abu Dhabi	Interviewee 5: Strategic Planning & Performance Management Director Interviewee 6: Project Manager
General Authority of Youth and Sports Welfare	Interviewee 7: Quality Control Manager Interviewee 8: Policy Control Director
Zayed Foundation	Interviewee 9: Director General of the Strategic Planning Department Interviewee 10: Assistant Manager of Finance
Tawam Hospital	Interviewee 11: Senior Operations Officer Interviewee 12: Cost Control Manager
Department of Municipal Affairs	Interviewee 13: Strategic Planning Director Interviewee 14: Human Resources Manager

3.3. The Analysis Approach

All interviews were recorded, transcribed and coded. The theoretical framework used to analyse the collected identified categories in accordance with Grounded Theory (Glaser & Strauss, 1967). Grounded Theory is one of the main analytic methods used to understand a certain phenomenon and to develop a research hypothesis. Urquhart and Fernandez (2006) maintain that the use of Grounded Theory as a research method has gradually increased in information systems over the past few years. This is based on the fact that the qualitative approach has widely spread in most management disciplines (Urquhart & Fernandez, 2006). The grounded analysis was used as the main theoretical framework for data analysis. This type of analysis is a form of inductive qualitative analysis. Grounded analysis is 'grounded' in the way that it implicates listening to the collected data. It is based on key aspects of Grounded Theory, but deviates from Grounded Theory in some important respects, which all aim to provide a more practical approach to attaining the goal of understanding data in hand.

3.4. Summary

This chapter described the methods used by the research. The research used the case study strategy for collecting primary data using semi-structured interviews. Case studies of seven public organisations in Abu Dhabi Emirate were considered. The case studies provided a rich understanding of the context of the research and the processes being enacted. Later chapters in this thesis provide the details of the findings from the collected data and a discussion of their relevance. The discussion chapter shows how the data led to a data quality framework intended to identify areas of improvement for of quality data flowing in the different channels in Abu Dhabi's public organisations.

Chapter 4: Data Mining for Data Quality

4.1. Introduction

It is not an exaggeration to claim that the current age is overwhelmingly characterised by an unprecedented flow of information, which is largely boosted by the advancements in digital technologies and smart devices. There are enormous amounts of data that are generated and communicated on a daily basis worldwide. These data need to be stored, maintained, and highly possibly retrieved in the future. “Big Data” is the term which essentially refers to the large datasets that pose extreme challenges in data processing using traditional methods. Big Data have incited scientists in different areas into studying and understanding particular relevance and interests in them (Boyd & Crawford, 2011).

Due to the high need for storage and maintenance of large amounts of data, special processes have been proposed and implemented to perform efficient data handling. Data mining is a process in which an individual or an organisation employs certain methodologies that use various techniques to explore the large datasets and from them extract information of interest. Such information could lead to knowledge or significant information that may identify certain patterns and trends from the database. The obtained patterns or trends can further be used by the organisation in its decision making process (Larose, 2014).

There are several tools and techniques used in the field of data mining. Figure 17 gives a broad classification of data mining techniques into two classes: *Classical techniques* and *Modern techniques*. A few examples of data mining techniques in both the classes are also illustrated.

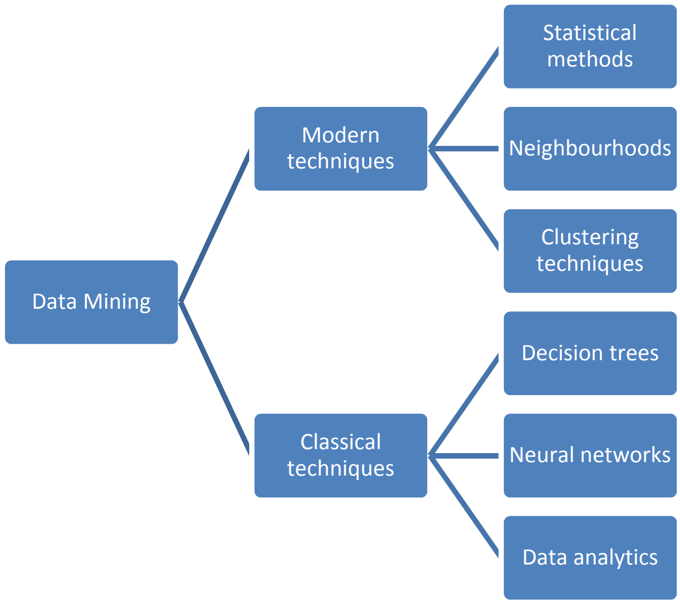

Figure 17: Classification of data mining techniques

In the following sections, some of the most commonly employed tools and techniques of data mining by organisations are introduced and their efficiency is analysed.

4.2. Tools for Storage, Retrieval, and Management of Data

An important and essential feature of database management systems is the storage and retrieval of data. Depending on the field of application, data are stored for different reasons and for different lengths of time. With increasing adoption of real-time systems where data need to be stored and retrieved at extremely short intervals, and when the amounts of data stored are huge, it is important that appropriate tools be used. Some of the common tools are described below.

4.2.1. Hadoop systems

One of the most popular tools for the storage and batch processing of large datasets is the Hadoop Distributed File System. Hadoop is an open-source software application which is widely used. Functionality-wise, Hadoop breaks the data and distributes it across various clusters of servers. When particular data is to be retrieved, it uses the MapReduce framework which maps the query to the appropriate data cluster. The MapReduce framework was originally developed by Google and adapted in the Hadoop systems.

Despite its remarkable performance, open-source availability and wide adoption, some of the issues associated with Hadoop are: it cannot be implemented in real-time systems, a lack of expertise on Hadoop systems, and other related challenges. Nevertheless, it is a powerful tool and organisations can adopt it depending on their requirements. Large organisations such as Google, IBM, and Yahoo are using Hadoop systems for their search engines and advertising (Rouse, 2012).

4.2.2. NoSQL tools

Traditional databases were not designed or equipped with technologies to meet the demands of recent database-driven devices. NoSQL databases are recent types of databases which are designed to handle faster query speeds for large volumes of data. One of the popular tools for managing NoSQL database is the Apache Hbase. It, again, is an open-source framework with wide support from governmental and research communities. Organisations requiring to meet real-time demands could adopt such a tool for their database handling. For instance, Facebook adopted Apache Hbase over MySQL in order to better manage their email, chat, and text messaging systems. It is also used to store, manage and retrieve the incredible amounts of data and metadata it generates (Metz, 2010).

4.3. Data Mining Techniques

4.3.1. Genetic algorithms

Genetic Algorithms (GAs) are based on the theory of natural evolution. This technique uses stochastic methods for exploring large databases, which simulates the evolutionary processes such as inheritance, mutation, selection and crossover. To briefly explain the GA process, the entities which can be a potential solution are iteratively evolved towards better solutions. During this evolution process, an entity can be mixed with other entities (simulating mutation) and eventually converges to a solution which is considered as optimal (Gordini, 2014). Figure 18 illustrates the GA technique.

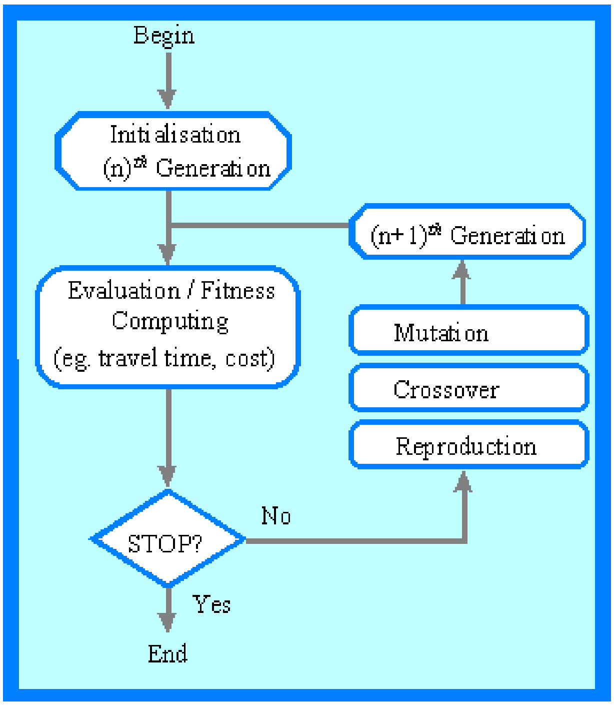

Figure 18: Illustration of the Genetic Algorithm technique (Kar, 2011).

A significant feature of GAs is that they provide a solution from a host of potential solutions. The GA process initiates potential solutions and chooses the best from them over several iterations. This makes the obtained solution optimal and more robust. GAs can also be easily implemented without a pre-requisite of high mathematical knowledge and are easily portable. In scenarios where it is not possible to define gradient or derivatives, GAs can potentially offer solutions.

Despite some of the above advantages, a drawback of GAs is that the solutions provided by them are optimal for a local scenario, i.e., a local optimum. The same solution cannot be emulated as a global solution. Therefore, in situations where the data is continuously changing or highly dynamic, especially in real time, the implementation of GA might not give efficient solutions. Furthermore, defining the entities of the GA process (referred to as chromosomes) might be a difficult process in some scenarios (Sivanandam, 2008).

Today, GA-based data mining is used in various applications, one being the stock markets. Several companies have implemented GA-based data mining techniques in stock market analysis and financial market predictions. Companies such as Ultragem have specialised in providing data mining technologies based on genetic algorithms (Lapointe & Desieno, 2003).

4.3.2. Artificial neural networks

Artificial Neural Networks (ANNs) are a set of models that can be used for data mining. As the name suggests, ANNs are based on neural networks of the brain. Essentially, ANN methodologies detect pattern in the dataset and based on those patterns, prediction models are built. An important aspect of ANNs is that they require training sets which are used further for pattern detection and prediction. With this type of working process, ANNs are perceived to be highly efficient data mining solutions with the ability to provide high accuracy for different types of problems (Berson, 2004).

A challenge of using ANNs for data mining is that they require an adequate training set to train the prediction model. Quite often, deciding the nature and scope of the training set may turn out to be an issue. The training process must be stopped at the right time in order to avoid overtraining or over fitting of the prediction model. Further, a neural network once trained cannot be retrained with a new training set. Therefore, in dynamic datasets using neural networks, this might pose challenges in efficiency. However, current research in ANNs is developing efficient algorithms towards addressing these issues.

4.3.3. Decision trees

Decision trees are one of the oldest and most widely used data mining approaches for various applications. Essentially, a decision tree is a predictive model in the shape of a tree where each leaf or branch represents a classification of the data. Different types of algorithms are used for data classification for making the decision tree. Figure 19 depicts the decision tree process.

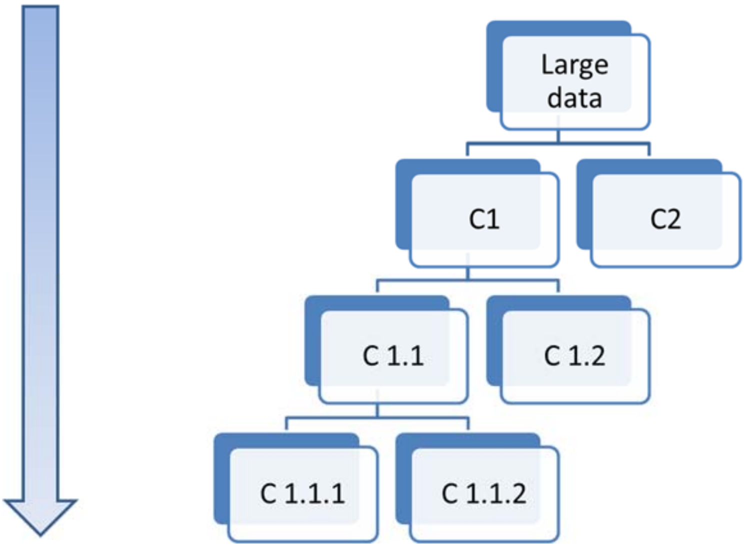

Figure 19: Illustration of the decision tree process

Effectively, decision trees put the data into different segments by predicting the segments form the entire dataset. The most attractive feature of the decision tree is that it represents the data in an easy to understand and interpretable manner. Because of this simplistic way of representing the data, decision trees are widely preferred in business and government organisations. Also, decision tree approaches do not necessarily require pre-processing of the data and thus work on raw data (Maimon & Rokach, 2005). However, decision trees are also instable in terms of the precision of the data used. Minor changes in the data can sometimes potentially lead to major changes in the associated decision tree. Therefore, in contexts where the data changes dynamically or in case of presence of noisy data, decision trees may not perform adequately. Furthermore, for very large datasets, the decision trees can become too complex to understand (Nayab, 2011). Therefore, application of decision trees must be done in line with the nature of the datasets being analysed. For instance, decision trees can be efficient for pre-defined data or for managing relatively small amounts of data of an organisation.

4.3.4. Big data analytics

In the past few years, the term "Big Data" has drawn considerable attention in different avenues, such as industry, media and research community. Essentially, Big Data is the process of analysing extremely large amounts of data to identify hidden patterns, correlations within the dataset, and other relevant information which could shed light on new types of information. Popularly, Big Data is associated with the so-called “petabyte” size data, which is equal to 10^{15} bytes of digital information (Srinivasa & Bhatnagar, 2012).

The Big Data phenomenon gained momentum when researchers from Google were able to demonstrate that their data analytics were able to identify more accurately than the government the spread of influenza. It is also believed that Big Data analysis was used in the presidential elections of President Obama in order to understand individual voters and hence target them individually (Issenberg, 2012). Today, large companies such as IBM are developing advanced tools for Big Data analysis for various applications.

There is some scepticism about the term Big Data, especially in the research community. Scholars consider the term to be a more “glamorous” way of expressing existing data mining and processing technologies. Even though Big Data is essentially data mining and processing, it aims further at merging several disciplines of predictive analysis, data analytics and others for analysing huge amounts of data which are believed not explored to their maximum potential by existing business intelligence programmes (Rouse, 2012). Business organisations are increasingly accepting the belief that Big Data analytics is essential to businesses and can potentially redefine their market competencies. There has been an increasing adoption of Big Data analytics and it has gained interest from different business fields and government organisations. Big Data analytics is hence expected to play an important role in the near future in different respects (Columbus, 2014).

4.3.5. Social media data mining

The ubiquity of social media is so obvious that the need to emphasise its presence in our daily lives is minimal. Huge amounts of text (as well as other multimedia formats) are generated on an hourly basis on social media across the globe, rendering data mining in social media as an enormous interest for governmental and business organisations. For instance, Asur, and Huberman (2010) show that discussions on Twitter could be used to predict box office performance of a movie using their prediction model. Furthermore, their prediction model could also be adapted to predict the business prospects of various other businesses. Some other social media websites such as LinkedIn are used by organisations to identify potential employees for their businesses. Furthermore, social media is today an effective tool for marketing. For instance, the largest social network, Facebook, can be used in a variety of ways to target individuals for marketing. Several Facebook-related tools such as "Like Button", "Booshaka" and "Social Graph" provide data visualisation and analysis which can be used to target individual users for marketing purposes. For instance, the "like button" tool can be an indicator of what people "like" on Facebook (Odden, 2010).

Generally speaking, social medial can be a mine of data for understanding the general public disposition about certain events, better understand human interaction, and reshape business models, among other things (Tang & Liu, 2010).

Different types of data mining approach are tested on social medial platforms. For instance, Cugliari and Guille (2014) use Event Detection with Clustering of Wavelet-based signals (EdCoW) to analyse the temporal dynamics of word frequency on Twitter. A graph-based structure mining methodology was used by Corley et al. (2010) to develop a disease surveillance model for diseases such as influenza from the web and social media. The report showed that social media could be used to identify a public disease outbreak and hence used for targeted public health communications.

Social media data mining is an emerging field and is currently a subject of keen research interest. Nevertheless, the nature of social media poses several challenges. A major one is that social media is very big, unstructured and messy. Therefore, to identify the data of interest is a major challenge. Further, social media can give rise to questions of authenticity, reliability, consistency, and are inherently noisy. Also, in many data mining techniques, training data is required, which in many cases might not be possible to obtain from social media. Therefore, it can be said that social media data mining is an emerging field which requires its own computational approaches for data mining (Zafarani et al., 2014).

4.4. Data Mining Tools

In order to cater to the demands of the data mining (DM) requirements of various organisations, different types of data mining tools are being introduced onto the market. Today, there are several commercial and open-source data mining tools available.

Table 6 lists some of the popular data mining tools widely used and also lists the DM techniques they support.

Table 6: Some of the commonly used data mining tools and the techniques they support

Tools	DM techniques supported
IBM SPSS Modeller	Statistics, Clustering, Neural Networks, Data modelling
SAS data mining	Predictive and descriptive data modelling, data visualisation
Orange	Decision trees, machine learning, data visualisation
Viscovery	Predictive data analytics
Statistica	Statistical data mining techniques
WEKA	Classification, clustering, statistics, data visualisation
Apache Mahout	Machine learning, clustering and classification
R	Machine learning, support vector machine, statistical analysis
NetMiner	Social media data analysis

Some of the most popularly used data mining tools are classified in the following sections.

4.4.1. Open-source data mining tools

There are several open-source data mining tools available which are largely used by researchers and academicians. The types of data mining support provided by the open-source tools varies depending on the developers, however, the open-source tools are quicker to update with the latest data mining algorithms. Some of the most popular ones are RapidMiner, Python, Orange, R, QDA Miner, among several others. The R software is the most common and is described below.

4.4.1.1. R

The R open-source software is one of the most popular data mining tools currently in use by individuals. According to a survey of 700 respondents conducted by KDNuggets.com in 2013, 61% of the participants use the R environment. Interestingly, R environment is not offered as a data mining tool yet is widely used as one.

The R environment is open-source software which is strongly supported by an active group of developers. It supports a variety of software packages and is also highly flexible for integration of different modules, graphical user interfaces, and libraries. There are over 2,000 libraries available for the R environment. Therefore, it is more likely that the R software would include the latest data mining algorithms faster than the commercial ones (Team, 2014).

Despite the above benefits, some limitations with the R environment may arise. It is mainly a command line based tool, therefore does not offer the convenience of user interfaces. However, other tools such as "Rattle" provide user interfaces to R. Further, since it is an open-source tool, it might pose challenges or limitations and security vulnerability for applications in large organisations, and also there might be inadequate technical support available from the developers (Cortez, 2010). Nevertheless, the R environment is still widely popular and can be used by the research community and small and medium-sized enterprises for their general data mining purposes.

4.4.1.2. WEKA

WEKA is a software application that provides implementations of learning algorithms to datasets. WEKA stands for the Waikato Environment for Knowledge Analysis. The original aim of WEKA when it was introduced in 1992 was to attain a unified work-bench to allow researchers easy access to modern techniques in data mining and machine learning. At that time, learning algorithms were used on different platforms, available in various languages, and operated on various data formats (Hall et al., 2009).

WEKA provides not only a toolbox of learning algorithms, but also a framework inside which researchers can implement new algorithms without having to be concerned with supporting infrastructure for data manipulation and scheme evaluation. Currently, WEKA is recognised as a pioneer software application in data mining and machine learning. It has attained widespread acceptance within academia and business circles, and has become a widely used tool for data mining research.

WEKA gives users free access to the source code, which has improved its development and facilitated the creation of many projects that incorporate or extend WEKA (Hall et al., 2009).

4.4.2. Proprietary data mining tools

The wider use of data mining is concentrated in the business environment and specifically under what is known as Customer Relationship Management (CRM). CRM focuses on building a long-term profitable relationship by the business with its customers and is an important approach in almost all businesses. Data mining tools can be used for developing CRM in a business. Data mining can be used to identify customer interests and target them for marketing products accordingly. It can also be used to maintain retention of the customer base. In CRM, data mining tools can aim to perform one or several types of data modelling, such as: classification, clustering, forecasting, regression and visualisation. These tools can help identify the customer relation patterns, optimise business processes according to changing customer needs in order to maximise profit, and get a better understanding of customer behaviour and preference. These tools are especially relevant in retail industries (Bhate & Patil, 2014).

There are numerous commercial software applications and tools which do data mining for commercial use. Some of the most popular ones are Oracle, SAS, SPSS, Insightful Miner, and others. Each application caters to different demands and mostly offers customisation to suit the organisation needs. The most common targeted aspects of data mining for business applications are (Ranjan & Bhatnagar, 2008):

- Customer profiling
- Targeted marketing
- Market basket analysis
- Managing relationship with the customer
- Fraud detection
- Avoid customer attrition
- Mine unstructured data.

4.4.3. Social network analysis tools

The increasing importance of data mining in social networks is described in the previous section. Despite data mining being a relatively new phenomenon in the context of social media, there are several tools available in particular for social network analysis. According to Combe et al. (2010), a social network analysis software application is expected to have basic functionalities such as: data representation where the data can be represented in graphs and nodes, visualisation where the graph can be represented with different approaches, characterisation using indicators where indicators can be used to represent the network at various levels, and community detection where a set of points having strong connections can be represented as a community of data.

There are several open-source and commercial tools for social network analysis. Some of these are InfiniteGraph, GraphStream, Gephi, Netminer and AllegroGraph. The NetMiner tool is very common and is briefly described below.

4.4.3.1. NetMiner

NetMiner is a social network analysis tool which facilitates network data exploration and visualisation interactively. It performs various types of statistical analysis on the networks and visualisation of its data. It also uses the Python scripting language which is an open-source programming language. NetMiner is also available in a non-licensed academic version (NetMiner, 2014).

4.5. Data Mining as a Tool in Public Sectors

Data collection, storage, maintenance, and retrieval processes by government agencies have always been executed with different approaches in accordance to the times and technologies. Like other organisations, government agencies nowadays are widely adopting digital approaches for data handling. The emergence of recent allegations of breach of privacy, for example, by Edward Snowden in the U.S. government, poses a complex question of legitimacy of data collection and mining by the public sector. Nevertheless, government agencies have been implementing data mining tools for various applications.

The U.S. government uses data mining tools for identifying patterns related to criminal activities. This was adopted more intensively after the 9/11 attacks. The U.S. government applies various surveillance programmes that use data mining tools. Some of the programmes are the domestic surveillance programme, total information awareness, terrorist surveillance programme, among others. All of these programmes focus on identifying detectable patterns related to criminal activities (Cate, 2008). The U.S. government has defined its principles for data mining in its report "*Principles for*

Government Data Mining: Preserving Civil Liberties in the Information Age" as part of "*The Constitution Project*" (Sloan & Sharon, 2010). The United Kingdom government uses data mining tools for not just detection of patterns of criminal activities but also for assistance in providing social benefits to individual groups, identification of people who are at risk of harm, or who pose risk to others (Parliament, 2009). The government of Ireland has used data mining tools to better target selection for auditing purposes. The Irish government used data mining techniques to identify targets who are potentially non-compliant or tax-evading. They used SAS Enterprise Miner and SAS Enterprise Guide to identify potential targets (Cleary, 2011). The research finds no indications that data mining is currently in use by any public organisations in the UAE.

4.6. Summary

Despite the spectacular accomplishments of current data mining methodologies, there still exists an enormous challenge in data management that needs to be addressed. As the Internet and digital communication have become widespread over the past two decades, it is important that more efficient data handling solutions be developed.

The applications of data mining technologies are varied. The major applications of data mining are still in the retail industry. The majority of the data collected are required by business organisations to target specific groups of customers. However, data mining can also play a significant role in the public sector where data patterns could be used to identify crime, diseases, etc., at earlier stages. Furthermore, public sector agencies can also use data mining techniques to plan and provide appropriate social services to individuals, groups or communities.

The risks to data privacy and security have also increased with the increasing transformation to digital platform. The data mining tools and technologies should also focus on minimising these risks. There is a need to develop a generic framework wherein public organisations and government agencies could ethically share and manage data.

Chapter 5: Data Mining based Experiments for Data Quality Enhancement: Using Classifiers to Predict Missing Data Values

5.1. Introduction

This chapter provides an implementation of data mining classification on datasets provided by one of the Abu Dhabi public organisations studied, namely, ADPO. The aim is to use classification to predict missing data values in the datasets. The experiments reveal that different success rates are attained by using various classifiers.

Related Publication:

Al-Ketbi, O., & Conrad, M. (2013). Supervised ANN vs. Unsupervised SOM to Classify EEG Data for BCI: Why can GMDH do better?. *International Journal of Computer Applications*, *74*.

5.2. Implementing Data Mining: Classification of ADPO Traffic Data

5.2.1. Applying data mining on traffic data: Choice of classification

A data mining classification method is implemented in this section based on real data obtained from ADPO. The choice of classification was based on the insight into the data problems the organisation faces gained from an interview with the manager of Information and Systems. The datasets obtained from the department embedded a number of quality issues in terms of completeness, consistency and accuracy. These issues may be addressed by classification, given that robust pre-processing data cleansing and preparation are conducted (Blake & Mangiameli, 2011).

In the interview conducted with the manager of Information and Systems at ADPO, he stated that "*electronic formats of data are the most dynamic and vital information the organisation possesses since we are working towards minimising our paper use to* *minimal.*” The manager added that “*we are looking for ways to deploy and benefit from our data yet currently the uses of the data are limited to storage and retrieval*.”

The data and information held in the organisations are described in Table 7.

Table 7: Types of data held in ADPO

How are data stored?	What are the data types?	How are data handled?	What operations do data undergo?
Databases	Numeric, textual	Database management systems. Electronically stored and backed up.	Basic data entries, retrieval, and queries
Spreadsheets	Numeric, textual	MS Excel. Electronically stored and backed up.	Basic data entries, retrieval, searches and sorts
Documents	Textual	Word processors, PDF. Electronically stored and backed up.	Word processing, operating system searches and retrieval
Hard copies	Textual, numeric	Physically stored.	Basic archiving techniques.

A typical example of datasets that ADPO obtains and processes on a daily basis is traffic data. Huge amounts of data are received in a continuous manner. These data include details of traffic offences, such as *type of offence*, *date*, *time and place of offence*, *nationality of the offender*, *age of the offender*, *profession of the offender*, among others. These data are stored in spreadsheets and processed by basic operations, such as editing, sorting, searching, and some other similar operations. Currently, these data are not saved in databases nor managed with a database management system.

The Information and Systems department in ADPO, upon the researcher’s request, provided three spreadsheet files. Each file contains up to 5,000 records as detailed above. The department manager maintained that these data come to the department incomplete and missing certain details due to several factors, essentially suffering the above highlighted data quality issues. The reasons vary by issue, but can be summarised in three major aspects: simply unavailability of the missing details, misinterpretation by the law enforcement officer of the received information, human errors at the collection and/or entry levels.

A number of traffic datasets were obtained from ADPO in MS Excel format. The datasets contained a relatively large amount of records with several quality issues as detailed above. The datasets underwent pre-processing and cleansing in order to generate suitable training sets to generate classification models.

The experiments were conducted on WEKA software and five classifiers were used to generate models that aimed to classify the data fed and hence use the models for prediction. The generated models were tested with test datasets extracted from the original dataset received from ADPO. Running the model with the test data revealed the accuracy of the model generated by the specific classifier. Figure 20 depicts the process.

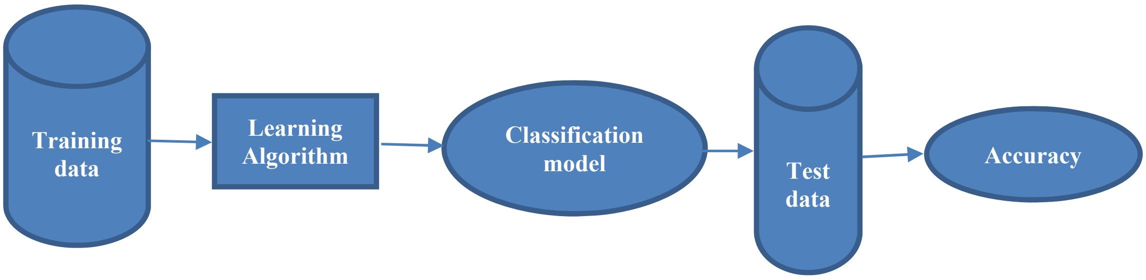

Figure 20: Classification was conducted using different classifiers. Each generated model was tested with the same test dataset in order to evaluate the model's accuracy

5.2.2. Classification in WEKA

WEKA provides a large number of classifiers for different purposes, ranging from simple to complex. The classifiers are divided into categories and each category has a number of classifiers (e.g., as in Figure 21).

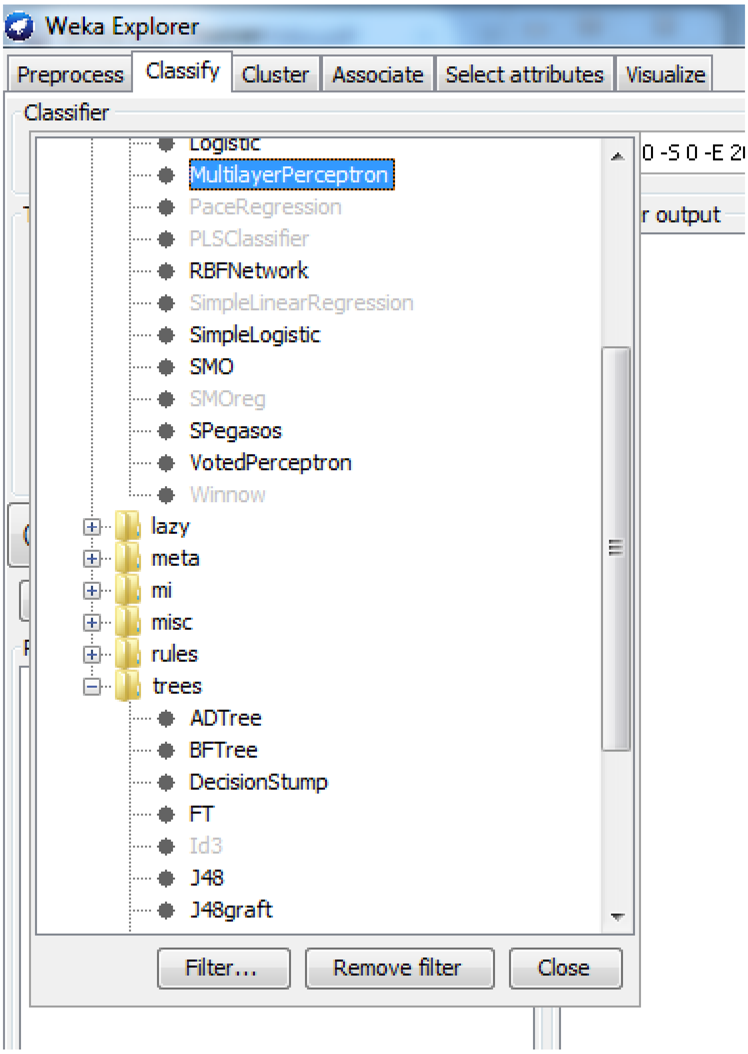

Figure 21: WEKA provides a considerable number of classifiers for different purposes

WEKA provides not only a toolbox of learning algorithms, but also a framework inside which researchers can implement new algorithms without having to be concerned with supporting infrastructure for data manipulation and scheme evaluation. This approach is particularly used in this experiment for the Support Vector Machine (SVM) classifier.

Currently, WEKA is recognised as a pioneer software application in data mining and machine learning. It has attained widespread acceptance within academia and business circles, and has become a widely used tool for data mining research.

WEKA is free and gives users free access to the source code, which has improved its development and facilitated the creation of many projects that incorporate or extend WEKA (Hall et al., 2009).

The following is a brief description of some of the WEKA classifiers that are made available to users. One is introduced in each category apart from the miscellaneous category which has 3 classifiers. It is not the purpose of this section to provide a full account of the WEKA classifiers. It rather aims to provide some examples of these classifiers in order to introduce the reader to these classifiers and their uses. For a detailed description of each classifier available for WEKA along with references to publications, the interested reader may want to consult Theofilis (2013).

1. **Bayes:** Let $U= \{x_1, x_2, \dots, x_n\}$ be a set of n variables. A Bayesian network structure B_S over U is a Directed Acyclic Graph (DAG) along with a set of probability tables $B_P = \{p(u \mid pa(u)) \mid u \in U\}$ where $pa(u)$ is the set of parents of u in B_S (Bouckaert, 2008).
   - Example: NaiveBayes: This classifier uses the Naive Bayes classifier with estimator classes. Training data are used to choose the values of the numeric estimator precision. Therefore, the classifier is not an Updateable Classifier, whereby data are initialised with zero training instances.
2. **Functions:** As the name indicates, the function classifiers in WEKA are those classifiers that model data inputs using certain functions. One of the popular classifier in this category is the multilayer artificial neural network given in the example here.
   - Example: MultilayerPerceptron: This classifier uses backpropagation to classify data instances. The network can be manually built or created by an algorithm or both. The network can also be supervised and altered during training time. The nodes in the network are all sigmoid: $S(t)=\frac{1}{1+e^{-t}}$. However, if the class is numeric, the output nodes become unthresholded linear units.
3. **Trees:** Decision trees form a simple and straightforward way to classify data. A decision tree classifier bases its classification decision on a series of questions about the attributes of the data. Each time an answer is received, a follow-up question is posed until a conclusion about a class label is reached.
   - Example: J48: The J48 algorithm is used for generating a pruned or unpruned C4.5 decision tree. J48 builds decision trees from a set of labelled training data using the information entropy concept based on the fact that

each data attribute can be used to make a decision by splitting the data into smaller subsets.

4. **Lazy:** The lazy classifiers differ from other classifiers. Lazy classifiers store all training samples instead of building a classifier until a new sample needs to be classified. They differ from eager classifiers, such as decision trees, which build a general model before receiving new samples.
   - Example: The IB1 Nearest-neighbour classifier: This classifier uses the normalised Euclidean distance $d = \frac{d_e}{\sqrt{n}}$, where d_e is the Euclidean distance, to find the training instance closest to the given test instance, and predicts the same class as the training instance. When multiple instances have the same smallest distance to the test instance, the first one found is the one used.
5. **Rules:** Classification based on rules is attained when data undergo a number of previously set and progressively established rules until a classification decision is made.
   - Example: ZeroR: This classifier determines the median (in the case of numeric values) or the most common class (in the case of numeric values). It essentially tests how well a class can be predicted without considering other attributes. It is mainly suitable as a lower bound on performance.
6. **Meta:** Meta classifiers evaluate the correctness of base classifiers. A Meta classifier is trained to predict the correctness of each classification of the base classifier and assesses whether the classification is reliable (Kaptein, 2005).
   - Example: Classification viaRegression: Using this classifier allows the user to do classification using regression methods. The class is converted to binary format and one regression model is created for each class value.
7. **Multi-instance:** Multi-instance classification is a supervised learning technique, but differs from normal supervised learning by allowing training of a classifier from ambiguously labelled data (Babenko, 2008).
   - Example: TLD Two-Level Distribution: Using this classifier requires changing the starting value of the searching algorithm, supplementing the cut-off modification and checking missing values.
8. **Miscellaneous:** The Miscellaneous category includes three classifiers (Witten et al., 2011):

- *HyperPipes*: This classifier records the range of values observed in the training data for each attribute working out which ranges have the attribute values of a test instance. It then chooses the category with the largest number of correct ranges.
- *Serialised Classifier*: This classifier loads a model that has been serialised to a file and uses it for prediction.
- *VFI (Voting Feature Intervals)*: With this classifier, intervals are constructed around each class by discretising each numeric attribute or using point intervals for nominal ones. It then records class counts for each interval on each attribute, and classifies test instances by voting.

5.2.3. Classification procedure

5.2.3.1. Training and model generation

Since building a classification model requires that the dataset be as accurate as possible, a pre-step of data mining is to prepare an accurate dataset by cleansing and refining records. The dataset provided as spreadsheets by ADPO was cleansed and records with missing information were discarded. The final dataset used consisted of 2,028 records of traffic offences with five fields, namely, *Date* of the traffic offence, and *Country*, *Age*, *Profession* and *Gender* of the offender. After the dataset was cleansed and prepared for classification, it needed to be converted to *.arff* format to be analysed by WEKA. *.arff* files are specially formatted files understood by WEKA. Figure 22 depicts the structure of the *.arff* file.

```
@relation datatraffic
@attribute Date
{24/06/2013,04/01/2013,05/01/2013,09/01/2013,16/01/2013,31/01/2013,05/05/2013,27/05/2013,07/06/2013,28/07/20
}
@attribute Country
{Jordan,UAE,Sudan,Philippines,Morocco,India,Pakistan,Oman,Nepal,Russia,Somalia,Bangladesh,Palestine,Egypt,Sy
rian,Tunisia,Iraq,Algeria,USA,Iran,Yemen,Comoros,Mauritania,Afghanistan,Lebanon,'Sri
Lanka',Ethiopia,'South Africa',Bulgaria,KSA,Qatar,Bahrain,' Ivory
Coast',Trinidad,Tanzania,Nepal,UK,China,Canada,'Not
Known',Italy,France,Uzbekistan,Libya,Eritrea,Germany,Tajikistan,Ukraine,Djibouti,Kazakhstan,Nigeria}
@attribute Age numeric
@attribute Profession {'Not identified','Specialist programming','Specialist insurance','Specialist risk
insurance','Marketing Specialist','Sales Specialist','Senior Specialist / Financial
Controller','Specialist third / Civil Engineer',Administrative,Professor,'University Assistant
Professor','Professor of Music','Professor of Information Systems','Sales Consultant','Consultant
Information Systems',Fireman,Freelancers,'Specialists in the
judiciary',Dynasty,Vendors,Barbers,'Maintenance in electronic exchanges',Nurses,'Professions and military
ranks','Engineers Fitters'}
@attribute Gender {Female,Male,'Not identified'}
@data
24/06/2013,Jordan,34,'Not identified',Female
04/01/2013,UAE,34,'Not identified',Male
05/01/2013,UAE,100,'Not identified',Male
09/01/2013,UAE,25,'Not identified',Male
16/01/2013,UAE,25,'Not identified',Male
31/01/2013,UAE,21,'Not identified',Male
05/05/2013,UAE,24,'Not identified',Male
27/05/2013,UAE,23,'Not identified',Male
07/06/2013,UAE,26,'Not identified',Male
```

Figure 22: An extract of an *.arff* file showing the attribute and data structures

Once that was done, the classification process was ready to start. Based on research (for example, Wu et al., 2008; Miller & Han, 2009) and among the available classifiers in WEKA, the following were chosen for classification of the traffic data:

1. **Bayesian Networks:** Bayesian networks provide a robust probabilistic representation, and their use in classification has received considerable attention by scholars (Miller & Han, 2009). Bayesian networks learn the conditional probability of each attribute from the training data. The result of running the algorithm on the traffic data is shown below.

```
=== Run information ===
Scheme:weka.classifiers.bayes.BayesNet          -D          -Q
weka.classifiers.bayes.net.search.local.K2 -- -P 1 -S BAYES -E
weka.classifiers.bayes.net.estimate.SimpleEstimator -- -A 0.5
Relation:     datatraffic
Instances:    2028
Attributes:   5
              Date
              Country
              Age
              Profession
              Gender
Test mode:evaluate on training data
=== Classifier model (full training set) ===
Bayes Network Classifier
not using ADTree
#attributes=5 #classindex=4
Network structure (nodes followed by parents)
Date(353): Gender
Country(49): Gender
```

```
Age(3): Gender
Profession(123): Gender
Gender(2):
LogScore Bayes: -25320.283616688437
LogScore BDeu: -33111.94596826987
LogScore MDL: -31285.227054287414
LogScore ENTROPY: -27291.261640496457
LogScore AIC: -28340.261640496457
Time taken to build model: 0.05 seconds
=== Evaluation on training set ===
=== Summary ===
Correctly Classified Instances        1957               96.499
%
Incorrectly Classified Instances        71                3.501
%
Kappa statistic                          0.8968
Mean absolute error                      0.0638
Root mean squared error                  0.1701
Relative absolute error                 18.0802 %
Root relative squared error             40.4833 %
Total Number of Instances             2028
=== Detailed Accuracy By Class ===
               TP Rate   FP Rate   Precision   Recall  F-Measure
ROC Area  Class
                 0.994     0.131      0.962     0.994     0.978
0.979    Male
                 0.869     0.006      0.976     0.869     0.919
0.979    Female
Weighted Avg.    0.965     0.103      0.965     0.965     0.964
0.979
=== Confusion Matrix ===
    a    b   <-- classified as
 1554   10 |    a = Male
   61  403 |    b = Female
```

2. **Support Vector Machines:** Support vector machines (SVM) are a must-try classifier regardless of the data structure (Wu et al., 2008). They can provide the most accurate and robust methods among all well-known classification algorithms. They require a relatively small number of training records and are generally insensitive to the number of dimensions. The result of running the classifier on the traffic data is shown below.

```
=== Run information ===
Scheme:weka.classifiers.functions.LibSVM -S 0 -K 2 -D 3 -G 0.0
-R 0.0 -N 0.5 -M 40.0 -C 1.0 -E 0.001 -P 0.1 -seed 1
Relation:     datatraffic
Instances:    2028
Attributes:   5
              Date
              Country
              Age
              Profession
              Gender
Test mode:evaluate on training data
=== Classifier model (full training set) ===
LibSVM wrapper, original code by Yasser EL-Manzalawy (= WLSVM)
```

```
Time taken to build model: 0.31 seconds
=== Evaluation on training set ===
=== Summary ===
Correctly Classified Instances       1662               81.9527
%
Incorrectly Classified Instances      366               18.0473
%
Kappa statistic                          0.2923
Mean absolute error                      0.1805
Root mean squared error                  0.4248
Relative absolute error                 51.1194 %
Root relative squared error            101.134  %
Total Number of Instances             2028
=== Detailed Accuracy By Class ===
               TP Rate   FP Rate   Precision   Recall  F-Measure
ROC Area  Class
                 1         0.789      0.81      1         0.895
0.606    Male
                 0.211     0          1         0.211     0.349
0.606    Female
Weighted Avg.    0.82      0.608      0.854     0.82      0.77
0.606
=== Confusion Matrix ===
    a    b   <-- classified as
 1564    0 |    a = Male
  366   98 |    b = Female
```

3. **C4.5 Decision Tree based on the J48 algorithm:** as details in the classifier description above, this classifier builds decision trees from a set of labelled training data using the information entropy concept based on data attributes. The result of running the classifier is shown below.

```
=== Run information ===
Scheme:weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:     datatraffic
Instances:    2028
Attributes:   5
              Date
              Country
              Age
              Profession
              Gender
Test mode:evaluate on training data
=== Classifier model (full training set) ===
J48 pruned tree
Number of Leaves  :     123
Size of the tree :      124

Time taken to build model: 0.08 seconds
=== Evaluation on training set ===
=== Summary ===
Correctly Classified Instances       1960               96.6469
%
Incorrectly Classified Instances       68                3.3531
%
Kappa statistic                          0.9
Mean absolute error                      0.06
Root mean squared error                  0.1733
```

```
Relative absolute error                  17.0057 %
Root relative squared error              41.2464 %
Total Number of Instances              2028
=== Detailed Accuracy By Class ===
               TP Rate   FP Rate   Precision   Recall  F-Measure
ROC Area  Class
                 0.999     0.144      0.959     0.999     0.979
0.976    Male
                 0.856     0.001      0.997     0.856     0.921
0.976    Female
Weighted Avg.    0.966     0.112      0.968     0.966     0.966
0.976
=== Confusion Matrix ===
    a    b   <-- classified as
 1563    1 |    a = Male
   67  397 |    b = Female
```

4. **Conjunctive rule:** The conjunctive rule consists of antecedents combined with logic AND's together to obtain the consequent of the classification. The result of running the classifier is shown below.

```
=== Run information ===
Scheme:weka.classifiers.rules.ConjunctiveRule -N 3 -M 2.0 -P -
1 -S 1
Relation:     datatraffic
Instances:    2028
Attributes:   5
              Date
              Country
              Age
              Profession
              Gender
Test mode:evaluate on training data
=== Classifier model (full training set) ===
Single conjunctive rule learner:
--------------------------------
(Profession = Housewife) => Gender = Female
Class distributions:
Covered by the rule:
Male  Female
0     1
Not covered by the rule:
Male  Female
0.894511    0.105489
Time taken to build model: 0.02 seconds
=== Evaluation on training set ===
=== Summary ===
Correctly Classified Instances       1845              90.9763
%
Incorrectly Classified Instances      183               9.0237
%
Kappa statistic                          0.7031
Mean absolute error                      0.1621
Root mean squared error                  0.2842
Relative absolute error                 45.9069 %
Root relative squared error             67.6636 %
Total Number of Instances              2028
=== Detailed Accuracy By Class ===
```

```
                 TP Rate   FP Rate   Precision   Recall  F-Measure
ROC Area  Class
                   1         0.394      0.895     1          0.945
0.803    Male
                   0.606     0          1         0.606      0.754
0.803    Female
Weighted Avg.    0.91      0.304      0.919     0.91       0.901
0.803
=== Confusion Matrix ===
    a    b   <-- classified as
 1564    0 |    a = Male
  183  281 |    b = Female
```

5. **Classification via regression:** This classifier uses regression methods whereby the class is binarised and for each class value one regression model is built. The result of running the classifier is shown below.

```
=== Run information ===
Scheme:weka.classifiers.meta.ClassificationViaRegression    -W
weka.classifiers.trees.M5P -- -M 4.0
Relation:     datatraffic
Instances:    2028
Attributes:   5
              Date
              Country
              Age
              Profession
              Gender
Test mode:evaluate on training data
=== Classifier model (full training set) ===
Classification via Regression
Time taken to build model: 4.41 seconds
=== Evaluation on training set ===
=== Summary ===
Correctly Classified Instances       1995              98.3728
%
Incorrectly Classified Instances       33               1.6272
%
Kappa statistic                          0.9529
Mean absolute error                      0.0333
Root mean squared error                  0.1111
Relative absolute error                  9.4382 %
Root relative squared error             26.4573 %
Total Number of Instances             2028
=== Detailed Accuracy By Class ===
                 TP Rate   FP Rate   Precision   Recall  F-Measure
ROC Area  Class
                   0.999     0.067      0.981     0.999      0.99
0.999    Male
                   0.933     0.001      0.995     0.933      0.963
0.999    Female
Weighted Avg.    0.984     0.052      0.984     0.984      0.984
0.999
=== Confusion Matrix ===
    a    b   <-- classified as
 1562    2 |    a = Male
   31  433 |    b = Female
```

The results of the above classifications are summarised in Table 8.

Table 8: Using different classifiers on the same dataset. The best result in terms of accuracy was obtained with classification via regression classifier with a rate of 98.4% executed in 4.41 seconds. The best execution time was achieved with Bayesian Networks and Conjunctive Rule classifiers with 0.02 second. However, both classifiers achieved around 96.5% accuracy.

Classifier	Bayesian Networks	SVM	C4.5 based on J48	Conjunctive Rule	Classification via regression
Number of instances	2,028	2,028	2,028	2,028	2,028
Classifier type	Bayes	Function	Tree	Rule	Meta
Learning type	Training	Training	Training	Training	Training
Basis variable	Gender	Gender	Gender	Gender	Gender
Execution Time (seconds)	0.02	0.31	0.08	0.02	4.41
Error rate (%)	3.501	18.0473	3.3531	9.0237	1.6272
Accuracy (%)	96.499	81.9527	96.6469	90.9763	98.3728

The dataset had 2,028 instances and the chosen basis variable was the binary variable *Gender*. This means that data are classified based on Gender and hence may be able to predict gender in a new dataset with the same data structure. All the classifiers used the dataset as a training set for learning. As Table 7 shows, when using various classifiers on the same dataset, different results were obtained. The best result in terms of accuracy was obtained with the classification via regression classifier with an accuracy rate of 98.4% and execution time of 4.41 seconds, being the slowest among all the tested classifiers. The best execution time was achieved with Bayesian Networks and Conjunctive Rule classifiers with 0.02 seconds each. However, both classifiers achieved around 96.5% accuracy. The worst classifier in terms of accuracy for the particular data set and basis variable was the support vector machines (SVM) classifier with an accuracy rate of 82.0% only. Neural network was run on the dataset on the same machine for around one hour but no results were obtained.

5.2.3.2. *Model verification*

The generated models were saved and used to verify the recorded accuracies of the classifiers with a test dataset. The test dataset consisted on 22 records with the same structure as the training set but with entries that were never fed to the classifiers. The model verification results attained are summarised in Table 9.

Table 9: Model verification results: As can be seen in the table, all classifiers apart from Bayesian Networks successfully classified 19 instances out of a total of 22 based on Gender. The Bayesian Networks successfully classified 18 instances out of a total of 22.

Classifier	Bayesian Networks	SVM	C4.5 based on J48	Conjunctive Rule	Classification via regression
Number of instances	22	22	22	22	22
Classifier type	Bayes	Function	Tree	Rule	Meta
Verification type	Test dataset	Test dataset	Test dataset	Test dataset	Test dataset
Basis variable	Gender	Gender	Gender	Gender	Gender
Correctly classified instances	18	19	19	19	19
Incorrectly classified instances	4	3	3	3	3
Error rate (%)	18.1818	13.6364	13.6364	13.6364	13.6364
Accuracy (%)	81.8182	86.3636	86.3636	86.3636	86.3636

Table 9 shows that all classifiers had similar results in classifying the test data, although at model generation the classifiers had different error and accuracy rates. Several reasons may be the cause of this, including the number of instances supplied in the test dataset and the percentage of data considered in the training datasets, among many others.

The significance of the results however lies specifically in the misleading conclusion that might be arrived at from considering the classification accuracy rates obtained without applying actual datasets. For example, the Bayesian Networks classifier achieved 96.5% classification accuracy rate at the model generation. However, when fed with the test dataset, the model achieved only 81.8% accuracy rate. Table 10 shows the differences in accuracy rates between the models generated when training the data and the models when were verified with the test data.

Table 10: The difference between model generation and model verification accuracy rates for each classifier

	Bayesian Networks	SVM	C4.5 based on J48	Conjunctive Rule	Classification via regression
Model Generation Accuracy (%)	96.499	81.9527	96.6469	90.9763	98.3728
Model Verification Accuracy (%)	81.8182	86.3636	86.3636	86.3636	86.3636

As the above results show, ADPO or other public organisations, may be able to predict values of missing records based on saved models of their datasets. This can contribute greatly to data quality of the data used later on for decision making. It is important, however, to note that although a certain rate of classification accuracy is attained with a classifier, this rate might be misleading. It is important, therefore, to test the attained models using real datasets in order to verify the obtained accuracy. As could be seen above, this was proven crucial in the traffic data from ADPO, and accordingly, better insights will be attained into the datasets.

The choice of the data mining method can be decisive for analysing certain data. As this chapter illustrated, certain techniques may not produce adequate results, which may lead to the false belief that the data analysed hold no useful information. Given the variety of processes and their interconnection in public organisations, such organisations would not be able to afford the "trial and error" mode of data mining as these organisations usually operate in "right first time" mode. Therefore, data mining as a choice for a public organisation such as ADPO must be integrated in other processes that the organisation undertakes in order to attain results. This necessitated that this research study be oriented towards suggesting a more comprehensive model of data analysis. This model is described in the following chapters.

Chapter 6: Integration of Decision Support Systems and Data Mining for Improving Data Quality for Decision Making

6.1. Introduction

Decision Support Systems (DSSs) comprise different aspects of software, hardware, and data, as well as human inputs in order to help decision makers improve and enhance their decisions based on analytical processing of the available information. DSSs have found uses in different types of organisation and wherever strategic decisions are to be made with high uncertainty involved. On the other hand, Data Mining (DM) allows robust analysis of huge amounts of data in order to discover useful relationships or patterns based on advanced statistical methods. Integration of these two paradigms is believed to improve data quality for decision makers as they both apply data refinement processes.

In this chapter, a data mining decision support (DM-DS) integrated system is suggested to improve data quality and a perception of the system is provided. The chapter looks into existing DM-DS integrated systems in the literature and conduct a comparative analysis to propose a system that is tailored against the particularity of Abu Dhabi public organisations to assist decision makers in the decision making process based on better knowledge of the stored data. The solution could be implemented by integrating a set of DM techniques in the already established DSS in the organisation. The new system is expected to increase the efficiency sought from the DSS by acquiring knowledge of the data fed to the system, thereby improving data quality. As seen in Chapters 4 and 5, that DM processes are demanding in terms of time and other resources. Hence, it is suggested that the process be reduced in order to accommodate to the particularity of public organisations in terms of size and current system performance. The system is perceived on this basis.

6.2. Background and Related Work

Public organisations face real challenges of using the correct analysis of huge amounts of data. These data are used for producing statistical analyses and forecasts on, for instance, economic, social, education and health issues, which are highly related to government planning in aspects where development of interest rates and inflation, economic growth, household income, crime trends, education standards, and climate change are a major input. Whereas business providers may be interested essentially in extensibility and automation and aim at obtaining fast results via combining simple analysis with human expertise, public organisations would have to be more scrutinising in their data analysis approaches (Ganguly & Gupta, 2004).

Decision support systems (DSSs) are rooted in business. They have been in wide used in businesses around the world for the main aim of helping executives to make better decisions based on advanced levels of data refinement and presentation.

Traditionally, DSSs belong to a class of computer-based systems that help in the process of decision making (Hardin & Chhieng, 2007). DSSs are commonly defined in the literature as interactive computer information systems designed to support solutions to problems with taking decisions (Liu et al., 2009).

Padhy et al. (2012) argue that the value of strategic information systems is easily recognised, but the efficiency and speed are not the only factors of competitiveness, rather the quality of the data that such systems can attain. Essentially, the large amounts of data have called for new methods to analyse and understand the relationships in these data. Conclusions and inferences from these data need special tools and techniques that are able to delve deeper than traditional decision support systems can. Moreover, the rapid development in data digitisation generated large amounts of data stored in organisations' data warehouses, which require efficient exploitation and knowledge extraction. Consequently, traditional problem solving DSSs became less efficient and started to decline in the late 1990s (Liu et al., 2009). Liu et al. (2009) classify the main challenges facing DSSs in supporting decision making. These challenges include:

- Changes in technology from database to data warehouse and on-line analysis processing (OLAP), from single user model to World Wide Web access, and from mainframe to client/server architecture;

- Increasing business interconnections in a more dynamic business environment and intelligence. For this, a variety of other information systems have been proposed, such as supply chain management (SCM), enterprise resource planning (ERP), and customer relationship management (CRM);
- The continuous increase in complexity in decision making which requires executives to consider a vast number of inputs and a considerable amount of knowledge.

Han and Kamber (2006) argue that organisations are usually data rich but information poor. Extracting information from data is not only important but also necessary to use these data for decision making. Data mining techniques help to analyse data and uncover important data patterns, which may contribute to better, knowledge-based strategies. In doing so, data mining helps to bridge the gap between data and information, which usually prevents realising knowledge.

Given the uses and nature of solutions based on data mining and those based on decision support systems, there is no doubt that such solutions can be integrated in order to offer optimal solutions to improve the quality of data reaching decision makers. However, only a few attempts of using data mining to support management decisions are found in the literature. For example, Abu-Naser et al. (2011) suggest a DSS based on data mining techniques for optimising e-learning in educational institutions (Figure 23). The suggested system was not implemented, but according to the authors, integrating data mining functionality into a single DSS will be promising. The authors believe that such a system will enable educational institutions to realise the importance of the DSS-produced information in optimising their adopted learning strategies.

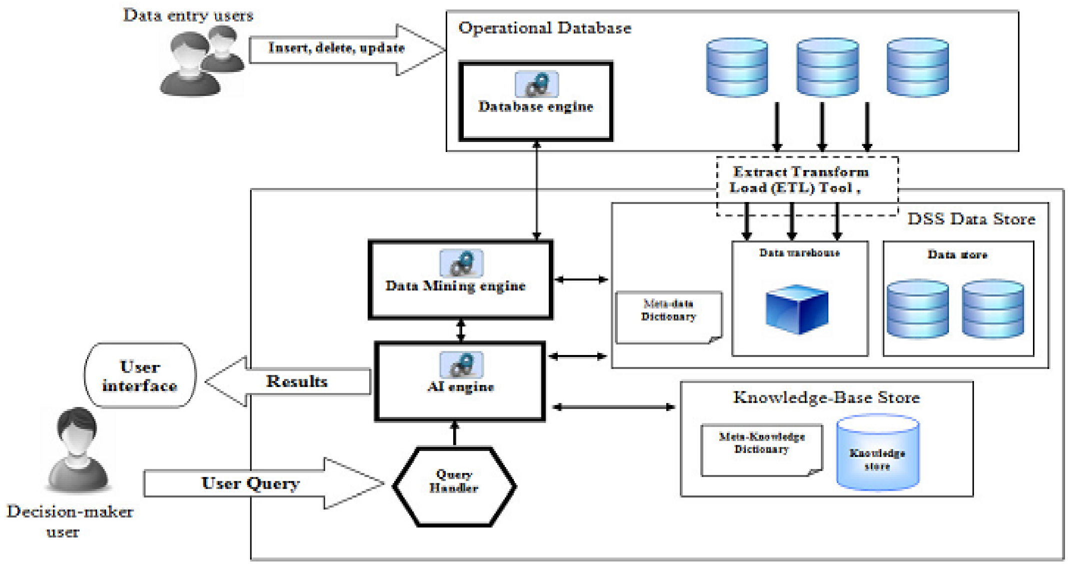

Figure 23: DM-based DSS for optimising e-learning suggested by Abu-Naser et al. (2011)

El Seddawy et al. (2012) propose a DM-based DSS to support top-level management to make a good decision at any time (Figure 24). According to the authors, the proposed system can help decision makers in the banking sector to address decisions related to new investments by providing refined data.

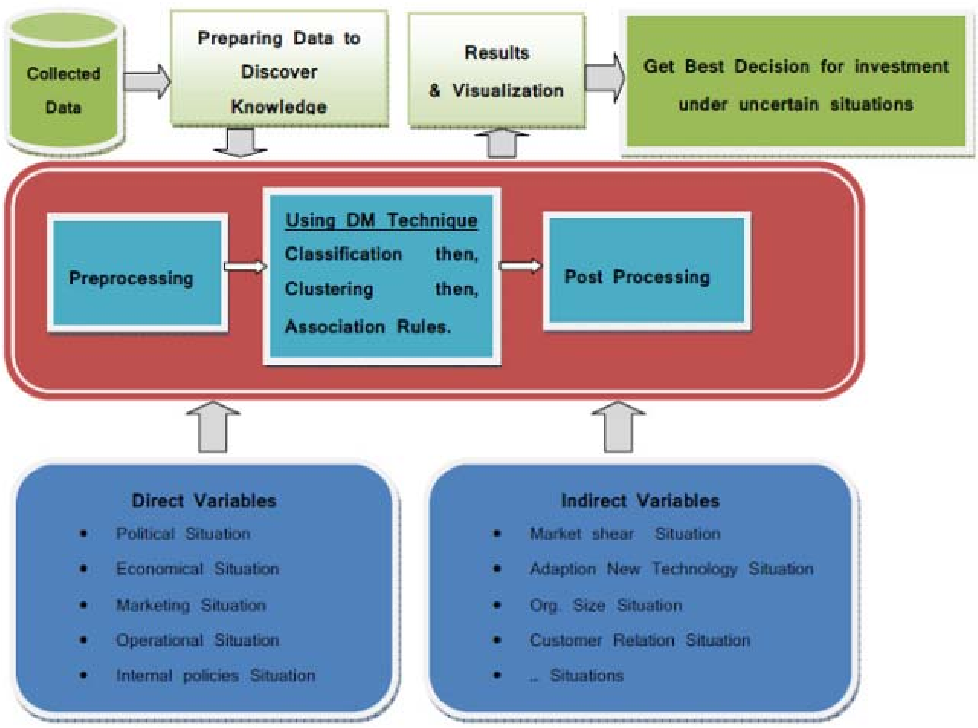

Figure 24: DM-based DSS suggested by (El Seddawy et al., 2012) to help decision makers make decisions regarding new investments

In the application area of health services, Kumar et al. (2011) suggest the use of DM based on decision tree algorithms to classify certain diseases and compare the effectiveness and correction rate among them in order to support decisions on the diagnostic process. According to Kumar et al. (2011), traditional decision support systems developed to assist doctors in the diagnostic process are usually based on static data which may be dated. Therefore, a decision support system which can learn the relationships among certain parameters would be very useful to doctors and hospitals.

Mohemad et al. (2010) argue that traditional support systems that are widely used in the construction industry are not optimal. In spite of the efforts to integrate and transform the whole construction tendering processes into electronic or digital forms, the use of unstructured documents either in hard copy or digital are still widely present. The authors stress the need to extract and represent information in machine-readable formats, attained by integrating data mining in DSS model, which they believe to be a promising approach.

Liu et al. (2010) have conducted research into integrated decision support systems (IDSS) including DM agent-enhanced integrated DSS to improve decision support performance. The researchers conclude that the main challenges in developing an integrated DSS are the trade-off between tight and loose integration strategies within the integration frameworks and the seamless integration across data, models and processes within the integration frameworks (Figure 25).

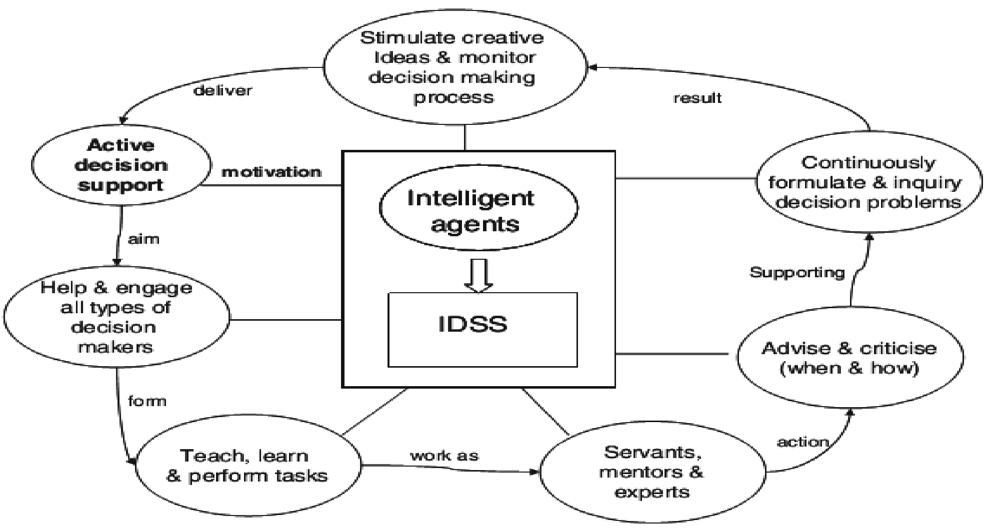

Figure 25: DM agent-enhanced integrated DSS to improve decision support performance (Liu et al., 2010)

Srinivasan et al. (2011) also suggest using DM intelligent agents in DSS for achieving higher work efficiency. They suggest that such a system can provide mobility, autonomy and collaboration of different agents in order to provide a simple and fast solution. Srinivasan et al. (2011) maintain that using agents as data mining techniques, can help decision makers by providing a more robust and quick DSS in resolving issues in any complex situation.

Mladenić et al. (2003) maintain that there has been no systematic attempt to integrate DM and DSS. Reasons behind that are many but mainly include the nature of data mining processes that combine computer science and statistics, which create some confusion on what implementation aspects may be suitable for managerial decisions.

In an initiative to address the drawbacks of decision support systems the EU sponsored the SolEuNet project from 1999 over a 39-month period, which comprised a network of expert teams from business and academia to meet client's Data Mining and Decision Support needs (Mladenić, 2001). The outcomes of the project were promising. The project identified the main objectives to improve collaboration and communication, promote awareness of organisational resources and achievements, and enable organisational learning and dissemination of such knowledge. However, certain difficulties were encountered as detailed below.

According to the final report at the project closure, the project team maintained that collecting and managing knowledge is a very hard task that can never be fully accomplished. The project team considered that the set of knowledge management tools to support e-collaboration in data mining and decision support can be considerably extended, improved and further integrated. The team suggested including in the near future, a data mining advisor that, given a dataset, suggests data mining algorithms, and a central model evaluation service, automatically built activity logs for data mining and decision support, and database and data transformation services. The team also suggest adopting a standardised description of the knowledge produced in all phases, which would significantly simplify communication among distributed cooperating groups.

Table 10 summarises the different approaches to linking DM to DSS.

Table 11: Approaches to DM-DS systems found in the literature

Author	Context	Approach
El Seddawy et al. (2012)	The banking sector	DM-based DSS to support top level management to make decisions in any time
Abu-Naser et al. (2011)	E-Learning in educational institutions	DM-based DSS for optimised results
Kumar et al. (2011)	Disease classification	DSS that can learn the relationships between certain parameters
Srinivasan et al. (2011)	Higher work efficiency	DM intelligent agents in DSS
Liu et al. (2010)	Research on different DSS's	Integrated Decision Support Systems (IDSS) and DM agent-enhanced IDSS to improve decision support performance
Mohemad et al. (2010)	Construction industry	Optimising DSS in the construction industry through DM
Mladenić (2001)	SolEuNet	Enabling organisational learning and dissemination of knowledge

Related Publication:

Al-Ketbi, O., & Conrad, M. (2013). Integration of Decision Support Systems and Data Mining for Improved Decision Making. *ICEIS 2013*, 503.

6.3. Relevance of the Integration to Data Quality

As previously stared, the aim of retaining data is no longer a significant concern of an organisation as it is common that large amounts of data are regularly stored in its warehouses, with only a few of real relevance to the organisation's decision making process. The amount of data stored in databases increases daily and goes beyond the technical skills and human capacity to interpret these data.

Database management systems have advanced at a faster rate than the techniques used for extracting and utilising data to be used in making decisions (Power, 2007) in the sense of extracting useful information (Lv & Li, 2009). Obtaining, storing and managing information in larger organisations are now ordinary operations and usually performed automatically by electronic data repositories (Saxena & Rajpoot, 2009).

One of the efficient techniques used for this aim (i.e., extracting useful information) is data mining. Some of the organisation's data may be in a textual format described in natural languages or does not have a structure as opposed to data present in data tables and structured relational databases. This type of data, found mainly in the form of electronic documents and emails for instance particularly in organisations with limited affordability (technical and financial), cannot be used directly with traditional DSSs, and thus minimises their potential. A particular data mining set f techniques, known as Artificial Intelligence (AI), such as Neural Networks and Fuzzy Inference algorithms, would be appropriate for finding a solution to this problem by extracting information from data of various formats.

The integration of DM and DSS are particularly interesting for public organisations. Public organisations can benefit greatly from a DM-DS integrated system because of the vast and loosely related information these organisations deal with. According to McCue (2003) one of the most significant challenges in using decision support systems in law enforcement is that most, if not all, data are not intended for the decision making process. This issue applies as well to public organisations in Abu Dhabi Emirate. Despite the state-of-the-art DSS implemented in some of the public organisations in the Emirate, little success has been seen in improving decisions. Some of the challenges faced are related to data formats, content, validity and reliability. Hence, the biggest advantage data mining would provide is preparing data for this undertaking. For example, law enforcement organisations can use computer technologies that support criminal investigations, which include clustering and link analysis algorithms, geographical information systems displays, and the more complex use of data mining technology for profiling crimes or offenders and matching and predicting crimes (Oatley et al., 2006). Oatley et al. (2006) also argue that knowledge from disciplines such as forensic psychology, criminology and statistics are essential to the efficient design of operationally valid systems, thereby providing organisations' decision support systems with refined data quality that would improve the decision making potential.

6.4. Possible Data Mining Techniques to Use

As detailed in Chapter 2 (Section 2.5), 4 and 5, data mining can be used to improved data quality. As argued in previous sections of this chapter, when combined with DSSs, data mining can help decision makers obtain better data that help them make more effective decisions. Recent developments of information systems as well as the availability of extensive business data repositories and database management systems, accompanied by the advances in computer systems and algorithms, have provided a gateway to effective use of data mining (Hand et al., 2001).

Data mining consists of a set of techniques inferred from statistics and Artificial Intelligence (AI) with the specific aim of discovering new, useful, relevant and non-trivial knowledge that may be hidden in a large mass of data (Markov & Larose, 2007). There have been numerous examples of its uses in areas such as marketing, economics, engineering, medicine, among others.

Several techniques of data mining such as classification, neural network, genetic algorithm and others have long been known (Segall & Zhang 2006). What distinguishes recent perception of data mining is the development of techniques for data mining applications on a large scale databases. In addition, several techniques have emerged from the field of database management and are now an integral part of the process of data mining.

Depending on the application domain and user interest, various types of techniques can be identified and applied. Some of these techniques are briefly identified below in order of relevance to the DSS of public organisations in Abu Dhabi. For example, dimensionality reduction is useful when the number of involved variables exceeds the capacity of the DSS to perform better, whereas modelling helps improve the DSS operations by feeding it with a stripped-down, quality version of a collection of data which models the entire collection.

6.4.1. Dimensionality Reduction

Dimensionality reduction is a mathematical technique used to reduce the number of random variables involved in a dataset. It uses projection from one vector space onto another one of lower dimension.

6.4.2. Correlation

Correlation is a statistical method used to depict relationships between random variables through analysing potential links and inferring the degree of connectivity. Given the large number of loose data available for the DSS, correlation helps establish useful relationships among seemingly unrelated variable.

6.4.3. Modelling

Modelling or mathematical modelling is a description of observed behaviour, simplified by ignoring certain details to simulate the behaviour of a phenomenon. Modelling allows complex systems to be understood and their behaviour to be predicted within the scope of the generated model.

6.4.4. Association

Finding association is a data mining technique that allows searching for simultaneously occurring items that occur in the transactions database. Algorithms such as DHP (Dual Heuristic Programming), GSP (Generalised Sequential Pattern) and Apriori (Han et al., 2006) among others are examples of tools that implement the task of discovery of association.

6.4.5. Classification

Classification is a method that consists of defining a mathematical function that maps a set of records to one another in a predefined set of categorical labels, called classes. This function is applied to predict the class new records fall under. Some classification methods were experimented on collected data from one of the case study organisations in Abu Dhabi in Chapter 5.

6.4.6. Regression

Regression includes a search for a function that maps the records from a database to actual values. This method is similar to classification, but being restricted only to numeric values.

6.4.7. Clustering

Clustering is used to separate records in a database into subsets or clusters, such that the elements of a cluster share common distinguishing characteristics from other clusters. The objective of this task is to maximise intra-cluster similarity and minimise inter-cluster similarity. Unlike a classification task, which has predefined labels, clustering needs to automatically identify the data groups to which the user must assign labels. Some algorithms used for implementing this method are K-Modes, K-means, K -Prototypes, K-Medoids, among others. (Han et al., 2006).

6.4.8. Summarisation

This task, very common in KDD (Knowledge Discovery in Databases), is to seek, identify and indicate common features among data sets. Inductive logic and genetic algorithms are some examples of technologies that can be applied in summarisation.

6.4.9. Detection of Deviations

This technique helps to identify records in the database whose characteristics do not meet the normal standards. Statistics is the main resource provider used by this technique.

6.4.10. Discovery Sequence

It is an extension of the technique of finding associations that are sought frequently by considering several transactions occurring over a period. The technique of association can be adjusted to engage the generalised mining association rules. The post-processing step includes processing the knowledge gained in data mining. Among the main tasks of the post-processing step are preparation and organisation, and may include the simplification of charts, diagrams or reports demonstrating, in addition to the conversion of the representation of knowledge gained.

6.5. Data Mining-Based DSS Solution

Lessons learnt from science and business applications of data mining identified in this chapter can be transferred to a certain degree to the situation of public organisations in Abu Dhabi. Practitioners have devised methods for obtaining relevant and quality data from a large set of data and feed them to the organisation's DSSs.

Optimising the decision making process requires targeting different stages of the process in order to improve the quality of the supplied data. This study suggests a DM-DS integrated system to improve the decisions made in public organisations in Abu Dhabi. For the operations of public organisations in Abu Dhabi, the process needs to be tweaked in order to accommodate the particularity of these organisations in terms of size and current system performance. Data mining and DSS techniques can be used to design and implement custom applications that help improve data quality, for example by pattern recognition, predictive model design and human-error reduction (Ripley, 2008).

What is proposed is improved implementation of DS system based on utilisation of data mining techniques to attain optimal results gained from improved data quality fed to the DS system. The proposed DSS system will benefit from data mining processes prior to having to deal with huge amounts of data acquired from the different departments involved. The techniques involved will be selectively deployed on the different sets of data acquired. The system will require a set of techniques to be able to deliver the desired outcomes. Those techniques consist of the database system, the business logic (the integrated DM-DS system) and a user interface model. Several components will make up the system. The collected data will selectively undergo several processes prior to be sent to the DSS. These processes range from data preparation into classification and cluster analysis. Once done, the data enter into an anomaly-detection process where anomalies in the processed data will be detected and flagged up for further investigation. The system will contribute to the development and enhancement of data acquisition and analysis processes through defining a development approach and setting up all the needed variables, resulting in improved data quality at the hands of decision makers.

6.6. Chapter Summary

To study an integrated solution of data analysis in public organisations rather than rely on data analysis techniques for reasons described in this chapter, the chapter looked at the integration of DM-DSS (Data Mining-Decision Support System). It was shown that the integration of DM and DSS, given the particular case of Abu Dhabi public organisations, could improve the effectiveness of decisions by improving the quality of the data fed to DSSs. The suggested system uses DM as a pre-process to the DSS implementation. The suggested system could improve the quality of data when fed to the DSSs used.

Chapter 7: Findings

This chapter describes the findings obtained from the interviews conducted with senior managers at seven public (i.e., state) organisations in the Abu Dhabi Emirate. The organisations are: the Abu Dhabi Police Organisation (ADPO), Al Ain Hospital, the Department of Economic Development of Abu Dhabi, the General Authority of Youth and Sports Welfare, Zayed Foundation, Tawam Hospital, and the Department of Municipal Affairs. The choice was based on several considerations but most importantly the involvement of the organisations in different federal bodies of the public sector in the country, such as law enforcement, traffic regulations, the healthcare sector, and civil defence, among others.

Two senior managers from each organisation were interviewed. Posts of the managers in each organisation were detailed in the methodology chapter.

The interviews asked senior managers at the studied organisations questions as to how the decision making process is carried out in the organisation, who the staff involved are, and what barriers there are to effectively carry out the process. The interviews provided a wealth of information on how decision making takes place. This provided insights into the process and the senior managers' beliefs about the decision making process. The key findings are provided below. The results aggregate common issues identified in the studied organisations. These issues are organised into main themes.

Using grounded analysis as described in the methodology chapter, the findings were categorised in themes in order to identify designate groups of instances with common characteristics (such as processes, events and occurrences, see for example Charmaz, 2003). It is important to notice that category identification in grounded analysis is different from content analysis, which this research was careful not to confuse them. Content analysis makes use of categories defined prior to starting data analysis, which are designed to be mutually exclusive. Therefore, the same data cannot be assigned to more than one category. Conversely, categories in grounded analysis emerge from the data, they are not mutually exclusive and they evolve throughout the research process. With a low level of abstraction for categories, Table 12 summarises the findings in the organisations.

Table 12: Summary of the findings in the 7 organisations. The column on the left defines a specific aspect of the findings and the other column details this aspect

Finding	Details
1. Known challenges to decision making	1. Decisions come from headquarters without close attention to the specific needs of the organisation 2. The organisation's decisions need to comply with headquarters' decisions 3. Unavailability of information in a timely and appropriate manner and format 4. Lack of efficient data storage and retrieval procedures 5. Lack of regulation and standardisation of information storage and format 6. Lack of information and data integration with various organisations, whether federal or local 7. Lack of accurate data for decision-making 8. Inadequate spread of strategic thinking and long-term planning 9. Lack of commitment to carry out plans due to insufficient funds allocated to the implementation of the plans 10. Lack of ability to verify quality of information received 11. Lack of creditability and proficiency of staff 12. Different perspectives on priorities 13. Lack of power/authority coverage of some of the decision-making levels 14. Inadequate time available to make efficient decisions
2. Potential reasons for known data quality issues	1. Lack of data validation and cross-referencing 2. Decision makers do not have the resources to make sure of the quality of information they receive 3. Anomalies in values are settled in the headquarters' favour 4. Remedies to poor quality are based on the experience of the decision maker 5. Incompetence of staff in data handling and data quality 6. Lack of staff training on data handling and data quality

Finding	Details
3. Data generation, handling and storage	1. Data come from various loosely coupled systems 2. System data are validated against data from questionnaires, e-mail correspondence, field visits, forums, and polls 3. Storage is done in central systems 4. Reports are limited to certain staff to ensure confidentiality of information 5. Strategic data creation and dissemination is done by one department 6. Some data are stored in the internal servers and do not reach decision makers 7. Use of other company's services to store and assess information 8. Significant amounts of data is still being stored in a paper format
4. Realising the importance of data quality	1. Recognise that poor data can lead to many problems from inefficient performance to poor services to citizens 2. Understand the multitude of sources of poor data 3. Smarter investments benefits from quality data 4. Some sources of poor data quality include: Human errors at the entry levels, system failure, and misinterpretation at the strategic level 5. Associated poor data quality to increased costs 6. Recognise that preparation of strategic plans, processes and methodologies such as the ISO standard require significant and reliable data 7. Recognise that improvement of data quality from the technical and human sides is required

Finding	Details
5. Data quality assessment	1. Decision makers rely on the feedback from affected organisations, which is not a regular or specific mechanism 2. Data received from central authorities are usually considered valid 3. Statistical data is reviewed by specialists and reviewed by officials to match consistency with mainstream data of the organisation 4. Central authorities may allow organisations to give counter proposals to their data reports based on evidence 5. Departmental managers are responsible for validating their data 6. Managers refer to procedures and policies provided by higher authorities in their validation 7. Decision makers are advised to follow best practices and refer to international standards whenever possible 8. There are no specific internal criteria of acceptance of data 9. Random sampling, checking of the sources and providing evidence upon request are used 10. Evaluation is also by specialists, analysts and strategic planners
6. The decision-making process	1. Central authorities set strategic planning for adopt and implement the organisational strategy 2. Decisions are usually taken by senior leaders 3. The strategic committee discuss and analyse data provided by the strategic planning department and make strategic decisions 4. Strategic analysis is conducted annually, which includes analyses of many sources of internal and external information 5. The analysis aims for the development or amendment of existing strategic objectives 6. Results are discussed with the various sectors of the organisation 7. Results may be moderated by the executive committee 8. Periodical reviews of the decisions are conducted by the executive committee

Finding	Details
7. Sources and nature of data used in decision making	1. Information systems are also accessible by internal or external decision makers 2. Two types of data sources: internal external 3. Internal data are mainly performance indicators 4. External data such as government policies, international standards and environmental criteria 5. Decisions are mainly based on the strategic orientations of the government and its strategic vision 6. Decisions rely on financial statements, operational data and customer related feedback. 7. Data are subject to studies (including analysis of economic indicators) and other references such as customer satisfaction questionnaires, strategic plans of the organisations, the annual and quarterly performance reports and annual economic policies 8. There has been an initiative Emirate-wide to establish a knowledge base to be used by decision-makers
8. Quality of data used in decision making	1. It is not known how often decisions are made based on quality data due to the complexity of the decisions 2. Decisions are taken by the central authorities and their effects are circulated to the relevant organisations 3. Decisions are either based on the data provided, or they are subsequent implementation of the authorities' mandates 4. Decisions are often made based on the strategic vision of the authorities 5. After implementation organisation work on anticipating and rectifying any negative outcomes of the decision 6. Decisions frequently rely on VIP opinion or view of certain influential people 7. Not all decision makers rely on data and figures to the same degree 8. It is common that the organisations use the quality of the outcomes of a certain decision as criteria for such assessment

Finding	Details
9. Data handling policies	1. External policies from the governing bodies or high authorities are usually referred to locally in the organisations 2. Local data handling policies of the organisations are mainly concerned with data security 3. The knowledge management strategy by central authorities attends to some aspects of data handling yet with some issues with the implementation of the strategy 4. The need for internal development of the information systems and electronic archiving, and improved skills in dealing with data and documents is evident 5. Paper forms of data still prevail 6. Department managers are responsible for validating their departments' data with sufficient information available 7. The management of each of the organisations restrict the vital role of data quality to the first stage data entry clerks and their direct managers
10. Reporting	1. Certain reports run periodically and are provided to key individuals and/or groups 2. Emails circulated among staff are a typical method of reporting in the organisation 3. Some of these emails contain effective decisions and even submissions of complaints and suggestions 4. It is common also that various reports are collected from the different department and sent to those concerned by e-mail or in print 5. The underway Emirate-wide knowledge base is believed to improve the reporting process inside and outside the organisations 6. Occasional and regular meetings are a reporting method, whereby vertical and horizontal communication takes place. 7. The authorities have access to electronic systems, where they can collect and audit data for their reports

As can be seen in Table 12, 10 categories were identified by the study's findings. The categories highlighted key aspects of data handling and potential causes of data quality issues in the organisations. The interviewees described known challenges to decision making in their organisations, potential reasons for known data quality issues, how data are generated, handled and stored, their organisations' perspectives to data quality, assessment used for data quality, the decision-making process that takes place in their organisations, the sources and nature of data used in decision making, the quality of data used in decision making, the policies and procedures for data handling imposed by the organisations, and the reporting adopted by the organisations.

The elements of the above categories are discussed in the next chapter in relation the relevant literature and well-established work on data quality.

Chapter 8: Discussion

8.1. Analysis of the Findings

The literature review conducted revealed that public organisations, in general, store vast amounts of data in different formats that are acquired on a daily basis. Stored raw data however may provide little benefit for example for the purpose of developing strategies and devising and implementing long term plans. Moreover, information is not often "right the first time", which necessitates further refinement and presentation when provided to decision makers. Such information is used for producing statistical analyses and forecasts on economic, social, health and education issues, which are highly related to government planning in aspects such as development of interest rates and inflation, economic growth, household income, crime trends, education standards, and climate change are all major inputs. This is also significantly related to the measurable and non-measurable aspects that public organisations aim to attain, in terms of efficiency, effectiveness and responsiveness (for example, Andrews & Van de Walle, 2012; Fountain, 2001). These organisations need "good" information to make good decisions on the above aspects. The goodness of the information is controversial and is subject to many, subjective and objective, considerations. Generally speaking, there are certain measures that can be adopted to assess the quality of the acquired information. The loose nature of definitions of standards and measures for the quality of information (or data), as well as the fact that most data quality measures are devised on an ad hoc basis (Pipino et al., 2002), renders those measures ever more individually determined. For example, data accuracy, reliability, timeliness, completeness and relevance, among others, are common dimensions in many considerations for data quality (Pipino et al., 2002).

Further development of the subject of data quality and several attempts to synthesise global metrics for it have been undertaken by scholars and researchers in the literature. For example, Pipino et al. (2002) use the above dimensions to develop subjective and objective assessments, simple ratio, min/max operators, and weighted average for data quality. However, the authors maintain that such an attempt to unify data quality metrics is limited as "one size fits all" cannot be a solution. Dedeke (2000) argues, however, that although data may be evaluated based on quality categories such as accessibility, context and representation, these data quality measures cannot be developed in isolation from the data generation processes and the data utilisation contexts.

Based on the above, empirical studies were conducted to contextualise data quality in a specific context. The findings from seven public organisations in Abu Dhabi Emirate, which were investigated for the aim of understanding potential issues with data quality and hence improving it. The findings were obtained from surveys and meetings held in the organisations and from interviews conducted with senior members in the organisations. The findings are discussed in this chapter in relation to existing work. The aim of the discussion is to identify aspects of improving data quality in the studied organisations and organise these aspects in a normative framework. The term "organisation" is used throughout to represent a public-sector structure and can be replaced by the commonly used term when needed (for example, public body, department, etc.).

8.1.1. Data management and systems

Each of the organisations studied acquires a large amount of information from various internal and external sources. The vast amount of information is stored and processed in several information systems. The organisations have a number of electronic systems for data storage, processing, and retrieval. Common systems mentioned by the interviewees are, a directorate system, penal system, congestion system, port security system, human resources system, financial resources system, and a civil defence system. These systems are all functional and used to support decision makers and help in several processes along their main functions, such as provide statistics, detect anomalies, and indicate trends, among others.

According to the interviewees, usually one department is responsible for strategic data creation and dissemination. Almost all data are generated by the business intelligence department; then reports are created and distributed to those who should receive them. Reports are also generated by other departments, such as information management, business performance, and quality departments. A few other departments run their own data from their own systems such as IT and HR. Data usually come from multiple systems. For example, Interviewee 3 said that Al Ain Hospital uses three different information systems: Oracle for HR-related data, Malaffi for health service related data, such as the number of visits, type of service, patient financials, etc., and risk assessment systems. At the creation of any new report, a validation process is conducted to ensure that the report is pulling correct and accurate data. Some reports are sent to different departments for validation. For example, Interviewee 3 stated that: "*one report at Al Ain Hospital is created to show the number of cases that had a DVT (deep vein thrombosis) after going through a surgery. At the hospital, we validate by cross-referencing the report for a short timeframe and checking the cases to see if they really meet the report criteria and if the results are consistent. However, data entries are usually not validated. So, although some HQ reports are validated by the hospital, entry level values that may have human errors are not usually validated*."

Generating reports is limited to certain staff within those departments to ensure confidentiality of information, and limit the load on reporting tools. Questionnaires, e-mail correspondence, field visits, forums, and polls compared with normative data after auditing are other ways mentioned by interviewees for validating data.

Storage of electronic data is done on enterprise servers where all data are hosted centrally, even applications and organisation-wide licensed software. Generated reports are stored, if needed to be stored, at the generating department in protected shared folders. Otherwise they can be regenerated whenever needed. Only authorised personnel may store data in the central servers. It was noted by interviewees that a significant amount of data is still being stored in a paper format. Paper-based data are stored locally for some time and later may be moved to secondary storage areas.

Proposed enhancement:

The following is suggested to address the issue and is reflected the "People and Systems" element of the framework as shown in Figure 30:

Management of information determines data quality policies and directs their implementation. This function should provide systematic and planned actions needed to attain adequate confidence that data meet a predefined set of quality requirements. In this respect, integration of the organisation's information system is a first step into data integrity which is a main factor of data quality. Each organisation should integrate its information systems in order to manage and compile a range of processes. The integration should consider staff, software and hardware deployment with the ultimate goal of improving the accuracy and quality of decisions. Moreover, data and information processing and interpretation techniques can be used according to their feasibility. Decision Support Systems (DSSs) may be used by decision makers to improve their decision making insights. New and emerging techniques of data quality enhancement and knowledge extraction may also be of significant importance for the organisation as shown in the experiments conducted, and should be considered as key aspects of improving data quality. In this sense, data mining techniques and their integration with decision support systems as detailed in previous Chapters 4, 5 and 6 will help improve data quality and provide decision makers with valuable information.

8.1.2. The decision making process and data quality

The interviews indicated that it is common in the studied organisations that decisions are usually taken by the senior leaders after reviewing all the required inputs, including data and views of expected outputs. The strategic committee – which consists of a working group of senior managers headed by the enterprise leader in the presence of members of the council, who are executives and directors of departments – discusses and analyses data provided by the strategic planning department and make strategic decisions. The data flow upwards from departments to the governing bodies up to the senior management. The central authorities hold a strategic planning session on how to adopt and implement the organisational strategy. Data from the organisations reaching the authorities are used to support the decision making process, and to ensure that initiatives and decisions serve the priorities and strategic objectives of the organisations.

The studied organisations face frequent difficulties in attaining data quality. The challenges start from data entry up to data extraction and analysis. The interviews revealed that unnecessary bureaucracy takes place in some cases. For example, Interviewee 3 maintained that most decisions at the strategic level come directly from the headquarters, and one of the challenges encountered is that those decisions sometimes have no consideration for the organisation's available resources. Interviewee 4 stressed the same point and asserted that when the headquarters decides that the budget will be reduced to a certain amount, the organisation has to face the challenge of finding places to make cuts or cost reduction with the minimum impact possible. Interviewees 2 and 3 both maintained that consequences of headquarters' decisions on the organisation require tweaking the organisation's own decision. They hence use their data to mitigate the influence of the decisions, such as data related to service size and staffing to see if cutting some positions would do. They mostly try to cut vacant positions or reduce the size of the service.

The interviewees also referred the challenges to decision making to the data available to decision making, their storage and retrieval procedures, and their quality. For example, Interviewee 4 said that unavailability of information in a timely and appropriate manner and format is one of the main problems facing the organisation. Interviewee 5 asserted that the lack of regulation and standardisation of information storage and format across public organisations is another problem facing an efficient decision making process. Interviewee 4 revealed that the Capability Maturity Model Integration (CMMI) Methodology has been initiated in their organisation.

Proposed enhancement:

The following is suggested to address the issue and is reflected the "Leadership" element of the framework as shown in Figure 30:

It is evident in several data quality studies (for example, Lin et al., 2007; Reid & Catterall, 2005) that data quality considerations and data quality awareness should be attained at an early stage, i.e., data entry. Data entry clerks should understand the data that they deal with and their importance for the organisation. The staff awareness of the importance of having minimal errors in the data entering the electronic systems is crucial. For example, they may need to pay attention to the appropriate fields and question any anomalies they may notice. What may help in this regard are data validation rules imposed by the electronic systems. Furthermore, database management systems should be configured with more queries. This would help better attain the information extraction and integration process, as well as the construction of statistical figures. This would eventually promote the organisation's capability for analysing different indicators in a fine-grained way leading to a more appropriate decision making process. Although these problems are still common in the organisations studied, the senior management recognise these as existing challenges and aim to deal with them.

The first step into improving data quality is recognising its importance at different levels of the organisation. According to the interviewees, importance of data quality is recognised in their organisations. They believe that poor data can lead to many problems, from inefficient performance to poor services to citizens. The sources of poor data quality are many, and can range from human errors at the entry levels, to system failure, up to misinterpretation at the strategic level. Interviewees also associated poor data quality with increased costs. This was illustrated by Interviewee 3 who maintained that, “*poor quality data can increase our cost and make the service not cover its costs. Also, the data we send to the insurance providers, if quality is not assured they can simply reject our claims and not pay us.*” Another interviewee claimed that the integrated strategy for knowledge management in the organisation promotes data quality. Furthermore, the organisations understand the importance of data quality during the preparation of strategic plans, processes and methodologies such as the ISO standard, all of which require significant and reliable data. Some interviewees also stated that their organisations started data quality programmes in the past few years and are seeking continuous development to improve data quality from the technical and human sides. Interviewee 13 said that, “*The municipality is aware of the importance of data quality and understands the challenges faced, and therefore seeks cooperation with the Municipal System institutions through a common information gateway. Also, the municipality through the establishment of competent management and technical planning oversees the application of information system standards in Abu Dhabi Emirate, which aims at data governance in the organisation and ensures their quality and relevance when reaching the decision makers*”. Interviewee 1 maintained that, “*if you want to invest in the right way, you have to have accurate data and to conduct accurate analysis*.”

Given the above, recognition of the importance of data quality should be also expanded vertically in the organisation. For that, leadership is an important factor in the decision making process and is consequently an aspect of the assessment of data quality. The findings indicated that leaders are well-perceived by employees, particularly in Arab culture, as an exemplar to fellow staff and subordinates. The UAE’s power-distance index is very high (90%), which explains the importance of leadership in the Emirate (Hofstede, 2015).

Leadership is required at all levels and is not exclusive to a certain level. Furthermore, research has shown that leadership is the number one talent issue facing organisations around the world. However, there is a significant gap in developing leaders at all levels, according to a study by Deloitte (Trapp, 2014). According to Betts (2004), leadership involves not only people but also the initiation and continuation of processes and proper decision making procedures. Billy et al. (2012) maintain that leadership serves as a stronger factor in affecting process approach in decision making. They also emphasise that more attention should be given to leadership in decision making. Furthermore, leadership has more important influence in tribal, collectivist and high power-distance cultures, which are features characterising Arab culture of the UAE. Therefore, corporate leadership of data quality is an essential factor. Organisations need to develop leadership pipelines at every level and accordingly data will be sieved in intermediate stages up to reaching senior management. Thus, senior management will act as a final judge on the information received after passing the different-level quality controlled leaderships (Figure 26). The intermediate stages should define the roles and responsibilities of staff for data handling. Performance, risk management, reward and recognition are all important leadership-considered measures when advancing data to a higher stage. Leaders must maintain clear insights about the purpose of certain data in order to make better judgements. The definition should be in line with the business objectives and strategies of the organisation. Leaders are also encouraged to liaise with international partners in this respect to add an extra and independent dimension to data quality.

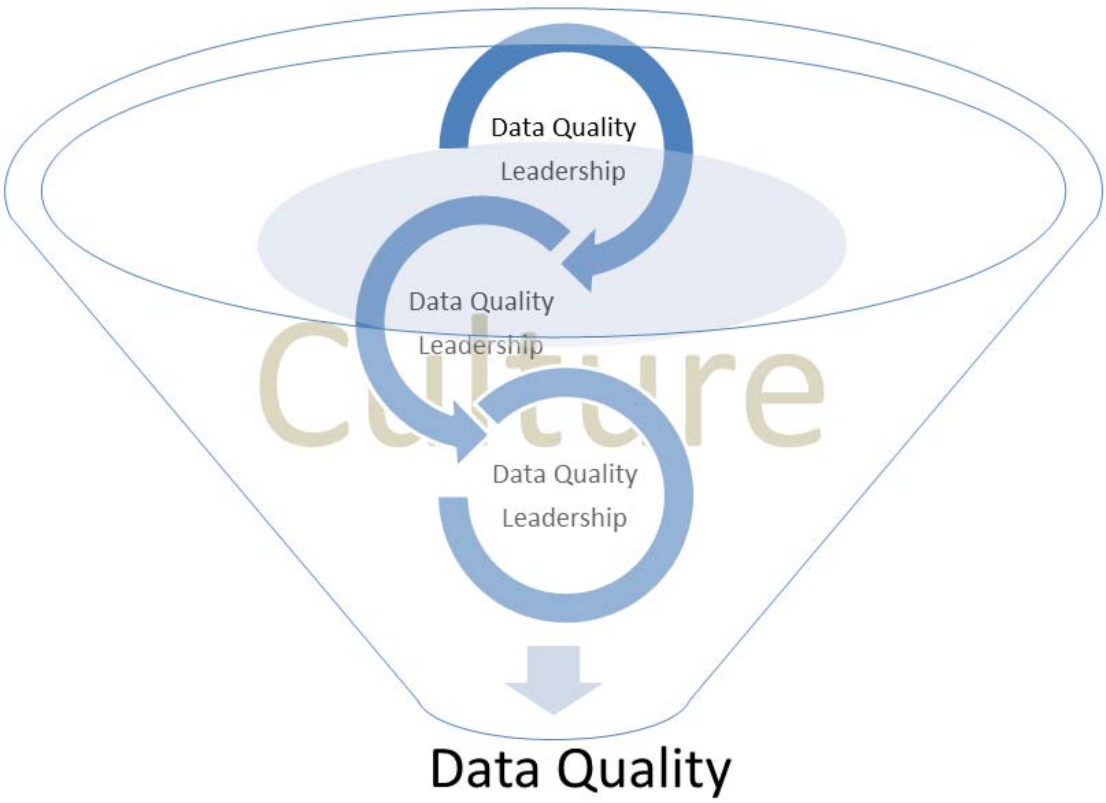

Figure 26: Leadership pipelines at every level within a given cultural context to maintain that data are sieved in intermediate stages up to reaching senior management

8.1.3. Data quality assessment

Organisations use several methods for assessing the quality of the data acquired, such as internal comparisons with the previous periods, cross-referencing of reports from various sources, periodic reviews of figures and analysis of figures. All these steps are used to make sure whether certain data is somewhat accurate. In the studied organisations, there is no specific way to evaluate the decisions taken, but decision makers rely on the feedback from affiliated organisations, which is not a regular or specific mechanism.

Interviewees maintained the difficulty of assessing the quality of the data obtained. They also referred to the presence of the data in different systems and formats, which requires the evaluation of these data in different ways by each department. For example, Interviewee 14 stated that, "*the statistical data is collected annually and reviewed by specialists and studied by municipal officials to match consistency with mainstream data of the organisation and with the current situation of the municipality policy*." The interviewees highlighted a rather important aspect of data quality. The interviewees maintained that data received from the central authorities are usually considered valid unless severe anomalies are spotted, if there are corresponding data internally, then an investigation request would be sent to the central authorities. The central authorities may allow organisations to give counter proposals to their data reports if the organisations think their data are wrong. If the organisations can provide evidence to their argument then the data from the central authorities would be set to change. However, this is quite an involved procedure and managers try to avoid doing it.

The common aspect of assessment and validation of data in the organisations is done by the departmental managers who also coordinate training of representatives from their department to double check the quality of the data. In these undertakings, they refer to procedures and policies provided by higher authorities. There is a certain autonomy given to local decision makers depending on the nature and influence of the decisions. These decision makers are advised to follow best practices and refer to international standards whenever possible.

There have also been several projects to review the statistical and various electronic systems in some of the organisations, aimed at auditing data stored in the databanks, as well as to test the robustness of storage media. This was noted by ADPO officials:

ADPO is not the only organisation in the Emirate that issues the annual road death rate statement countrywide; it is actually issued by other government and non-government bodies in the countries, such as the Federal Health Organisation. ADPO compares and contrasts these reports for any anomalies internally and with the issuing bodies. In this respect, Interviewee 2 maintained that, "*trends of periodically acquired data are studied for any abnormal leaps as such leaps are usually indicators of potential inaccuracies*."

It is worth noting however that there are no specific internal criteria of acceptance of data. For example, Interviewee 12 maintained that, "*for the data on organisational and operational performance, validation is done through random sampling, checking of the sources and providing evidence upon request. As for the data that is extracted from electronic systems, it is evaluated by specialists, analysts and strategic planners before being utilised.*"

Proposed enhancement:

The following is suggested to address the issue and is reflected the "Organisation" element of the framework as shown in Figure 30:

Organisational performance may act as an indicator for data quality as they (i.e. organisational performance and data quality) are logically in a positive relationship. Better performance significantly indicates higher quality of data used by decision makers. Improving data quality on the other hand will also yield improved organisational performance. Organisational performance can be identified by results of efforts undertaken to attain citizen-oriented service delivery. Accountability, transparency, responsiveness, efficiency, and effectiveness are common indicators of performance. These indicators can be compared and contrasted with previous values in order to evaluate improvement. At the same time, data quality should be evaluated based on previous values and compared to evaluation of performance. This can help understand the relationship and allow further tweaking of improvement parameters. Trend analysis is an important tool for performance evaluation. This will help strategic decision making in the organisation to estimate future trends and make such decisions in light of this analysis. Other criteria can also be used such as feedback from service-targeted citizens based on established evaluation methods.

8.1.4. Suppliers of data used in decision making

There are many types of data used for strategic decision making in organisations depending on the nature of the decision. For example, for human resource related decisions, the main source of information is the human resources department, and the data obtained are therefore of this nature. Financial decisions on the other hand are made based on extensive information received from relevant departments. Sometimes, and perhaps quite often, decisions require joint and cross-referenced information from several sources. For example, pure crime-fighting related decisions use information from crime-fighting authorities, prison authorities, traffic management, and so on.

Generally speaking, decisions are based on the strategic orientations of the government and its strategic vision. However, decisions rely on data provided from different sources. For example, data related to staffing plans, financial statements, and decisions related to capital expenditures provided by the organisations to decision makers require justification in terms of organisational performance and expected outcomes. These data are subject to studies (including analysis of economic indicators) and other references such as customer satisfaction questionnaires, strategic plans for the organisations, the annual and quarterly performance reports and annual economic policies. Financial statements make an important input in the decision making process as well as the impact of the decision on the public. In addition to relying strategic plans, views of the stakeholders both inside and outside the organisation are also considered. Interviewees 8, 9, 12 and 14 maintained that the data that they are required to provide to decision makers are financial statements, operational data and customer related feedback.

The organisations' information systems are also accessible by internal or external decision makers. Even if they do not have direct access, they can request the data and they will be provided to them. Moreover, decision makers have two types of data sources: internal to the organisation and external. Internal data are mainly performance indicators that are provided by the relevant departments, and external data are collected by various government agencies, such as government policies, international standards and environmental criteria, and are provided to decision makers whenever officially requested. Interviewee 5 noted that there has been an initiative emirate-wide to establish a knowledge base to be used by decision makers.

Proposed enhancement:

The following is suggested to address the issue and is reflected the "People and Systems" and "Culture" elements of the framework as shown in Figure 30:

Regardless of the source of the information, validity and quality are the main concerns of the studied organisations. One of the aspects of validation of information and understanding it is contextualising and drawing particular predictions of future trends by comparison and cross-referencing. Some parameters and figures obtained may be compared to those of other countries and relevant international reports, so that these organisations may evaluate certain aspects of their performance, although to a limited extent, based on international values. Strategic decision making in organisations is based on different types of information and relies extensively on trend analysis in order to estimate future trends and make such decisions accordingly. Decision makers attempt always to make strategic decisions by using a pool of information required to draw a clearer picture of the situation in hand. This is partly implemented in some of the studied organisations. For example, Interviewee 1 from ADPO pointed out that trend analysis is used for performance evaluation and hence information quality in ADPO. The same interviewee noted that the total number of deaths on the roads in a particular place would not reflect much unless compared with previous figures, and consequently, can provide estimates for the near future in this respect.

Therefore, it is crucial that data suppliers recognise the importance of data quality and be able to handle data accordingly. It is essential that an organisation should continuously develop its human resources. In particular, it is crucial to develop staff acquired with the adequate skills, competencies and knowledge necessary to fulfil their roles and responsibilities in relation to data quality. The skills should also include data capture, analysis and presentation. Training programmes must focus on data handling but also maintain a wider scope related to data quality, for example, cultural aspects regarding data handling and security aspects of data quality. This is particularly important in high-context cultures such as Arab culture (Alkaabi & Maple, 2013). According to Hofstede (2001), Arab culture is characterised by collectivism and high-power distance. Collectivism and high-power distance are two dimensions set by Hofstede in his attempt to establish quantifiable aspects of culture. Collectivism refers to a higher extent to which the group is prioritised over the self-interest as opposed to individualism. Hofstede asserts that cultures with high power distance, such as Arab culture, tend to be more collectivist.

This has implications into how training and raising awareness are implemented to lead to effective methods of system development. Accordingly, indicators of awareness, commitment and improvement of staff should be inputs in data quality. Furthermore, staff are to be encouraged to contribute to data quality by adding their inputs such as suggestions and ideas, and their innovations and contributions should be supported and possibly implemented. Training programmes should be conducted on a regular basis and be evaluated, whereby the organisation has to determine training and development needs and to response to change. The roles and responsibilities should hence be well-defined and incorporated in the job descriptions. This aspect also includes development of a human resource plan that aims to meet the objectives set by the organisation in relation to data quality. The plan must contain assessment criteria for staff regarding data quality and include measures for commitment to achieving its aims and objectives. Therefore, there should be methods established for staff selection and recruitment, and consequently for managing staff performance. There, hence, must be certain procedures defined to contain potential detrimental effects of restructuring.

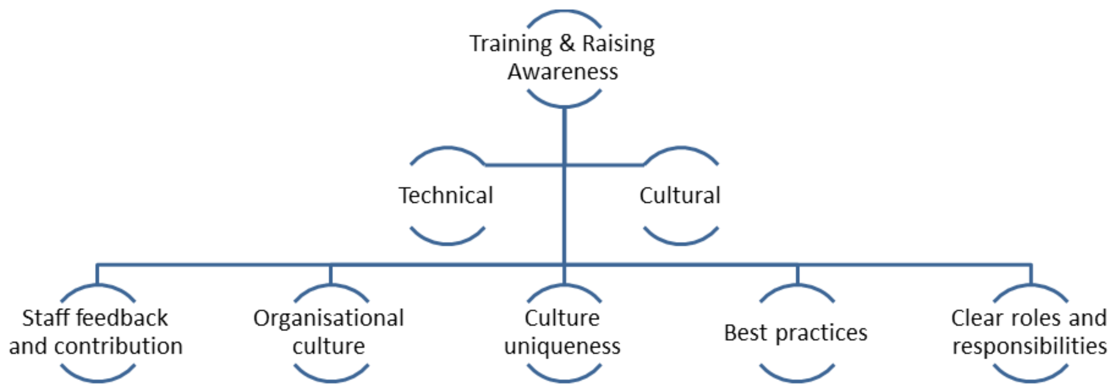

Figure 27: Training and cultural awareness of employees as an aspect of data quality attainment. Furthermore, staff feedback is a key input in improving data quality

Furthermore, data quality projects aimed at identifying data and system problems and defining certain processes that help mitigate the encountered problems usually consider data quality starting at the entry level and impose preventive and corrective measures (Eppler & Helfert, 2004). Accordingly, a data quality project should set certain mechanisms for data verification whereby these mechanisms must be tested against expected results to conclude the optimal mechanism to be used. The project is to enable the management to begin to know where their problems lie in order for it to deal with them. The project, however, must be unique to each organisation. Furthermore, the models obtained by individual projects may not be comparable with those of similar organisations worldwide. The particularity of the models to the organisations might thwart any efforts to compare and contrast them. Even some of the quantitative values obtained are incomparable due to other particularities considered when computing such values. This poses a real problem to the organisations as worldwide standards cannot be regarded with adequate significance to be used for local measures. Therefore, decision makers do not enjoy the use of global estimates that would perhaps give a clearer vision but have to stick mostly to local measures. Therefore, the outcomes of the data quality projects need to overcome the challenges of accuracy and the global perspective of quality.

8.1.5. Measures of effective decisions based on quality data

As was noted in the answers, interviewees considered the poor quality or unavailability of information as an important and major influential factor in making all decisions, including strategic ones. The interviewees could establish a direct link between data quality and decision quality/efficiency. Some of the interviewees referred data quality issues as mainly related to lack of validation and cross-referencing. For example, Interviewee 3 maintained that "…*some of the problems are due to data quality, but not because it's insufficient, but the problem is the validation. A lot of data are collected at the organisation's level but some are also at the enterprise level, which the organisation is part of. The process of collecting data from two different sources creates inconsistency*."

This issue of receiving data from the headquarters is rather common in the studied organisations. Interviewees supported the claim that a lot of figures reported by the headquarters do not align with internal reports of the organisations. To illustrate the issue, Interviewee 4 from Al Ain Hospital provided the following example: *"'patient days' is a data field that we calculate based on the number of admission days, length of stay, and other inputs. The HQ calculates the same field as well. We both use the same calculation methodology that the HQ provided, however, most of the time the* *figure for the same time frame that we calculate comes out different than the one provided by the HQ. It takes about one to two weeks every quarter just to try and liaise with the HQ what the reason could be behind the difference. Sometimes we find it's the cut off dates on time frames and sometimes other reasons and sometimes we just cannot pinpoint the reason because everything seems in alignment. At the end of the day, if we cannot align to the HQ figure we have to accept it as it is. Consequently, even though it is one figure that has the problem, many areas are affected. The 'patient days' figure is used for the calculation of many financial and non-financial Key Performance Indicators (KPIs), we may fail to meet some of our KPIs due to such issues when we know that in reality we compromised the figures*."

As could be seen in the above example, quality issues could be inevitable and decision makers have to consider figures for which they suspect their validity as invalid. The same point is stressed by Interviewee 14 who asserted that decision makers do not have the resources to ensure of the quality of information they receive, so the challenge is essentially the difficult access to quality information by decision makers. In this respect, Interviewee 5 noted that remedies are based on the experience of the decision maker. Interviewee 8 considered, however, that the movement of low-quality information between branches greatly affects the outcomes of strategic decisions, and maintained that staff training based on citizen/customer satisfaction and the overall welfare of the organisation will eventually improve strategic decision-making.

There are no metrics available or devisable to accurately measure how many of the decisions made (over a certain period of time, for example) are built on quality data. This is based on the difficulty of the calculation process of such measures. The decision makers of the organisations studied take on different responsibilities, and decisions are made on a daily basis. Furthermore, decisions are taken continuously in sprawling branches over the Emirate. The decentralisation aspect of the organisations, for example ADPO, allows decisions to be made by other subsidiary organisations with different specialties and services countrywide.

Personal opinion may influence the decision making process in any organisation. In fact, the Delphi method relies on expert opinions in critical decision making. However, decisions generally require solid data in order to constitute educated opinion on certain matters.

For the case of the studied organisations, decisions frequently rely on VIP opinion or views of certain influential people. Consequently, not all decision makers rely on data and figures to the same degree. Even when provided with precise information, the individual decision maker may have their own interpretation and evaluation of this information. Therefore, the personal side has a significant influence on interpretation of data and hence decision-making. This aspect of decision making may render the process extremely complicated. Accordingly, the provision of accurate or quality information may not be sufficient to generate robust decisions. The matter of personal judgment regardless of presence of information is a real phenomenon in Arab culture and similar cultures characterised by collectivism and high power-distance (Hofstede, 2001). The common belief in the organisations studied is that leaders retain their abilities, experience and education to the best of making the right decision. This aligns with Hofstede's (2001) cultural dimensions, whereby Arab culture is characterised by collectivism and high power-distance. Collectivism and high power-distance are two dimensions set by Hofstede in an attempt to establish quantifiable aspects of culture. Collectivism refers to a higher extent to which the group is prioritised over self-interest, as opposed to individualism. Hofstede asserts that cultures with high power-distance, such as Arab culture, tend to be more collectivist.

Proposed enhancement:

The following is suggested to address the issue and is reflected the "Service Quality" element of the framework as shown in Figure 30:

Decisions differ in their importance and requirements for data, and vary in their nature, for example, some decisions are related to traffic, others to security, others to public health, others are mainly financial, and so on. Decisions based on quality data and highly accurate figures may not be mainstream, however, no statistics have been gathered in relation to the estimation of such decisions. In this regard, Interviewee 2 of ADPO asserted that, "*knowing what decisions have been made based on quality data is a complex process that would consider a huge number of variables, given no tools are available to our organisation at this stage.*"

Decisions are taken by the headquarters or central authorities and their effects are circulated to the relevant organisations. Decisions are either taken based on the data provided as above, or they are subsequent implementation of the authorities' decisions. Interviewee 3 provided the following example of decisions' effects on their organisation: "*The hospital's outpatient clinics are currently opened at the evening. There was a data review to make a decision on whether or not it is worth to continue having them open. The statistics show that some of the clinics are not as active as they should be; therefore, the facility is now in the process of finalising a decision to close them. On the other hand, there was a mandate from the headquarters to see only certain insurance holder outpatients in certain hours within the day, this project had to be implemented as mandated even though our current stats show that we will suffer a drop on volume due to this decision but there was nothing much to do about it. This decision was not made based on our data, but rather on the strategic vision of the authorities. After implementation, we are now monitoring the data to make decisions on resources allocation and service quality to anticipate any negative outcomes of the decision*."

However, it is not known how often decisions are made based on quality data due to the complexity of the decisions. For example, one interviewee mentioned that about 80% of decisions are based on quality data. When asked to elaborate more about how that number was calculated, the interviewee claimed that it was a personal estimate based on the criteria cited in the "Quality of data used in decision making" section above. It is common that these organisations use the quality of the outcomes of a certain decision as criteria for such assessment.

A key aspect of examining data quality based decisions is linking service quality to data quality thereby transcending the importance of data quality to the proposed level of decision making. Generally speaking, it is important for a public organisation to be oriented towards the segment of citizens it serves, as well as to strive to improve its services to them. A first step for improving the service is to define the service stakeholders. Data quality will hence act as part of service quality and each can be a positive indicator of the other. Future and growing needs of stakeholders can anticipate, and be anticipated by, the quality of data in hand, and hence consensus in the organisation on the importance of attaining stakeholder satisfaction and therefore data quality standards. Accordingly, the organisation should have defined methods of acquiring and using stakeholder inputs as a means of attaining quality. Public feedback should be captured, analysed, and used as an input to quality assurance in the organisation. There should also be tools for measuring stakeholder satisfaction with the service given and comparing it with previous records. Ultimately, the organisation should aim at continuously improving this satisfaction as well as its approach to citizen orientation. Figure 28 depicts the relationship between service quality and data quality.

Figure 28: Service quality linked to data quality and hence the importance of data quality transcends to the decision making level

8.1.6. Policies and procedures for data handling

Organisations strive to achieve standards for data quality and to assure that information received by decision makers is accurate and precise. For that, members of staff with direct involvement with relevant data must be emphatically informed about the value of the information they deal with. Accordingly, every employee must have a certain role fully described which he or she is held responsible for. Training courses should be conducted regularly either inside or outside the organisation, aiming at updating and improving skills in information security and the value of information in different units of the organisation. Leaders and managers must always emphasise roles and responsibilities to staff. Managers and leaders can also contribute to the accuracy of information by using methods of strategic management.

It must be emphasised that the organisations studied retain large amounts of data and some of them are of high importance and sensitivity. These data are acquired continuously, and are related to different affairs such as traffic, crime, human resources, patients, citizens, and infrastructure, among many others. Data entries are a constant operation in almost all of the organisations' systems. Some of the systems are in service around the clock seven days a week to receive instant and live data, while others are fed with data at later times. Being service-oriented organisations, they have some of the systems alert 24/7 to retrieve and process information. For example, the information security department of ADPO stores and deals with information according to a specific methodology for storing information permanently and continuously.

No explicit policies adopted or devised by the organisations for data handling have been noted. By data handling, it is meant the policies of handling of data in different forms, which includes security, storage, processing, archiving, and destruction, among other aspects, of data. The officials' opinions about data handling policies vary in the organisations: some see them as bureaucratic overheads whereas others consider them as necessities missing in the organisations. The local data handling policies of the organisations are mainly concerned with data security. For example, there is a data integrity and confidentiality policy present in the studied organisations. However, Interviewee 3 noted that the way reporting and accessibility to reports is done, there are no policies currently involved in data handling. Interviewees noted that the knowledge management strategy attends to some aspects of data handling with some issues with the implementation of the strategy. The interviewees raised some issues in this regard, such as the need for internal development of the information systems and electronic archiving, and improved skills in dealing with data and documents. They also noted that electronic forms of data should prevail over the paper form, which will improve accessibility and quality of data. Interviewee 7 noted that there is a corporate performance and governance policy that deals with institutional performance data handling. External policies from the governing bodies or high authorities are usually referred to locally in the organisations, such as the information technology policies of the Abu Dhabi Centre for Systems and Electronic Information, policies of the executive board regarding human resources data, and policies of the department of finance on the financial statements, financial planning and financial reporting.

Proposed enhancement:

The following is suggested to address the issue and is reflected the "Policies and Procedures" element of the framework as shown in Figure 30:

The roles and responsibilities in information and information management may not be significantly recognised in the studied organisations. It is observed that department managers are responsible for validating their departments' data although sufficient information about the data might not be available. The management of each of the organisations restricts the vital role of data quality to the first stage data entry clerks and their direct managers, with the help of the information security department.

Policies and procedures set must embed the overall direction and culture of the organisation in relation to data quality. Accordingly, the organisation should set policies and procedures for data handling that clearly describe how data are dealt with in different units of the organisation. The policies must be accessible by all staff and able to provide guidance on data collection, storage, analysis and reporting for the aim of attaining data quality. Policies should be formally expressed by senior management, whereby relevant country laws and best practices can be incorporated if necessary. The policies should be reviewed and updated regularly and promptly in response to change. The policies and procedures may also be supported whenever possible by the organisation's information systems, such as user restricted access based on credentials, demilitarised zones, and virtualisation.

8.1.7. Reporting

Canonical reporting is common in information communicated within the studied organisations. It is typically reported the usual ways: weekly meetings in departments, monthly meetings in districts, and quarterly meetings at the level of general administrations. It is observed that other meetings are held whenever necessary such as for conducting a specific study or dealing with emerging issues.

There are certain reports run periodically and provided to certain key individuals and/or groups in the organisations, for example, the KPI reports that run on a monthly or quarterly basis. There are other reports created based on request and provided to the requester. An example is reports related to specific studies, research or projects requested by the authorities. Emails circulated among staff are also a typical method of reporting in the organisation. Some of these emails contain effective decisions and even submissions of complaints and suggestions. The interviewees noted that the emirate-wide knowledge base which is under development will help improve the reporting process inside and outside the organisations. Occasional and regular meetings are another way used by the organisations as a reporting method, whereby vertical and horizontal communication takes place. It is common also that various reports are collected from different departments and sent to those concerned by e-mail or in print. The authorities have access to electronic systems, where they can collect and audit data for their reports, and then post them on several levels, inside and outside the organisation.

Proposed enhancement:

The following is suggested to address the issue and is reflected the "Reporting" element of the framework as shown in Figure 30:

Reporting is of paramount importance in data quality. Different units of the organisation must be able to prepare robust reports aimed at internal and external communication. The data supporting reported information will play a significant role in decision making and can thus be considered by decision makers for refinement of their decisions. Financial and non-financial reporting should also consider data validation and control, which are prerequisites for data quality. Reports may include importance, accuracy levels and timeliness. There also should be a formal, well-defined process for report preparation for more accuracy and consistency. The report should highlight how much data quality expectations and requirements were met based on assessment criteria provided and when possible indicate how reliable information can be attained based on such data. In this respect, the organisation can benefit from the International Auditing and Assurance Standards Board's work on reporting, such as the *International Standard on Assurance Engagements (ISAE)*. Furthermore, organisations are encouraged to follow an integrated reporting approach as suggested by the Committee of Sponsoring Organizations of the Treadway Commission (COSO) (Accountant Magazine, 2013) (Figure 29). Clarity and conciseness are generally rules of thumb, but the main importance lies in the accuracy of information extracted.

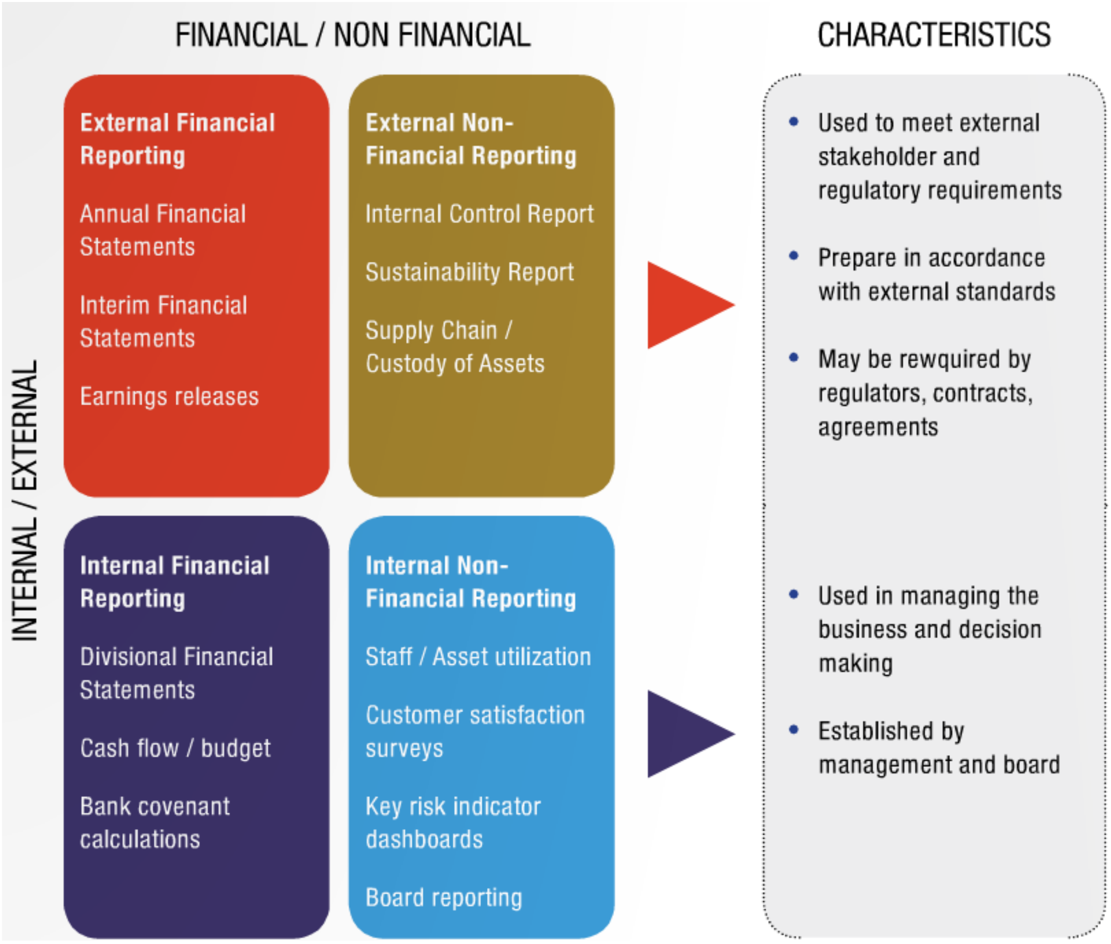

Figure 29: The integrated reporting framework suggested by COSO (Accountant Magazine, 2013)

8.2. Suggestion of a Data Quality Framework for Abu Dhabi Public Organisations

Based on the findings and discussion, a data quality framework for the public sector is suggested based on the data problems encountered in the target public organisations in Abu Dhabi and substantiated with extensive research into data quality requirements.

Data quality issues in a public organisation may be exhibited, among others, in unsatisfactory results, declined revenues or/and squandering of public funds. The direct causes of such problems are inefficient strategic decisions made in the organisation and incompetency of the decision makers. The problems have direct and indirect effects on citizens and the economy of the country given decisions made by public organisations are related to different aspects of economic, social and political issues in any country. The underlying roots of the problem are often related to the quality of data reaching decision makers. Poor data quality results in inefficient decisions and consequently affects the performance of the entire organisation. Enhancing the quality of data will ultimately improve decisions and hence contribute to solving these problems. The data quality framework should introduce the necessary elements needed to achieve data quality enhancement by identifying and resolving the root causes.

Based on the discussion of the findings as details above, seven aspects of data quality were identified on which data received by decision makers can be rated in terms of their quality. These aspects are: developing all-level leadership within the cultural context, mapping data quality to service quality, initiating and enhancing staff training and development programmes (cultural and technical), establishing robust internal reporting, enacting effective policies and procedures for data handling, adopting information management and data refinement techniques, and continuously measuring and evaluating organisational performance. The framework is depicted in Figure 30.

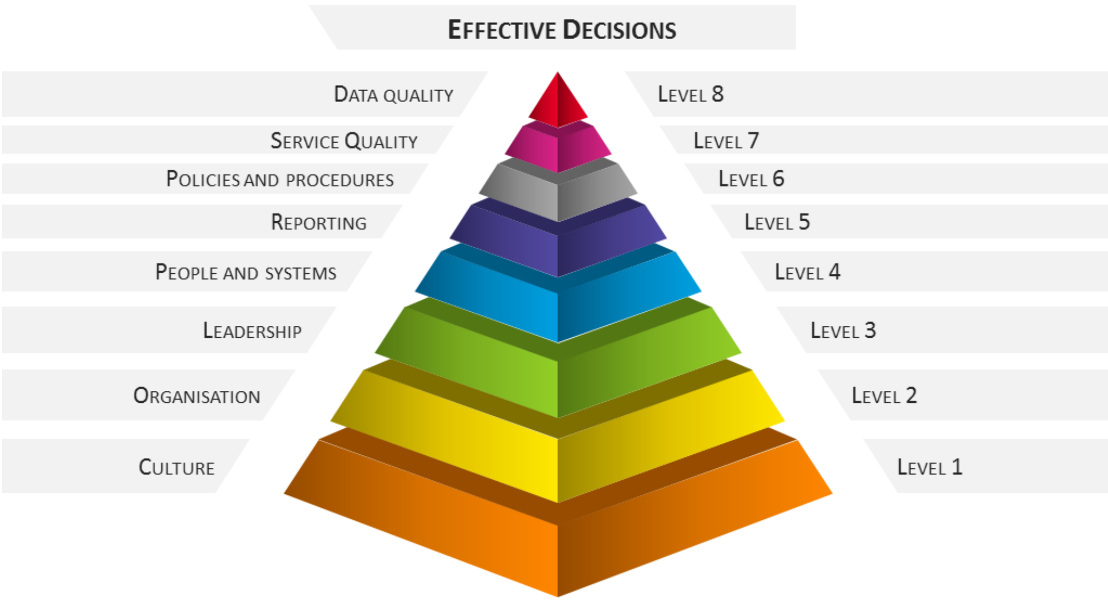

Figure 30: The proposed data quality framework for Abu Dhabi Emirate (ADDQF). The framework comprises 8 levels up to effective decision making

The results obtained emphasise the integrity of the framework and the intertwining nature of its processes, with further attention to data refinement methods to data quality such as data mining. These factors have been overlooked by other frameworks aimed at addressing the same aspects. This is further discussed in the next section where the framework is compared to other similar frameworks.

8.3. Framework Comparison

In this section, the developed framework is compared to existing data quality framework in order to show the aspects of similarity and difference with other well-established frameworks.

8.3.1. Comparison with the Zachman-based information quality framework

The first framework considered is the Framework for Information Quality (FIQ) based on the Zachman Framework for Enterprise Architecture (McGilvray, 2010) shown in Figure 31.

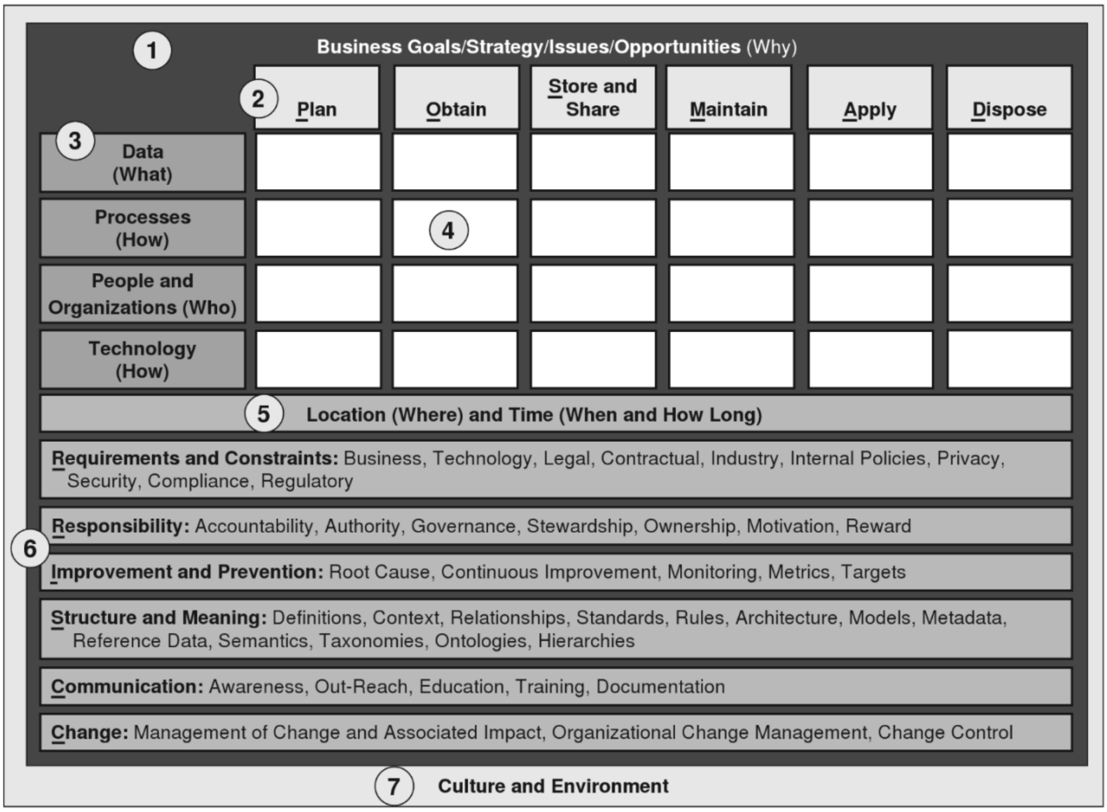

Figure 31: The Framework for Information Quality is a conceptual framework for data quality diagnosis, planning and design. The framework is suggested by McGilvray (2010) independently of the Zachman Framework for Enterprise Architecture although both frameworks comprise the same interrogatives

In order to compare FIQ with the proposed ADDQF, FIQ's main elements are first discussed:

1. Business Goals/Strategy: FIQ maintains the organisation's strategy should be the main drive of actions and decisions.
2. The information lifecycle: which comprises the stages information undergoes in its management process.
3. Key components affecting information lifecycle: these are: the data being processed, the processes being undertaken, the organisation of people involved in the processes, and the technology required for information processing.
4. The interaction between Element 2 and Element 3: The interaction between the information lifecycle and the key components.
5. Location and Time: Indicating where the information is at a given point of time and when it is available.
6. Broad-impact components: these include other factors that affect information quality, namely: information requirements and constraints, responsibility, improvement and prevention, structure and meaning, communication, and change.
7. Culture and environment: the organisational culture of the company and the environment it operates in.

The above elements of FIQ can be mapped onto ADDQF as follows:

- Element 1 of FIQ is mapped to leadership in ADDQF as leaders must have a clear definition of the purpose of certain data in order to make better judgements. The definition should be in line with the business objectives and strategies of the organisation.
- Element 2 of FIQ is mapped to People & Systems in ADDQF as information lifecycle as defined in FIQ is mainly concerned with people and systems.
- Element 3 of FIQ is mapped to People & Systems and Policies & Procedures in ADDQF as data, processes, people and technology are defined in those two levels of ADDQF.
- Element 4 in FIQ is mapped to Reporting in ADDQF as the interaction between the information lifecycle and the key components occurs via reporting in ADDQF.
- Element 5 in FIQ is mapped to Organisation in ADDQF, which defines data location and time.

- Element 6 in FIQ is mapped to Service Quality in ADDQF as those broad-impact components are defined under the comprehension of linking Service Quality to Data Quality in ADDQF.
- Element 7 in FIQ is clearly mapped to Culture in ADDQF.

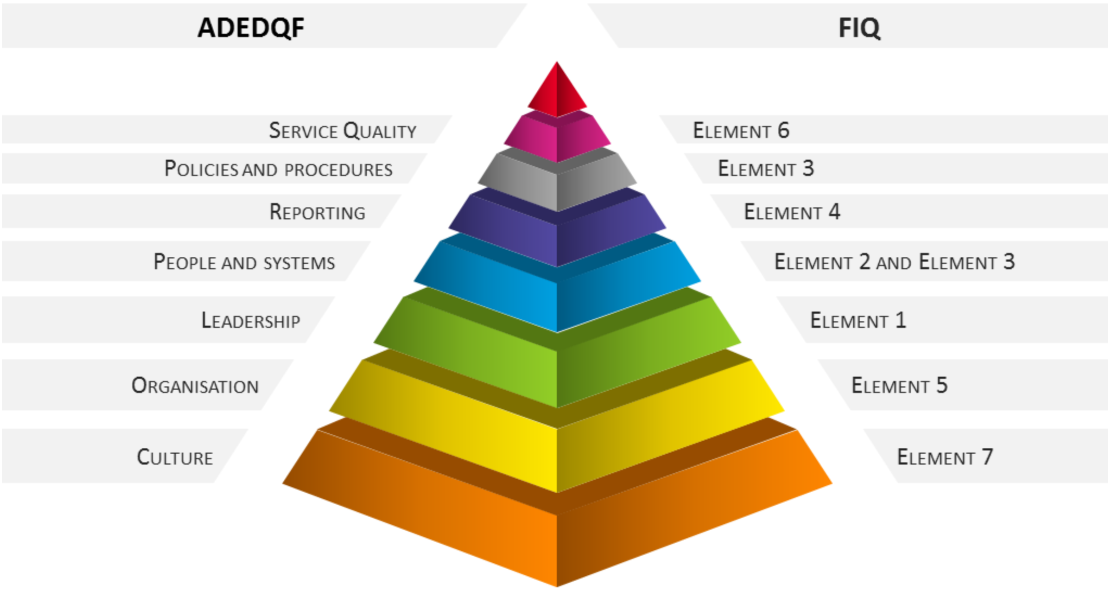

Figure 32: Mapping FIQ elements to ADDQF elements

8.3.2. Comparison with the UK's public sector data quality framework

The Audit Commission of the UK was "*an independent body responsible for ensuring that public money is spent economically, efficiently and effectively, to achieve high-quality local services for the public*" (Audit Commission, 2007). The body's remit extended to over 11,000 public organisations in England, with a total expenditure of over £180 billion of public money each year. The Commission's work covered local government, health, community safety, fire and rescue services and housing. The Commission closed on 31 March 2015 and its work is now carried out by several bodies as part of the then government's cost-cutting measures.

The Audit Commission of the UK developed a data quality framework to be used by the UK's public sector. The framework covers management arrangements for public organisations in the UK to put in place to ensure the quality of the data they use to manage and report on their activities. The framework distils the principles and practices identified in existing guidance, advice and good practice. The framework comprises the following standards:

1. Governance and leadership: Accountability of data quality with a commitment to secure a culture of data quality throughout the organisation.
2. Policies: Appropriate policies and procedures to secure the quality of data.
3. Systems and processes: Systems and processes to secure the quality of data as part of the normal business activity of the organisation.
4. People and skills: Ensuring that staff have the appropriate knowledge, competencies and capacity for their roles in relation to data quality.
5. Data use and reporting: Ensuring that data supporting reported information is actively used in the decision-making process, and is subject to a system of internal control and validation.

The above dimensions of the UK's public sector data quality framework are closely related to the elements of ADDQF.

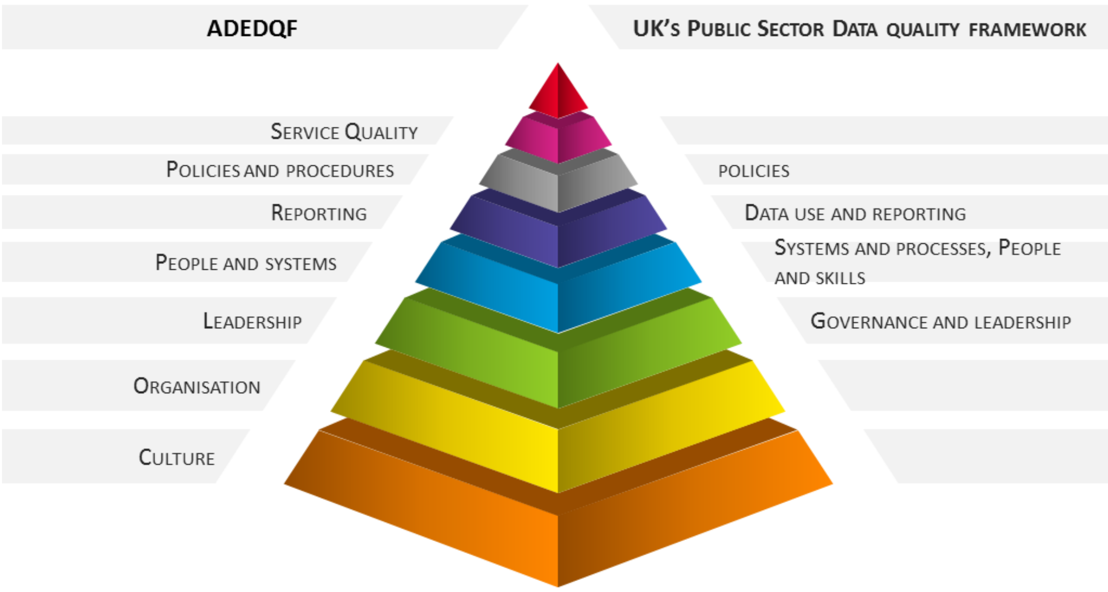

Figure 33: Mapping of the UK's Public Sector Data Quality Framework to ADDQF

As can be seen in Figure 33, all dimensions of the UK's public sector data quality framework are elements of ADDQF.

8.4. Need, Relevance and Uniqueness of ADDQF

It must be noted again that implementation of the framework must consider all its elements, and that it can be adopted by any public organisation in Abu Dhabi regardless of the data quality state it has. The framework aims to improve data quality in public organisations and will certainly provide improvement. This is based on the fact that the framework provides a formal approach to enhancing data quality and also benefits from other frameworks that are used for the same aim. Data quality is defined as "*the totality of features and characteristics of a dataset that bear on its ability to satisfy the needs that result from the intended use of the data*" (Arts et al., 2002). Redman (2001 cited by Angelsen et al., 2011 p.192) maintains that high quality data are those which fit their intended uses in operations, decision making, and planning. Furthermore, such data are supposedly free from defects and they acquire desired characteristics. Batini et al. (2009) studied common aspects of data quality. They considered and compared thirteen data quality methodologies and identified three directions that data quality is currently taking, namely: (1) considering more data types and moving from data quality to information quality, (2) approaching data quality issues in accordance with business process issues, and (3) focusing on new types of information systems, such as P2P and Web information systems. Batini et al. (2009) further maintain that there are open issues in data quality measures that require addressing, such as the identification of more robust statistical, probabilistic, and functional correlations between data quality and process quality.

Digitisation and storing of data in electronic format has made data more pervasive than ever. It has however raised many questions on its quality and hence efficiency of storing large amounts of data over a long time. Data quality plays a significant role in many if not all business and governmental applications (Batini et al., 2009). According to the Audit Commission and the National Health Service in England (2007), the risk of not identifying and addressing weaknesses in data quality is that information may be misleading, decision making may be flawed, policies may be ill-founded, resources may be wasted and poor services may not be improved. There is also a risk that good performance may not be recognised and rewarded.

Several initiatives have been launched worldwide by both the public and private sectors to promote the importance of data quality and to provide further guidance on its attainment, for example, the Data Quality Act passed by the U.S. government in 2002 (Batini et al., 2009), the Data Quality Initiative Framework by the Government of Wales in 2004 to improve the information quality for general medical practices

(HIQA, 2011), and a framework to support improvement in data quality in the public sector in the UK by the Audit Commission in 2007, among many others.

Public bodies and governmental institutions are confronted with real challenges of choosing the correct data from huge amounts of data produced by the operations of these organisations. Moreover, the rapid development of data digitisation produced data stored in organisations' data warehouses, which required efficient exploitation and knowledge extraction. Padhy et al. (2012) argue that the value of strategic information systems is easily recognised; however, efficiency and speed are not the only factors of competitiveness. Padhy et al. maintain that the need for quality data is crucial given several reasons, including cost and the likely sensitivity of decisions taken by governmental bodies. Typically, these data are used for producing statistical analyses and forecasts on economic, social, health and education issues. These facets are highly related to government planning in areas like economic growth, development of interest rates and inflation, household income, education standards, crime trends and climate change, among others. Furthermore, the continuous increase in complexity in decision making requires executives to consider a vast number of inputs and a considerable amount of knowledge (Liu et al., 2010), rendering the necessity of acquiring tools for assessing the quality of the data at hand inevitable. Decision makers in the public sectors vary in their needs for the quality of data. At high levels, leaders often need to take complex decisions on priorities and resource allocation. Lawmakers and regulators must acquire correct information to help them make judgements on governance and performance. Even the general public and the service end-users need accessible information for their informed decisions, which is the responsibility of the competent body to provide them with such information.

Therefore, the considerations of the difficulties of attaining data quality measures as described above is somewhat lessened by the importance of assessing the quality of the data they use for making decisions for public organisations. International relevant standards, such as the ISO 9000 suite, can help benchmark data quality performance but cannot replace customised and tweaked frameworks of data quality based on particular settings of the public sector in general and a certain country's public sector in particular. Although some few successful cases of adoption of international data quality standards are recorded in the literature (for example in China (To et al., 2011), India (Srivastav, 2010), Australia (Singh & Mansour-Nahra, 2006) and Finland

(Moreland & Clark, 1998)), governments worldwide are still reluctant to do so (Singh & Mansour-Nahra, 2006). Moreover, the application of the ISO system to the public sector may even be counter-effective. For example, Abdullah et al. (2013) identified five common barriers in the literature to implementing the ISO 9000 standard in local government organisations, as well as two additional barriers identified in the researcher's study findings. These barriers relate to aspects of an individual public organisation, its resources, behaviour, culture, administrative practices and changes within the ISO 9000 version itself.

Governments worldwide have attempted to approach data quality issues and develop their own guidelines or frameworks for data quality assessment. Regardless of the difficulties of attaining such frameworks, the examples of government data quality guidelines are many in practice. The Health Information and Quality Authority (HIQA, 2011) provide information on data and information quality frameworks for the health sector in England, Wales, Canada and New Zealand. Moreover, the Data Quality Act passed by the U.S. government in 2002 aims to ensure and maximise the quality of information, the Data Quality Initiative Framework by the Government of Wales in 2004 to improve the information quality for general medical practices (HIQA, 2011), and a framework to support improvement in data quality in the public sector in the UK by the Audit Commission in 2007. Therefore, it can be seen that the need for data quality assessment criteria has already been addressed for the public sector in several countries either by law or by initiatives of competent bodies. Hence, the question of choosing the appropriate data quality for certain public organisations, which can be generalised to other organisations, is a task governments usually undertake. As the above examples demonstrate, governments realise the importance of data quality and have hence sought to develop standards for assessing the quality of the data they acquire.

Following this trend, and given the considerations mentioned above, a framework for data quality for the public sector in Abu Dhabi Emirate will bring about similar benefits as its counterparts. More importantly however, the unique factors of the framework to Abu Dhabi that were devised from empirical data where the framework would be implemented will ensure it will work as intended.

The undertaken comparison established a mapping between ADDQF and two other data quality frameworks, namely the UK's public sector data quality framework and the Framework for Information Quality (FIQ) based on the Zachman Framework for Enterprise Architecture. The elements of FIQ and the UK's public sector data quality framework were all matched by elements of ADDQF. However, ADDQF has unique features that the other two frameworks do not have as highlighted above. In particular, the integrity and interaction among the elements were emphasised in the discussion. The elements of ADDQF have mutual and transitive relationships and therefore require that the framework be considered as a whole not as individual elements. The elements could be thought of as levels, each of which should be attained in order to obtain data quality and eventually effective decisions. For example, embracing cultural values is the basis of the other levels. Relationships between levels are transitive and hence extend forwards and backwards across levels. This means that, for example, the relationship between decision making and data quality extends to leadership and so on. Other frameworks suggest using the framework elements without considering the layered approach or viewing the framework as an integral system but rather suggest that elements could be adopted individually. The other frameworks do not emphasise the concept of layers or levels as required by ADDQF. Furthermore, no transitive relationships among elements are regarded in the other frameworks. These relationships were found essential in this study in order to improve data quality in the adopting organisation. Another aspect unique to the ADDQF is adoption of knowledge extraction and data mining techniques. The use of data refinement techniques, such as data mining and integration of data mining and decision support systems is another unique feature of ADDQF. New and emerging techniques of data quality enhancement and knowledge extraction may also be of significant importance for the organisation as shown in the experiments conducted, and should be considered as key aspects of improving data quality. In this sense, data mining techniques and their integration with decision support systems as detailed in previous chapters 4, 5 and 6 will help improve data quality and provide decision makers with valuable information.

The importance of those distinguishing factors of ADDQF as the findings revealed as well as shown in the discussion is that attaining data quality requires a chain of processes and elements that interact with one another. Existence of those elements is necessary but not sufficient as detailed in the discussion since, for example, leadership should extend to all levels, reporting must follow a vertical and horizontal approach in the organisation, and measuring organisational performance requires inputs from all levels, and so on.

8.5. Implementation, Recommendations and Expert Reviews

Implementation of the framework (ADDQF) can be suggested based on the findings and discussion. Implementation procedures of each of the elements of ADDQF were provided in the framework element sections under the element description. Further implementation recommendations are provided hereby.

The findings coupled with the discussion in particular the comparison of existing frameworks with ADDQF, provide strong evidence of the ability to adopt and use the framework by Abu Dhabi’s public organisations. ADDQF is suggested as an outcome of the conducted research on the data quality issue encountered in the target public organisations in Abu Dhabi and substantiated by extensive research into data quality requirements. ADDQF should provide the necessary elements needed for achieving data quality enhancement up to attaining effective decisions by identifying and resolving the root causes. As the discussion above maintains, adoption of the framework must consider it as an integral system with connected elements, and therefore implementation should be undertaken as an ongoing project in the adopting organisation. Each element has a mutual positive relationship with the lower level’s element. Data quality has a mutual positive relationship with the highest level, decision making. As can be noticed, embracing cultural values is the basis of the other levels. Relationships between levels are transitive and hence extend forwards and backwards across levels. This means that, for example, the relationship between decision making and data quality extends to leadership.

The implementation project by the adopting organisation would start by understanding and embracing cultural values of the organisation. This is a key aspect in shaping organisational practice and performance in public sector organisations (O'Donnell & Boyle, 2008), which ultimately helps undertake change in the organisation. Based on findings from the interviews and the literature reviewed in this work, developing organisational culture that admires data quality requires addressing key issues represented as the elements of the framework. This directly leads to the next level of the framework which is leadership. Leadership is obviously important for determining the effectiveness of culture change. Leaders are key personnel who are responsible for understanding and managing culture in their organisation. Leaders need to adopt a methodological basis to support the decision making process. Leadership acts as the centre of all the elements required for improving data quality, which provides direction and control. For that, leaders must resort to training programmes, technical and interpersonal, that also consider the cultural values that the organisation preaches. This includes more training courses and supportive bodies of the organisational units, such as decision support centres, information security and strategic management. These all act as organisational units to provide decision makers with quality data. It is also important to acknowledge human and cultural factors involved in the decision making process. Such factors have implications into how training and raising awareness are implemented to lead to effective methods of system development. The programmes can be reinforced by policies and procedures that align with the organisation's cultural aspects. Reporting should therefore be effective in order to act as evaluation of improvement. Reporting requires that information systems should be robust and centralised, with data consistency being verified and enhanced by using advanced data refinement techniques. Since the framework targets public organisations, service quality can be used as a tracking mechanism for organisational performance. These all contribute to improving data quality, which ultimately serves attaining effective decisions. Each element should be revisited at various intervals to ensure continuous improvement of data quality in the organisation. The mutual and transitive relationships among elements described earlier mean that attaining each element will have a positive impact on improving the other elements.

The implementation process is depicted in Figure 34.

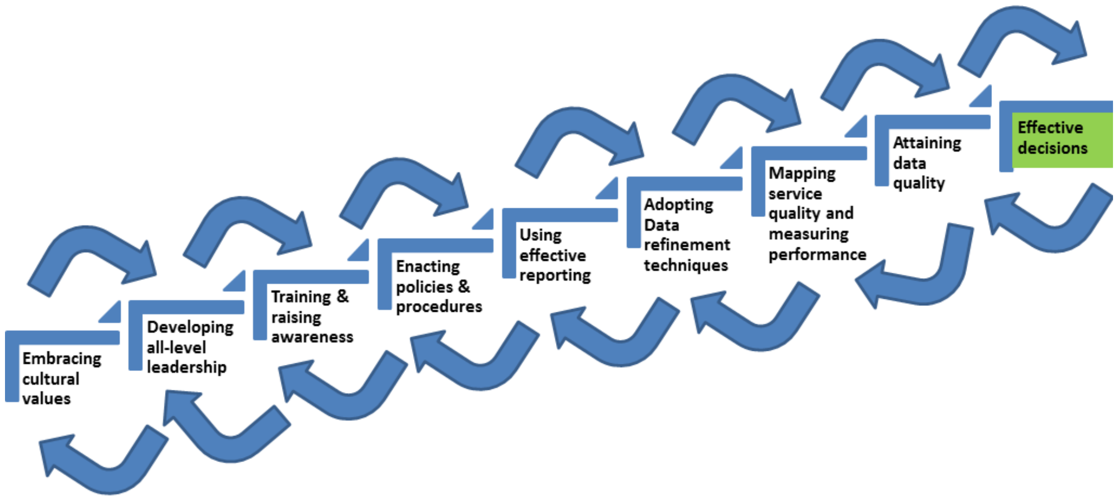

Figure 34: ADDQF implementation processes with transitive positive relationships existing between them

The framework and the implementation aspects were also reviewed by the 14 executives interviewed earlier in the 7 organisations in Abu Dhabi Emirate. Initial feedback received included comments such as "*the framework may provide a systematic approach to identifying and rectifying quality issues at ADPO*" by Interviewee 3 from ADPO and "*the framework is well-thought-of in terms of applicable elements and can be tested and evaluated with little to no change*" by Interviewee 4. As seen in the comments below, the executives also pointed out the normative nature of the framework in the sense of its possible implementation in their organisations:

> "*The framework gives importance to the continuous development of the organisation*" (Interviewee 6).
>
> "*It provides strategies for redefining and analysing data*" (Interviewee 12).
>
> "*The aspect on training with culture in mind is a very important point to notice*" (Interviewee 9).
>
> "*It establishes a basis for improving data quality and contains executable elements*" (Interviewee 5).

The framework simply reflects some of the known problems and identifies the areas which result in current data quality problems, it unfolds the underlying roots of the problems which may be overlooked or are not easily spotted by ordinary auditing. Furthermore, the executives maintained that the framework is normative in the sense of the possibility of implementation in their organisations. The elements that the framework embeds provide a basis for refining several relevant factors that add up to improving the quality of data reaching decision makers.

The theoretical basis of the framework, supported by the mapping onto other existing, well-established frameworks, in addition to the distinctive features of the framework detailed above, all provide strong evidence for the ability of the framework to improve data quality. Moreover, given the multitude of elements encapsulated in the framework, for example, policies, systems, humans, culture and leadership, it is highly likely that the framework can form a basis for changed organisational practices of Abu Dhabi public organisations towards improved data quality. The argument of forming a basis for organisational change is reinforced by the integration and interaction of elements, and culture as an influencing factor. The framework embeds inclusion of key features of Arab culture being drawn from Arabic settings of the investigated organisations. The role of Arab culture may hence facilitate making decisions in the absence of quality data. Given concerns with previous attempts to address the data quality issue in Abu Dhabi Emirate, the framework is likely to provide an effective response to these concerns. As realised from the framework development, the literature and other established frameworks, common elements exist in many frameworks with the same objectives. However, the distinctive features of ADDQF are believed to be significant as a basis for change for Abu Dhabi's organisations into a data quality adopting organisations.

Chapter 9: Conclusion

9.1. Conclusion and Recommendations

Data quality is an important aspect of organisations' strategies. Providing decision makers with the most accurate data possible for reaching the best decisions possible and consequently attaining their objectives is an ever-sought aim. In the case of public organisations, decisions ultimately concern the public and hence, diligence is required to make sure that these decisions do, for instance, preserve the economic resources, maintain public health, and provide national security. This research comprised a study of several aspects of data quality and its decisive role in providing decision makers with appropriate insights into making efficient decisions. It was evident from the results obtained that there is a positive relationship between the quality of data reaching decision makers and the efficiency of the decisions made. The focus of the study was on strategic decisions in public organisations.

Generally speaking, strategic decisions made in public organisations are not efficient if the results obtained are unsatisfactory, such as waste of public resources and squandering. The decision making process requires a wealth of information in order to achieve efficient results. Public organisations typically acquire great amounts of data generated by public services. However, the vast amount of data stored in public organisations' databases may be one of the main reasons for inefficient decisions made by public organisations. Processing vast amounts of data and extracting accurate information is not an easy task. Although technology helps in this respect, it cannot improve decisions to a significant level of assurance. Technology is not sufficient to furnish data quality for public organisations' decision makers to make efficient decisions. For example, Decision Support Systems (DSSs) have been widely used in the private and public sectors to help decision makers improve their decisions. However, DSSs are associated with some drawbacks, such as rising costs and inaccuracy. Data mining techniques can help improve accuracy of data provided by DSSs. This research proposed certain aspects of using data mining to improve results obtained by DSSs. It was shown, however, that further considerations and requirements are needed for improving data quality.

Seven public organisations in Abu Dhabi Emirate were considered to discuss how public organisations can attain a better decision making process.

Research shows that there is a positive relationship between data quality and efficiency of decisions: higher data quality promotes better decisions. Therefore, in order to improve the decision making process, it is necessary to provide decision makers with quality data. The findings in the organisations revealed inefficiency of decisions, although certain countermeasures were considered. For example, DSSs have been in use in the organisations but they have failed to improve the quality of decisions to a confident level of sufficiency. This study argued that technological considerations would not be adequate to improve the decisions. The research maintained, based on the findings of studies conducted in these Abu Dhabi public organisations and supported by a literature review, that a complete data quality framework is needed in order to improve data quality and consequently the decision making process in public organisations. The studies conducted in the public organisations in Abu Dhabi contributed to the design of a data quality framework. The framework comprises seven elements ranging from technical to human-based, which are found important to attain data quality in public organisations taking Abu Dhabi public organisations as the case. The interaction and integration of these elements contributes to the quality of data reaching decision makers and hence to the efficiency of decisions made by public organisations. Sufficiency and generalisation of the framework to any public organisation are maintained by the flexibility of the sub-elements which may be modified according to the different needs of individual public organisations, given that all elements of the framework are considered in the implementation project. By adopting the framework to incorporate data quality in the strategic endeavours, the decision making process in public organisations in Abu Dhabi will be more efficient and decision will be more effective. The quality of data will directly improve the overall quality of decisions.

9.2. Achievement of the Aim and Objectives

The general aim stated at the initial stage is satisfied by conducting this study. The aim was "*to understand data quality issues in public organisations in Abu Dhabi Emirate and accordingly suggest appropriate ways for overcoming these issues up to improving data quality used for strategic decisions of these organisations*." This aim was attained by the devised data quality framework for Abu Dhabi public organisations. The framework was realised by a combination of three elements:

1. Findings from interviews at seven public organisations in Abu Dhabi;
2. Outcomes of a review of existing literature; and
3. Comparison with existing frameworks for data quality.

Implementation guidance of the framework was provided and the framework was reviewed by the 14 executives interviewed earlier in the 7 organisations of Abu Dhabi Emirate. The executives maintained the normative nature of the framework in the sense of its possible implementation in their organisations. The framework does not only reflect some of the known problems in the studied organisations, but also identifies the areas which result in current data quality problems. Essentially, the framework targets the underlying roots of the problems which may be overlooked by ordinary auditing. The elements that the framework integrates provide a basis for refining several relevant factors that add up to improving the overall quality of data reaching decision makers.

The framework has high potential to provide an effective response to data quality issue in public organisations in Abu Dhabi. The framework has distinctive features, such as culture, all-level leadership, system integration, and data refinement techniques. Moreover, the integrity of the framework and the intertwining nature of its elements are distinctive features of the framework. These factors have been overlooked by other frameworks aimed at addressing the same aspects.

The research questions initiated at early stages of the study can hence be answered now:

1. *Is data quality an issue in public organisations in general and in Abu Dhabi's public organisations in particular? How is this issue manifested?*
   Answer: Yes, it is. Improving the quality of data reaching decision makers is an ongoing aim of any organisation regardless of its type or nature. The study provided evidence that data quality is a main issue in public and private organisations (Chapters 2, 6 and 7). However, because strategic decisions in public organisations are directly related to the public and these decisions have social, economic and political consequences in the countries, improving data quality is more important and necessary to address in public organisations than

in private organisations. The need for quality data is crucial given several reasons, including cost and likely sensitivity of decisions taken by governmental bodies (Chapter 7 and 8).

2. *What are the main problems facing improving data quality in public organisations?*

   Answer: Public organisations face challenges of choosing the correct data for strategic decisions from huge amounts of data produced by the operations of these organisations (Chapters 2 and 7). The complex and bureaucratic nature of public bodies and governmental institutions requires careful attention to several matters when undertaking a change project. Furthermore, the continuous increase in complexity in decision making requires executives to consider a vast number of inputs and a considerable amount of knowledge, rendering the necessity of acquiring tools for assessing the quality of the data at hand inevitable (Chapter 4). Moreover, decision makers in the public sectors vary in their needs for the quality of data (Chapters 2 and 4).

3. *How can data quality contribute to better decision making?*

   Answer: This relationship is better understood by looking at the proposed data quality framework. The framework consists of elements contributing to data quality. Each element has a mutual positive relationship with the lower level's element. Relationships between levels are transitive and hence extend forwards and backwards across levels. This means that data quality has a mutual positive relationship with the highest level, decision making (Chapter 8 the implementation recommendations section).

4. *Are there specific methods to be followed for improving data quality in Abu Dhabi's public organisations? If so, what are these methods? How can these methods assure better data quality?*

Answer: The data quality framework encapsulates the possible areas of improvement of the quality of data reaching the decision makers in public organisations in Abu Dhabi The theoretical basis of the framework, supported by the mapping onto other existing, well-established frameworks, in addition to the distinctive features of the framework detailed above, all provide strong evidence for the ability of the framework to improve data quality. Moreover, given the multitude of elements encapsulated in the framework, for example, policies, systems, humans, culture and leadership, it is highly likely that the framework can form a basis for changed organisational practices of Abu Dhabi public organisations towards improved data quality. The argument of forming a basis for organisational change is reinforced by the integration and interaction of elements, and culture as an influencing factor. The framework embeds inclusion of key features of Arab culture being drawn from Arabic settings of the investigated organisations. The role of Arab culture may hence facilitate making decisions in the absence of quality data. Given concerns with previous attempts to address the data quality issue in Abu Dhabi Emirate, the framework is likely to provide an effective response to these concerns. As realised from the framework development, the literature and other established frameworks, common elements exist in many frameworks with the same objectives. However, the distinctive features of ADDQF are believed to be significant as a basis for change for Abu Dhabi's organisations into a data quality adopting organisations (Chapter 8).

9.3. Contribution to Knowledge

Data quality plays a significant role in many if not all business and governmental applications (Batini et al., 2009). According to the Audit Commission and the National Health Service in England (2007), the risk of not identifying and addressing weaknesses in data quality is that information may be misleading, decision making may be flawed, policies may be ill-founded, resources may be wasted and poor services may not be improved. There is also a risk that good performance may not be recognised and rewarded.

This work makes a significant contribution to knowledge in several respects all related to data quality. These are detailed in the Discussion chapter and a brief description of the contribution is provided below.

This work comprised extensive research and an empirical study that collected primary from interviews in seven public organisations in Abu Dhabi. The outcome is a data quality framework, denoted ADDQF, for Abu Dhabi's public organisation. The framework aims to improve data quality in public organisations with proven potential to do so. This is based on the fact that the framework provides a formal approach to enhancing data quality, and also benefits from other frameworks that are used for the same aim.

A comparison was undertaken between ADDQF and other well-established data quality frameworks, namely the UK's public sector data quality framework and the Framework for Information Quality (FIQ) based on the Zachman Framework for Enterprise Architecture. The elements of FIQ and the UK's public sector data quality framework were all matched by elements of ADDQF. However, some facets which contribute to both data quality and to better implementation of the framework are overlooked by the other frameworks. These facets are unique to ADDQF. In particular, the integrity and interaction among the elements of the framework were found significant. The elements of ADDQF have mutual and transitive relationships and therefore require that the framework be considered as a whole not as individual elements. The elements could be thought of as levels, each of which should be attained in order to obtain data quality and eventually effective decisions. For example, embracing cultural values is the basis of the other levels. Relationships between levels are transitive and hence extend forwards and backwards across levels. This means that, for example, the relationship between decision making and data quality extends to leadership and so on. Other frameworks suggest using the framework elements without considering the layered approach or viewing the framework as an integral system but rather suggest that elements could be adopted individually. The other frameworks do not emphasise the concept of layers or levels as required by ADDQF, which are key aspects for the implementation of the framework. Furthermore, no transitive relationships among elements are recognised in the other frameworks. These relationships were found essential in this study in order to improve data quality in the adopting organisation. Another aspect unique to the ADDQF is adoption of knowledge extraction and data mining techniques. The use of data refinement techniques, such as data mining and integration of data mining and decision support systems is another unique feature of ADDQF. New and emerging techniques of data quality enhancement and knowledge extraction may also be of significant importance for the organisation as shown in the experiments conducted, and should be considered as key aspects of improving data quality. In this sense, data mining techniques and their integration with decision support systems will help improve data quality and provide decision makers with valuable information.

The importance of the distinguishing factors in ADDQF as the findings revealed and shown in the discussion is that attaining data quality requires a chain of processes and elements that interact with one another. Existence of those elements is necessary but not sufficient as detailed in the discussion since, for example, leadership should extend to all levels, reporting must follow a vertical and horizontal approach in the organisation, and measuring organisational performance requires inputs from all levels, and so on.

9.4. Limitations and Future Work

The work has certain limitations which are highlighted in this section. First, the study relied mainly on qualitative design, which may be regarded over-descriptive in some contexts. The study focussed on public organisations in Abu Dhabi and it may need further discussion on how results specific to Abu Dhabi's public organisations can be generalised to a broader spectrum of public organisations. Given the cultural context and particularity of that context of the chosen organisations, the proposed framework may need to be tailored in order to be adopted by other public organisations, although the flexibility of the sub-elements of the framework, which may be modified according to the different needs of individual public organisations, allows generalisation of the framework to other public organisations. Regardless of that, the results might have been more substantial had a wider range of organisations been investigated.

The choice of the data mining technique for the conducted experiments was limited to classification. The aim of the experiments was to show the use of the technique can improve data quality. Testing with more techniques, such as clustering for instance, may provide further insight into the uses and benefits of data mining in public organisations.

The framework's theoretical basis, supported by the review of the literature and the mapping onto other well-established frameworks, in addition to the executives' positive feedback on its applicability and normative nature, all provide high potential for the framework. However, implementation of the framework would be more insightful

Future work can use the abovementioned limitations as a basis for developing the work further. For example, investigating the aspects of data quality in several public organisations across the UAE, other Arab Gulf countries, or even across the Arab world would be desirable for generalising the results over Arab culture. Accordingly, the impact of culture would be better realised and understood. Complementary quantitative survey would also be a desirable addition in measuring recurrence of problems related to data quality, or for measuring integrity of data. Further research on this aspect would be necessary in order to identify aspects of how quantitative data may be benefited from.

References

Abdel-Aal, R. E. (2005). Improved classification of medical data using abductive network committees trained on different feature subsets. *Computer Methods and Programs in Biomedicine*, *80*(2), 141-153.

Abdullah, S., Razak, A. A., Hanafi, M. H., & Jaafar, M. (2013). Implementation Barriers of ISO 9000 within the Malaysian Local Government. *International Journal of Quality & Reliability Management*, *30*(8), 3-3.

Abu Dhabi e-Government (2015). Available at: https://www.abudhabi.ae/portal/public/en/gen_info_detail?docName=ADEGP_DF_13272_EN&_adf.ctrl-state=15jdyteh1v_4&_afrLoop=559502574986836. (Last accessed 08/04/2015).

Abu-Naser, S., Al-Masri, A., Sultan, Y. A., & Zaqout, I. (2011). A Prototype Decision Support System for Optimizing the Effectiveness of E-learning in Educational Institutions. *International Journal of Data Mining & Knowledge Management Process*, *1*(4).

Accountant Magazine Middle East - February 2013. Page 47.

Adlakha, A. (2002). *Single trial EEG classification*. Technical Report. Lausanne: Swiss Federal Institute of Technology.

ADSIC (2015). Abu Dhabi Systems and Information Centre. Available at: https://adsic.abudhabi.ae/adsic/faces/en/home?_afrLoop=3248155816092421#%40%3F_afrLoop%3D3248155816092421%26_adf.ctrl-state%3Dd8sgeddaa_49. (Last accessed 15/04/2015).

Aguilar-Saven, R. S. (2004). Business process modelling: Review and framework. *International Journal of Production Economics*, *90*(2), 129-149.

Al–Hujran, O. (2012). An assessment of Jordan's e–government maturity: a user–centric perceptive. *International Journal of Electronic Governance*, *5*(2), 134-150.

Alkaabi, A., & Maple, C. (2013). Cultural impact on user authentication systems. *International Journal of Business Continuity and Risk Management*, *4*(4), 323-343.

Andrews, R., & Van de Walle, S. (2012). New Public Management and Citizens' Perceptions of Local Service Efficiency, Responsiveness, Equity and Effectiveness. *Public Management Review*, (pre-print), 1-22.

Angelsen, A., Larsen, H. O., & Lund, J. (Eds.). (2011). *Measuring livelihoods and environmental dependence: Methods for research and fieldwork*. Routledge.

Antunes, P., & Carrio, L. (2003, October). Modeling the information structures of meetingware. In *Proc. of Workshop de Sistemas de Informao Multimdia e Cooperativos (COOP-MEDIA'03)*.

Arnott, D., & Pervan, G. (2005). A critical analysis of decision support systems research. *Journal of Information Technology*, *20*(2), 67-87.

Arnott, D., & Pervan, G. (2008). Eight key issues for the decision support systems discipline. *Decision Support Systems*, *44*(3), 657-672.

Arts, D. G., De Keizer, N. F., & Scheffer, G. J. (2002). Defining and improving data quality in medical registries: a literature review, case study, and generic framework. *Journal of the American Medical Informatics Association*, *9*(6), 600-611.

Asur, S., & Huberman, B. A. (2010, August). Predicting the future with social media. In *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on* (Vol. 1, pp. 492-499). IEEE.

Audit Commission (2007). Improving information to support decision making: standards for better quality data. *Audit Commission*, *20*.

Babenko, B. (2008). Multiple instance learning: algorithms and applications. *View Article PubMed/NCBI Google Scholar*.

Balutis, P. (2001). E-government 2001, Part 1: Understanding the challenge and evolving strategies, *The Public Manager*, *2001*(Spring), pp. 33-37.

Batini, C., Cappiello, C., Francalanci, C., & Maurino, A. (2009). Methodologies for data quality assessment and improvement. *ACM Computing Surveys (CSUR)*, *41*(3), 16.

Berson, A., Smith, S., & Thearling, K. (2004). An Overview of Data Mining Techniques. *Building Data Mining Application for CRM*.

Betts, S. C. (2004). Resolving a paradox between mentoring, LMX and charisma: a process approach to leadership development. *Journal of Organizational Culture, Communications and Conflict*, *8*(1), 111-25.

Bhargava, H. K., Power, D. J., & Sun, D. (2007). Progress in Web-based decision support technologies. *Decision Support Systems*, *43*(4), 1083-1095.

Bhate, P. M. & Patil, S. B. (2014). Data Mining Techniques for Customer Relationship Management in Organized Retail Industry. In *Proceedings of National Conference on* (p. 35).

Billy, T. W., To, W. M., & Lee, P. K. (2012). Quality management framework for public management decision making. *Management Decision*, *50*(3), 420-438.

Blake, R., & Mangiameli, P. (2011). The effects and interactions of data quality and problem complexity on classification. *Journal of Data and Information Quality (JDIQ)*, *2*(2), 8.

Bogue, R. (2010). Brain-computer interfaces: control by thought. *Industrial Robot: An International Journal*, *37*(2), 126-132.

Bos, D. P. O., Reuderink, B., van de Laar, B., Gürkök, H., Mühl, C., Poel, M., ... & Heylen, D. (2010). Brain-computer interfacing and games. In *Brain-Computer Interfaces* (pp. 149-178). London: Springer.

Bouckaert, R. R. (2008). Bayesian network classifiers in Weka for version 3-5-7. *Artificial Intelligence Tools*, *11*(3), 369-387.

Boyd, D. & Crawford, K. (2011, September). Six Provocations for Big Data. *A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society*.

Bujak, A., Carvalho, W., & Sriramulu, R. (2012). Lean management and operations in the global professional services industry. In U. Baumer, P. Kreutter & W. Messner (Eds.), *Globalisation of professional services* (pp. 95-104). Berlin: Springer-Verlag.

Calderon, T. G., Cheh, J. J., & Kim, I. W. (2003). How large corporations use data mining to create value. *Management Accounting Quarterly*, *4*(2), 1.

Calman, L. (2012). What is Grounded Theory. UK: *University of Manchester*.

Carter, L., & Bélanger, F. (2005). The utilization of e-government services: citizen trust, innovation and acceptance factors*. *Information Systems Journal*, *15*(1), 5-25.

Cate, F. H. (2008). Government data mining: The need for a legal framework. *Harvard Civil Rights-Civil Liberties Law Review (CR-CL)*, *43*(2).

Chapman, A. D. (2005). *Principles and methods of data cleaning*. GBIF.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). *CRISP-DM 1.0 Step-by-step data mining guide*.

Charmaz, K. (2003). Grounded theory. *Strategies of Qualitative Inquiry*, *2*, 249.

Chumerin, N., Manyakov, N. V., Combaz, A., Suykens, J. A., & Van Hulle, M. M. (2009, September). An application of feature selection to on-line P300 detection in brain-computer interface. In *Machine Learning for Signal Processing, 2009. MLSP 2009. IEEE International Workshop on* (pp. 1-6). IEEE.

Clarke, S. (2001). *Information Systems Strategic Management: An Integrated Approach*. London: Routledge.

Cleary, D. (2011). *Predictive Analytics in the Public Sector: Using Data Mining to Assist Better Target Selection for Audit*. Ireland: Revenue Irish Tax and Customs: .

Collin, C. (2008). *Introduction to e-Business.* New Delhi: Elsevier.

Columbus, L. (2014). *84% Of Enterprises See Big Data Analytics Changing Their Industries' Competitive Landscapes In The Next Year*. [Online] Forbes.com. Retrieved 15/09/2014: http://www.forbes.com/sites/louiscolumbus/2014/10/19/84-of-enterprises-see-big-data-analytics-changing-their-industries-competitive-landscapes-in-the-next-year/.

Combe, D., Largeron, C., Egyed-Zsigmond, E., & Géry, M. (2010, October). A comparative study of social network analysis tools. In *WEB INTELLIGENCE & VIRTUAL ENTERPRISES*.

Corley, C. D., Cook, D. J., Mikler, A. R., & Singh, K. P. (2010). Text and structural data mining of influenza mentions in web and social media. *International Journal of Environmental Research and Public Health*, *7*(2), 596-615.

Cortez, P. (2010). Data mining with neural networks and support vector machines using the R/rminer tool. In *Advances in Data Mining. Applications and Theoretical Aspects* (pp. 572-583). Springer Berlin Heidelberg.

Cugliari, J. & Guille, A. (2014). *Social Media Mining.* France: Université Lumière Lyon 2.

Cuong, B. C., Lanzi, P. L., & Thong, N. T. (2012). A novel intuitionistic fuzzy clustering method for geo-demographic analysis. *Expert Systems with Applications*, *39*(10), 9848-9859.

Databeacon Staff (2004). East of England Observatory adopts hosted services decision support solution. Available at http://www.dssresources.com/cases/eastofengland/index.html. (Accessed on 26/09/2014).

Davidrajuh, R. (2004). Planning e-government start-up: a case study on e-Sri Lanka. *Electronic Government, an International Journal*, *1*(1), 92-106.

De Oliveira, J. V. and Pedrycz, W. (2007): *Advances in Fuzzy Clustering and Its Applications.* John Wiley & Sons.

Dedeke, A. (2000, October). A Conceptual Framework for Developing Quality Measures for Information Systems. In *IQ* (pp. 126-128).

Delone, W. H., & McLean, E. R. (2003). The DeLone and McLean model of information systems success: a ten-year update. *Journal of Management Information Systems*, *19*(4), 9-30.

Du, H. (2010). *Data Mining Techniques and Application: An Introduction*, Hampshire, UK: Course Technology Cengage Learning.

Dwivedi, Y. Wade, M. and Schneberger, S. (2012). *Information System Theory: explaining and predicting our digital society*, New York, London: Springer.

Edmunds, A., & Morris, A. (2000). The problem of information overload in business organisations: a review of the literature. *International Journal of Information Management*, *20*(1), 17-28.

El Seddawy, A. B., Khedr, A., & Sultan, T. Adapted Framework for Data Mining Technique to Improve Decision Support System in an Uncertain Situation.

Eppler, M., & Helfert, M. (2004, November). A classification and analysis of data quality costs. In *International Conference on Information Quality*.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine, 17(3).*

Fetaji, M., Loskoska, S., Fetaji, B., & Ebibi, M. (2007). Investigating human computer interaction issues in designing efficient virtual learning environments. In *Balkan Conference in Informatics (BCI 2007), vol. 2* (pp. 313-324).

Filip, F. G. (2008). Decision support and control for large-scale complex systems. *Annual Reviews in Control*, *32*(1), 61-70.

Forrest, J. (2005). *The Space Shuttle Challenger Disaster: A failure in decision support system and human factors management*. Originally prepared November 26, 1996, published October 7, 2005 at http://dssresources.com/cases/spaceshuttlechallenger/.

Fountain, J. E. (2001). Paradoxes of public sector customer service. *Governance*, *14*(1), 55-73.

Friston, K. J., & Dolan, R. J. (2010). Computational and dynamic models in neuroimaging. *Neuroimage*, *52*(3), 752-765.

Gaidos, S. (2011). *Mind Controlled*. An article in Science News Magazine, pp. 26-29, 2nd July.

Ganguly, A. R., Gupta, A., & Khan, S. (2005). Data mining technologies and decision support systems for business and scientific applications. *Encyclopaedia of Data Warehousing and Mining, Blackwell Publishing*.

Georgieva, P., Silva, F., Milanova, M., & Kasabov, N. (2014). EEG Signal Processing for Brain–Computer Interfaces. In *Springer Handbook of Bio-/Neuroinformatics* (pp. 797-812). Berlin Heidelberg: Springer.

Gilbert, D. T., Morewedge, C. K., Risen, J. L., & Wilson, T. D. (2004). Looking forward to looking backward the misprediction of regret. *Psychological Science*, *15*(5), 346-350.

Gil-García, J. R., & Pardo, T. A. (2005). E-government success factors: Mapping practical tools to theoretical foundations. *Government Information Quarterly*, *22*(2), 187-216.

Glaser, B. G. & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. New York: Aldine de Gruyter.

Gordini, N. (2014). A genetic algorithm approach for SMEs bankruptcy prediction: Empirical evidence from Italy. *Expert Systems with Applications*.

Gordon, S. and Gordon, J. (2004). *Information Systems: A Management Approach* (3rd Edition), NJ: Leyh Publishing, LLC.

Grad, B. (2007). The creation and the demise of VisiCalc. *IEEE Annals of the History of Computing*, *29*(3), 20-31.

Haag, S. and Cummings, M. (2013). *Management Information Systems for the Information Age*, NY: McGraw-Hill Irwin.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: an update. *ACM SIGKDD explorations newsletter*, *11*(1), 10-18.

Han, J., Altman, R. B., Kumar, V., Mannila, H., & Pregibon, D. (2002). Emerging scientific applications in data mining. *Communications of the ACM*, *45*(8), 54-58.

Han, J., Kamber M. and Pei, J. (2006). *Data Mining: Concepts and Techniques*, 2nd Edition. Morgan Kaufmann.

Hand, D. J., Mannila, H., & Smyth, P. (2001). *Principles of data mining.* MIT Press.

Hardin, J. M. and Chhieng, D. C. (2007): *Data Mining and Clinical Decision Support Systems*. Clinical Decision Support Systems, *VOL*, 44-63

Harrison, M. I., Koppel, R., & Bar-Lev, S. (2007). Unintended consequences of information technologies in health care—an interactive sociotechnical analysis. *Journal of the American Medical Informatics Association*, *14*(5), 542-549.

Health Information and Quality Authority (HIQA). (2011). International Review of Data Quality. Available at www.hiqa.ie/system/files/HI-International-Review-Data-Quality.pdf.

Hedgebeth, D. (2007). Data-driven decision making for the enterprise: an overview of business intelligence applications. *VINE*, *37*(4), 414-420.

Heeks, R. (2000). *Information Systems for Public Sector Management*. Institute for Development Policy and Management.

Hofer, C.W. and Schendel, D.E.. (1978). *Strategy formulation: Analytical concepts*. St. Paul, MN: West Publishing.

Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions and organizations across nations*. Sage Publications.

Hofstede, G. (2015). *Cultural Dimensions: The United Arab Emirates*. Available online at lihttp://geert-hofstede.com/arab-emirates.html. Last accessed 22nd April 2015.

Huang, Z. (1997, February). Clustering large data sets with mixed numeric and categorical values. In *Proceedings of the 1st Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)* (pp. 21-34).

Hung, S. Y. (2003). Expert versus novice use of the executive support systems: an empirical study. *Information & Management*, *40*(3), 177-189.

Huotari, M. L., & Wilson, T. D. (2001). Determining organizational information needs: the critical success factors approach. *Information Research*, *6*(3), 6-3.

Issenberg, S. (2012). How President Obama's campaign used big data to rally individual voters. *MIT Technology Review*.

Jabr, F. (2011). Mind-controlled prosthetics to help amputees. *New Scientist*, *210*(2810), 11.

Jessup, L. and Valacich, J. (2003). *Information Systems Today*, New Jersey: Prentice Hall.

Jovanovic, P., Romero, O., Simitsis, A., & Abelló, A. (2012, November). ORE: an iterative approach to the design and evolution of multi-dimensional schemas. In

Proceedings of the fifteenth international workshop on Data warehousing and OLAP (pp. 1-8). ACM.

Junhua, Z. (2002). Will the government serve the people? the development of Chinese e-Government, *New Media and Society*,*4*(2, June), .

Kaptein, A. M. (2005). *Meta-classifier approaches to reliable text classification* (Doctoral dissertation, Universiteit Maastricht).

Kar, A. (2011). *Genetic Algorithm*. Business fundas. Retrieved on 14/10/2014: http://business-fundas.com/2011/genetic-algorithm/.

KD Nuggets (2013). Top Languages for analytics, data mining, data science. Available at http://www.kdnuggets.com/2013/08/languages-for-analytics-data-mining-data-science.html (accessed 15/09/2014).

Keim, D. A. (2002): *Information visualization and visual data mining*, IEEE Transactions on Visualization and Computer Graphics, *7*(2), 100–107.

Kerby, A. (2009): Spatial Clustering Using the Likelihood Function. PhD Thesis submitted to University of Nebraska at Lincoln.

Kohavi, R.(2000): *Data Mining and Visualisation*, National Academy of Engineering (NAE) US Frontiers of Engineering 2000.

Kopáčková, H., & Škrobáčková, M. (2006). Decision support systems or business intelligence: what can help in decision making?. *Scientific Papers of the University of Pardubice. Series D, Faculty of Economics and Administration*, *10*.

Krauledat, J. M. (2008). Analysis of nonstationarities in EEG signals for improving brain-computer interface performance (Doctoral dissertation,).

Kriegel, H. P., Kröger, P. and Zimek, A. (2009): Clustering high-dimensional data: A survey on subspace clustering, pattern-based clustering, and correlation clustering. ACM, *Knowledge Discovery*, Data 3, 1, Article 1, March

Kumar, D. S., Sathyadevi, G., & Sivanesh, S. Decision Support System for Medical Diagnosis Using Data Mining. *International Journal of Computer Science*, *8*.

Lam, S. S., & Schaubroeck, J. (2000). Improving group decisions by better pooling information: a comparative advantage of group decision support systems. *Journal of Applied Psychology*, *85*(4), 565.

Laperrouze, A., Autio, E., Bucar, M., Licht, G., Molero, J., & Tsipouri, L. (2010). *Eurostars Programme Interim Evaluation*.

Lapointe, J., & Desieno, D. (2003). *U.S. Patent No. 20,030,004,906*. Washington, DC: U.S. Patent and Trademark Office.

Larose, D. T. (2014). *Discovering knowledge in data: an introduction to data mining*. John Wiley & Sons.

Laudon, K., & Laudon, J. (2010). Management Information Systems: International Edition, 11.

Law, M. H. C. (2006): *Clustering, Dimensionality Reduction and Side Information*, PhD thesis, Michigan State University.

Lin, S., Gao, J., Koronios, A., & Chanana, V. (2007). Developing a data quality framework for asset management in engineering organisations. *International Journal of Information Quality*, *1*(1), 100-126.

Lin, W. K., Chiu, C. K., & Tsai, Y. H. (2008). Modeling relationship quality and consumer loyalty in virtual communities. *CyberPsychology & Behavior*, *11*(5), 561-564.

Liu, S., Duffy, A. H., Whitfield, R. I., & Boyle, I. M. (2010). Integration of decision support systems to improve decision support performance. *Knowledge and Information Systems*, *22*(3), 261-286.

Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., & Arnaldi, B. (2007). A review of classification algorithms for EEG-based brain–computer interfaces. *Journal of Neural Engineering*, *4*.

Lucas, J.R. & Henry, C. (1994). *Information Systems Concepts for Management* (Fifth Edition), London: Mitchell Publishing.

Lv, J., & Li, Z. (2009, August). Research on architecture of decision support system. In *Computer Science and Information Technology, 2009. ICCSIT 2009. 2nd IEEE International Conference on* (pp. 199-202). IEEE.

Mahdi, I. M., & Alreshaid, K. (2005). Decision support system for selecting the proper project delivery method using analytical hierarchy process (AHP). *International Journal of Project Management*, *23*(7), 564-572.

Maimon, O. Z., & Rokach, L. (Eds.). (2005). *Data mining and knowledge discovery handbook* (Vol. 1). New York: Springer.

March, S. T., & Hevner, A. R. (2007). Integrated decision support systems: A data warehousing perspective. *Decision Support Systems*, *43*(3), 1031-1043.

Markov, Z., & Larose, D. T. (2007). Data mining the Web: uncovering patterns in Web content, structure, and usage. Wiley-Interscience.

McCue, C., & Parker, A. (2003). Connecting the Dots: Data Mining and Predictive Analytics in Law Enforcement and Intelligence Analysis. *Police Chief*, *70*(10), 115-124.

McGilvray, D. (2010). *Executing data quality projects: ten steps to quality data and trusted information (TM)*. Morgan Kaufmann.

McLean, E.R. and Tuban, E. (2005). *Information Technology for Management,* 5th Editio), Edison: Wiley.

McLeod, R. & Schell, G.P. (2007). *Management Information Systems*. 9th Edition. New Jersey: Prentice-Hall. Inc.

Metz, C. (2010). Facebook: Why our 'next-gen' comms ditched MySQL. *The Register.* Retrieved 20/10/2014 from: http://www.theregister.co.uk/2010/12/17/facebook_messages_tech/.

Miller, H. J., & Han, J. (Eds.). (2009). *Geographic data mining and knowledge discovery*. CRC Press.

Mishra, P.; Padhy, N. and Panigrahi R. (2012). The Survey of Data Mining Applications and Feature Scope. *Asian Journal of Computer Science and Information Technology, 2*, 468-77

Mladenić, D. (2001). EU project: Data mining and decision support for business competitiveness: a European virtual enterprise (Sol-Eu-Net). *OES-SEO 2001: Open enterprise solutions: Systems, experiences and organizations*, 172-173.

Mladenic, D.; Lavrač, N.; Bohanec, M. and Moyle, S. (2003): *Data Mining and Decision Support: Integration and Collaboration*. The Springer International Series in Engineering and Computer Science, Vol. 745, 304, p.32

Mohemad, R., Hamdan, A. R., Othman, Z. A., & Noor, N. M. M. (2010). Decision Support Systems (DSS) in Construction Tendering Processes. *arXiv preprint arXiv:1004.3260*.

Moreland, N., & Clark, M. (1998). Quality and ISO 9000 in educational organizations. *Total Quality Management*, *9*(2-3), 311-320.

Moss, L. T., & Atre, S. (2003). *Business intelligence roadmap: the complete project lifecycle for decision-support applications*. Addison-Wesley Professional.

Nai-Jen, H., & Palaniappan, R. (2004, September). Classification of mental tasks using fixed and adaptive autoregressive models of EEG signals. In *Engineering in Medicine and Biology Society, 2004. IEMBS'04. 26th Annual International Conference of the IEEE* (Vol. 1, pp. 507-510). IEEE.

Nayab, N. (2011). *Disadvantages to Using Decision Trees*. Retrieved 17/09/2014, from Bright Hub PM at http://www.brighthubpm.com/project-planning/106005-disadvantages-to-using-decision-trees/.

Nelson, G., & Wright, J. (2005). Real time decision support: creating a flexible architecture for real time analytics. *DSSResources. COM*, *11*, 18.

Nemati, H. R., & Barko, C. D. (2003). Key factors for achieving organizational data-mining success. *Industrial Management & Data Systems*, *103*(4), 282-292.

NetMiner. (2014). Website. http://www.netminer.com/index.php

Neville, J. (2010): *Data Mining*, Purdue University, September 28

Nijboer, F., Sellers, E. W., Mellinger, J., Jordan, M. A., Matuz, T., Furdea, A., ... & Kübler, A. (2008). A P300-based brain–computer interface for people with amyotrophic lateral sclerosis. *Clinical Neurophysiology*, *119*(8), 1909-1916.

O'Brien, J. and Marakas, G. (2010). *Introduction to Information Systems* (Fifteenth Edition), NY: McGraw-Hill Irwin.

Odden, L. (2010). *6 Facebook Search Engine & Data Visualization Tools*. Retrieved on 18/10/2014 from TopRank blog: http://www.toprankblog.com/2010/08/6-facebook-search-engine-data-visualization-tools/.

O'Donnell, O., & Boyle, R. (2008). Understanding and managing organisational culture. Institute of Public Administration. Dublin, Ireland.

Oracle (2014). Oracle and Crystal Ball. Available at http://www.oracle.com/us/products/applications/crystalball/acquisition-066567.html. (Accessed on 18/09/2014).

Padhy, N., Mishra, P., & Panigrahi, R. (2012). The Survey of Data Mining Applications and Feature Scope. *International Journal of Computer Science*.

Pan, Y. and Li, T. (2007). The data mining of the e-government on the basis of Fuzzy Logic, in *International Conference on Integration Technology*, IEEE.

Parliament. (2009). CHAPTER 2: Overview of surveillance and data collection. In: *Constitution Committee - Second Report Surveillance: Citizens and the State*. House of Lords. www.parliament.uk. Available at http://www.publications.parliament.uk/pa/ld200809/ldselect/ldconst/18/1804.htm

Pearlson, K. and Saunders, C. (2009). *Strategic Management of Information Systems* (Fourth Edition), NJ: John Wiley and Sons (Asia).

Peignot, J., Peneranda, A., Amabile, S., & Marcel, G. U. E. N. O. U. N. (2012). Strategic Decision Support Systems for Local Government: A Performance Management Issue? The Use of Information Systems on the Decision-making and Performance Management of Local Government. *International Business Research*, *6*(2), 92.

Perrizo, W., Ding, Q., & Denton, A. (2002, November). Lazy Classifiers Using P-trees. In *CAINE* (pp. 176-179).

Pipino, L. L., Lee, Y. W., & Wang, R. Y. (2002). Data quality assessment. *Communications of the ACM*, *45*(4), 211-218.

Poley, M. J., Brouwer, W. B., Busschbach, J. J., Hazebroek, F. W., Tibboel, D., Rutten, F. F., & Molenaar, J. C. (2008). Cost-effectiveness of neonatal surgery: first greeted with scepticism, now increasingly accepted. *Pediatric Surgery International*, *24*(2), 119-127.

Porter, M.E. (2001). Strategy and Internet, *Harvard Business Review*,*57*(137).

Power, D. J. (2002): *DSS News*, A Bi-Weekly Publication of dssresources.com, Vol. 3(23). Available at http://dssresources.com/newsletters/66.php. (Accessed on 26/09/2014).

Power, D. J. (2007). A brief history of decision support systems. *DSSResources. COM, World Wide Web,* http://DSSResources. COM/history/dsshistory.html*, version*, *4*.

Power, D. J. (2008): Understanding Data-Driven Decision Support Systems, *Information Systems Management*, *25*(2, March), 149–154.

Power, D. J., & Sharda, R. (2007). Model-driven decision support systems: Concepts and research directions. *Decision Support Systems*, *43*(3), 1044-1061.

Raghunathan, S. (1999). Impact of information quality and decision-maker quality on decision quality: a theoretical model and simulation analysis. *Decision Support Systems*, *26*(4), 275-286.

Rahm, E., & Do, H. H. (2000). Data cleaning: Problems and current approaches. *IEEE Data Eng. Bull.*, *23*(4), 3-13.

Rainer, K. and Cegielski, C. (2013). *Introduction to Information Systems* (Fourth Edition), NJ: John Wiley and Sons.

Ranjan, J., & Bhatnagar, V. (2008). A review of data mining tools in customer relationship management. *Journal of Knowledge Management Practice*, *9*(1).

Reid, A., & Catterall, M. (2005). Invisible data quality issues in a CRM implementation. *The Journal of Database Marketing & Customer Strategy Management*, *12*(4), 305-314.

Remenyi, D. (1991). *Introducing strategic information systems planning*, Oxford: NCC Blackwell.

Riad, A. M., El-Bakry, H. M., & El-Adl, G. H. (2010). A novel DSS framework for e-government. *IJCSI*, 32.

Riad, A. M., El-Bakry, H. M., & El-Adl, G. H. (2011, December). A new developed DSS framework for complicated services of E-government. In *Proceedings of the 10th WSEAS international conference on E-Activities* (pp. 124-129). World Scientific and Engineering Academy and Society (WSEAS).

Ripley, B. D. (2008). *Pattern Recognition and Neural Networks*. 1st edition. Cambridge University Press.

Rockart, F. and Scott, S. (1984). Implications of changes in Information Technology for corporate strategy, *Interfaces*,*14*(1) 84-95.

Rouse, M. (2012). *Big data analytics*. TechTarget.com Retrieved 15/09/2014 from: http://searchbusinessanalytics.techtarget.com/definition/big-data-analytics.

Saaty, T. L. (2006). Rank from comparisons and from ratings in the analytic hierarchy/network processes. *European Journal of Operational Research*, *168*(2), 557-570.

Saunders, M. Thornhill, A. & Lewis, P. (2009). *Research Methods for Business Students,* 5th edition. Pearson Education;.

Saxena, K., & Rajpoot, D. S. (2009). A Way to Understand Various Patterns of Data Mining Techniques for Selected Domains. *arXiv preprint arXiv:0911.0781*.

Schwartz, K. D. (2005). Decisions at the touch of a button. *Coca-Cola Japan puts the fizz back in vending machine sales*.

Shakespeare, S. (2012). *Shakespeare review: An independent review of public sector information*. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/198752/13-744-shakespeare-review-of-public-sector-information.pdf.

Shannon, T. and Zwick, M. (2002). Directed Extended Dependency Analysis for Data Mining, In *Proceedings of 12th International World Organization of Systems and Cybernetics and 4th International Institute for General Systems Studies Workshop*, Pittsburgh.

Shim, J. P., Warkentin, M., Courtney, J. F., Power, D. J., Sharda, R., & Carlsson, C. (2002). Past, present, and future of decision support technology. *Decision Support Systems*, *33*(2), 111-126.

Singh, P. J., & Mansour-Nahra, P. (2006). ISO 9000 in the public sector: a successful case from Australia. *The TQM Magazine*, *18*(2), 131-142.

Singh, R., & Singh, K. (2010). A descriptive classification of causes of data quality problems in data warehousing. *International Journal of Computer Science Issues*, *7*(3), 41-50.

Sivanandam, S. N. & Deepa, S. N. (2008). *Genetic Algorithm Optimization Problems* (pp. 165-209). Berlin, Heidelberg: Springer.

Sloan, V. E. & Sharon B. F. (2010). *Principles for Government Data Mining Preserving Civil Liberties in the Information Age*.

SOA & LL Global. (2011). Experience data quality: How to clean and validate your data. Society of Actuaries. Retrieved from https://www.soa.org/research/research-projects/life-insurance/research-2011-12-data-quality.aspx.

Srinivasa, S., & Bhatnagar, V. (2012). Big Data Analytics. In *Proceedings of the First International Conference on Big Data Analytics BDA* (pp. 24-26).

Srinivasan, S., Singh, J., & Kumar, V. (2011). Multi-agent based decision Support System using Data Mining and Case Based Reasoning. *International Journal of Computer Science Issues (IJCSI)*, *8*(4).

Srivardhana, T., & Pawlowski, S. D. (2007). ERP systems as an enabler of sustained business process innovation: A knowledge-based view. *The Journal of Strategic Information Systems*, *16*(1), 51-69.

Srivastav, A. K. (2010). Impact of ISO 9000 implementation on the organisation. *International Journal of Quality & Reliability Management*, *27*(4), 438-450.

Stellent (2004). University of Alberta increases timely access to policies and procedures. Available at http://dssresources.com/faq/index.php?action=artikel&id=136/. (Accessed on 26/06/2014).

Tan, P.-N., Steinbach, M. and Kumar, V. (2005): *Introduction to Data Mining*, 1st edition. Pearson Education.

Tang, L., & Liu, H. (2010). Community detection and mining in social media. *Synthesis Lectures on Data Mining and Knowledge Discovery*, *2*(1), 1-137.

Team, R. (2014). *The R Project for Statistical Computing*. Vienna: R Foundation.

Theofilis, G. N. (2013). *WEKA Classifiers Summary*. Academia.edu. Available at http://www.academia.edu/5167325/Weka_Classifiers_Summary.(Accessed on 15/12/2014).

To, W. M., Lee, P. K., & Billy, T. W. (2011). ISO 9001: 2000 implementation in the public sector: a survey in Macao SAR, the People's Republic of China. *The TQM Journal*, *23*(1), 59-72

Trapp, R. (2014). *Successful Organizations Need Leaders at All Levels*. Forbes Magazine. Forbes.com. Available online at http://www.forbes.com/sites/rogertrapp/2014/03/23/organizations-need-leaders-at-all-levels/. Last accessed 22nd April 2015.

UK Government Cabinet Office (1999). *Modernising Government*. White Paper. London: HMSO.

Urquhart, C., & Fernandez, W. (2006). *Grounded theory method: the researcher as blank slate and other myths*.

Vassilakis et al. (2004), Integrating e-government public transaction services into public authority workflows, *Electronic Government*, *1*(1) 50

Wadhwa, R., Fridsma, D. B., Saul, M. I., Penrod, L. E., Visweswaran, S., Cooper, G. F., & Chapman, W. (2008). Analysis of a failed clinical decision support system for management of congestive heart failure. In *AMIA Annual Symposium Proceedings* (Vol. 2008, p. 773). American Medical Informatics Association.

Wang, R. Y., Ziad, M., & Lee, Y. W. (2001). *Data quality*. New York: Kluwer.

Watson, I. (2001, May). A Decision Support System for Local Government Regulatory Advice. In *FLAIRS Conference* (pp. 329-333).

Wen, W., Wang, W. K., & Wang, T. H. (2005). A hybrid knowledge-based decision support system for enterprise mergers and acquisitions. *Expert Systems with Applications*, *28*(3), 569-582.

Westland, D., & Al-Khouri, A. M. (2010). Supporting e-government progress in the United Arab Emirates. *Journal of E-Government Studies and Best Practices*, *2010*, 1-9.

Witten, I. H.; Frank, E., & Hall, M. A. (2011). *Data Mining: Practical Machine Learning Tools and Techniques*. 3rd edition. The Morgan Kaufmann Series in Data Management Systems. Morgan Kaufmann.

Witten, I.H. & Frank, E. (1999): *Data Mining: Practical machine learning tools with Java implementations.* Morgan Kaufmann.

Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G. & Vaughan, T. M. (2002). Brain-Computer Interfaces for Communication and Control. *Clinical Neurophysiology, 113*, 767–791.

Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H., ... & Steinberg, D. (2008). Top 10 algorithms in data mining. *Knowledge and Information Systems*, *14*(1), 1-37.

Yan, H., Yu, Z., & Cheng, T. E. (2003). A strategic model for supply chain design with logical constraints: formulation and solution. *Computers & Operations Research*, *30*(14), 2135-2155.

Yin, R. K. (2013). *Case study research: Design and methods*. Sage publications.

Yiu, C. (2012). The Big Data opportunity: Making government faster, smarter and more personal. Policy Exchange. Retrieved from http://www.policyexchange.org.uk/images/publications/the%20big%20data%20opportunity.pdf.

Zafarani, R., Abbasi, M. A., & Liu, H. (2014). *Social Media Mining: An Introduction*. Cambridge University Press.

Zeng, H. and Cheung, Y. M. (2008): *Feature Selection for Clustering on High Dimensional Data*, In PRICAI, pages 913 – 922.

Appendix I: Using Data Mining in Organisations: A Pilot Questionnaire

1. How is data required by decision makers stored and accessed?
   [ ] Totally in database management systems (MS Access, Oracle, etc.)
   [ ] Partly in database management systems
   [ ] Totally manual/hard archive
2. Do you apply on the data stored any regular statistical analysis such as regression analysis or decision support techniques?
   [ ] Yes
   [ ] No
   [ ] Not sure
3. Does the organisation use any forms of Decision Support System (DSS) to support decision makers
   [ ] Yes
   [ ] No
   [ ] Not sure
4. Do you believe, based on knowledge or insight, that systems used by decision makers, whether DSS or others, must be improved?
   [ ] Yes
   [ ] No
   [ ] Not sure
5. Do you believe, based on knowledge or insight, that some of the organisational data quality can be of significant improved if there exists any technique of extracting it?
   [ ] Yes
   [ ] No
   [ ] Not sure
6. Does the organisation have any perspective on data mining uses in the present or the future?
   [ ] Yes
   [ ] No
   [ ] Not sure
7. Would your organisation be part of integrating current systems with data mining techniques for optimising systems used by decision makers?

[ ] Yes

[ ] No

[ ] It needs further studies to decide

Appendix II: Interview Questions

Forward

The risk in not identifying and addressing weaknesses in data quality, or the arrangements that underpin data collection and reporting activities, is that information may be misleading, decision making may be flawed, resources may be wasted, poor services may not be improved, and policy may be ill-founded. There is also a risk that good performance may not be recognised and rewarded.

Questions:

1. What are the major challenges and problems the organisation faces (or has faced) at the decision making level and how do you deal (or have you dealt) with them?
2. Do you refer any of the above problems to insufficient quality of the data received by decision makers?
3. How are data generated, evaluated and stored in your organisation?
4. How does the organisation realise the importance of data quality?
5. How do you assess the quality of the data? In other words, what criteria do you use to make sure that certain data received by decision makers are of quality?
6. How is the process of strategic decision-making undertaken in the organisation?
7. What are the types and nature of the information used in strategic decision-making (For example financial statements, figures, tips, ideas, etc.)?
8. What are the sources of data that reach the decision-makers?
9. How often are decisions taken based on quality data?
10. Are there any policies established or adopted by the organisation for data handling at all levels?
11. How does the reporting process take place in the organisation?

مقابلة حول جودة البيانات في المنظمة

تهدف هذا المقابلات إلى فهم جودة البيانات في المؤسسات العامة وستكون مفيدة لاقتراح أطر تطوير جودة البيانات في هذه المؤسسات.

1. ما هي التحديات والمشاكل الرئيسية التي تواجهها المنظمة (أو قد واجهتها) على مستوى صنع القرار وكيف تتعاملون (أو قد تعاملتم) معها؟
2. هل تعيدون أي من المشاكل المذكورة أعلاه لسبب نقص جودة البيانات التي تصلإلى صناع القرار؟
3. كيف يتم إنشاء البيانات، تقيمها وتخزينهافي المنظمة ؟

4. كيف تقوم المنظمة بإدراك أهمية جودة البيانات؟
5. كيف تقيمون جودة البيانات؟ بعبارة أخرى، ما هي المعايير التي تستخدمونهاللتأكد من أن بيانات معينة هي ذات جودة؟
6. كيف تتم عملية اتخاذ القرارات الاستراتيجية في المنظمة؟
7. ما هي أنواع وطبيعة المعلومات المستخدمة في اتخاذ القرارات الاستراتيجية؟ (مثل البيانات المالية، الأرقام، النصائح، الأفكار، الخ.)
8. ما هي مصادر البيانات التي تؤخذ من قبل صانعي القرار؟
9. كم من القرارات تبنى على بيانات ذات جودة؟
10. 6. هل هناك أي سياسات أنشأتها أو اتخذتها المنظمة للتعامل مع البيانات؟
11. كيف تتم عمليات الإبلاغ، بمعنى آخر تداول المعلومات (Reporting) في المنظمة؟

Appendix III: Sample Interview Transcripts

Al Ain Hospital (22 October 2014)

1. What are the major challenges and problems the organisation faces (or has faced) at the decision making level and how do you deal (or have you dealt) with them?

As a business entity of an enterprise most decisions at strategic level are coming directly from the HQ. for the facility that by itself is a challenge as those decisions sometimes has no consideration for the facility resources availability. For example when the HQ decides that the budget will be reduced to a certain amount, the facility have to face the challenge of finding places to make cuts or cost reduction with the minimum impact possible. Difficult in decision making for the facility comes here. We often use data as productivity compared to service size and staffing to see if cutting some positions would do "mostly we try to cut vacant positions" or reduce the size of the service.

2. Do you refer any of the above problems to insufficient quality of the data received by decision makers?

Some of the problems are due to data quality, but not because its insufficient. As an enterprise we collect huge amounts of data, but the problem is the validation. A lot of data elements are collected at the facility level and also at the enterprise level. Aside from being efforts duplication, this process of collecting same data at 2 different location creates inconsistency. We have always had and still have lots of figures that are reported by the HQ totally different that our in-facility reports. One example is "patient days" it is a data field that we calculate based on the number of omissions, length of stay, and other inputs. And the HQ calculates the same field. We both use the same calculation methodology that the HQ created, however more than often the figure for the same time frame that we calculate comes out different than the one from HQ. it takes about 1 to 2 weeks every quarter just to try and reconcile with the HQ what could be the reason behind the variance. Sometimes we find it's the cut off dates on timeframes and sometimes it is other reasons and sometimes we just cannot pinpoint the reason because everything seems in alignment. At the end of the day if we can't align to the HQ figure we have to accept it as it is. Now, even though it is one figure that has the problem, many areas are affected. The "patient days" figure is used for calculation of many financial and non-financial KPIs, we can fail to meet some of our KPIs due to such issues when we know that in reality we are meeting the KPI.

3. How are data generated, evaluated and stored in your organisation?

At the HQ level, almost all data are generated by the Business intelligence department, then certain reports and dashboards are created and distributed to those who should receive it. Few other departments run their own data from their own systems such as IT and HR. data usefully comes from 3 to 4 different systems for example, Oracle for manpower related data, Malaffi for service related data "number of visits, type of service, patients financials, etc." and risk assessment systems. At the creation of any new report a validation process is conducted to ensure that report is pulling correct data accurately. For reports that can be run at the facility as well as the HQ, the facility is asked to validate the report as well. For example if the report is created to pull the number of cases that had a DVT after going through surgery, at the facility we validate by running the report for a short timeframe then checking the cases to see if they really meet the report criteria, then we run a different time frame and repeat to see if the results are consistent. However data extracts "taking the data fields out of the system raw as they are usually not validated. They should be inserted correctly from the beginning and they have entry criteria therefore they should be accurate "there are some errors due to human errors "

At facilities level reports are generated at few different locations, but mostly from Health Information Management department, business performance/performance management departments, IT, quality. Departments that have Mallafi Modules can run their own reports as well, such as Lab and Radiology. Usually running reports is limited to certain staff within those departments to ensure confidentiality of information, and limit the load on reporting tools. Similar systems to the HQ are used to generate reports at the facility, and very rarely a facility will have different reports than the HQ.

Storage of electronic data is on an enterprise servers, everything is hosted centrally, even applications and facility wide licensed software. Generated reports are stored "if needed to be stored" at each generating department in protected share folders. Otherwise they can be regenerated whenever needed.

As patient records, we still have paper based records "hybrids" and the paper based data is stored in the facility for specific time frames, then the can be moved to secondary storage areas.

4. How does the organisation realise the importance of data quality?

As previously mentioned data quality is very important, and inaccurate data can lead to many problems from losing performance KPIs to providing under standard service for our patients. Simple example, data entry for the patient information into mallafi system, although it's governed by many system restrictions and criteria, error is possible due to human errors or system malfunction. If we enter the wrong patient name and fail to match the information in the system with the actual patient we are treating, there is a risk of administering the wrong treatment and endangering the patient life. Even though this scenario is really rare to happen as we have a patient identification process, but such a simple data quality issue can lead to a huge problem. From business performance perspective, if the data accuracy is not up to standard we could be operating services that are pure cost without income value. For example we have to ensure that the data telling us the number of visits to radiology is actually reliable. Because if the number of visits is less than a certain number, then the service its not covering its costs. Also the data we push to the insurance providers, if quality is not assured they can simply reject our claims and not pay us.

5. How do you assess the quality of the data? In other words, what criteria do you use to make sure that certain data received by decision makers are of quality?

Any data received from the decision makers at the HQ are usually considered valid unless we see that it doesn't align with our data at the facility "if we have corresponding data internally". Usually the HQ allows us to give counter proposals to their data reports if we think their data is wrong. If we can provide evidence to our argument then the data at the HQ level is changed. The criteria of assessment depend on each data report provided.

6. How is the process of strategic decision-making undertaken in the organisation?

Once the HQ set the enterprise strategy, the facility set a strategic planning session on how to adapt and implement the HQ strategy. Data of facility performance and manpower is used to ensure decision making is supported with availability of requirements

7. What are the types and nature of the information used in strategic decision-making (For example financial statements, figures, tips, ideas, etc.)?

Staffing plans, financial statements, for decisions related to capex justification are required in terms of facility performance "number and trend of visits" and expected outcomes is required as well

8. What are the sources of data that reach the decision-makers?

As previously mentioned there are certain data sources in the enterprise such as the Oracle system and the EMR "malaffi" system and other data creating systems such as Patient Safety Net which produce data about incident reporting. All these sources are accessible for decision makers at both the enterprise and the facility level. Even if they don't have a direct reporting access, they can request the data and its provided to them.

9. How often are decisions taken based on quality data?

At facility level, decisions are either taken based on data or they are implementation to the HQ decisions. For example, the hospital outpatient clinics are currently opened at the evening, there was a data review to make a decision on whether or not it is worth continue having them open. The data statistics shows that some of the clinics are not as active as they should be; therefore the facility is now in the process of finalizing a decision to close them. in the other hand, there was a mandate from the HQ to see only "Thiqa" insurance holders in certain hours within the day, this project had to be implemented as mandated even though our current stat shows that we will suffer a drop on volume due to this decision but there is nothing much to do about it. From the HQ level this mad ate was not done based on our data, but rather based on a strategic vision. After the implementation however, we are now monitoring the data to make decisions on resources allocation and service quality like waiting times

10. Are there any policies established or adopted by the organisation for data handling at all levels?

there are few policies adopted regarding data handling. we have a data validity policy from the HQ, we also have data integrity policy and confidentiality policy. But as far as how reporting is done and accessibility to reports there are no policies currently. For the communication of information within the facility and with the HQ we have a communication plan.

11. How does the reporting process take place in the organisation?

There are certain reports that are run periodically and provided to certain key individuals and/or groups in the facility. an example is the KPI reports on monthly or quarterly bases.

There are other reports created based on request and provided to the requester. an example is diagnosis related reports for specific studies or research or projects like the annual SEHA Transformational Event project.

Abu Dhabi Police (5 June 2013)

1. كيف تقوم المنظمة بإدراك أهمية جودة البيانات؟

مشكلة البيانات عندنا مشكلة قديمة اي متخذ قرار يحتاج الي بيانات في الزمن الحالي اصبحت المعلومات من الشغلات الداعمة لمتخذ القرار يعني يمكن الموارد بدات تقل فاستثماراتك لازم تكون ذكية جدا يعني هل انا بستثمر في التدريب؟ في بناء مركز؟ على اساس انته تستثمر بشكل صحيح لازم يكون عندك ارقام دقيقة وتحليل ادق فالقيادة في السنوات الاخيره بشكل عام ولما اتكلم عن القيادة اقصد فيها من مدير فرع الى اعلى متخذ قرار في الموسسة بدات تدرك ان العملية عندنا الوقت محدود التحديات كبيرة الامور اصعب عدم الوضوح في المتغيرات المستقبلية اصعب فاصبحت المعلومات عامل جذري لمحاولت التنبا بالمستقبل كتحليل أصبحت المعلومات والبيانات والاحصائيات عامل مهم جدا في تحديد اولوياتنا كاهداف استراتيجية وايضا اهم المشاريع الي عندنا فالقيادة في الفترة الي طافت بسبب كل المتغيرات الي قاعدة تستوي ادركت ان المعومات لها اهمية كبيرة ويمكن هذا اصبح من اهم المواضيع الساخنة الي تناقش في اجتماعاتنا دقة البيانات في الفترة الي طافت.

2. ما هي المعايير التي تستخدمونها لتقييم جودة البيانات؟ بعبارة أخرى، كيف تقيمون جودة البيانات؟

القيادة تستخدم عده طرق لتقييم جوده البيانات مثل المقارنات الداخلية المقارنات بالفترة السابقة المقارنة بين التقارير الصادرة من جهات مختلفة المراجعة الدورية للارقام هذي تحليل الارقام هذي كلها خطوات نستخدمها لتاكد هل هذا الرقم صحيح او لا بمعنى لو انا اتكلم عن الوفيات على الطرق انا لست الجهه الوحيدة التي تصدر البيان هذا في عندنا جهات شريكة لنا مثل هئه الصحة على المستوى الاتحادي والمحلي فدائما نقارن ارقامنا بارقامهم ونحاول نعرف وين الاخطاء ونقارن الاتمدات في السنوات القليلة الماضية على اساس لو انشوف قفزات نوعية تبدا تسال ليش لو كان منطقي تعرف ان هذي الارقام فيها جزء من الصحة فهذي الادوات الي ممكن محم نستخدمها في التدقيق على بياناتنا الحالية بالاضافة الى انه كان فيه عده مشاريع لمراجعة الانظمة الاحصائية والاكترونية المختلفة الي عندنا لتدقيق مخزن البيانات نفسة والمعلومات نفسها.

3. هل يوجد اي صعوبات اوتحديات في جودة البيانات؟

الصعوبات كثرة والصعوبات تبدا من مدخلين البيانات فسهم ودراكهم لاهمية البيانات هذي وادراكهم للادخال الصحيح من الانظمة الاكترونية اذا ما كان فيها الحقول المناسبة وطلب المعومات المناسبة من على سبيل المثال ترابط قواعد البيانات الي تسهل لي اساخراج الاحصائيات بشكل دقيق ومتكامل من قدرات الموسسة من تحليل الموشرات بشكلها الدقيق الى اتخاذ القرار المناسب من المعومات والاحصائيات المشار اليها هذه كلها تحديات كانت موجوده و لا اقول انها انتهت ولاكن القيادة ادركت وبدات تعرف اين مشاكلها حتى ابسط شي تعريف الموشر نفسه كبطاقة تعريف واليه احتساب واليه معادلة بدينا نخوض فيها ونشوف هل هذه المعادلة الاصح؟ وهل لو قارناها بموسسات مرجعية اخرى هل هذه المعادلة تقارن؟ مثل يوم اقول جرائم مغلقة الية تعريف الجرائم المغلقة كتعريف ومعادلة رياضية كاحتساب تختلف من دولة الى اخرى وانا عشان يعطيني الرقم قرائه حقيقية

لازم اعمله مقارنات مرجعية المقارنات هذه تكون اصعب لو اليه الاحتساب ونطاق كلمه جريمة مغلقة يعني اليوم جريمة المخدرات عندي انا تعتبر جريمة مغلقة ولاكن خارج الدولة في بريطانيا وامريكا لا تعتبر جريمة مغلقة وهنا لما احسب اجمالي الجرائم المغلقة مقابل امريكا وبريطانيا فالناتج لا يكون صحيحا وبتالي متخذ القرار لا يقدر يعطي تصور واضح تعال الى اليه الاحتساب كمعادلة رياضية انا اقسمها على عدد السكان على عدد السكان في بداية السنه على عدد السكان في اخر السنه او كل شهر بشهر هذي كلها تفاصيل دقيقه وكلها اضطرينا نخوض فيها بحيث نجود من البيانات والمعلومات المتوفرة لدينا بتداء من الادخال الى الاستخراج والتحليل الى اخ.

4. كم من القرارات تبنى على بيانات ذات جودة؟

صعبه عملية الاحتساب القيادة العامة لشرطة ابوظبي ضخمة جدا القرارات تاخذ فيها بشكل مستمر داء وادائم من فروع مترامية الاطراف على مستوى الامارة بتخصصات وخدمات ضخمة جدا يعنى مرورية امنية جنائية موراد بشرية مالية الى اعلى متخذ قرار فاتخاذ القرار على الاحصائيات عملية صعة جدا للاكتساب فما بالك الدقيقة منها لانى انا اساسا لا اعرف نسبه الدقيقة منها من عدمه و70% القرارات ذات جوده ولاكن هذا تقدير شخصي ويمكن ان اكون قد اخطات فيه ولاكنه غير مبنى على الي قاعدة من قواعد المعاير.ولاذالك لايمكن احتسابه او العمل به.

5؟ ما هي مصادر البيانات التي تؤخذ من قبل صانعي القرار

نحن عندنا انظمة رئيسية الكترونية وانظمة ادارية النظام الجنائي النظام المروري نظام امن المنافذ نظام الموارد البشرية نظام المالية نظام الدفاع المدنى كموسسة هذي كلها انظمة حقيقية موجوده عندنت لدعم متخذ القرار وطلع إحصائيات وموشرات وبيانات هذه البرامج الموجوده عندنا بالاضافة الى العدد الضخم من المعلومات الي هيه مختلفة من مصادر مختلفة سواء كانت داخل او خارج الموسسة

6. ما هي أنواع وطبيعة المعلومات المستخدمة في اتخاذ القرارات الاستراتيجية؟ (مثل البيانات المالية، الأرقام، النصائح، الأفكار، الخ.)

أنواع كثيره على حسب طبيعة الموضوع اذا نتكلم عن الموارد البشرية بتاخذ معومات مواد بشرية واذا تتكلم عن المالية بتاخذ معلومات مالية الجريمة غير عن مرورية غير عن سجون على حسب القرار الموتخذ ولاكن نحن دائما بغض النظر عن المعلومات ومصدرها نحن نحاول نرسم صور واضحة لهذه المعلومات يعني معلومات ثلاث او اربع سنوات عشان نرسم ترنت معين تنبات مستقبلية مقارنات مرجعية للارقام هذي مع دول مختلفة تقارير دولية بحيث انك حيث انك فعلا تحصل على معلومة الان الرقم نفسة ليس له قيمة ولاكن الرقم مقابل مستهدف الرقم مقابل الثلاث السنوات الماضية مقابل مقارنات مرجعية مجموعه ارقام تودي معلومة

مثل مجموع الوفيات على الطرق في مكان معين بالاضافه على الجريمة في مكان معين يمكن يعطيك قراءه مختلفه او المخالفين والوفيات والجرائم يعني انا جبت المخالفين من نظام والجرائم من نظام ولاكن عطاني صوره اكبر لانه فيه صورة تكاملية في الاحصائيات وهنيه عملية معقده جدا. وجواب السوال هوه على حسب الموضوع ولاكم نحن دائما نسعى لرسم صوره واضحة بستخدام كل المعلومات المطلوبة لرسم الصورة الواضحة

هل الراي الشخصي او الحليل الشخصي يوثر على اتخاذ القرار؟.7

هذا شي طبيعي نحن بشر ليس كل القادة يعتمد على الارقام بنفس الدقة ويصدقها يعني مثل استطلاعات الراي اليومي وهي الادراك وعملية انطباعات معينة لمتخذ القرار و لناس معينين وهل هذي يعني كل القادة يومنوا فيها بنفس الدرجة ؟ لا البعض يومن بها ايمان كبير والبعض الاخر لا طيب ارتفاع جريمة معينة هل معناه ان هذه ظاهرة حقيقية؟ البعض منهم يقول لا والبعض الاخر يقول اتوقع انها سوف تتغير ما اظن تستمر فاذا مسالة الحكم الشخصي على البيانات والارقام تختلف بختلاف اطباع القادة بختلاف قدراتهم وتجاربهم في الحياة ودرجة تعليمهم ولذلك فان الجانب الشخصي له تاكثر كبير على تفسير البيانات وبتالي على اتخاذ القرار

هل هناك أي سياسات أنشأتها أو اتخذتها المنظمة للتعامل مع البيانات؟.8

فيه منهجية اتخاذ قرار حاولنا ندعم فيها الموضوع هذا فيه دورات تدريبيه كثيره فيه اليوم استحدثنا وحدات تنظيمية داعمة لتخاذ القرار مثل مركز دعم اتخاذ القرار مثل المعلومات الامنية مثل ادارة الاستراتيجية وموشرات استراتيجية هذي كلها هياكل تنظيمية ووحدات تنظيمية .

كيف يتم إنشاء البيانات، تقيمها وتخزينها؟.9

نحن موسسة كبيرة بيانات شغالة على مدار الساعة مروري جنائي منافذ كل الادخالات مستمرة بغض النظر عن الانظمة الداخلية مثل المالية والموارد البشرية ولاكن هذه الانظمة الميدانية شغالة على مدار الساعة طيلة ايام الاسبوع موسسة خدمية تدعم المجتمع فيتم ادخال البيانات على حسب الجرائم وغيره بشكل امن وغيره ويوجد لدينا انظمة كثيره مثل امن المعلومات وغيره تخزينها ومساندتها على حسب منهجية معينة تخزن فيها المعلومات بشكل دائم ومستمر

ما هي الأدوار والمسؤوليات المناطة بالموظف فيما يتعلق بجودة البيانات؟.10

كل انسان له دور معين يعني انته عندك الوحدات التنظيمية مدخلين البيانات لهم ايضا دور يتم تدريبهم وتاهيلهم وغيره المعلومات الامنية الموجودة في الوحدات التنظيمية لهم دور ادارة المعلومات الامنية المركزية لها دور قيادات الشرطة كمدراء عامين لهم دور تختلف الادوار من مستوى للاخر ولاكن هذه الادوار المفوض ان تسهم

في دقه البيانات هذي أدارة الاستراتيجية لها دور ولاكن الدور الاهم والاكبر هوه دوم مدخلين البيانات نفسهم المدراء المباشرين وايضا المعلومات الامنية .

12. كيف تتم عمليات الإبلاغ، بمعنى آخر تداول المعلومات (Reporting) في المنظمة ؟

طبعا نحن منظمة سرية ومعلوماتنا معتمدة بشكل كبير على السرية فيه معلاومات متعارف عليها تقارير توصل نحن عندنا مثلا اليوم المعلومات العامة والموشرات العامة مثل اجماعات اسيوعية في مراكز الشرطة اجتماعات شهرية في مديريات اجتماعات ربع سنوية على مستوى الادارات العامة والقائد العام وسمو الوزير هذا التسلسل المنطقي ولابعدين على حسب احاجة والطلب اوموضوع دراسة معينة او مشكلة قائمة وغيره يمكن تستحدث طلب البيانات